Journal of Official Statistics, Vol. 26, No. 2, 2010, pp. 317–339

Issues in Survey Measurement of Chronic Disability: An Example from the National Long Term Care Survey

Elena A. Erosheva[1] *and Toby A. White*[2]

The National Long Term Care Survey, a longitudinal study that consisted of the screen and detailed interviews, provides a unique opportunity to illustrate issues in survey measurement of chronic disability. The original intent of the survey was to use a shorter measure of disability status in the screen interview to maximize the yield of disabled cases and then examine those cases with a longer measure in the detailed interview, identifying possible "false positive" disability cases from the screen. In this article, we show empirically that the intended relationship between the screen and detailed NLTCS disability measures does not hold uniformly – the detailed measure provides significantly higher ADL disability estimates than the screen measure, contrary to expectation. We examine whether discrepancies between the two disability measures are associated with certain features of the survey design and respondent-level characteristics, discuss implications of our findings for prevalence estimation, and provide recommendations for disability survey design.

Key words: Activities of daily living; instrumental activities of daily living; longitudinal survey; screen interview; survey design; two-phase data collection.

1. Introduction

Core disability questions in national surveys often focus on basic and instrumental activities of daily living (ADL and IADL) (Katz et al. 1963; Lawton and Brody 1969). Because ADL and IADL tasks reflect the ability to perform personal care and self-maintenance activities, they are widely applicable for describing the health status and service needs of the elderly individuals. However, when it comes to survey disability measurement, methodology researchers have questioned the validity and reliability of ADL and IADL survey items (Rodgers and Miller 1997; Mathiowetz 2001; Gill and Gahbauer 2005; Wolf et al. 2005; Freedman et al. 2002), including the issue of differences in asking "difficulty" type or "gets help" type of questions (Weil et al. 2001; Freedman 2000). This article provides an examination of disability measures administered with the two linked interviews in the National Long Term Care Survey (NLTCS).

[1] Department of Statistics and School of Social Work, University of Washington, Box 354322, Seattle, WA, U.S.A. 98195-4322. Email: elena@stat.washington.edu
[2] Department of Actuarial Science and Finance, Drake University, Des Moines, IA, U.S.A. 50311-4505. Email: toby.white@drake.edu
Acknowledgments: Work on this project was supported in part by grants No. 1 R03 AG030605-01 and No. 1 ROl AG023141-01 from the National Institutes of Health. The authors thank Stephen E. Fienberg for helpful discussions, Nancy A. Mathiowetz for her comments on an earlier version of this article, and Eric Stallard for his suggestion to combine self and proxy responses in a cognitive impairment indicator. An earlier version of this work was presented at the NLTCS Research Meeting held at the National Institutes of Health, Bethesda, MD, in May 2009.

The NLTCS is known as one of the best surveys for assessing national trends in disability among Americans 65 years of age and older, 1982–2004. Attractive features of the survey include its longitudinal structure, long-term coverage, inclusion of the institutional populations, identical field procedures, detailed disability questions, and low loss-to-follow-up rates (Freedman et al. 2002). The NLTCS data on basic and instrumental activities of daily living has been used to generate such major findings as the persistent decline in chronic disability among the elderly Americans (Manton et al. 1997; Manton and Gu 2001; Manton et al. 2006).

While the NLTCS represents a rich source of longitudinal data on disability, there is a concern that the survey data has been underused, especially with regard to longitudinal features (Committee on National Statistics and Committee on Population 2005, 2006).

Secondary NLTCS data users need to decide whether to use standard chronic disability measures, derived by the Center for Demographic Studies at Duke University (The center has closed as of September 30, 2006.) and available in so-called analytic files, or to devise their own measures, using responses to questions available directly from the questionnaire. Making the right choice for a given secondary analysis is difficult because published studies provide little detail on the standard NLTCS disability measurement (see Manton et al. 2006, for example). Although the NLTCS contains two analogous yet different sets of ADL and IADL questions, one for the screen interview and one for the detailed interview, secondary researchers appear to consider disability measurements from the screen interview only (Rodgers and Miller 1997; Wolf et al. 2005; Gill and Gahbauer 2005).

The goals of this article are threefold. First, to explain the NLTCS screen and detailed interview disability assessments in a clear fashion. Second, to compare the screen and detailed disability assessments empirically. Third, to analyze whether discrepancies in disability measurement between the screen and detailed questionnaires are associated with time-lags between the two interviews, demographic and interview characteristics, and medical conditions of the respondent.

In Section 2, we describe the NLTCS survey design as it relates to ADL and IADL disability measurements, separately for the screen and the community detailed interviews. In Section 3, we examine compatibility of screen-based and detailed-based NLTCS disability measures, focusing mainly on ADL items. We show empirically that the detailed-based measure provides significantly higher ADL disability estimates than the screen-based measure, contrary to expectation. Using multivariate models in Section 4, we find that certain demographic characteristics, health conditions, and proxy response indicators are associated with higher discrepancies between the screen and detailed ADL disability measures. Finally, in Section 5, we discuss implications of our findings for estimating disability prevalence and provide recommendations for disability survey design.

2. National Long Term Care Survey

2.1. Background: Survey Design

The NLTCS was designed to assess changes in disability among elderly people in the United States through a longitudinal sampling mechanism. The first wave of data

collection was in 1982, followed by a second wave in 1984 and subsequent waves in 1989, 1994, 1999, and 2004. An important feature of the survey is the fact that questions on chronic disability, including those pertaining to basic and instrumental activities of daily living, remained unchanged over all survey waves (Manton et al. 2006).

The initial 1982 NLTCS sample of persons aged 65 or over from Medicare Eligibility files was stratified geographically into 173 primary sampling units which were then grouped into 101 strata. It was estimated that 66 strata needed to be retained in the sample in order to identify approximately 6,000 chronically disabled community residents. After 1982, the NLTCS sample was replenished at each wave in order to reflect the current U.S. population 65 and older. Details of the NLTCS sampling mechanism are described elsewhere (Manton et al. 1993; Manton and Gu 2001; Manton et al. 2006; Manton 2007).

