GETTING BETTER?
CHANGES OR ERRORS IN THE MEASUREMENT OF FUNCTIONAL LIMITATIONS

Nancy A. Mathiowetz and Tamra J. Lair
Nancy Mathiowetz, Bureau of the Census, Washington, D.C. 20233

Key Words: Response error, Interviewer characteristics

Introduction

Activities of Daily Living (ADL) have come to refer to a set of tasks generally necessary to function independently and to attend to personal care needs such as bathing, dressing and eating. An individual's ability to perform these activities is often used to measure his or her level of functional health. However, the measurement of individuals' ability to perform these activities, either alone or with the assistance of a person or equipment, is complex in definition, application and estimation. In addition, measures have shown substantial variability over time, most surprisingly in the direction of "improvement" or change in a positive direction (Branch and Ku, 1989; Katz, et al., 1983; Manton, 1988).

The working hypothesis of this paper is that the amount of improvement in functional status observed in surveys of the elderly, is theoretically unexpected. It is further hypothesized that change, specifically in the form of improvement, is largely a result of measurement error rather than health and demographic characteristics of the individual. Using longitudinal data from the 1987 National Medical Expenditure Survey (NMES), this paper describes and then attempts to explain patterns of change in ADLs over time in an elderly population.

Researchers have outlined the measurement properties which theoretically underlie most ADL measures. Two notable ones among these are the hierarchical nature of the activities and the generally unidirectional process of decline. Several investigators (e.g., Spector, 1986), have suggested that the ADL items form a hierarchy of limitation following the progressive trend of physical decline with advancing years. Difficulty bathing reflects the least severe level of disability followed, in order, by dressing, toileting, transferring from a bed or chair, and feeding. In addition, ADLs are generally viewed as linked directly to the aging process. Change in ADLs over time, when it occurs, is expected to be unidirectional, progressive, and deleterious. However, research on aggregated ADLs have also showed varying proportions of "improvement" over time (e.g., Manton, 1988).

There are several methodological factors that influence the measurement of ADLs at any one point in time, (Corder and Manton, 1991). One overarching dilemma in measuring ADLs in survey research is between the information being sought and the fact that, as a result, respondents' may be less able to provide accurate data -- the most functionally limited may have difficulty as a respondent. Proxy respondents typically report more activity limitations and more severe degrees of limitation than self-respondents. Self-respondents report fewer functional limitations because of the sensitivity of the questions and the social comparisons elderly people often make with respect to the aging process and their own health (Lair, 1987). In addition, the interpretation of the ADL questions often makes the meaning of the findings nebulous. Personal help with tasks can be defined in a variety of ways by the respondent. Adaptation to the difficulty through the use of assistive devices or alternative living arrangements may persuade the respondent into thinking that no problem exists.

Design of the National Medical Expenditure Survey

The 1987 National Medical Expenditure Survey (NMES) household component was a one-year multi-wave study, which collected information on health care status, utilization of services and associated expenditures for the noninstitutionalized household population. The findings presented here are limited to the data collected for those individuals 65 years of age and older. Respondents to the household

survey were interviewed five times between February 1987 and July 1988. Details concerning the design of the study can be found in Cohen, et al. (1987).

The Activity of Daily Living (ADL) measures were asked as part of a Long Term Care Supplement administered at the end of the Round 1 and Round 4 interviews (both in-person interviews). The tasks of interest included bathing, dressing, toileting, transfer from beds and/or chairs and feeding. Question wording for the specific items can be found in Edwards and Berlin (1989).

As with other panel designs, the quality of the NMES household data may be affected by the design of the study. Some of these factors, such as changes in the respondent, changes in the interviewer, questionnaire changes and changes in the mode of data collection can be measured and factored into the examination of gross change. Other changes between rounds are far more subtle and for the most part, were not measured; for example, simple response variability, panel conditioning, or changes in the interpretation of the meaning of the question.

The analysis presented is limited to those individuals for whom data were collected in both Round 1 and Round 4. Individuals who died or were institutionalized (N=226) between Round 1 and Round 4 were excluded as were individuals who responded for only part of the year due to refusals or other forms of nonresponse.

