VALIDATION AND SCALING OF QUASI-SUBJECTIVE SURVEY MEASUREMENTS

Table of Contents

This proposal is submitted by the National Academy of Sciences which assumes responsibility for handling and accounting for funds. The individual investigator assumes full responsibility for the scientific work to be carried out under any grant resulting from this proposal.

NATIONAL SCIENCE FOUNDATION

PROJECT SUMMARY

FOR NSF USE ONLY			
DIRECTORATE/DIVISION	PROGRAM OR SECTION	PROPOSAL NO.	F.Y.

NAME OF INSTITUTION (INCLUDE BRANCH/CAMPUS AND SCHOOL OR DIVISION)

National Academy of Sciences
National Research Council

ADDRESS (INCLUDE DEPARTMENT)

Commission on Behavioral and Social Sciences and Education
2101 Constitution Avenue
Washington D.C. 20418

PRINCIPAL INVESTIGATOR(S)

Charles F. Turner

TITLE OF PROJECT

Validation and Scaling of Quasi-Subjective Survey Measurements

TECHNICAL ABSTRACT (LIMIT TO 22 PICA OR 18 ELITE TYPEWRITTEN LINES)

The proposed research exploits a unique body of data in which quasi-subjective measurements made in large-scale health surveys were supplemented by physical examinations of the survey respondents. The resulting combination of objective and subjective measurements affords us a unique vantage point from which to examine the measurement qualities, usefulness, and limits of quasi-subjective survey measurements.

The proposed research attempts the validation and scaling of this important body of survey measurements. Specific aims of the proposed research are: (1) to demonstrate, if possible, that quasi-subjective data obtained in large-scale surveys can satisfy the rigorous measurement standard proposed by Rasch, (2) to specify the measurement conditions under which this is most likely to occur, and (3) to show, if possible, that objective measurements are related in a consistent and orderly manner both to responses to specific quasi-subjective survey questions and also to latent-trait scalings derived from those responses.

VALIDATION AND SCALING OF QUASI-SUBJECTIVE SURVEY MEASUREMENTS

> One must ask whether the tangibility or meaningfulness of ... subjective measurements can be substantiated. Too often it is assumed that anything to which people will respond constitutes a measurement. This, however, begs the question of whether there exists something to be measured. It is surely difficult to demonstrate that such measurements fit into some coherent pattern of orderly (theoretical or empirical) relationships. It is not, however, beyond the reach of our current abilities.
>
> Validation provides a fundamental challenge to all who would take subjective measurements seriously.
>
> \- Surveying Subjective Phenomena[1]

---

[1] Report of the Panel on Survey Measurement of Subjective Phenomena, Committee on National Statistics, National Academy of Sciences (Turner and Martin, 1984, Volume 1, p. 325).

The proposed research attempts the validation and scaling of an important body of quasi-subjective[2] survey measurements. This work continues the study of the validity and reliability of subjective and quasi-subjective measurements begun in 1980 with NSF's support and resulting in the recent publication of the two-volume report, Surveying Subjective Phenomena. A major recommendation of that report was that greater attention should be focused on the meaningfulness of survey responses and that researchers should abandon the practice of assuming that any answer obtained in a survey constituted a measurement.[3]

---

2 We distinguish between survey reports of factual matters (objective phenomena) and reports of subjective phenomena. Subjective phenomena are those that, in principle, can be directly known (if at all) only through the self report of the respondent. Objective phenomena are those that can be known by evidence that is, in principle, directly accessible to external observers. So, for example, respondents' reports of their feelings, opinions, attitudes, beliefs, and so forth would be called subjective while reports of their age, income, etc. would not.

This theoretical division does admit to considerable blurring in practice, hence the term quasi-subjective (or quasi-factual). We classify the the measurements of health we propose studying are termed quasi-subjective because they involve both subjective and objective elements and thus seem intermediate between reports of indisputably objective phenomena (e.g., chronological age) and indisputably subjective phenomena (e.g., attitude toward a political candidate). (For further consideration of this matter, see discussions of the quasi-subjective aspects of ethnicity, housing, employment, and crime victimization measurements (Smith, 1984; Bailar and Rothwell, 1984; Newman, 1984; Martin, 1981)).

3 The resurgence of interest in what are called "non-sampling errors" or "response effects" in surveys has also served to rekindle appreciation for such "validation" studies (see, for example, Bishop et al., 1978; Bradburn and Sudman, 1979; DeMaio, 1983; Duncan and Schuman, 1980); Kalton et al., 1978; Martin, 1984; Schuman and Presser, 1981; Turner and Krauss, 1978; Turner and Martin, 1981, 1984).

STRATEGIES OF VALIDATION

Studies of the meaningfulness (validity) of responses to survey questions differ fundamentally for questions seeking to assess objective versus subjective phenomena. Validation of survey measurements of objective (i.e., factual) phenomena is typically done in what are called "record-check" studies. In such studies, a record that is presumed to be less error-prone than the survey report is obtained, and errors are defined as the deviation between the value obtained from the record and that reported in the survey. Although such studies can be quite costly and do involve some knotty problems,[4] the procedure is conceptually straightforward. Objective phenomena, by definition, are (potentially) observable by persons other than the respondent. Thus it is often possible to gather evidence from other witnesses in order to substantiate the validity of survey measurements of such phenomena. For example, responses to survey questions about age might be checked against birth certificates (which record the testimony of witnesses of the birth). Other examples of the use of such records to validate survey responses include Sudman and Bradburn's (1979) clever use of court and other records to study

---

4 E.g., the convenient assumption that, for example, a birth certificate records the "true value" for date of birth is compromised by the fact that such administrative records are not entirely error-free and, even more crucially, the process of matching survey respondents to such records admits to considerable error.

the errors that occur in obtaining survey reports of sensitive information (e.g., being charged with driving while drunk, filing for bankruptcy, etc.), Ferber's (1966) study comparing bank records to survey reports of savings, and Andersen et al.'s (1979) use of physicians' and hospital records to gauge the accuracy of survey reports of use of and expenditures for medical care.

Validating survey measurements of subjective and quasi-subjective phenomena poses a more difficult theoretical problem. Subjective phenomena are not directly accessible to anyone but the subjects themselves. Thus there are no records or independent observers that can provide evidence of equal standing to that provided by the respondents themselves.[5]

As part of its work, the National Academy of Sciences' Panel on Survey Measurement of Subjective Phenomena presented three studies that demonstrated one approach to the validation of survey reports of subjective phenomena (Beniger, 1985; MacKuen, 1985; MacKuen and Turner, 1985). These studies used time-series data from repeated cross-sectional surveys and attempted to provide evidence of validity by showing that changes over time in the aggregate distributions of subjective measurements were meaningfully associated with cross-temporal changes time in

---

5 Observation of a subject's behavior can only provide an indirect measure of that subject's attitudes, beliefs, opinions, etc. Conceptually we can distinguish between attitudes (and other subjective phenomena) and behavior; in practice, experience affirms that an individual's behavior need not accord with his or her beliefs, feelings, etc. (See also the literature on attitude-behavior inconsistency reviewed by Schuman and Johnson, 1976).

independently-measured objective indicators of the variables thought to be generating changes in the subjective state of the population. (For example, one study demonstrated that over time in repeated cross-sectional surveys of the national population, women's behaviors and attitudes about the safety of various contraceptives changed in predictable directions over time with the quantity of news coverage given to stories about contraceptive safety.)

The examples of validation presented by the panel shared four common characteristics. First, they all involved indisputably subjective phenomena. Second, they all relied on a time-series approach to provide validation at an aggregate level. Third, their survey questions involved the subjective assessment of an external object or event.[6] And, finally, their statistical analyses focused on response to individual survey questions (taken one at a time). The research we now propose expands the domain of these validational studies. It differs from that reported in Surveying Subjective Phenomena in that it focuses on quasi-subjective measurements for which we have respondent-level "objective" data. Thus we will be analyzing validity in the context of single cross-sectional surveys (rather than time-series) and we will focus on the structure of responses to a number of related questions (rather than tracking response to single question).

---

6 I.e., approval of a president, safety of a contraceptive, or importance of a national problem.

This research will use data from a series of national surveys designed to monitor the health of the American population. In addition to their use in monitoring public health, these and similar surveys provide basic research data for a large number of investigators in public health, medicine, epidemiology, psychiatry and psychology. The subjective and quasi-subjective measurements made in these surveys are particularly appropriate for study because

(1) physical measurements and medical diagnoses are readily available for respondents in these surveys, thus making it possible to analyze the relationship between response to survey questions and putatively "objective" measures of physical health,

(2) the subjective measurements made in these surveys are repeated in a large number of other surveys conducted both in the U.S.A. and abroad and so improvements in survey practice could have widespread effects, and

(3) our validation effort, if successful, may help develop important new areas of research involving the secondary analysis of archives of past health survey and environmental data. For example, it may permit linkage of subjective health measurements made in large-scale national surveys with independent data on the geographical distribution of environmental pollutants and other factors thought to affect the public's health.[7]

DESCRIPTION OF DATA

The proposed research will use data from a series of surveys begun in 1959 as the Health Examination Survey (HES) program.

