Iris M. Shimizu, National Center for Health Statistics

1. INTRODUCTION

During the fall of 1975, a longitudinal study of nursing home residents was included in a pilot test for the National Nursing Home Survey (NNHS) in order to measure changes that occur in the health status and the activities of residents. Since changes are bound to occur if enough time is allowed to elaspe, the primary object of the test was to determine whether changes occur rapidly enough to be detected within an 8 week period of time. Thus, data for the test was collected at 8 week intervals by repeating the same questions verbatim about a sample of residents. Since it is possible that changes implied by the data collected could be due to error, a reconciliation study was conducted during the second survey. A consistency study was also done on the data collected.

This paper deals with lessons learned from the pilot test about the conduct of longitudinal surveys. While some of the observations in this study have possible implications on the quality of data from the NNHS itself, those implications are ignored, here, due to space limitations for the present paper.

The basic survey design for the pilot test is described in sections 2 and 3. The methodology for the reconcilliation study and results of the pilot test are discussed in section 4.

2. BASIC DESIGN OF THE PILOT STUDY

2.1 Sample of Facilities

Since the project was a pilot study, the sample of facilities was restricted for convenience to $\frac{1}{4}$ cities, one in each of the Census Regions to allow geographical differences, if any, to surface during the study.

The sampling frame used for the first stage consisted of facilities listed in the 1973 Master Facility Inventory (MFI) of nursing homes. To reduce respondent burden, homes known to be in other surveys just prior to our pilot study were eliminated. These were homes in the 1975 pretest of the MFI, the 1975 Pretest Study of Institutionalized Persons done by the Social Security Administration, and the pilot study for the Survey on Head and Spinal Cord Injuries sponsored by the National Institute of Neurological Diseases and Stroke. Also deleted from the frame were homes with 300 or more beds since these are included in the NNHS with certainty or near certainty.

In order to have a variety of facilities represented in the test, 24 strata were defined and at least one facility was selected from each non-empty stratum. The variables used for defining the strata were:

> a. Certification status as listed in 1975 by the Social Security Administration:
> (1) Certified for Medicare, with or

without Medicaid (2) Certified for Medicaid only, or (3) Not certified.

- b. Bed Size: (1) Less than 25 beds,
 (2) 25 49 beds, (3) 50 99 beds
 and (4) 100 299 beds.
- c. Ownership:(1) proprietary or (2) non-proprietary.

Some of the strata described were empty. Hence, more than one home was selected from some strata with the restrictions that 8 homes had to come from each certification class and 6 homes had to be located in each of the 4 cities. (Health status of the resident was deemed to be more closely related to the certification of the home than any of the other stratifying variables.)

Of the 2^{k} homes selected for the longitudinal study, 20 participated in both the initial survey and the resurvey. Of these 20, nine were chosen at random, with at least two per city for a reconciliation study.

2.2 Procedures Used Within Facilities

The homes selected for the study were visited by survey interviewers twice with about 8 weeks between visits. (Eight weeks appeared to be the maximum length of time feasible for a longitudinal study in the 2 to 3 months of data collection planned for the full NNHS.) On the first visit to each home, the interviewer selected a sample of residents by using a systematic random sampling procedure. This yielded a total of 197 sample residents for the study.

The resurvey in each facility was done by an interviewer other than the one who conducted the initial survey in the home. This was done to prevent the possibility that an interviewer might be biased due to memory of answers given on the prior visit to the facility.

Data were collected on sample residents on both visits. This meant that in addition to the usual practice of keeping the identity of sampled residents confidential, the residents sampled in the initial visit had to be identified during the revisit. Where permitted, residents' names were used as the link between the initial survey and the resurvey. If the administrator of a home objected to the use of residents' names, a code was used which permitted facility personnel to uniquely identify sampled residents during the resurvey but which prevented meaningful identification by any one not connected with either the home or the survey.

During the first survey, the staff person present who was most responsible in the facility for a sample resident was asked a series of questions about the health and activities of that resident. When possible the same staff member was interviewed concerning the sampled resident during both visits to the home. Otherwise during the second survey that staff member present who was most responsible for the particular resident was interviewed. For both interviews, the respondent was asked to consult the resident's records for answers to the questions.