The NLTCS consisted of two linked interviews: a screen interview followed by a detailed interview. With the exception of the healthy and oldest-old NLTCS supplements that were introduced in the later waves of the survey, only those individuals who had been classified as chronically (90 + days) disabled by the screen interview were eligible to participate in the more extensive detailed interview. Persons who received a detailed interview in one survey wave were automatically eligible for a detailed interview in all subsequent survey waves until death. Approximately 80% of the NLTCS screen interviews in each wave were administered by telephone, with the rest done in person, whereas all detailed interviews were administered in person; the same field procedures were used for all waves of the NLTCS (Manton et al. 1993).

Manton et al. (1997) described chronic disability measurement by the NLTCS as follows: "Disability was defined as the inability to perform ≥ 1 IADL (e.g., cooking, doing laundry) due to health or aging, or the inability to perform at least one ADL (e.g., bathing, dressing) without using personal assistance or special equipment . . . To be identified as chronically disabled when initially selected for a detailed interview, a sample person had to have at least one ADL or IADL disability that had lasted, or was expected to last, >90 days." Although factually correct, this description does not allow one to fully understand how chronic disability status was measured in the NLTCS. Consequently, some of the most knowledgeable researchers in the field have often oversimplified disability measurement in the NLTCS by relying only on the 90-day lower bound of disability duration and ignoring the complex interplay between screen and detailed NLTCS interviews (see Gill and Gahbauer 2005; Freedman et al. 2002; Freedman and Soldo 1994, for examples).

In actuality, the NLTCS contains two measures of disability status that appear in the screen and detailed interviews.[3] The original intent of the survey designers was to use a shorter measure of disability status in the screen interview to "cast a wide net," maximizing the yield of disabled cases. A longer measure in the detailed interview was then used to examine disabled cases from the screen, identifying possible "false positives", relative to the detailed interview disability measure, and providing more information on disability (Manton 2007). Up to 2004, due to budget constraints,

[3] There were two versions of the detailed interviews, one for community dwelling and one for institutionalized elderly people. In this article, we focus on interviews with community dwelling elderly only.

individuals automatically eligible for detailed interviews (such as those who participated in a detailed interview at a prior wave) had been administered abbreviated screen interviews that did not include ADL and IADL questions. This practice changed in 2004 when all persons automatically eligible for detailed interviews were re-administered complete screen interviews.

Since the inception of the survey in 1982, the set of disability questions administered in the screen interview was meant to be the "gold standard" for the NLTCS measurement of chronic disability prevalence (Manton 2007). For returning cases that did not have disability status assessed at the screen interview, "disability status was approximately (in an effort to match the screen) determined only from the community detailed interview. If no eligible chronic disability state (i.e., defined by ADL and IADL items selected to be as consistent as possible with those on the screener) lasting more than three months was identified for such persons on the detailed interview, those persons were counted as non-disabled" (Manton 2007). The NLTCS chronic disability prevalence estimates therefore relied on a mixture of disability assessments, using the screen interview for those entering the survey and using the detailed interview for those continuing on from the previous wave(s) (Manton et al. 2006).

Besides data from the screen and detailed interviews, there is a third NLTCS data source that contains ADL and IADL measurements, known to secondary researchers as the *analytic* data file. The Duke Center for Demographic Studies describes the analytic file as containing "the products of specific analyses conducted by the Center for Demographic Studies" and "various correction factors and consistency checking not included in the standard Census data product" (Manton 1997).

2.2. *Disability Assessment by the Screen and Detailed Interviews*

In our investigation of the operational definition of disability by the NLTCS, we used the NLTCS data files, obtained under a Data Use Agreement from the Center for Demographic Studies at Duke University, http://www.nltcs.aas.duke.edu. We relied on original data from the screen and detailed interviews. We found that detailed binary ADL and IADL outcomes matched corresponding records in the analytic file for the majority of subjects in all waves, and matched perfectly for all subjects in the 2004 wave.

Table 1 summarizes two sets of ADL and IADL impairments assessed by the NLTCS screen and detailed interviews and provides respective SAS variable names. While the screen interview asked questions about 9 ADL and 7 IADL tasks, the detailed community interview included 6 ADL and 10 IADL tasks. Note that the screen interview considered "outside mobility" as an ADL, while the detailed interview listed this item as an IADL.

Each interview determined a binary disability outcome (1 = presence or 0 = absence of functional limitation) using sequences of triggering questions for each ADL and IADL task. To thoroughly understand the binary outcome determination process in the NLTCS (i.e., the process by which ADL and IADL outcomes are coded into yes/no categories), we reproduced logical paths of triggering questions using the community detailed and screen interview questionnaires (Appendix A provides web addresses). The Center for Demographic Studies (CDS) at Duke University also provides a version of triggering questions for the detailed community interview (Unicon Research Corporation 2002),

Table 1. ADL and IADL impairments in the screen and detailed NLTCS interviews with corresponding SAS variable names (binary outcomes)

	Screen	SAS Name	Detailed	SAS Name
ADL				
1	Eating	SCN_15_A	Eating	ADL_EAT
2	Getting in/out of bed	SCN_15_B	Getting in/out of bed	ADL_BED
3	Getting around inside	SCN_15_C	Getting around inside	ADL_INS
4	Dressing	SCN_15_D	Dressing	ADL_DRS
5	Bathing	SCN_15_E	Bathing	ADL_BTH
6	Toileting	SCN_15_F	Toileting	ADL_TOI
7	Getting in/out of chair	SCN_15_C		
8	Getting about outside	SCN_15_E		
9	Continence	SCN_15_I		
IADL				
1	Prepare meals	SCN_18_A	Prepare meals	IDL_MLS
2	Laundry	SCN_18_B	Laundry	IDL_LND
3	Light housework	SCN_18_C	Light housework	IDL_LTW
4	Grocery shopping	SCN_18_D	Grocery shopping	IDL_SHP
5	Managing money	SCN_18_E	Managing money	IDL_MON
6	Taking medicine	SCN_18_F	Taking medicine	IDL_10A
7	Telephoning	SCN_18_G	Telephoning	IDL_TEL
8			Heavy housework	IDL_HVW
9			Getting about outside	IDL_OUT
10			Traveling	IDL_WLK

available on-line at http://www.nltcs.aas.duke.edu/pdf/ADL_Measures.pdf and /IADL_Measures.pdf.

Although both ours and the CDS versions produce the same ADL and IADL binary outcomes, they appear to list different triggering questions. The difference is that the triggering questions from CDS include some nonessential questions at the end stage of the decision process while our version provides only the actual pathways on the decision tree.