---

7 While merging of such datasets is always possible, the small sample sizes that are typically available for any local area make it impossible to generate reliable estimates for the geographical areas for which pollution data are relevant.

This program was mandated by the National Health Survey Act of 1956 in order "to monitor the health status of the American population" (cited in Engel et al., 1978:1). Three Health Examination Surveys (aimed at different age groups) were conducted between 1959 and 1970 (see NCHS, 1965, 1967, 1969). Subsequently this survey program was expanded to collect nutrition data and was renamed as the National Health and Nutrition Examination Surveys (NHANES) program. Two national surveys (NHANES-1 and NHANES-2) drew samples from the entire population aged 1 to 74 during the 1970s (see NCHS, 1977, 1978, 1981), and a survey restricted to the Hispanic population was completed in 1984.[8]

The HES and NHANES survey programs were designed to obtain a very wide range of health data from large probability samples of the American population. Extensive self-report data were obtained in personal interviews. Basic interview data on health status, disease history, medical treatments, and various symptoms and health-related behaviors were obtained during household interviews by Census Bureau interviewers. Appointments were then made for examinations that were conducted in specially designed mobile examination centers.

Collection of the initial health interview data from adult respondents required a 27-page questionnaire in NHANES-2. At the time of their examination, respondents were also administered a (24-page) supplemental health history interview together with

---

8 NHANES-3 is scheduled to survey the entire population beginning in 1986.

additional questionnaires on diet, medications used, vitamins taken, and health related behaviors. Figure 1 (from NCHS, 1981) presents a more detailed statement of the interview data collected in these surveys.

In addition to these personal interviews, most respondents underwent extensive physical and medical examinations. NHANES-2, for example, conducted household interviews with a probability sample of 27,803 individuals, and physical examinations (lasting two to three hours) were subsequently performed on 20,325 of these respondents. These examinations included: a comprehensive medical exam by a physician; anthropometric measurements; blood tests; urine analyses; hearing tests; speech tests; allergy tests; spirometry tests of lung function; electrocardiograms; X-rays of chest, neck, and back; liver tests (bile acid); and glucose tolerance tests.

The NHANES surveys are the primary source of data for our research. There are two supplemental sources of relevant data: (1) a split-ballot health survey[9] of a probability sample of adult residents of the Boston SMSA (N = 2,991; Turner et al., 1984), and (2) the National Health Interview Surveys (NHIS) which have collected data annually since 1957 from probability samples of the American population (current samples include approximately 42,000 households).

---

9 In a split-ballot survey, randomly selected subset(s) of the sample are administered different questionnaires. This technique is commonly used to study the effect of question wording, questionnaire context, and other non-sampling factors on survey response (see, for example, Schuman and Presser, 1981:313).

FIGURE 1

INTERVIEW DATA COLLECTED IN NATIONAL HEALTH AND NUTRITION SURVEYS

Household questionnaire.—For each household member, this questionnaire included the family relationships; certain demographic items such as age, sex, and race; selected housing information; items such as occupation, income, veteran status; and an indication of participation in food stamp programs.

Medical history questionnaires.—For each sample person at ages 6 months to 11 years a questionnaire included items on birth weight, prematurity, developmental congenital conditions, medication, neurological conditions, lead poisoning, accidents, hospital care, disability, diarrhea, pica, vision, and a variety of chronic conditions. In addition, there were major sections on allergies, kidney and bladder disease, anemia, speech and hearing, lung and chest conditions, and participation in food programs.

Two questionnaires for each sample person at ages 12-74 years included items on medication; hospital care and tuberculosis; nutrition; a variety of acute and chronic diseases; tobacco, tea, and coffee usage; physical activity; weight; height; vision disability; exposure to pesticides; gastrointestinal problems; and for females, a menstrual and pregnancy history. In addition, there were major sections on anemia, diabetes, respiratory condition, hearing and speech, liver and gallbladder conditions, kidney and bladder disease, allergies, hypertension, cardiovascular conditions, stroke, arthritis (stressing middle and upper back and neck problems), and participation in food programs.

Two dietary questionnaires.—For each sample person, a dietitian recorded the quantity of every item of food or drink consumed during the previous day, so that after computer calculation, the data yielded measures of calories, cholesterol, fat, unsaturated fats, protein, carbohydrates, and specific vitamins and minerals consumed during the recall period.

A food frequency interview ascertained the usual pattern of food consumption, recording whether or not it included any foods in various groupings, including milk, meat, fish, eggs, fats and oils, legumes and nuts, cereals, fruits, vegetables, and alcoholic beverages. It also showed reported daily and/or weekly number of times each food was consumed and noted the use of salt and vitamin and mineral supplements.

Medications and vitamin usage.—This elicited a history of the preceding week's usage of any medicines, vitamins, or minerals, for all examined persons.

Dietary supplement interview form.—This form recorded the history of special diets, prior medications, and barriers to purchasing groceries or eating foods for examined persons aged 12-74 years.

Behavior questionnaire.—This questionnaire elicited data on behavior possibly associated with coronary heart disease for examined persons 25-74 years of age.

Source: NCHS, Plan and Operation of the Second National Health and Nutrition Survey, 1976-80. Vital and Health Statistics, 1981, Series 1, No. 15.

The Boston split-ballot survey was primarily concerned with pulmonary health, although some other questions were asked. The pulmonary health questions asked in this survey were identical to those asked in the NHANES, and physical measurements of lung function are also available for a small subsample (n=200) of these respondents. This survey is of interest since it incorporated an experimental variation of questionnaire context. In one-half of the surveys, respondents were first asked to characterize their overall health status[10] and then they were asked detailed questions about any pulmonary disorders they might be experiencing. The other half of the respondents in the survey were first asked the questions about pulmonary disorders and then they were asked to characterize their general health. Preliminary analyses of the results obtained from this experiment suggest that responses to such health survey questions are vulnerable to some theoretically interesting response effects. For example, as suggested by the data displayed in Appendix Figure 1, age-graded expectations about what is "good health" (i.e., good for someone my age) appear to moderate the context effect created by the format of the questionnaire.[11]

---

[10] The question wording was same as used in NHANES and NHIS surveys: "Would you say that your health in general is excellent, very good, fair, or poor?"

[11] Thus far we have done only some crude analyses using the categories shown in Appendix Figure 1 for Age (A), Disorders reported (D), Context (C), and self-assessed Health (H). A minimally constrained model that provided an acceptable fit to the counts in the 4-way crosstabulation (HxAxDxC) is specified by fitting the margins: {CHA}{HD}{CAD}. For this model, L^2 is 19.6 with 20 degrees of freedom. (Fitting the {CHA} marginals

(Cont'd)

The NHIS data may also be of interest to us because they cover some similar topics and include a few identical questions. However, the NHIS had generally sought to restrict its data collection to reports of hospital or physician's visits, doctor's diagnosed of disease, and similar information, so its usefulness for our purposes will be limited.

INITIAL FOCUS OF ANALYSIS

The range of topics for which the NHANES provides both subjective and quasi-subjective survey data, physical measurements, and medical opinions is extraordinarily wide. Therefore, choosing a promising starting point is an important task. Our work with these data will concentrate initially on a series of quasi-subjective measures of lung disorders. These measures provide a particularly promising area for study because there have been several attempts over the last two decades to develop a standard set of survey questions for use in pulmonary epidemiology.[12] These measures are also of importance due to the increasing reliance placed on these and similar quasi-subjective

---

significantly improves fit over a model that fits only the 2-way marginals for the Health response: {CH} {AH} {DH} plus {CAD}.)

12 See historical review by Samet (1979) and results of standardization project supported by the American Thoracic Society and the Division of Lung Diseases of the National Heart, Lung, and Blood Institute (Ferris et al., 1979).

survey measurements in regulatory decisionmaking[13] and in basic research on effects of air pollution on public health.[14] These measures include items such as the following series on shortness of breath:[15]

1. Are you troubled by shortness of breath when hurrying on level ground or walking up a slight hill?
2. When walking on level ground do you have to walk more slowly than other people your own age because of breathlessness?
3. When you are walking on level ground at your own pace do you ever have to stop for breath?
4. Do you ever have to stop for breath after walking a few minutes on level ground?
5. Are you too breathless to leave the house or breathless on dressing or undressing?