Among the questions asked about the sampled residents, about 30 questions were identical on the two surveys. The concept assumed, here, was that a change in response to any of these particular questions about a resident would indicate a change in the resident's status.

3. QUALITY IN DATA PROCESSING AND COLLECTION

A major concern in the resurvey study was the presence of errors in the data which might result in changes being indicated for an individual resident when indeed no change occurred. The errors could be due to the respondents, the interviewers, or the data processing. The errors due to respondents are dealt with only in the next section. This section deals with the quality of the data as it is affected by data processing and collection. Since, the study was only a pilot test, all the quality control procedures usually established for a full fledged NNHS were not instituted for the study and, hence, the data quality is not expected to be the same as that of a full survey. However, efforts were made to minimize the possibility of errors which could affect the number of changes in residents that would be detected in the pilot.

For the pilot test, keypunching was verified 100% and then the data was subjected to simple computer edits for such things as illegal codes and improper skip patterns. All errors detected in the computer edit were corrected manually after a review of both the error and the original questionnaire. The editing of the record was repeated until all edits were passed. Hence, it is expected the data processing has little effect on the counts of changes that resulted from the test.

Another factor in the data quality is the interviewer. It is conceivable that pilot test data can have proportionally more interviewer errors over all data collection than a full survey since interviewers are not as familiar with the survey or the data collection forms as they would become over a full survey. The interviewers did receive as much training as is usually given in the full survey. That is, they were asked to read the interviewer's manual during the two weeks prior to their participation in a training session of several days length.

Interviewers could possibly have transmitted their biases to the data recorded or to the cues which they gave the respondent. This type of error is not easy to detect and, indeed, no effort was made to measure it in the pilot. However, since the respondents for residents were for the most part nurses, it is assumed that the respondents had the ability to choose the correct answers from the options offered for each question with a minimum of influential cues from the interviewers. Furthermore, the respondent had to be in agreement with the answers recorded except for accidental recording errors. It is assumed for purposes of analysis that any accidential recording errors are random and not due to consistent bias on the part of the interviewer.

Interviewer error did affect the count of residents for which usable data could be tabulated for each question included in the pilot test questionnaire. When the interviewer failed to mark any answer for a question and when the interviewer marked an illegal number of options for a question during either the first survey or the resurvey, a zero was coded for the question and data for that question about the resident was then omitted from all tabulations made for the pilot study. Thus the resident was not counted among those for which data was available for the question. This error probably occurred most often because several answer options applied to the resident and the respondent, having difficulty choosing only one answer for the question, changed the answer originally given and then the interviewer forgot to draw a line through the original answer. Subsequently it was impossible to determine which answer was intended for the question since only one was allowed. The result is that the counts of residents included in tabulations ranged from 197 (the total sampled) on one question to 134 on another. For the majority of the questions, data were available for at least 170 residents. Analysis on any question must, hence, be restricted to those residents for which data could be tabulated.

4. QUALITY OF RECORDED DATA

Since it was desired to know whether changes indicated by the recorded data were real rather than due to error, two studies were conducted. A consistency check was done of the responses given on the two surveys for individual sample residents and a reconciliation study was conducted in a subsample of the facilities.

4.1. Results of Consistency Check

Among the questions that were repeated verbatim on the two surveys there were four questions for which the second answer should have been implicated by the first answer. Analysis of the consistency in responses to each of these questions was restricted to only those residents for which answers were available from both surveys, as mentioned above.

The questions used in the consistency study are:

- 1. What was this resident's primary diagnosis at admission?
- 2. Has this resident lived in this facility one full month or longer?
- 3. What was the primary source of payment when he/she was admitted to the home?

4. Does he/she have any of the following conditions or impairments?

The correct answers to the first and third of the above questions for any one resident must be the same on the two surveys. The correct answer to the second question was "yes" on the resurvey except for those residents who were discharged during the 8 week interval between the surveys and who had been in the home less than a full month at the time of the discharge.

On the fourth question the interviewer was to mark all answer options which applied to the resident. Among conditions listed as options were the following eight which are generally considered incurable once a person acquires them:

- a. Mental retardation
- b. Heart trouble
- c. Arthritis or rheumatism
- d. Parkinson's disease
- e. Chronic respiratory disease
- f. Diabetes
- g. Permanent stiffness or deformity of back or extremities
- h. missing extremities.