Change in Functional Status

Table 1 looks at the stability in the number of functional limitations reported by the noninstitutionalized elderly population across time for two surveys, NMES and the National Long Term Care Survey (NLTCS). In 1982, the NLTCS screened approximately 36,000 elderly individuals selected from the Medicare Master beneficiary file and interviewed more than 6,000 individuals who reported at least one ADL or IADL impairment. Reinterviews with all survivors were attempted in 1984.

Manton's (1988) findings are presented in the top panel of Table 1. As discussed by Manton, we see that a significant number of individuals change in status over the two year period. Among elderly remaining in the community, 25 percent of those with one or two functional limitations in 1982 report no ADL limitation in 1984, 35 percent of those with difficulties in 3 or 4 ADL areas improve during the same time period, and over 40 percent of the most severely limited showed improvement. Deleterious changes were also evident, with almost 27 percent of those with one or two difficulties declining in health status as compared to almost 30 percent of those with 3 to 4 reported difficulties.

The second panel of Table 2 presents similar findings based on the NMES data. Although the overall probabilities of improvement are different in the NMES data, none of the transitional probabilities represent magnitudes of change different from those presented in the Manton findings. The intent of lining up the two sets of transitional probabilities is not to focus on the differences between the two tables. Rather, it is important to note that the NMES data, which will be used to explore the question of factors which contribute to changes in functional status over time, *do not suggest transitional probabilities radically different from those already evident in the literature.*

Predicted Probabilities of Improvement and Decline

Of particular interest in our research is attempting to understand the substantive and methodological factors which contribute to atheoretical improvement over time and how these factors contrast to those which predict decline over time. To address these questions, multivariate logistic models were fitted to the NMES data. The two binary dependent variables for the models indicate (1) whether an individual "improved" in functional status between Round 1 and Round 4, that is reported fewer ADL difficulties in the later interview; and (2) whether an individual "declined" in functional status between Round 1 and Round 4, that is, reported more ADL difficulties in the later round. The universe for the first model includes only those individuals who could improve - people who reported at least one ADL difficulty at the time of the first interview (N=618), whereas the second model excludes only those who, according to the ADL scale used in NMES, could not decline, specifically, those reporting difficulty with five

ADLs at the time of the first interview (N=5,441).

The probability that a person improves or declines according to the two models is given by:

$$P = \frac{e^{B_0 + \Sigma B_i X_i}}{1 + e^{B_0 + \Sigma B_i X_i}}$$

All tests of statistical significance for these models were based on the RTILOGIT software which adjusts the estimated standard errors of the logistic regression coefficients for the complex NMES sample design.

Looking first at characteristics of the sample person related to improvement, we see that number of ADLs reported at Round 1 is positively related to improvement. Two possible hypotheses are offered as interpretations of this finding. First, the more ADL limitations an individual reported at the first interview, the more potential there is for improvement. From a methodological perspective, we would hypothesize that the improvement noted in Round 4 was a function of panel conditioning; respondents learned in Round 1 that for every reported difficulty there was a series of follow-up questions and therefore altered their behavior in Round 4.

Among demographic characteristics, only age and marital status are related to functional improvement. Older individuals are less likely to improve than younger elderly as are divorced individuals. None of the health service utilization measures (number of hospitalization, prescribed medicines, or medical provider visits) was indicative of improvement.

Methodological factors indicative of functional improvement over time were also included in the models. Self response in both rounds was positively related to improvement in functional status. This finding may provide further support for the conditioning affect noted above.

Interviewer characteristics included in the models describe the Round 4 interviewer. Differences in the characteristics of the Round 1 and Round 4 interviewers are captured in a global measure, "different interviewer". Respondents interviewed by interviewers who had completed four years of college were less likely to report improvement than was evident among other interviewer's respondents. The effect of changing interviewers between rounds had a significant, positive effect on overall functional improvement.