For example, many CDS triggering questions focus on disability duration. However, note that if duration questions are being asked, the respondent is nagged as being disabled on a particular task, regardless of the response. This is because binary outcomes are actually determined from earlier questions, not listed in CDS triggering tables, that themselves lead to the duration questions being asked in the first place.

The binary outcome determination process in the NLTCS screen interview can be summarized as follows. An individual was considered to have an ADL difficulty if he/she needed human help or special equipment to perform the activity, or if he/she was unable or did not perform the activity at all. An individual was considered to have an IADL difficulty if he/she was unable to perform the activity without help because of a disability or a health problem.

In order to be classified as screened-in, an individual must have had a difficulty with at least one ADL or with at least one IADL that had lasted, or was expected to last, 3 months or longer. Operationally, the screen first determined whether there was a problem

performing each ADL at the present time. Then, if at least one ADL difficulty was found, disability duration was assessed with one set of generic ADL-problem duration questions.[4] The screen interview IADL duration assessment process was similar in that there was only one set of duration questions asked about a generic IADL problem.

The detailed survey employed a more complex scheme of triggering questions to determine binary outcomes on ADL and IADL tasks. Questions referred to difficulties occurring "during the past week" for ADL tasks and "usually" for IADL tasks. The detailed survey did record the length of impairment for each ADL, using response categories of "<3 months," "3–6 months," "6–12 months," "1–5 years," and ">5 years," although that information did not factor into the determination of detailed binary ADL outcomes. The detailed survey did not record the length of impairment for IADL tasks.

To illustrate question design and wording differences between the screen- and detailed-based ADL and IADL questions, we compare binary outcome determination for one ADL task ("eating") and one IADL task ("managing money") in Table 2. First, we observe that the screen interview has ADL triggering questions of a "difficulty" type while the detailed survey has ADL triggering questions of a "gets help" type. Noticing this inconsistency, Wolf et al. (2005) pointed out that this type of mix may prompt "false negative" disability outcomes for the screen (i.e., when functionally limited respondents adapt to their limitation and are able to perform tasks without perceived difficulty) and "false positive" outcomes for the detailed interview (i.e., when respondents receive help for tasks they could perform themselves possibly with some difficulty).

Second, there are differences in the number of triggering questions used by the screen and the detailed survey. In the NLTCS, consistent with the idea of having a brief screen disability assessment and a more in-depth assessment on the detailed interview, the detailed interview included more triggering questions per ADL or IADL task than did the screen interview. It has been shown, however, that such differences in question design may impact disability measurement. For example, in an experimental study, Lee et al. (2007) found that finer decomposition influenced disability reports; disability rates were higher when respondents answered a single question than when they had to answer multiple questions about each activity.

Third, there are differences in the reference periods built into the questions. Thus, the screen questions implicitly refer to the present ("now"), the ADL detailed questions refer to "the past week," while the IADL detailed questions ask whether respondents "usually" do those tasks. In addition, newly sampled individuals, who had at least one problem with either basic or instrumental ADL tasks, were included in the detailed interviews only when the actual or expected duration of the problem was 90 days or longer. In the detailed interview, on the other hand, disability duration information was not incorporated in the binary outcomes for ADL and IADL tasks.

[4] "You said you have a problem with (one activity/some activities). Have you had (this problem/any of these problems) for three months or longer? Do you expect (this problem/any of these problems) will last for the next three months or longer? Altogether, from beginning to end, will (this problem/any of these problems) have lasted three months or longer?"

Table 2. Examples of ADL and IADL binary outcome determination

NLTCS screen interview	NLTCS detailed interview
ADL-Eating	*ADL-Eating*
1. Do you have any problem eating without the help of another person or special equipment?	1. During the past week. . . did anyone help you eat? 2. Did you use special equipment to help you eat? 3. Did someone usually stay by just in case you might need help with eating?
If Question 1 was answered "yes", or "can't do/don't do at all," the individual is ADL-disabled on eating. Otherwise, no disability is recorded.	If any of Questions 1–3 were answered "yes," the individual is ADL-disabled on eating. In addition, if the individual did not eat at all, the individual is ADL-disabled on eating. Otherwise, no disability is recorded.
IADL-Managing Money	*IADL-Managing Money*
1. Are you able to manage money without the help of another person or special equipment? 2. Does a disability or health problem keep you from managing money?	1. Do you usually manage money by yourself? 2. If you had to manage money on your own, could you do it? 3. Is the reason you cannot manage your own money because of disability or a health problem?
If Question 1 was answered "no," and Question 2 was answered "yes," the individual is IADL-disabled on managing money. Otherwise, no disability is recorded.	If Questions 1 and 2 were answered "no," and Question 3 was answered "yes," the individual is IADL-disabled on managing money. Otherwise, no disability is recorded.

Other factors that may impact differences in disability measurement between the screen and the detailed interviews are related to features of survey design not evident from Table 2. For example, time lags between interviews may contribute to differences in disability measurements between these two survey components. Even a few days gap between the interviews may lead to real changes in disability status. Because disability is more often characterized as a dynamic process with multiple recurrent episodes rather than a progressively worsening conditioning (Manton 1988; Gill et al. 2002; Gill and Kurland 2003; Lynch et al. 2003), the direction of the net effect of these time gaps is not clear.

Finally, contextual differences between ADL/IADL questions in the screen and detailed interviews may be another influential factor. It is known that the context of preceding questions is likely to have an effect on answers to questions that are coming next, especially if the latter are somewhat open to interpretation. For example, studying responses to a question on "serious difficulty seeing" in the National Health Interview Survey, Todorov (2000) showed that when preceding questions covered six medical conditions related to vision,[5] respondents were less likely to report "serious difficulty seeing" than others who were asked preceding questions unrelated to vision. A suggested explanation for the observed difference was that respondents in the vision context were more likely to interpret "serious difficulty seeing" as linked to medical conditions. In the NLTCS, disability questions in the detailed interview came after questions on more than two dozen medical conditions, including some with high population prevalence, whereas disability questions in the screen interview came after a section on basic demographic information. Assuming a similar cognitive model of question understanding as in Todorov (2000), we expect that placing disability questions after questions on medical conditions may have increased detailed-based disability measures, especially with respect to more serious ADL disability.