---

[13] Surveys of pulmonary health in local communities are frequently commissioned when anticipated changes in the environment are expected to affect the levels of pollutants to which the communities will be exposed. The Boston split-ballot study, for example, was mandated by the Massachusetts Department of Public Health and the Environmental Protection Agency as part of an environmental impact report for a proposed conversion (from oil to coal) of a power plant in Boston.

[14] When asked why lung disorders receive so much attention in environmental health studies, physicians are quick to point out that (besides the empirical evidence of a linkage) the lung is the organ with the greatest surface exposure to the external environment. Its exposed area exceeds the surface area of the skin by a very large margin.

[15] Response categories are "yes" and "no" (and, of course a small number of "don't know," etc.). Respondents who answer "no" to question 1 are not asked questions 2 through 5, however those who answer "yes" are asked <u>all</u> of the remaining questions.

Similar series of questions cover symptoms such as chronic coughing, phlegm production, and wheezing.

IMPORTANCE OF SUBSTANTIVE AREA

The survey topic chosen for initial study is not merely methodologically convenient. It involves a very serious national health problem. Between 50,000 and 60,000 Americans die each year of chronic obstructive pulmonary diseases, and it is estimated that as many as 20 million Americans may suffer from such disorders.

Although there is a relatively large literature on the epidemiology of such disorders, the research methodology used in past studies has been less than optimal. Almost all studies of living populations either collected physical measurements from convenience samples of restricted populations (e.g., residents of Tucson, participants in a particular Health Maintenance Organization, etc.) or they limited their data collection to subjective symptom reports and other information that can be readily gathered in a national sample survey. Taken alone, either of these research designs faces serious problems. Use of restricted populations (often obtained without probability sampling) raises questions about the extent to which research findings will generalize to the rest of the American population. Use of subjective survey data alone inevitably raises questions about the validity and reliability of the resultant measurements of pulmonary disorder.

WORK COMPLETED OR IN-PROGRESS[16]

In the first stage of this research (which has not been supported by NSF) we developed improved baseline models for pulmonary functioning (as measured by standard spirometry procedures).[17] We believe these estimates are superior to previous estimates of "normal" pulmonary functioning for two reasons: (1) we used data derived from full-probability samples designed to be representative of the population of the continental U.S.A. rather than the geographically limited or non-probability samples used in past studies, and (2) we used more adequate techniques of statistical analysis.

---

[16] The first stage of this research (described in the present section of the proposal) is supported by a modest research grant of $20,840 in direct costs from the the Russell Sage Foundation. We presently have available sufficient funding to complete our estimation of the basic spirometry equations that will be used in the proposed (second stage) validity studies. Additional funding for analysis of the basic spirometry data may be requested from NIH (or EPA) to cover some special analyses, e.g. to develop norms for youth aged 6 to 24. The availability of this additional funding is not, however, a prerequisite for undertaking the second stage of this research (described in subsequent sections of the proposal). (Funds for this second stage of research are being requested from NSF.)

[17] What physicians call pulmonary or ventilatory function is a measure of the amount of air individuals can exchange with their surroundings either in one full breath or in some time interval. Measurements are made using a machine known as a spirometer which calculates the volumes of air inhaled and exhaled -- among other things. Two common indices of pulmonary function are Force Vital Capacity (FVC) and Forced Expiratory Volume in one second (FEV-1). FVC is the total volume of air in liters exhaled after a single full inspiration of air. FEV-1 is the total volume of air expired in one second (from the time the subject begins exhaling).

Additional improvements in estimates for the black population are anticipated because detailed anthropometric measurements in our nationally representative samples of the black population allow the calculation of more appropriate adjustments for across-race differences in the relation of height to lung capacity. (This relationship is the largest source of variation in spirometry measurements. Other things being equal, taller people have larger lung capacities.)

While the adult spirometry data from NHANES-1[18] have been publicly available since 1981, the spirometry measurements have not previously been intensively analyzed. The sole publication using these data is a set of tables published by National Center for Health Statistics listing the averages for spirometry measurements broken down by age, sex, and smoking history (O'Brien and Drizd, 1981). (The spirometry data from NHANES-2 have not yet been released.)

Our analyses of the NHANES-1 spirometry measurements have produced some interesting and potentially important results:

- Exploratory analyses indicate that the normative equations used for clinical diagnosis in pulmonary medicine[19] are deficient in their functional form, in their fit to the

---

18 In NHANES-1 questionnaire data were collected from 18,836 respondents. A subsample of 6,913 adults aged 25 to 74 were selected for spirometry testing; usable spirometry data were obtained from 5,544 of these subjects.
In NHANES-2, questionnaire data were obtained from 20,322 respondents. A subsample of approximately 6,000 persons aged 6 to 24 were selected for spirometry testing, and it is anticipated that usable data will be available from approximately 5,000 of these subjects.

19 See, for example, Discher and Seinbern, 1970; Cherniak and Raber, 1972; Knudson et al., 1976; Morris, 1976.

data, and in their ability to provide representative norms for the national population. Analyses indicate, for example, that the relationship between height (the dominant factor) and lung volume measurements is quite non-linear -- a result that is consistent with the geometry of the lung. (See Appendix Table 1 which assesses the functional form of the relationship using a log transformation[20] to estimate the equation: FVC = $c(\text{Height})^b$; see also the earlier theoretical arguments of Cole, 1975). As a result of these analyses, alternative "normal" equations for key spirometry parameters have been estimated.

- The norms commonly used in epidemiological research and medical practice may obscure pulmonary disorders in the black population. In particular, past norms have been derived by empirically estimating an equation where "black" was a dummy variable (see Appendix Table 2 for an example of the results of such a procedure). The resulting estimate of the lower lung capacities of the black population (approximately 10 to 15 percent less than whites) was assumed to arise from inherited differences in the ratio of limb to trunk size in the two populations. It is, however, possible that a significant portion of this estimated decrement in lung capacity does not reflect true (i.e., anatomical) differences in lung capacity and/or lung functioning but are due rather to environmental causes (e.g., due to the concentration of the black population in areas of the country, jobs, etc., that expose their lungs to greater environmental insult). This possibility is being empirically tested by re-estimating the normative equations using the extensive

---

20 OLS procedures are used to estimate

$$\log_e(\text{FVC}) = b \log_e(\text{Height}) + c$$

in order to assess the functional form of the relationship between FVC and height. For example, when b equals 1.0, the relation is linear; taking antilogs we would have:

$$\text{FVC} = c\ (\text{Height})^{1.0}$$

It should be realized, of course, that use of the logarithmic transformation changes the units of variance that are being used. Prediction errors are treated as equivalent if they reflect the same ratio of error (rather than the same amount). Besides solving a heteroscedasticity problem that appeared to exist in past analyses, this approach makes considerable theoretical sense. Clearly errors of 0.5 liters made with 5' tall women (whose FVCs at age 20 average about 3 liters) are more serious than the same 0.5 liter error made on 6'5" tall men (whose FVCs at age 20 average about 6 liters). By treating prediction errors as proportions, the logarithmic transformation provides a theoretically reasonable approach to fitting a function relating height to FVC.

anthropometric data on body size available in NHANES.

- We have studied the effects of smoking, aging, and other factors on pulmonary functioning. These analyses replicate and extend the work of other investigators (e.g., Ferris et al., 1979; Seltzer et al., 1974). An example of the results of these analyses is shown in Appendix Table 2. These analyses reveal effects similar in direction (but sometimes different in size) to those found by investigators using non-probability or geographically-restricted samples. These estimates have the advantage of being derived from nationally representative probability samples.

As a secondary benefit of this initial research, preparations for the proposed validational and scaling studies are quite advanced. In particular, a data file for the proposed analyses of NHANES-1 has been extracted from the public use tapes. This file presently contains information for each respondent extracted from four sources (1) the general household questionnaire tape, (2) the basic medical history questionnaire tape, (3) the supplemental medical history tape, and (4) the spirometry test results tape. (Data for the Boston split-ballot survey are also readily available and some data from a special smoking supplement to the 1979 NHIS have also been prepared for analysis.)

NEW WORK PROPOSED

Relating Subjective Reports of Pulmonary Symptoms to Objective Measurements of Pulmonary Function

The second stage of this research (for which NSF funding is now being requested) seeks to test the validity of the quasi-subjective measurements used in this and other epidemiological surveys. Quasi-subjective measurements are an important tool in such research because they can be inexpensively and routinely collected (see, for example, Turner et al., 1984; Helsing et al., 1979; Comstock et al., 1981; Samet, 1978; Samet et al., 1982). But in the medical and other research communities there is considerable skepticism about the validity and reliability of subjective measurements.