A resident having one of these particular conditions during the first survey would still have it during the resurvey. The reverse is not necessarily true since a resident could just begin to show the symptoms for some of the conditions for the first time during the 8 week interval between surveys.

To determine a lower limit on the frequency of errors which occurred for these questions, an error was counted each time the data recorded in the resurvey was inconsistent with that recorded for the same resident during the first survey. That is, errors were counted if changes other than the possible ones described above occurred to the resident according to the data recorded on the two surveys. The errors that did not yield inconsistencies in the data could not be detected and, hence, were not counted here. Changes that resulted because the response on one survey was "Don't know" were also not counted as due to errors. The error counts obtained are given in Table 1 together with the percent of residents for which errors were found.

Since it was thought that the number of residents for which change occurred may be affected by changes in respondents between the two surveys or the resident's discharge from the facility during the 8 week interval, tabulations were also made according to the respondent during the two surveys and according to the discharge status of the resident. These counts are also shown in Table 1.

It can be seen that errors occurred for at least 37% of the sample residents for which data on primary diagnosis at admission was available from both surveys. The error percentages shown for the other items range from 0 to 20 percent. If one defines 5 percent as the maximum amount of error that is permitted before data are labeled as being unreliable, then data for 6 of the 12 question items considered here would be labeled as unreliable on the basis of inconsistencies alone.

The error percentages for the sampled residents are two or more percentage points higher on 10 out of 12 items for those discharged than for those not discharged. Likewise, it is noted that errors occurred relatively more often on 9 of 11 items when the respondent was different between the two surveys. On the first 3 questions above, it is possible that the differences between the groups of residents could be affected by the lack of data for some of the 197 sample residents. However, in view of the relationships between error percentages shown for the condition items, where data are available for all the sampled residents, it appears that differences would probably still exist, and the tendency toward higher error percentages would likely continue for the residents who were discharged between surveys and the residents with different respondents.

These observations from the pilot study suggest that answers to questions about residents can vary with the date of interview. Part of the variation can result from a difference in the respondent that would be interviewed on different dates. But inconsistencies such as those found in the pilot test for the primary diagnosis question studied here suggest that variation in answers are possible even though the respondent were to remain the same.

During field observations made in the pilot test it was noted in regard to the primary diagnosis question, that several diagnoses could be recorded in a resident's file with no indication about which is the primary one. Hence, the respondent used judgement to pick the diagnosis most likely to be primary for the resident. On the basis of the pilot test data, it is evident that one's judgement of what should be primary can vary with time even though the correct answer remains the same.

4.2 Reconciliation Study

As indicated above, error could occur in the data reported during the two surveys for a resident without any indication of impossible changes such as those described above. A reconciliation study was conducted in a subsample of 9 facilities which contained 84 of the sampled residents. In these homes a copy of the questionnaire completed for each resident during the first survey was given to the interviewer for the second survey. After the second interview for each resident was completed, the two forms were reviewed and the respondent was asked to explain any differences in answers. Since there was a concern that the questionnaire design or some other item in the survey could be responsible for erroneous changes in responses, the respondent was also asked to identify the source of the error when an error was reported. The errors are tabulated in Table 2 by sources of errors.

In the reconciliation homes, a total of 1077 changes were noted over all the approximately 30 questions asked. Of these 527 or 48.9 percent were reportedly due to errors. Seventy of the erroneous changes were blamed on the questionnaire. A review of these cases revealed that at most 7 erroneous changes were blamed on any one questionnaire item. This implies that the questionnaire itself did not appear to be much of a problem to the respondents.

Forty percent of the erroneous changes were blamed on the "unknown". These changes probably include many of those that resulted because the respondent could not find the requested data in the resident's file or if the data were found, the information was not adequate. That is the respondent(s) had to rely on memory or judge which of the data that were in the file best described the resident's status. An example of such a question is the diagnosis question discussed earlier. Certainly, when "concrete" data is not available for the record, it is possible for a change of respondents or other circumstances present at the time of the survey to influence the responses given and, thus, any changes in responses that would be detected between two points in time.