The effects of differences in wording and ordering of ADL and IADL questions between the screen and detailed interviews cannot be disentangled as all individuals followed the same survey protocol. With this caveat in mind, we attempt to examine: (1) the extent of discrepancies in disability measures between the screen and the detailed interviews, (2) whether the discrepancies are associated with time-lag between the two interviews, demographic and interview characteristics of the respondent, and the number of medical conditions reported, and (3) whether the discrepancies can be accounted for by using duration-adjusted disability outcomes. Our main analysis will focus on ADL outcomes because detailed interviews did not record duration of IADL impairments, preventing us from carrying out duration adjustments for IADLs.

To examine consistency between screen-based and detailed-based disability measures empirically, we rely on data from the 2004 wave. Apart from the very first data collection wave in 1982 that was special in many ways, 2004 was the only wave in which all detailed interview participants were also administered the screen interviews, irrespective of their eligibility to be automatically included in the detailed survey.

[5] The conditions were legal blindness, cataracts, glaucoma, color blindness, a medical retina condition, and trouble seeing even when wearing glasses.

We present unweighted analyses that also do not account for clustering[6], working in the framework of model-based, as opposed to design-based, inference in survey sampling (Särndal 1978). Hence, reported estimates and standard errors should be interpreted as those of the model parameters and not of population quantities. Rederiving population estimates and attempting to assess validity of the measures is beyond the scope of this article.

3. Comparison of the Screen and Detailed Disability Measures: Preliminary Analysis

To compare disability assessments in the screen and detailed survey interviews in 2004, we focus on binary outcomes from the 6 ADL and 7 IADL tasks that are present in both survey components (see Table 1). The 6 ADL tasks are: eating, dressing, toileting, getting in and out of bed, inside mobility, and bathing. The 7 IADL tasks are: telephoning, managing money, light housework, cooking, taking medicine, laundry, and grocery shopping.

There were 5,201 individuals who received the community detailed survey in 2004; all of these individuals were administered the screen interview as well. While 46% (2,396 individuals) received the screen and detailed interviews on the same day, the mean (median) time lag between interviews for others was 26 (18) days.[7] We extracted ADL and IADL data from the screen and detailed interviews for all 5,201 individuals. There were no missing data on ADL and IADL variables; all records were either 0 or 1. All 5,201 individuals have completed the screen interview.[8]

To examine overall consistency in core disability outcomes, we calculate mismatches on the 6 ADLs and 7 IADLs that the screen and detailed interviews had in common. On average, respondents had 1.66 mismatches among the 13 binary disability outcomes. Table 3 illustrates observed mismatches by each ADL and IADL variable. We note that observed mismatches were much more frequent among the ADLs than among the IADLs. For ADL variables, mismatches of the type when the detailed interview indicated disability presence while the screen interview did not were more common than vice versa; this relationship is reversed for the IADLs.

To examine the direction of discrepancies further, we compare differences in ADL and IADL counts for all 5,201 individuals. The average number of ADL impairments on the detailed interview (1.62) was significantly higher than the corresponding average on the screen (0.93); the 95% confidence interval for the difference of 0.69 was (0.65,0.73), based on a 2-sample t-test for paired data. However, the average number of IADL impairments on the detailed interview (1.35) was significantly lower than the corresponding average on the screen (1.55); the 95% confidence interval for the difference of -0.20 was $(-0.23, -0.16)$.

Table 4 shows a classification of individuals by their disability records from the screen and detailed interviews. Following Manton and Gu (2001), we use five disability categories: "no ADL/IADL impairment," "IADL impairment," "1–2 ADL impairments," "3–4 ADL impairments," and "5–6 ADL impairments." We consider two disability

[6] To the best of our knowledge, clustering information is not available in the public use NLTCS data files.
[7] There were 111 individuals who had an interview date missing for one of the two survey components.
[8] As indicated by the SC_COMPLETE variable in SAS.

Table 3. NLTCS 2004 mismatch counts between the screen and detailed interviews, by ADL and IADL. N = 5,201

	Count (%)	Scr = N, Det = Y	Scr = Y, Det = N
ADL			
Eating	472 (.091)	362	110
Dressing	524 (.101)	347	177
Toileting	1,105 (.212)	968	137
Getting in/out of bed	953 (.183)	831	122
Getting around inside	893 (.172)	690	203
Bathing	1,337 (.257)	1,227	110
Average ADL	1.016 (.169)	0.851	0.165
IADL			
Telephoning	333 (.064)	73	260
Managing money	418 (.080)	144	274
Light housework	486 (.093)	145	341
Prepare meals	455 (.087)	135	320
Taking medicine	459 (.088)	324	135
Laundry	490 (.094)	160	330
Grocery shopping	710 (.136)	184	526
Average IADL	0.644 (.092)	0.224	0.420
Average ADL and IADL	1.660 (.127)	1.075	0.585

classifications based on detailed interviews: using the set of 6 ADL and 7 IADL variables that overlap with those from the screen, and using the full set of 6 ADL and 10 IADL variables from the original NLTCS analyses (Manton and Gu 2001). Detailed interviews classify about twice as many individuals in the two most disabled groups than the screen interviews. On the contrary, the screen measure classifies about three times as many people in the IADL-only group based on ADL and IADL tasks that were common between two survey components; that difference reduces substantially if we take the remaining 3 IADLs that were not available on the screen into account. We also note that the screen measure appears to place more individuals into the nondisabled category than either of the detailed measures.

Because the detailed interview did not record duration of IADL impairments, preventing us from adjusting detailed IADL outcomes for duration, we limit the rest of the analysis in this article to ADL outcomes. Note that the IADL-ADL hierarchy in Manton and Gu's (2001) disability classifications implies that the number of IADL impairments does not affect ADL disability levels. Thus, focusing on ADL outcomes still allows us to

Table 4. NLTCS 2004 screen and detailed interviews disability classifications, N = 5,201

	Nondisabled	IADL-disabled	1–2 ADL	3–4 ADL	5–6 ADL
Screen (6 ADL, 7 IADL)	2,550	707	1,205	378	361
Detailed (6 ADL, 7 IADL)	2,242	223	1,257	779	700
Detailed (6 ADL, 10 IADL)	1,870	595	1,257	779	700

Table 5. Cross-classification of ADL disability categories by the screen (Y_0) and detailed (Y_1) 2004 interviews; in parentheses, transition probabilities from disability categories on the screen (Y_0) into disability categories on the detailed interview (Y_1)

	$Y_1 = 0$ No ADL	$Y_1 = 1$ 1–2 ADL	$Y_1 = 2$ 2–3 ADL	$Y_1 = 3$ 3–4 ADL	Total
$Y_0 = 0$	1,601 (.633)	693 (.274)	162 (.064)	75 (.030)	2,531 (1.000)
$Y_0 = 1$	107 (.089)	490 (.407)	449 (.373)	158 (.131)	1,204 (1.000)
$Y_0 = 2$	9 (.024)	48 (.127)	132 (.349)	189 (.500)	378 (1.000)
$Y_0 = 3$	19 (.053)	26 (.072)	36 (.100)	278 (.774)	359 (1.000)
Total	1,736	1,257	779	700	4,472

make precise statements about more severe levels of disability. Consistent with our focus on ADL disability, we remove 729 individuals who constituted the newly sampled healthy supplement component in the 2004 NLTCS, reducing the sample size to 4,472.