Validation of such relatively inexpensive subjective measures on a national sample would provide a basis for more confident use of such easily obtained data to monitor changes over time in the pulmonary health of the national population and to study the variations that occur between different local areas or different subpopulations. On the other hand, a demonstration that some subjective measures do not validate (or have low levels of validity) will allow investigators to make better choices of subjective measurements to be incorporated in future epidemiological surveys.

A variety of exploratory analyses have been performed using subjective report data. These analyses have studied the association between quasi-subjective symptom reports and deviations in observed lung function. These deviations are defined as the difference between the observed FVC or FEV-1 and that predicted by an equation including height, age, sex, race,[21]

---

[21] In future work we hope to be able to use anatomical data on torso size -- which is the more appropriate control.

and body mass (an indicator of obesity).[22] Appendix Tables 3 and 4 illustrate some results of this work. It will be seen that even such amorphous subjective notions as "shortness of breath" and "wheezing" appear to have noteworthy associations with diminished lung function. (Appendix Figure 2 displays the associations between subjective reports of pulmonary disorders and reductions in pulmonary function found in Appendix Tables 3 and 4.)

The foregoing analyses, however, are properly termed exploratory. We have not yet attempted either to scale the various subjective measurements that are available nor to take adequate account of the effects of the complex sample design of the NHANES on our estimates. These are a major part of the work we propose to undertake.

TREATMENT OF COMPLEX SAMPLE DESIGN

The need to examine respondents at a central location led the NHANES designers to use a stratified and highly clustered, multistage, probability sample design. The use of such a complex sampling design affects all variance calculations. In particular, the convenient algorithms for computing variances for simple random samples produce biased estimates of variances when applied

---

[22] In the exploratory analyses presented in Appendix Tables 3 and 4 we also included controls for smoking history. Use of these controls should result in an underestimate of the net association between the quasi-subjective reports and decrements in lung function. This will occur because smoking is known to cause both increased reporting of pulmonary symptoms and actual impairments of lung functioning.

to such complex designs.[23] Since variances and their roots figure crucially in inferential statistics, the analysis of data from such samples is particularly problematic. At a practical level, one implication of the NHANES sample design is that common statistical packages (e.g., SPSS-X, BMD) cannot be used to obtain variance estimates.

In analyzing data from the NHANES surveys we will follow a heuristic strategy of analysis recommended by statisticians at NCHS (see, Landis et al., 1982). This strategy involves the initial use of relatively inexpensive computational procedures that ignore the complex sample design of the NHANES surveys and treat the data as if they were derived from a simple random sampling of the national population. In initial analyses we will employ inexpensive computational procedures, subsequently we will recalculate statistics of particular interest using (more costly) algorithms that take account of the sample clustering and stratification used in the NHANES surveys (see, Landis et al., 1982). (Appropriate computational procedures are available in OSIRIS to produce unbiased estimates of the standard errors of means and regression coefficients when the data are derived from complex sample designs. Appropriate jackknifed Chi-squared tests for log-linear models can be computed using the procedures developed by Fay, 1983, 1985.)

---

23 See Hansen et al.(1953) and Kish (1965). See also Frankel (1971) for some examples of design effects in common federal surveys. For an elementary introduction, see Blalock (1975:520-530).

APPROACHES TO SCALING SUBJECTIVE SURVEY DATA

Survey data have frequently been "scaled" by assigning arbitrary numerical values to response categories (e.g., "agree strongly" = 4; "agree somewhat" = 3, etc.) and then subjecting the resultant data matrix to some form of factor analysis. Such practices suffer from a number of defects. In a careful study, Clogg (1984a, pp. 337-348) shows that the required assumptions of equal spacing of response categories and multivariate normality are often seriously wrong. Crucial properties of the data structure are thus lost when polytomous response distributions are arbitrarily coded and it is then assumed that first and second moments of the resulting distributions are sufficient for analysis. In Notes on Social Measurement, Duncan (1984a) levels even more fundamental charges against such "scalings." He argues that "multiple factor analysis is not really a method of measuring at all (p. 210)" but rather (when its assumptions are met) it is a method of obtaining a minimum estimate of how many distinct hypothetical variables are required to account for a set of observed variables. The procedure provides no basis for inferring that any of the hypothetical variables or the estimates of scores derived for those variables are truly "measurements" in the strict sense of that word.

The fundamental advances that have been made in the last decades in the statistical treatment of cross-classified data have produced a wealth of new alternatives to such questionable procedures. Important statistical research has, for example, used log-linear techniques to adapt latent class and latent trait models to the analysis of categorical self-report data (see, for example, Goodman, 1981; Clogg, 1979; Duncan, 1984a,b). This research has also begun to provide new and statistically rigorous reformulations of traditional models. For example, models have been proposed that incorporate basic notions of the Guttman scale but eschew the deterministic elements of the original formulation and incorporate an accounting for scale errors (Clogg and Sawyer, 1981). Similarly, practical methods for estimating and testing latent-structure models have been developed thus renewing interest in the research applications of models proposed by Lazarsfeld over thirty years ago.

Duncan (1984a; 1984b) has argued persuasively that the measurement model proposed by Rasch (1960/1980) is particularly deserving of exploration by those interested in the scaling of subjective survey data. This model is unique in its separation of the probability of response into separate parameters representing (1) the "difficulty" of the item or question used to measure a given trait, and (2) the position of respondents on that trait.

The basic postulate of the Rasch model is

$$\Lambda_{Ai} = \underline{a}\ \underline{x}_i$$

where Λ_{Ai} represents the odds on a favorable response to item $\underline{a}$ by respondent i, and $\underline{x}_i$ is the value of the latent trait $\underline{x}$ for respondent i.[24] The probability of favorable response is given by:

$$p_{Ai} = \lambda_{Ai} / (1 + \lambda_{Ai})$$

For a test measuring achievement in arithmetic, for example, the item parameter, a, could be called the "easiness" of item a and x_i represents the level of competence of respondent i in arithmetic. For the shortness-of-breath questions in the NHANES, a might represent the "easiness" of the physical task described by the question (perhaps representing the level of physical exertion required) and x_i would represent the respondent's degree of shortness-of-breath.

Item parameters ("easiness") for different items (a, b, c, etc.) are assumed to be invariant across respondents. Similarly, person parameters ("competence" or "shortness of breath") are assumed to be independent of the item used. In this respect, the Rasch model is unique as Wright observes in meeting the requirements set down by Thurstone in the 1920s that

> A measuring instrument must not be affected in its measuring function by the object of measurement. To the extent that its measuring function is so affected, the validity of the instrument is impaired or limited. If a yardstick measured differently because of the fact it was a rug, a picture, or a piece of paper that was being measured, than to that extent the trustworthiness of that yardstick as a measuring device would be impaired. (Thurstone, 1928, cited in Wright, 1980a)

---

24 Note too that the basic postulate implies a unidimensional scale since the probability of a favorable response is affected by a single latent trait.

Duncan notes that not only is it true that the Rasch model achieves the separability of parameters for the measuring instrument (i.e., the items) and the object being measured (the person's competence, shortness of breath, etc.), but the Rasch model is the only model of its general type that does so (in Turner and Martin, Volume 1, p. 224)

Among its other properties, the Rasch model, where it holds, provides numerical values on an interval scale with a common unit for both the latent trait value ascribed to the person and the parameter measuring question difficulty (Duncan, 1984b drawing on Perline et al., 1979). Moreover, the Rasch measurement model can be shown to be a special case of additive conjoint measurement when that theory is modified to accommodate stochastic response structure (see, Duncan, 1984a, p. 216; Perline et al., 1979).

Explicit statistical tests are available to assess whether a given set of items meets the measurement standard of the Rasch model. These tests derive from the postulate of separability which requires, among other things, that estimates of the ratio of two item parameters (e.g., a/b) have the same expected value across all strata of the population (see Duncan, 1984b; and related exposition in Turner and Martin, 1984: Section 6.4). This ratio is estimated by the ratio of the observed frequencies

$$F_{+-}/F_{-+}$$

for the two items. The numerator of this ratio is the count of persons who respond positively to item a and negatively to item b and the denominator represents the reverse pattern. Log-linear techniques for the analysis of contingency tables can be applied to test the invariance of this ratio across strata in the sample.[25]

Application to NHANES data. Besides its other attractive characterictics, we believe the Rasch model's unidimensional scaling of an inferred latent trait provides an appropriate initial scaling model for our work with the NHANES data. In this regard, we would note that the vocabulary used by physicians, medical researchers, and patients often suggests that they conceive of many disorders as varying along a single dimension. Thus they speak of the degree of breathlessness, the extent of impairment, and so forth. We recognize, of course, that this conceptual vocabulary need not correspond to the manner in which symptoms are perceived by those suffering them. We would nonetheless observe that a scaling procedure that captured the undimensionality implicit in the lexicon might also capture an analogous facet of the cognitive structures that survey respondents employ in responding to questions about those symptoms.