In contrast to the model predicting improvement between the two interviews, several characteristics of the individual and his or her health utilization are significant in predicting functional decline over time. Age is positively related to decline. Decline in functional status does not appear to be a function of the number of ADL difficulties reported in the first interview. Marital status, specifically being widowed or divorced, is positively related to decline in functional status. Unemployed individuals were more likely than the employed elderly to report an increase in the number of ADL difficulties.

All three measures of health care utilization are significantly related to an increase in the number of ADL difficulties over the year. The probability of decline in functional health increased with the number of hospitalizations, the number of prescribed medicines and the number of medical provider visits during 1987. Not unexpectedly, those reporting being in fair or poor health status had a higher probability of functional decline than those reporting excellent or good health status.

Those who lost functional capabilities over the course of the year were also less likely to have reported for themselves in the two interviews. In contrast to the model predicting improvement, neither education of the interviewer nor change in interviewer were significant in predicting decline in ADL status.

Estimated Probabilities of Improvement and Decline

The logistic regression coefficients for the independent variables discussed above represent the rate of change as a function of the dependent variable per unit of change in the independent variable. Unlike linear regression, in a logistic regression model, the coefficients represent the change in the **logit** of the dependent variable, given a change in the independent variable. Interpretation of the coefficients in a logistic regression model requires a transformation, so that one is looking at the change in the dependent variable, not the logit of the dependent

variable. The net effect on the dependent variable of a change in the independent variable for a logistic regression coefficient depends upon whether the change in the independent variable is from, for example, 65 to 66 years of age or 75 to 76 years of age, and, in a multivariate logistic regression model, the effects of any one factor can only be specified by making explicit assumptions about the value of all of the other factors which are included in the model.

Using the transformation equation noted above, estimated probabilities of improvement and decline in functional status are presented in Tables 2 and 3, respectively. Note, once again, that interpreting the marginal effect of each of the independent factors is conditional upon all of the other factors in the model. Therefore, the magnitude of the changes noted in deviations from the described base case would be different for a different base case, but the **relative** effects, that is, whether one factor has greater or less influence on the predicted probabilities, would be the same.

Table 2 examines the predicted probability of functional improvement. The base cased as described in Table 2 had a 75 percent probability of improving between the Round 1 and Round 4 interviews. Changes in either the characteristics of the interview circumstances or the interviewer as well as substantive changes, result in shifts in the predicted probability of improvement. For example, had the interviews been completed by a proxy respondent in both Round 1 and Round 4, the probability of "improvement" would have declined to 64 percent. The "base case" female's probability of improvement drops to 33 percent when the interviewer has a college education and less than one year interviewing experience, but increases to 82 percent if the Round 1 interviewer is different than the Round 4 interviewer. Major changes in health care utilization or chronic conditions, have a much less dramatic affect on predicted probabilities. Even the multiple effects of reporting having ever had a stroke, having been hospitalized twice between the first and fourth interview, and having reported her health status as "poor", does not result in a near zero probability of improvement. Table 3 presents the predicted probabilities for declining functional status between the two interviews. The base case has been changed from that presented in Table 2 to more closely depict someone likely to decline over time.

As indicated in Table 3, the predicted probabilities indicate that the base case person has approximately a 47 percent probability of declining between the two interviews. If all data were collected by proxy, that probability rises to 55 percent, and if the interviewer has a college education and less than one year of experience, the probability increases to 49 percent. An increase from one to two hospitalizations for this person increases the probability of decline by almost 10 percentage points, as does reporting ever having had a stroke. In the opposite direction, the same individual having no hospitalizations in 1987 and reporting to be in good health has only an 11 percent predicted probability of declining between the time of the two interviews. In contrast to the probabilities associated with improvement, we see a smaller effect for interview and interviewing characteristics and a larger impact of substantive factors on the probability of decline.