Let Y_0 denote an ADL disability category as classified by the screen and Y_1 denote an ADL disability category as classified by the detailed interview. The disability categories that we consider in the rest of the article are: 0 = "no ADL impairment," 1 = "1–2 ADL impairments," 2 = "3–4 ADL impairments," and 3 = "5–6 ADL impairments."

Table 5 presents the cross-classification of ADL disability categories from the screen and detailed interviews for all individuals in the 2004 detailed survey. Examining transition probabilities from disability categories on the screen into disability categories on the detailed interview, we observe that a large proportion of transitions belongs to off-diagonal entries, illustrating a lack of consistency between the screen-based and detailed-based disability classifications. Transition probabilities above the main diagonal are larger than those below the main diagonal, indicating a shift towards greater amounts of ADL disability observed on the detailed survey. While transition probabilities below the main diagonal are smaller, they still show a substantial number of elderly survey participants whose disability status was improved by the detailed survey classification.

Overall, the preliminary analysis in this section illustrates the extent of discrepancies between the screen and detailed NLTCS disability measures.

4. Multivariate Models

Sources of ADL and IADL disability measurement error examined in the literature include demographic characteristics (Mathiowetz and Lair 1994), duration of disability, and interview mode: telephone versus in person (Rodgers and Millers 1997). Multiple studies showed that proxy respondents tend to overestimate functional disability compared to self-respondents (Hardy et al. 2006; Mathiowetz and Lair 1994; Rodgers and Miller 1997; Wolf et al. 2005). Respondents in face-to-face interviews tend to report significantly higher disability levels as compared to respondents in phone interviews (Rodgers and Miller 1997). The physical or mental health of the respondent can also induce measurement error. For example, depressed individuals tend to overestimate their functional limitations, whereas those with high self-efficacy may underestimate them (Kempen et al. 1996; Mathiowetz and Lair 1994). Finally, those in poorer health and with more severe disability are more likely to make errors in their reports of ADL limitations (Rodgers and Miller 1997).

Next, we analyze whether discrepancies between disability measurement in the screen and detailed NLTCS interviews are associated with time-lags between the interviews, demographic and interview characteristics of the respondent, and the number of medical conditions reported. We first assess effects of covariates on differences in the marginal distributions of disability between the two survey components. We then use transition models to assess the effects of covariates on transition probabilities (Agresti 2003).

4.1. Individual-level Covariates

Demographic covariates include age (five-year age groups above 65), education, marital status, gender and race. Interview characteristics include proxy indicators and the time lag between the screen and detailed interviews. Medical conditions that preceeded ADL and IADL questions in the detailed interviews consist of three blocks: 17 current medical conditions[9], 3 conditions related to dementia[10] that were asked only of proxy respondents, and 12 conditions during the past year[11]. Appendix Tables 8 and 9 summarize variables used, and provide both their locations in the NLTCS questionnaire and identification names in the 2004 NLTCS SAS data files. Briefly, demographic information is located in part 1 of the detailed survey, questions relating to ADL and IADL tasks come in parts 2 and 3, and cognitive functioning questions are found in part 9.

We use the total number of current and past year medical conditions as indicators of a respondent's present and recent health status, respectively. We define a cognitive impairment indicator to be 1 if at least one of the following holds: (a) an individual has answered 3 or more questions wrong on the Short Portable Mental Status Questionnaire (SPMSQ), (b) a proxy respondent has reported that the sample person has Alzheimer's disease, mental retardation, or dementia or (c) a proxy respondent has reported that cognitive impairment, as measured by the presence of Alzheimer's disease, was the reason for the sampled person to not answer SPMSQ questions; otherwise, we define the cognitive impairment to be 0. When a sampled person responded to the SPMSQ battery, our cognitive impairment indicator separates the SPMSQ mild, moderate and severe categories of cognitive functioning from the normal functioning category (Pfeiffer 1975).

Because the NLTCS screen interviews typically used a mixture of telephone and in person contacts (Manton et al. 1993; Wolf et al. 2005), it would be important to account for the interview administration mode (Rodgers and Miller 1997). Unfortunately, the 2004 NLTCS public use data files do not contain interview mode data, which precludes us from assessing the impact of interview mode in our analysis.

Demographic variables of age, sex, and race, as well as all ADLs had no missing data. Out of 4,472 individuals, 437 (9.9%) had missing data on some covariates. Given

[9] "Does. . . now have any of the following: Rheumatism or arthritis? Paralysis? Other permanent numbness or stiffness (besides paralysis, rheumatism or arthritis)? Multiple sclerosis? Cerebral palsy? Epilepsy? Parkinson's disease? Glaucoma? Diabetes? Cancer? Frequent constipation? Frequent trouble sleeping? Frequent severe headaches? Obesity or is. . . overweight? Arteriosclerosis or hardening of the arteries? Chronic pain? Pressure sores or skin ulcers?"

[10] "Does. . . now have Alzheimer's disease? Mental retardation? Dementia?"

[11] "Has. . . had any of the following in the past 12 months: A heart attack? Any other heart problems? Hypertension or high blood pressure? A stroke? Circulation trouble in arms or legs? Pneumonia? Bronchitis? Flu? Emphysema? Asthma? A broken hip? Other broken bones?"

this modest percentage of missing data, we take the approach of case-wise deletion; all analyses in the following sections use the sample size of 4,035 individuals with complete individual-level covariates. Limitations of this approach include a possibility of biased results when missing data are not missing completely at random (Little and Rubin 2002).