Using the NHANES data we will have the opportunity to use the Rasch measurement model to see if the questions on pulmonary disorders provide satisfactory unidimensional scalings of inferred disease traits. To assess the adequacy of the resultant scalings, we can make use of the other self-report data collected in the survey, e.g., reports of physician's diagnoses of lung disorders.

---

[25] The power of this test to reject the Rasch model when it is does not hold for the data increases as the strata become more closely associated with the latent variable (Duncan, 1984b).

In our case, the model postulates that the ratios of item parameters be equivalent, for example, for persons diagnosed as having a lung disease and those diagnosed as "normal." Thus, tests could be performed across strata defined by whether or not the respondents reported having emphysema -- a severe chronic obstructive pulmonary disease that causes extreme shortness of breath. Similarly a larger number of respondents in the NHANES report having chronic bronchitis, a disease that causes a moderate degree of breathlessness. A similar test can be done using this item to stratify the sample. Other tests might use various combinations of diseases and environmental factors (e.g., smoking, working in a dusty job, etc.). We can use the Boston split-ballot data in an analogous manner. In this case the questionnaire manipulation would provide the sample stratification.

In addition to the tests available using strata defined by the other survey data, NHANES affords us a further test typically unavailable to survey researchers. Besides the self-reports of our survey respondents, we have physical measures of lung function derived from the spirometry tests. These measures are the preferred medical tests used in the diagnosis of obstructive lung disorders; they provide an independent standard measurement (made outside of the survey) which we can be used to assess the adequacy of our scaling of the quasi-subjective survey items. If the scale items we obtain from our attempt to fit a Rasch model to the breathlessness data provide a reasonable proxy for the respondents' true degree of lung dysfunction, it should follow that the degree of lung dysfunction estimated using spirometry should be closely associated with $\underline{x_i}$ (the respondent's score under the Rasch model on the latent trait of breathlessness).[26]

The NHANES data will thus permit us to conduct a wide range of strong tests of the applicability and usefulness of the Rasch measurement model in the context of an important national survey program. (While we will concentrate initially on quasi-subjective measures of lung disorders, the available data offers a variety of other topics for further work.)

POTENTIAL BENEFITS OF RESEARCH

At a general level, this research may expand our knowledge of the qualities, uses, and limits of the survey method as a research tool.[27] Such improvements may be important to scientists

---

[26] Since the the symptom of breathlessness admits to three classes of diagnosis, the analysis will require some minor refinements. In particular, while lung disease is the common source of breathlessness, heart disease also produces this symptom. Similarly, low levels of physical conditioning and obesity may also induce breathlessness among subjects who otherwise have normal lung function. For this reason we will conduct separate analyses for respondents who are free of all known heart disorders, and we will take account of "muscle tone" by including a measure of body mass in the normative equation used to estimate "expected" lung function.

[27] Of the need for such improved knowledge, Schuman and Presser (1981:316) observed in the conclusion of Questions and Answers:

> ...asking questions and giving answers are ancient ways of exchanging information. But they take on new meanings in the context of a large-scale survey .... Always, the challenge is to allow ourselves to learn more from the answers gathered so readily today by large-scale surveys. In a positive as well as a negative sense there are many unanswered questions about both questions and answers.

in a wide range of disciplines because social research in many areas has come to depend increasingly on survey data (Presser, 1984).[28] This research may also contribute to practical improvements in the quality and usefulness of survey data in epidemiology, environmental studies, medicine, and public health.[29]

This research may have several other important consequences. One of its most practical involves our ability to make inferences about the objective health status of the national population using "cheap" subjective data. If we are successful in scaling and validating these subjective measurements, it should be possible to append a small number of subjective items to large-scale epidemiological surveys (e.g. the annual Health Interview Survey conducted by the federal government) and then to recalibrate the resulting data to generate estimates of lung dysfunction.[30] Over

---

[28] In 1980, for example, over 25 percent of the articles published in the major American economics and political science journals relied on survey data, and, in sociology, this figure exceeded 50 percent. {Journals were American Economics Review, Journal of Political Economy, and Review of Economics and Statistics (Economics); American Journal of Sociology, American Sociological Review, and Social Forces (Sociology); American Political Science Review, Journal of Politics, American Journal of Political Science).}

[29] With luck, it might even provide agnostic scientists in those disciplines with an example of the "useful" knowledge and practical benefits derived from basic research in the social and behavioral sciences.

[30] In a sweeping statement of the possibilities of such an "item banking" approaches using the Rasch model, Wright (1980b, p. 194) observes that: "When a family of test items are constructed so that they can be calibrated along a single common dimension, and when they are employed so that they retain these calibrations over a useful realm of application,

(Cont'd)

time, the accumulation of data from the annual NHIS surveys would allow small area estimates of health conditions to be made with reasonable statistical precision. It might thus provide a basis for the linking of national databases containing subjective report data on pulmonary health to other national databases containing measurements of ambient air quality for small geographical areas (e.g., the National Aerometric Databank maintained by the Environmental Protection Agency)[31]. Such linkages would provide a unique opportunity for detailed nationwide studies of the effects upon pulmonary health of variations over time and between localities in air quality.

---

then a scientific tool of great simplicity and far-reaching potential becomes available. The "bank" of calibrated items can serve the composition of a wide variety of measuring tests .... Whatever the test, its measures will be expressed on the one common variable defined by the bank."

[31] EPA's National Aerometric Databank contains time-series measurements stretching back to 1970 on the observed levels of a wide variety of air pollutants. This measurement program presently uses reports from monitors at over 8,000 sites throughout the country. While these measurements are not without problems (see National Research Council, 1978) they do nonetheless provide a unique opportunity for innovative research.

REFERENCES

Andersen, R., Kasper, J., Frankel, M.R., and Associates (1979) Total Survey Error. San Francisco: Jossey-Bass.

Beniger, J. (1984) Mass media, contraceptive behavior, and attitudes on abortion: toward a comprehensive model of subjective social change. In Turner, C.F. and Martin, E., Eds. (1984) Surveying Subjective Phenomena. Volume 2. New York: Russell Sage Foundation and Basic Books.

Bishop, G.F., Oldendick, R.W., Tuchfarber, A.J. (1978) Effects of question wording and format on political attitude consistency. Public Opinion Quarterly, 43:81-92.

Bradburn, N.M. and Sudman, S., with associates (1979) Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.

Cherniak, R.M. and Raber, M.B. (1972) Normal standards for ventilatory function using an automated wedge spirometer. American Review of Respiratory Disease, 106, 38-46.

Clogg, C.C. (1984) Some statistical models for analyzing why surveys disagree. In Turner, C.F. and Martin, E., Eds. (1984) Surveying Subjective Phenomena. Volume 2. New York: Russell Sage Foundation and Basic Books.

Clogg, C.C. and Sawyer, D.O. (1981) A comparison of alternative models for analyzing the scalability of response patterns. In Leinhardt, S., Ed., Sociological Methodology, 1981. San Francisco: Jossey-Bass.

Cole, T.J. (1975) Linear and proportional regression models in the prediction of ventilatory function. Journal of the Royal Statistical Society (Series A), 138, 297-337.

Comstock, G.W. et al. (1981) Respiratory survey findings as predictors of disability from respiratory disease. American Review of Respiratory Diseases, 124, 367-371.

DeMaio, T. (1983) Approaches to Developing Questionnaires. (Statistical Policy Working Paper 10) Washington D.C.: Office of Management and Budget.

Discher, D.P. and Seinbern, J. (1970) LS/MFT: Lung screening by meaningful function testing: Part II. American Journal of Public Health, 60, 2361-2385.

Duncan, O.D. and Schuman, H. (1980) Effects of question wording and context: an experiment with religious indicators. Journal of the American Statistical Association, 75:269-275.

Duncan, O.D. (1984a) Notes on Social Measurement: Historical and Critical. New York: Russell Sage Foundation and Basic Books.

Duncan, O.D. (1984b) Rasch measurement in survey research: further examples and discussion. In Turner, C.F. and Martin, E., Eds. (1984) Surveying Subjective Phenomena. Volume 2. New York: Russell Sage Foundation and Basic Books.

Fay, R.E. (1983) CPLX: Contingency Table Analysis for Complex Sample Designs. Program documentation. Washington D.C.: U.S. Bureau of the Census.

Fay, R.E. (1985) A jackknifed Chi-squared test for complex samples. Journal of the American Statistical Association, 80:148-157.

Ferber, R. (1966) The Reliability of Consumer Reports of Financial Assets and Debts. Urbana, Ill.: Bureau of Economic and Business Research, University of Illinois.

Ferris, B.G. et al. (1979) Effects of sulfur oxide and respirable particles on human health. American Review of Respiratory Diseases, 120, 767-779.