Conclusion

This paper considered changes in ADL status over a one year period and speculated that the amount of change observed, particularly in the form of improvement, was theoretically unexpected and more likely due to measurement error than the observation of "true" change in this population. While "no functional limitations" was the prevailing condition among the elderly, given difficulty with ADLs, stability in the number of difficulties was not the norm. The research indicates that this pattern is not an anomaly in survey research nor is it simply a function of summative scales or the level of difficulty considered. In predictive models of ADL change over time, the probabilities suggest that improvement in ADL status is overestimated. Although this cannot be definitely supported from these findings, it is highly unlikely, given expectations regarding declining health with age, that we would expect someone with two hospitalizations, in poor health, with both arthritis and the experience of a stroke, and 70 years of age, to have almost a 50 percent probability of improving over a one year period. Clearly, this implies that there may be

"noise" in the cross-sectional estimates of ADL limitations. This paper suggests that estimates of ADL difficulties and changes in ADL status over time may have a significant error component. [References and complete text and tables available upon request].

Table 1. Comparisons of Transitional Probabilities for ADL Difficulty Between the 1984 NLTCS and the 1987 NMES for Persons 65 and Older[1]

National Long Term Care Survey:

1982	1984				
	Not Disabled	1 to 2 ADLs	3 to 4 ADLs	5 to 6 ADLs	Total
1 to 2 ADLs	*25.5%*[2]	48.0%	17.6%	9.0%	100.0%
3 to 4 ADLS	*9.0*	*26.7*	34.4	29.8	100.0
5 to 6 ADLs	*10.3*	*14.7*	*16.8*	58.1	100.0

Source: Recalculated from Manton (1988)

National Medical Expenditure Survey:

Round 1	Round 4				
	Not Disabled	1 to 2 ADLs	3 to 4 ADLs	5 ADLs	Total
1 to 2 ADLs	*41.6%*	39.9%	16.1%	2.4%	100.0%
3 to 4 ADLs	*17.3*	*22.0*	54.7	6.0	100.0
5 ADLs	*2.2*	*7.8*	*43.8*	46.5	100.0

Source: National Medical Expenditure Survey - Household Survey (1987)

[1] Estimates are limited to respondents for whom data were collected in both Rounds 1 and 4. Individuals who died or were institutionalized between the two interviews (for each survey) are deleted from the table.
[2] Numbers in *italic* show percent "improving" between occasions.

Table 2a. Predicted Probabilities of Functional Improvement[a]

Base case: Married, white female (not hispanic) with 2 ADLs at time of Round 1, age 70, income between 200%-399% of poverty line, less than high school graduate, not employed, no hospitalizations, 15 prescriptions during 1987, 10 doctor visits, in good health with arthritis, self response in both rounds, completed SAQ, R4 interviewer was white, age 40-49, high school graduate with 5-9 years experience, from organization "X" and was the same as the R1 interviewer.	Probability of Improving 75.3%
Deviations from Base Case	
(1) All Proxy Responses	63.8%
(2) Interviewer with College Education and less than 1 year interviewing experience	33.3%
(3) Different Interviewer	82.3%
(4) Hospitalized twice in 1987	75.2%
(5) Ever had Stroke	63.3%
(6) Poor Health, Stroke, and 2 Hospitalizations	43.5%

Table 2b. Predicted Probabilities of Functional Decline[a]

Base Case: Widowed, White (non-hispanic) female, with 2 ADLs at time of R1, age 80 years old, income between 200% and 399% poverty line, less than HSG, not employed, 1 hospitalization, 20 prescriptions, 10 doctor visits, in poor health, with arthritis all self responses, completed the SAQ, and R4 interviewer was white, age 40-49, HSG, with 5-9 years experience, from organization "X", and was the same as the R1 interviewer.	Probability of Decline 47.2%
Deviations from Base Case	
(1) All proxy response	55.2%
(2) Interviewer with College Education and less than 1 year interviewing experience	49.1%
(3) Different Interviewer	48.2%
(4) Two hospitalizations	57.2%
(5) Ever had Stroke	57.8%
(5) No Hospitalizations, in good health	10.9%

[a] Estimated as follows using logit coefficients:

$$P_i = \frac{e^{B_0 + \Sigma B_i X_i}}{1 + e^{B_0 + \Sigma B_i X_i}}$$