4.2. Modeling Changes in Ordinal Disability Classification

In this section, we analyze differences in marginal distributions of disability categories between the screen and detailed interviews. Let Y_j, $j = 0, 1$, be the disability

Table 6. Parameter estimates for matched pairs marginal models, $N = 4,035$

Variable	Parameter estimate	Standard error	Odds ratio	Confidence interval
Model (a)				
Intercept ($Y = 0$)	0.348	.043	1.42	(1.30, 1.54)
Intercept ($Y = 1$)	1.602	.048	4.96	(4.52, 5.46)
Intercept ($Y = 2$)	2.551	.060	12.82	(11.41, 14.41)
Detailed interview (yes = l, no = 0)	− .783	.025	0.46	(0.43, 0.48)
Lag (months)				
lag = 0	–	–	–	–
0 < lag ≤ 30 days	.023	.061	1.02	(0.91, 1.15)
lag > 30 days	.048	.073	1.05	(0.91, 1.21)
Model (b)				
Intercept ($Y = 0$)	1.967	.159	7.15	(5.23, 9.76)
Intercept ($Y = 1$)	3.560	.165	35.16	(25.47, 48.54)
Intercept ($Y = 2$)	4.736	.170	113.99	(81.69, 159.06)
Detailed interview (yes = l, no = 0)	− .970	.032	0.38	(0.36, 0.40)
Lag (months)				
lag = 0	–	–	–	–
0 < lag ≤ 30 days	− .023	.063	0.98	(0.86, 1.11)
lag > 30 days	− .082	.077	0.92	(0.79, 1.07)
Age (years over 65)				
65–69	–	–	–	–
70–74	.487	.127	1.63	(1.27, 2.09)
75–79	.250	.124	1.28	(1.01, 1.64)
80–84	.007	.118	1.01	(0.80, 1.27)
85 +	− .607	.112	0.54	(0.44, 0.68)
Proxy (yes = l, no = 0)	− 1.390	.076	0.25	(0.21, 0.29)
Education (coll. = l, HS/less = 0)	− .065	.060	0.94	(0.83, 1.05)
Marital status (yes = l, no = 0)	.214	.066	1.24	(1.09, 1.41)
Gender (female = l, male = 0)	− .260	.064	0.77	(0.68, 0.87)
Race (white = 1, other = 0)	.125	.095	1.13	(0.94, 1.36)
Cognitive impairment (yes = l, no = 0)	− .621	.084	0.54	(0.46, 0.63)
Medical conditions (current)	− .299	.018	0.74	(0.72, 0.77)
Medical conditions (last year)	− .170	.023	0.84	(0.81, 0.88)

classification outcome with categories $i = 0, 1, 2, 3$ (from "no ADL impairment" to "5–6 ADL impairments"), where Y_0 and Y_1 denote the disability classifications by the screen and detailed interviews, respectively. A cumulative logit model for matched pairs can be written as:

$$logit[P(Y_j \leq i|X)] = \alpha_i + \gamma_j + \sum_{k=1}^{K} \beta_k X_k, \quad i = 0, 1, 2, 3, j = 0, 1 \tag{1}$$

where X_k is the kth covariate. We fit two versions of the model: version (a) includes the time lag as the only additional covariate and version (b) controls for demographic and medical conditions variables in addition to the time lag.

Model (a), Table 6, includes effects of the detailed interview and the time lag. The effects of differences in wording and ordering of ADL and IADL questions, along with other design differences between the screen and detailed interviews that we may not have discovered, combine to produce the main detailed interview effect, given that the time lag is 0. The odds of the detailed disability classification $Y_1 \leq i$ equal $\exp(-0.783) = 0.46$ times the odds of the screen disability classification $Y_0 \leq i$. This implies that the marginal distributions of disability classifications are stochastically ordered so that Y_1 tends to be higher than Y_0, i.e., the detailed measure tends to assign higher disability categories than the screen measure. This ordering becomes stronger when we control for other covariates (Model (b), Table 6). To illustrate the effect of the detailed interview graphically, we provide cumulative response probabilities for an individual who is at the "median" level of all covariates in Figure 1. The "median" individual has a time lag less than 1 month, is between 80 and 84 years of age, is a

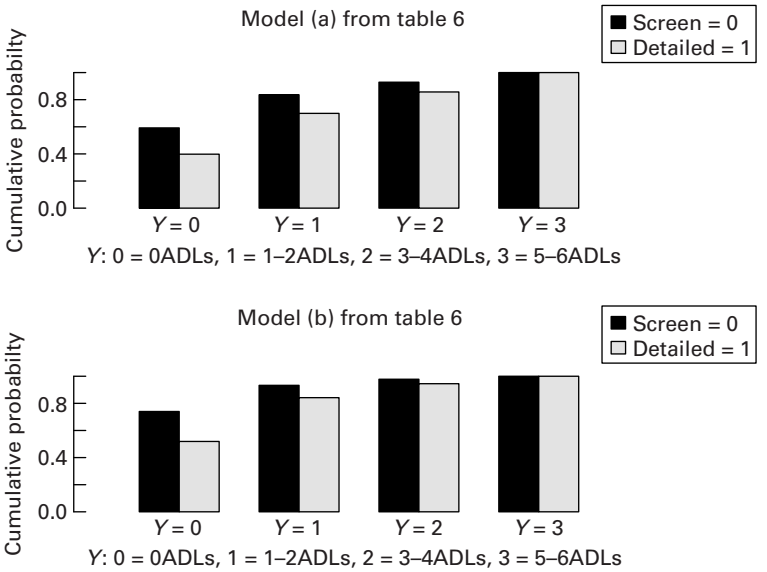

Fig. 1. Cumulative probabilities for matched pairs marginal models

self-respondent (no proxy), does not have college education, is nonmarried, female, white, has 2 current and 1 past year medical conditions and does not have cognitive impairment. Note that the screen has higher cumulative ($\leq i$) probabilities at $Y = 0$, $Y = 1$, and $Y = 2$, suggesting that the detailed measure tends to classify individuals into higher categories than the screen.

Examining Table 6, we observe that neither short nor long time lags between two interviews have an effect on the marginal distributions of disability categories, whether we control for other covariates or not. Older people ($85+$), those with proxy respondents, women and cognitively impaired individuals also tend to be classified into higher disability categories by the detailed than by the screen interview, controlling for other covariates. On the contrary, all other characteristics being equal, younger and married elderly people tend to fall into a less disabled status based on the detailed interview compared to their screen-based classification. The effects of reported numbers of current and past medical conditions are the most striking, considering that the estimated coefficients are on a per condition basis: with increased numbers of past and current medical conditions reported, detailed interview disability measurement tends to be more severe.