Ferris, B.G. et al. (1979) Epidemiology standardization project II: recommended respiratory disease questionnaires for use with adults and children in epidemiological research. American Review of Respiratory Diseases, 1978, 118 (Supplement) 7-53.

Ferris, B.G. et al. (1979) Epidemiology standardization project III: recommended standardized procedures for pulmonary function testing. American Review of Respiratory Diseases, 1978, 118 (Supplement) 54-88.

Frankel, M.R. (1971) Inference from Sample Surveys: An Empirical Investigation. Ann Arbor, Mich.:Institute for Social Research, University of Michigan.

Hansen, M.H., Hurwitz, M.H. and Madow, W.G. (1953) Sample Survey Methods and Theory. New York: Wiley.

Helsing, K. et al. (1979) Comparison of three standardized questionnaires on respiratory symptoms. American Review of Respiratory Diseases, 120, 1221-1231.

Higgins, M. and Keller, J.B. (1975) Seven measures of ventilatory function. American Review of Respiratory Disease, 108, 258-271.

Kalton, G., Collins, M., and Brook, L. (1978) Experiments in wording opinion questions. Journal of the Royal Statistical Society, Series C, 27:149-161.

Kish, L. (1965) Survey Sampling. New York: Wiley.

Knudson, R.J., Slatin, R.C., Lebowitz, M.D., and Burrows, B. (1976) Maximal expiratory flow volume curve: normal standards, variability, and effects of age. American Review of Respiratory Disease, 113, 587-600.

Landis, J.R., Lepkowski, J.M., Eklund, S.A., and Stehouver, S.A. (1982) A statistical methodology for analyzing data from a complex survey: The first National Health and Nutrition Examination Survey. Vital and Health Statistics, Series 2, No. 92, September , 1982.

Lazarsfeld, P.F. (1954) A conceptual introduction to latent structure analysis. In Lazarsfeld, P.F., Ed., Mathematical Thinking in the Social Sciences. New York: Free Press.

MacKuen, M. (1981) Social communication and the mass policy agenda. In MacKuen, M. and Coombs, S.L., Eds., More Than News. Beverly Hills, Ca.: Sage Publications.

MacKuen, M. (1984) Reality, the press and citizens' political agendas. In Turner, C.F. and Martin, E., Eds. (1984) Surveying Subjective Phenomena. Volume 2. New York: Russell Sage Foundation and Basic Books.

MacKuen, M. and Turner, C.F. (1984) The popularity of presidents: 1963-80. In Turner, C.F. and Martin, E., Eds. (1984) Surveying Subjective Phenomena. Volume 2. New York: Russell Sage Foundation and Basic Books.

Martin, E. (1978) Trends in crime victimization: problems of measurement. Paper presented at the 86th annual meeting of the American Psychological Association, August 28 - September 1, 1978.

Martin, E. (1981) A twist on the Heisenberg principle: or how crime affects its reporting. Social Indicators Research, 9:197-223.

Martin, E. (1983) Surveys as social indicators: problems in monitoring trends. In Rossi, P., Wright, J., and Anderson, A. Handbook of Survey Research. New York: Academic Press.

Morris, J.F. (1976) Spirometry in the evaluation of pulmonary function. Western Journal of Medicine, 125, 110-118.

National Center for Health Statistics (1965) Plan, and initial program of the Health Examination Survey. Vital and Health Statistics, Series 1, No. 4.

National Center for Health Statistics (1967) Plan, and operation and response results of a program of children's examinations. Vital and Health Statistics, Series 1, No. 5.

National Center for Health Statistics (1969) Plan and Operation of a Health Examination Survey of U.S. youths 12 - 17 years of age. Vital and Health Statistics, Series 1, No. 8.

National Center for Health Statistics (1977) Plan and Operation of the Health and Nutrition Examination Survey, United States, 1971-1973. Vital and Health Statistics, Series 1, Nos. 10a and 10b.

National Center for Health Statistics (1978) Plan and Operation of the HANES-1 Augmentation Survey of adults 25-74 years, United States, 1974-1975. Vital and Health Statistics, Series 1, Nos. 14.

National Center for Health Statistics (1981) Plan and Operation of the Second National Health and Nutrition Survey, 1976-80. Vital and Health Statistics, Series 1, Nos. 15.

National Research Council (1978) Environmental Monitoring: Volume 4. Washington D.C.: National Academy of Sciences Press.

Newman, S. (1984) Housing research: conceptual and measurement issues. In Turner, C.F. and Martin, E., Eds. (1984) Surveying Subjective Phenomena. Volume 2. New York: Russell Sage Foundation and Basic Books.

O'Brien, R.J. and Drizd, T.A. (1981) Basic data on spirometry in adults 25-74 years of age. Vital and Health Statistics, Series 11, No. 222.

Perline, R., Wright, B.D., and Wainer, H. (1979) The Rasch model as additive conjoint measurement. Applied Psychological Measurement, 3:237-256.

Presser, S. (1984) The use of survey data in basic research in the social sciences. In Turner, C.F. and Martin, E., Eds. (1984) Surveying Subjective Phenomena. Volume 2. New York: Russell Sage Foundation and Basic Books.

Rasch, G. (1960/[1980]) Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Denmarks Paedagogiske Institut. [Expanded edition, Chicago: University of Chicago Press, 1980.]

Samet, J.M. (1979) A historical and epidemiological perspective on respiratory symptoms questionnaires. American Journal of Epidemiology, 108, 435-446.

Samet, J.M. et al. (1982| Respiratory disease in a New Mexico population sample of Hispanic and non-Hispanic whites. American Review of Respiratory Disease, 125, 152-157.

Schuman, H. and Johnson, M. (1976) Attitudes and behavior. In A. Inkeles, Ed., Annual Review of Sociology, 2:161-207.

Schuman, H. and Presser, S. (1981) Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. New York: Academic Press.

Seltzer, C.C., Siegelaub, A.B., Friedman, G.D., and Collen, M.F. (1974) Differences in pulmonary function related to smoking habits and race. American Review of Respiratory Disease, 110, 598-608.

Turner, C.F., Colten, M.E., and Collin, P. (1984) Estimates of Pulmonary Disorder in South Boston and the Boston Metropolitan Area. (Boston Area Pulmonary Health Survey, 1983) Boston: Center for Survey, Research University of Massachusetts and Joint Center for Urban Studies, MIT and Harvard University.

Turner, C.F. and Krauss, E. (1978) Fallible indicators of the subjective state of the nation. American Psychologist, 33, 456-470.

Turner, C.F. and Martin, E., Eds. (1981) Surveys of Subjective Phenomena: Summary Report. Washington D.C.: National Academy of Sciences Press.

Turner, C.F. and Martin, E., Eds. (1984) Surveying Subjective Phenomena. Two volumes. New York: Russell Sage Foundation and Basic Books.

Wright, B.D. (1980a) Foreword to G. Rasch, Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press.

Wright, B.D. (1980b) Afterword to G. Rasch, Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press.

Wright, B.D. and Masters, G.N. (1981) Rating Scale Analysis: Rasch Measurement. Chicago: MESA Press.

TABLE 1

Estimates for Relationship of Height
and Spirometry Measurements Incorporating Race as a Variable
Using a Representative National Sample Aged 25-74
(Data from National Health and Nutrition Examination Survey)

FORCED VITAL CAPACITY (FVC)

EQUATION: $Ln(FVC) = b_1\ (Ht) + b_2\ (Black) + c$ (Log form)

$FVC = b'_2(Race)\ \ c'(Ht)^{b_1}$ (Antilog form where, $c' = e^c$ and $b_2' = e^{b_2}$)

	Estimates							
POPULATION	LOG HEIGHT (b_1)	(se)	BLACK (b_2)	(se)	CONSTANT (c)	R^2	N	BLACK DEFICIT (a)
All Men	3.246	(.098)	-.109	(.005)	-5.345	.303	2648	-10.3%
"Healthy" Men	2.547	(.293)	-.134	(.036)	-2.624	.172	415	-12.5%
All Women	2.771	(.107)	-.148	(.014)	-3.446	.209	2960	-17.7%
"Healthy" Women	2.756	(.204)	-.130	(.026)	-3.357	.243	650	-12.0%

NOTES: "White" is non-black in this analysis (i.e. it includes the "other" category which is mainly composed of persons reporting themselves to be "Hispanic"). Height is measured in inches; Forced Vital Capacity (FVC) is measured in milliliters. "Healthy" includes only those who never smoked and who report no chronic cough, phlegm, shortness of breath, chest trouble, chest pain, heaviness in chest, or asthma. Black is a dummy variable coded 1 if respondent is black, coded zero otherwise.