About half (260) of the erroneous changes identified were blamed on the respondents and/or the interviewers.) This suggests that in a longitudinal study where questions are reasked verbatim about nursing home residents, if it were possible to design the survey and questionnaire in such a way that the only possible sources of error would be the participants in the interview, then it could be that as much as a third $(260 \div [1077 - 267] = .32)$ of the changes detected would be due to error.

5. SUMMARY AND CONCLUSIONS

On the basis of the pilot test for a longitudinal study of residents in nursing homes, it is evident that when questions are simply repeated verbatim at two points in time, the result is that the changes detected in individual sample units are as likely to be due to error as not. The experience indicates that the attention of the respondent for the second survey should be focused more specifically on change. One could simply ask whether a change had occurred since the last interview. In the NNHS such a question would require that the respondent know the resident's status at the time of the prior interview. Based on observations in the pilot test, the respondent may not always have that information.

It appears at least for the NNHS that a better procedure for measuring change would be to first tell the respondent what was recorded in response to the question during the first survey and then to ask what answer would apply at the time of the resurvey. A bias in responses may be introduced by informing the respondent about the past answer in that it may encourage some resurvey respondents to repeat the response that was recorded during the first survey. This would especially be true in the event that the respondent, who for some reason, is not absolutely sure of the correct answer.

The proposed procedure would at least force the respondent to think about whether a change has occurred since the earlier survey. That is, if no change has occurred, then the answer in the resurvey would have to be the same and if a change has occurred then the resurvey answer must be consistent with the answer given during the first survey.

Admittedly, the response recorded during the first survey may not be correct. In this case, results from the reconciliation study indicate that resurvey respondents may identify errors in the data recorded during the first survey so that changes will not be erroneously implied by the resurvey answer which they supply.

In any event it is expected that the proposed procedure would yield fewer erroneous changes in the resulting data. TABLE 1. Error Counts for Each Question According to Whether Resident was Discharged and Whether Respondent for the Resident was the Same

(Numerators of ratios are error counts and denominators are number of residents for which useable data was available from both surveys for the question.)

Abreviated Question		uted	Total	Resident Was		Respondent Was		
que	.5010			charged	Discusi geu	Dane	Different	
1.	Pri at	mary Diagnosis Admission	56/152= 37%	50/133= 38%	6/19= 32%	34/103= 33%	22/49= 45%	
2.	In	Home One Month	9/180= 5%	5/163= 3%	4/17= 24%	6/133= 5%	3/47= 6%	
3.	Pri at	mary Payment Source Admission	25/127= 20%	19/115= 17%	6/12= 50%	*		
4.	Impairments or Condition							
	a.	Mental Retardation	6/197= 3%	3/172= 2%	3/25= 12%	6/142= 4%	0/55= 0%	
	ъ.	Heart Trouble	13/197= 7%	10/172= 6%	3/25= 12%	6/142= 4%	7/55= 13%	
	c.	Arthritis	26/197= 13%	22/172= 13%	4/25= 16%	15/142= 11%	11/55= 20%	
	đ.	Parkinson's Disease	4/197= 2%	3/173= 2%	1/25= 4%	2/142= 1%	2/55= 4%	
	e.	Chronic Respiratory Disease	5/197= 3%	4/172= 2%	1/25= 4%	2/142= 1%	3/55= 5%	
	f.	Diabetes	5/197= 3%	3/172= 5%	2/25= 8%	3/142= 3%	2/55= 4%	
	g.	Permanent Stiffness	19/197= 10%	17/172= 10%	2/25= 8%	11/142= 8%	8/55= 15%	
	h.	Missing Extremities	0/197= 0%	-	-	-		

*The financial questions about many residents were answered by someone other than the respondent for the remainder of the questionnaire but no records were made on whether the respondent to financial questions was the same or different between the two surveys.)

TABLE 2. Counts of Changes Due to Errors Identified in Reconciliation Study by Source of Error and Survey in Which Error Occurred

Source of	Total	Survey in which Error Occurred				
Error		Initial Survey	Resurvey	Both		
Total Changes Due to Error	527	191	69	267		
Respondent only	145	92	52	l		
Interviewer only	111	94	17	0		
Both Respondent and Interviewer	4	4	0	· 0		
Questionnaire	70	l	0	69		
Unknown	197	0	0	197		