The effects of covariates can be explored further via interactions. For example, per suggestion from a reviewer, we fitted more general models assuming the detailed interview effect differs by observed disability category. The corresponding interactions indeed turned out to be significant (not shown).

Next, we specify a model for the marginal distribution of detailed-based disability categories, conditional on a screen-based disability category:

$$logit[P(Y_1 \leq i|X, Y_0)] = \alpha_i + \gamma_j I(Y_0 = j) + \sum_{k=1}^{K} \beta_k X_k, \quad i, j = 0, 1, 2, 3 \tag{2}$$

where X_k is the kth covariate. We fit two versions of the model, similar to the above.

For model (a), Table 7, the odds of detailed-based disability classification $Y_1 \leq i$, when the two interviews are more than one month apart, equal $\exp(0.177) = 1.19$ times the odds of detailed-based disability classification $Y_1 \leq i$, when the two interviews are administered on the same day, controlling for the screen-based disability category. This implies that, given screen-based disability classifications, the detailed survey tends to assign slightly lower disability categories when interviews are more than 1 month apart than it does when interviews are on the same day. This association becomes less significant when we control for other covariates. The directions and magnitudes of other covariate effects are consistent with those observed in the previous model without conditioning on screen-based severity of disability.[12]

[12] All analyses in this section were also carried out with an additional covariate, a depression indicator which we defined using responses to the following question: "During the last two weeks, have you felt so sad, blue, or depressed that you did not feel like doing the things you usually do?" Even though the addition of depression reduced the complete case sample size from 4,035 to 3,417, all of the covariate effects were consistent with those reported above and the depression indicator was not significant at the 0.05 level.

Table 7. Parameter estimates for cumulative marginal models, controlling for disability level, N = 4,035

Variable	Parameter estimate	Standard error	Odds ratio	Confidence interval
Model (a)				
Intercept ($Y = 0$)	0.606	.046	1.83	(1.68, 2.00)
Intercept ($Y = 1$)	2.515	.069	12.37	(10.81, 14.15)
Intercept ($Y = 2$)	4.157	.094	63.86	(53.13, 76.76)
Lag (months)				
lag = 0	–	–	–	–
0 < lag ≤ 30 days	−.057	.045	0.94	(0.86, 1.03)
lag > 30 days	.177	.056	1.19	(1.07, 1.33)
ADL-category$_{screen} = 1$	−2.487	.079	.077	(.065, .091)
ADL-category$_{screen} = 2$	−4.129	.131	.016	(.012, .021)
ADL-category$_{screen} = 3$	−5.199	.180	.011	(.007, .017)
Model (b)				
Intercept ($Y = 0$)	1.610	.135	5.00	(3.84, 6.52)
Intercept ($Y = 1$)	3.690	.148	40.06	(29.97, 53.55)
Intercept ($Y = 2$)	5.467	.166	236.82	(170.91, 328.14)
Proxy (yes = l, no = 0)	−.855	.089	0.43	(0.36, 0.51)
Lag (months)				
lag = 0	–	–	–	–
0 < lag ≤ 30 days	−.031	.046	0.97	(0.89, 1.06)
lag > 30 days	.123	.057	1.13	(1.01, 1.26)
Age (years over 65)				
65–69	–	–	–	–
70–74	.488	.079	1.63	(1.40, 1.90)
75–79	.076	.072	1.08	(0.94, 1.24)
80–84	−.170	.066	0.84	(0.74, 0.96)
85 +	−.492	.059	0.61	(0.54, 0.69)
Education (coll. = l, HS/less = 0)	−.089	.068	0.91	(0.80, 1.05)
Marital status (yes = l, no = 0)	.238	.076	1.27	(1.09, 1.47)
Gender (female = l, male = 0)	−.221	.074	0.80	(0.69, 0.93)
Race (white = 1, other = 0)	−.016	.106	0.98	(0.80, 1.21)
Cognitive impairment (yes = l, no = 0)	−.573	.094	0.56	(0.47, 0.68)
Medical conditions (current)	−.202	.020	0.82	(0.79, 0.85)
Medical conditions (last year)	−.172	.026	0.84	(0.80, 0.89)
ADL-category$_{screen} = 1$	−2.133	.082	.108	(.091, .129)
ADL-category$_{screen} = 2$	−3.615	.137	.024	(.017, .032)
ADL-category$_{screen} = 3$	−4.344	.190	.021	(.014, .033)

4.3. Adjusting for Disability Duration

To address whether the discrepancies can be accounted for by using duration-adjusted disability outcomes, we reestimate coefficients from models in Equations (1) and (2) with duration-adjusted binary disability outcomes from the detailed survey component.

Recall that information on duration was collected in different ways in the screen and detailed interviews. In the screen interview, if a newly sampled individual did not have an

ADL or an IADL problem that lasted or was expected to last 3 moths or longer, they were considered healthy (not chronically disabled), and were not administered the detailed interview. If an individual did have a *chronic* ADL or IADL problem on the screen, that individual went on to the detailed interview and their screen-based binary ADL (IADL) records remained unchanged (i.e., uncorrected for duration).

For ADL tasks, the detailed survey recorded the length of impairment with categories "<3 months," "3–6 months," etc. To mirror the use of disability duration information by the screen interview, one could redefine the detailed-based ADL outcomes as follows: If no chronic (>3 months) disability is present on any ADL task (that has a matching ADL in the screen), then set all ADL binary outcomes to 0; otherwise, keep ADL binary outcomes as recorded. Alternatively, one could redefine each ADL outcome separately: If no chronic ($>$ 3 months) disability is present on a particular ADL task, then set the binary outcome for that ADL to 0; otherwise, keep the ADL outcome as recorded.

We implemented the two approaches to duration adjustment for detailed-based ADL records (results not shown). The directions and magnitudes of all effects were consistent with those observed in the cummulative logit models (Tables 6 and 7), with and without conditioning on severity of disability as specified by the screen interview.