In these and other preliminary analyses we follow the strategy recommended by Landis et al. (1982) of initially using computational algorithms that assume simple random sampling in deriving our estimates. In final analyses, the effects of sample stratification and clustering in the survey will be explicitly taken into account. See J.R. Landis et al., A statistical methodology for analyzing data from a complex survey: the First National Health and Nutrition Examination Survey. Vital and Health Statistics, Series 2, No. 92, September, 1982.

(a) The estimate of expected deficit in lung function for the black samples can be computed from the equation estimates as: $100\ [1.0 - \text{antilog}\ (b_2)]$.

APPENDIX A

TABLE 2

Estimates for Equations Predicting Spirometry Measurements as a Function of Height, Race, Age, Body Mass and Cigarette Smoking Using a Representative National Sample Aged 25-74 (Data from National Health and Nutrition Examination Survey)

FORCED VITAL CAPACITY (FVC)
and
FORCED EXPIRATORY VOLUME IN 1 SECOND (FEV_1)

EQUATION: $\text{Ln}(\text{FVC or } FEV_1) = b_1 \text{Ln(Ht)} + b_2 \text{(Black)} + b_3 \text{(Age)} + b_4 \text{(Body Mass)} + b_5 \text{(Cigs)} + b_6 \text{(Ex-Smoker)} + c$

		FORCED VITAL CAPACITY		FORCED EXPIRATORY VOLUME IN 1 SECOND	
Coefficients for	b_i	Males	Females	Males	Females
Log_e Height (in inches)	b_1 (se)	2.596 (0.086)	1.999 (.092)	2.398 (.106)	1.753 (.099)
Black	b_2 (se)	-.117 (.012)	-.154 (.012)	-.111 (.015)	-.177 (.013)
Age (in years)	b_3 (se)	-.0085 (.00027)	-.0089 (.0003)	-.0124 (.00033)	-0.012 (.00035)
Body Mass	b_4 (se)	-1.649 (.592)	-2.061 (.452)	.826(ns) (.726)	.0276(ns) (.486)
No. Cigarettes per day	b_5 (se)	-.0019 (.00027)	-.0017 (.0003)	-.0035 (.00033)	-.0031 (.00035)
Ex-Smoker	b_6 (se)	.0019(ns) (.009)	.035 (.010)	-.0047(ns) (.010)	.0276 (.011)
	c	-2.130	+0.249	-1.476	+1.126
	R^2	.507	.453	.502	.487
	N	2620	2939	2620	2939
ESTIMATED DEFICITS FOR: [(a)]					
BLACK Respondents		-11.0%	-14.3%	-10.5%	-16.2%
AGE					
aging 20 years		-15.6%	-16.4%	-22.0%	-21.4%
aging 40 years		-28.8	-30.0	-39.1	-38.1
CIGARETTE SMOKING					
smoking 30 per day		-5.5%	-5.0%	-9.2%	-8.9%
smoking 60 per day		-10.8	-9.7	-18.9	-17.0

NOTES: Samples include both "healthy" and "not healthy" respondents (see notes to Table 1 for definitions). "White" is non-black in this analysis (i.e., it includes the "other" category which is mainly composed of persons reporting themselves to be "Hispanic"). Height is measured in inches; Forced Vital Capacity (FVC) and Forced Expiratory Volume in one second (FEV_1) are measured in milliliters. Black is a dummy variable coded 1 if respondent is black, coded zero otherwise. Ex-Smoker is a dummy variable coded 1 if respondent does not currently smoke but has smoked 100+ cigarettes in lifetime; otherwise it is coded as zero. Body Mass is Quetelet index of body mass, i.e., Weight divided by Height squared (height is in inches and weight in pounds). Cigarettes per day is average of numbers reported: (1) today, and (2) one year ago.

In these and other preliminary analyses we follow the strategy recommended by Landis et al. (1982) of initially using computational algorithms that assume simple random sampling in deriving our estimates. In final analyses, the effects of sample stratification and clustering in the survey will be explicitly taken into account. See J.R. Landis et al., A statistical methodology for analyzing data from a complex survey: the First National Health and Nutrition Examination Survey. Vital and Health Statistics, Series 2, No. 92, September, 1982.

(a) The estimate of expected deficit in lung function for the black samples can be computed from the equation estimates as: $100\ [1.0 - \text{antilog}\ (b_2)]$. Estimates of deficits for variables that take more than one value (e.g., years of age) are computed as: $100\ [1.0 - \text{antilog}\ (n\ b_i)]$, where n is value of variable (e.g., number of years of age). Estimated deficits are net of effects of other variables in equation.

TABLE 3
Estimates for Equation Predicting Spirometry Measurements as a Function of Height, Race, Age, Body Mass, Cigarette Smoking, Sex, Residence in Massachusetts, and Subjective Reports of Wheezing Using a Representative National Sample Aged 25-74 (Data from National Health and Nutrition Examination Survey)

FORCED VITAL CAPACITY (FVC)

Variable	b	s.e.	F-Ratio
Log_e Height (in inches)	2.320241	0.06246	1380.1
Black	-0.1353193	0.00852	252.3
Pipe or Cigar Smoker	-0.0002128	0.00846	0.0
Age (in years)	-0.0086749	0.00018	2204.7
Body Mass (Quetelet's Index)	-1.6776810	0.35222	22.6
Ex-Cigarette Smoker	0.0223850	0.00813	7.5
Years Smoked Cigarettes	-0.0011271	0.00036	9.9
Avg. Cigarettes per day	-0.0002472	0.00025	0.9
Female	-0.1497461	0.00716	437.7
Boston SMSA	-0.0761200	0.02813	7.3
Massachusetts	0.1000896	0.02032	24.2
Report Wheezing	-0.0719965	0.00720	99.9
(Constant)	-0.9534865		
R^2	.64834		
N	5,551		

ESTIMATED DEFICITS OR BENEFITS ASSOCIATED WITH:

RESIDENCE (a)	
living in Massachusetts	+10.5%
living in Boston	-7.7%
REPORTING WHEEZING	-6.9%

NOTES: Samples include both "healthy" and "not healthy" respondents (see notes to Table 1 for definitions). Female, Residence in Massachusetts, and Residence in Boston are dummy variables coded 1 if respondent has attribute (i.e., was female, lived in Massachusetts or Boston), coded zero otherwise). Wheezing is also a dummy variable; it is based on respondent's subjective reports in response to the questions: (1) Does your chest ever sound wheezy or whistling when you don't have a cold? (2) Would you say this happens only occassionally or most days or nights? Variable was coded 1 if repondent answered yes to both questions, it was coded zero otherwise.

See notes to Tables 2 for descriptions of other variables and statistical procedures used in analysis.

(a) Effects are those estimated in above table. Effect for Boston is net of effect for Massachusetts residence. Result indicates that Boston SMSA residents have a deficit of -7.7 percent compared to other Massachusetts residents, but they, in turn, benefit from the fact that on the whole residents of Massachusetts are somewhat healthier (+10.6% higher FVC) than the rest of the nation. Compared to the sample from the whole nation the Boston residents have an estimated overall effect of: +10.6 - 7.7 = +2.9.

TABLE 4

Estimates for Equations Predicting Spirometry Measurements as a Function of Height, Race, Age, Body Mass, Cigarette Smoking, Sex, Residence in Massachusetts, and Subjective Reports of Breathlessness Using a Representative National Sample Aged 25-74 (Data from National Health and Nutrition Examination Survey)

FORCED VITAL CAPACITY (FVC)

Variable	b	s.e.	F-Ratio
Log_e Height (in inches)	2.319176	0.06265	1370.278
Black	-0.1336759	0.00856	243.668
Pipe or Cigar Smoker	-0.0029865	0.00848	0.124
Age (in years)	-0.0086323	0.00019	2152.184
Body Mass (Quetelet's Index)	-1.6087460	0.35447	20.597
Ex-Cigarette Smoker	0.0239416	0.00815	8.622
Years Smoked Cigarettes	-0.0011984	0.00036	11.218
Avg. Cigarettes per day	-0.0003201	0.00025	1.581
Female	-0.1492068	0.00719	431.146
Boston SMSA	-0.0768871	0.02822	7.423
Massachusetts	0.1050879	0.02039	26.564
Breathless: Level 1	-0.0463132	0.01079	18.429
Breathless: Level 2	-0.0926276	0.01607	33.243
Breathless: Level 3	-0.1070939	0.02466	18.856
(Constant)	-0.9562541		
R^2	.64627		
N	5,551		

ESTIMATED DEFICITS FOR:

BREATHLESSNESS	
- when hurrying or walking up slight hill	-4.5%
- must walk more slowly than others same age	-9.2
- must stop for breath when walking at own pace	-10.2

NOTES: Samples include both "healthy" and "not healthy" respondents (see notes to Table 1 for definitions). Breathlessness is a set of three dummy variables; they are derived from respondent's subjective reports in response to the following questions: (1) Are you troubled by shortness of breath when hurrying on level ground or walking up a slight hill? (2) When walking on level ground do you have to walk more slowly than people your own age because of breathlessness? (3) When you are walking on level ground at your usual pace do you ever have to stop for breath? Dummy variables were coded 1 if respondent answered yes to indicated question, coded zero otherwise.

See notes to Tables 2 and 3 for descriptions of other variables and statistical procedures used in analysis.

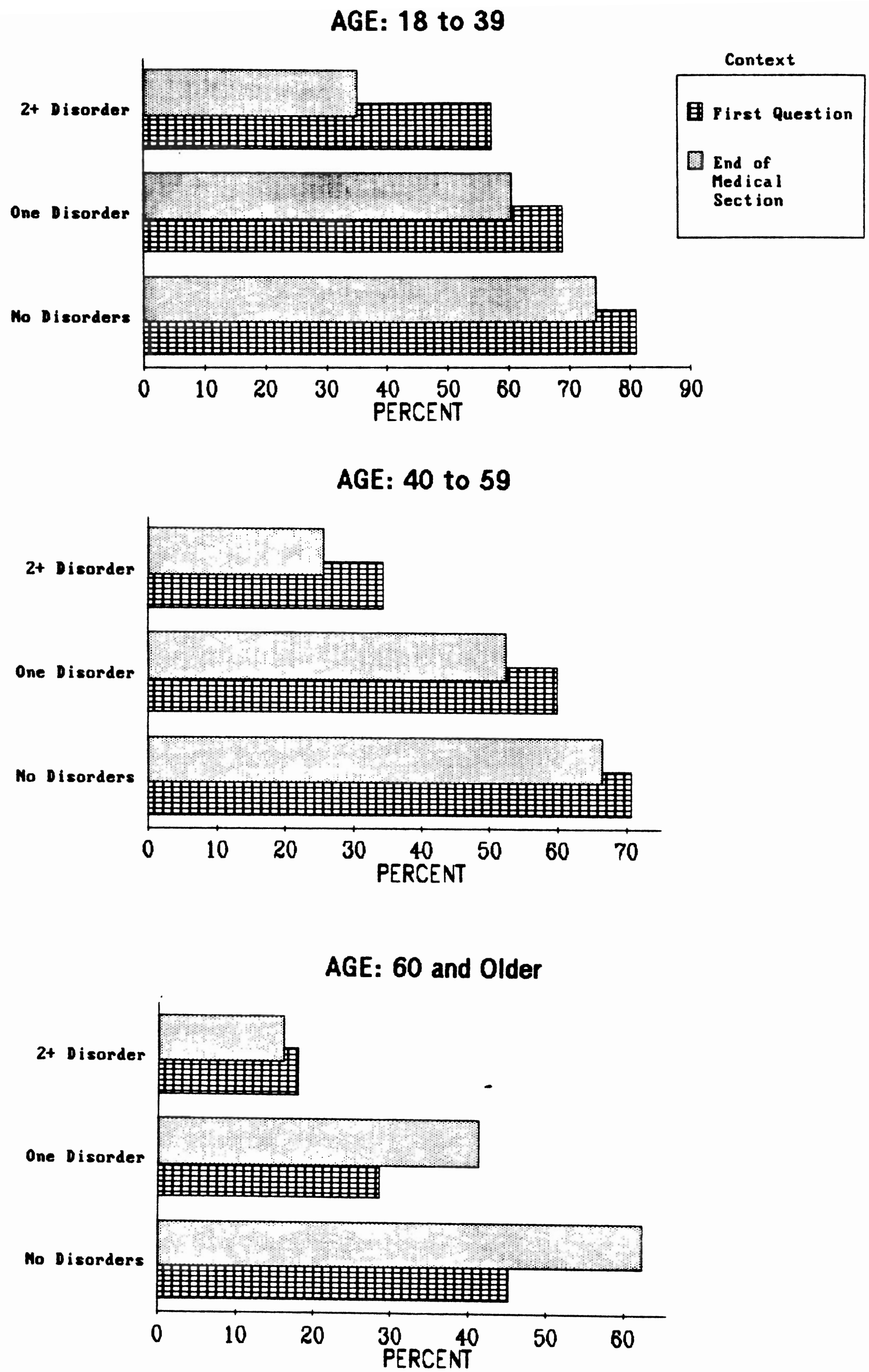
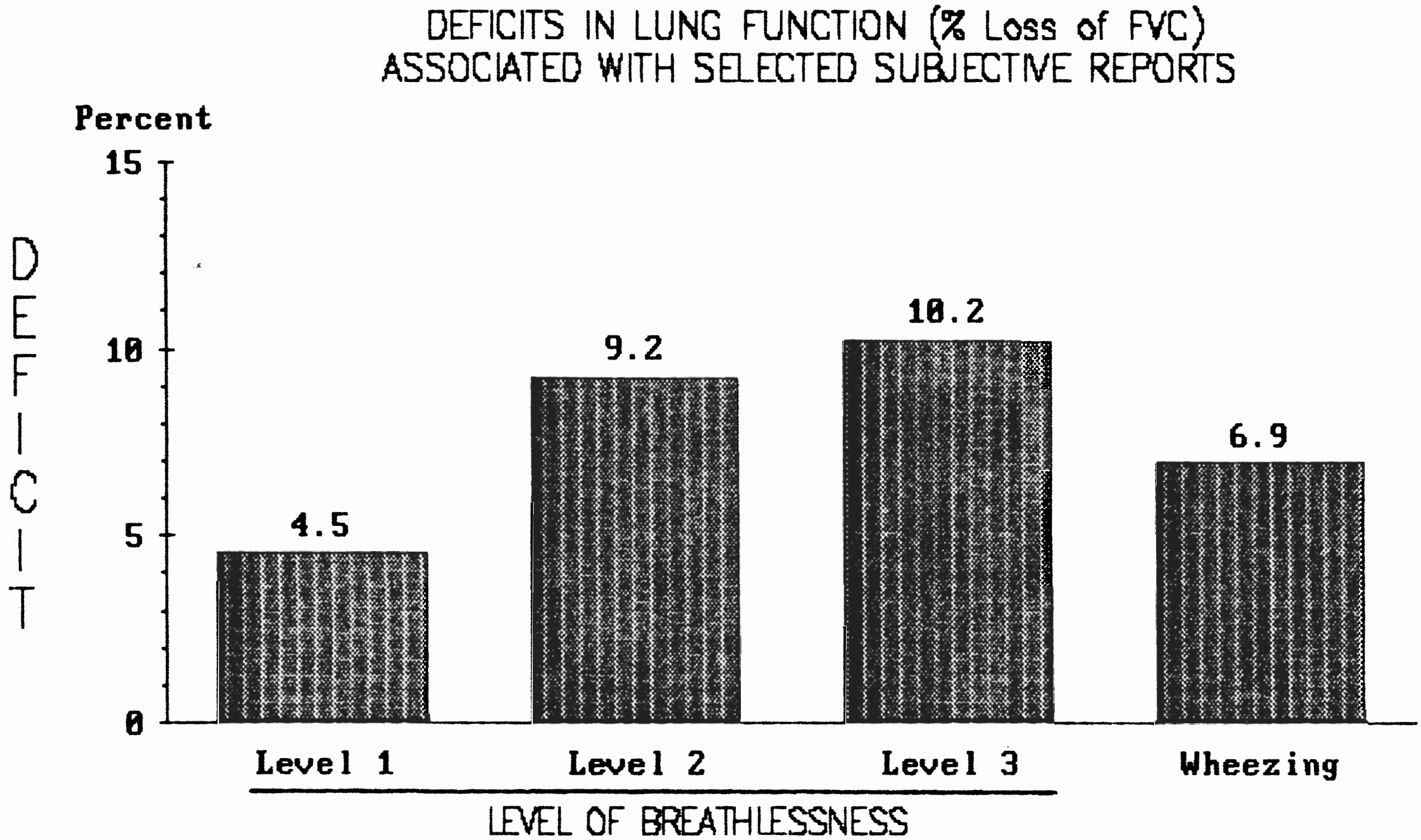


FIGURE 1: PERCENT DESCRIBING THEIR GENERAL HEALTH AS EXCELLENT OR VERY GOOD BY AGE, NUMBER OF REPORTED DISORDERS, AND QUESTIONNAIRE CONTEXT.

<u>Question</u>: Would you say that your health in general is excellent, very good, fair or poor?

APPENDIX A

FIGURE 2



SOURCE: Calculated from National Health and Nutrition Examination Survey (NHANES-1). N=5,551; see Tables 3 and 4.