To summarize, our analyses show that detailed-based NLTCS disability assessment produces significantly higher ADL disability than the screen-based assessment. This difference increases significantly for the older, the more cognitively impaired, those with proxy respondents and those who report higher numbers of current and past medical conditions. The discrepancies persist after incorporating additional ADL duration information from the detailed interview.

5. Conclusion

In this article we provide a comprehensive description of the operational definition of disability employed by the NLTCS, highlighting differences between disability measures from the two linked interviews of the survey. We should emphasize that the final NLTCS product is a legacy of the design that was put together in 1982 when people knew much less about chronic disability, and technological support for doing surveys was rudimentary compared to what it is today. Longitudinal surveys face a difficult choice between keeping the questionnaire constant, even if it becomes outdated, and changing the questionnaire and risking being unable to make meaningful comparisons over time.

Using statistical modeling, we observe that the NLTCS detailed measure of disability status produces significantly higher ADL disability than the screen-based measure. This discrepancy between the two measures remained after we redefined detailed-based ADL outcomes by using additional disability duration questions to adjust for inaccuracies in determining chronic disabilities in the detailed interview. This ordering implies that a number of people who have been identified by the screen interview as nondisabled on ADLs would have been identified as ADL-disabled by the detailed interview if that was administered. Similarly, a number of people who are identified as non-IADL disabled by the screen could have been identified as disabled by the detailed interview (Table 4).

Although the question of which measure is less biased in general is not appropriate due to the lack of a universal gold standard, we should note that the two measures reflect different underlying conceptions of disability. Thus, the screen measure appears to reflect the notion of underlying disability (i.e., disability in carrying out an activity without help from another person or the use of equipment), while the detailed measure reflects the use of human help and/or special equipment.

The discrepancy in the NLTCS disability measures is problematic because different categories of NLTCS participants (newly sampled, returning, healthy supplement, etc.) are subject to different combinations of disability measures. The mix of survey participants with certain combinations of longitudinal disability assessments by the screen and detailed NLTCS interviews is different at each survey wave. Due to the dynamic nature of disability, the possibility of correcting 1984–1999 cross-sectional estimation of disability prevalence with survey weights is questionable. It also remains to be seen whether the observed discrepancy in screen-based and detailed-based disability measures may be related to disability declines previously estimated with the NLTCS. These issues await a development of comprehensible weights for the NLTCS, as the current set of publicly available weights is being re-examined (Spillman 2004).

Although the NLTCS remains the key source of trend data on chronic disability of community-based and institutionalized elders in the 1980s and 1990s, other surveys (National Nursing Home Survey, National Home and Hospice Survey, National Health Interview Survey, Health and Retirement Study, Survey of Income and Program Participation) may provide a basis for comparison of disability levels for community and institutionalized persons separately. A comprehensive review of instruments from these surveys may also help in estimating the chronic disability rates more effectively.

In the NLTCS, we identified multiple sources that jointly contribute to the reversal of the intended direction of difference between the screen and detailed disability measures, obtaining significantly higher ADL detailed disability estimates as compared to the screen measure. This illustrates a strong influence of multiple sources of variability that could be due to survey implementation, which again raises the important issue of replicability in survey research, especially when researchers deal with complex latent constructs such as disability. While the impact of some design choices examined in this article may be reduced or eliminated through careful planning, dealing with other influences may not always be possible due to increasing demands on, and the associated complexity of, modern survey instruments.

Our analyses indicate that disability survey researchers may want to avoid selecting subsamples of disabled persons based on a screen interview, even when it is believed that the screen selection process will "cast a wider net." When a negative screen assessment does not exclude the possibility of a positive detailed outcome, it becomes crucial that sampling procedures follow best practices of two-phase sampling (Deming 1977; Cochran 1977), making sure that the second-phase sample contains sufficient numbers of respondents who are not disabled on the screen to allow for estimation of the disability prevalence among that group. Two-phase sampling design was originally introduced by Neyman (1938); the method consists of drawing a random

sample from a population of interest during the first phase, classifying subjects into strata based on observed characteristics, allocating the second-phase subsample to maximize efficiency, and collecting additional data only from subjects sampled in the second phase.

Although the NLTCS design relies on a two-phase data collection scheme, it lacks a probabilistic allocation at the second phase to satisfy the definition of two-phase sampling design. Thus, with the exception of the special supplements starting in 1994, detailed interviews were only administered to the chronically disabled stratum as identified by the screen. It is not clear whether the sizes of the NLTCS healthy supplements and prevalence estimates for the waves for which the healthy supplements are available were based on two-phase design considerations.

In implementing disability surveys, researchers should consider relative efficiencies of multi-phase versus single-phase designs, based on the accuracy of the screen, disability prevalence, relative costs of interviews or tests in each phase (Shrout and Newman 1989), and the many practical and statistical complications arising with multi-phase designs. See Deming (1977) for a comparison of two-phase versus single-phase designs, and Clayton et al. (1998) for a recent data analysis example from a multi-phase study.

Appendices A. ADL and IADL Binary Outcome Determination

We reproduced logical paths of the triggering questions used in binary outcome determination processes from the 2004 NLTCS community detailed and screen interview

Table 8. Demographic and survey information variables

Variable description	Variable source	SAS name
Demographic information		
Age	Screen survey	CALCAGE
Gender	Analytic file	SEX
Race	Analytic file	RACE_WBO
Marital status	Analytic file	MARSTAT04
Education	Control file	EDUCA
Survey information		
Date of screen survey	Screen survey	SC_DATE
Screen complete (Y/N)	Screen survey	SC_COMPLETE
Date of detailed survey	Detailed survey (end)	COM_DATE
Proxy respondent status	Detailed survey (part 2)	ADLPROXY
ID (Person identifier)	Analytic file	SEQ
Disability duration		
Eating	Detailed survey – part 2	ADL_1H
Getting in/out of bed	Detailed survey – part 2	ADL_2H
Getting around inside	Detailed survey – part 2	ADL_3I
Dressing	Detailed survey – part 2	ADL_4G
Bathing	Detailed survey – part 2	ADL_5I
Toileting	Detailed survey – part 2	ADL_6K

Table 9. Medical conditions, dementia, depression, and cognitive impairment variables

Variable description	Part	SAS name	Variable description	Part	SAS name
Curr. Med. Conditions			*Past Med. Conditions*		
Arthritis	1	CND_1A01	Heart attack	1	CND_2_01
Paralysis	1	CND_1A02	Heart probs. (Other)	1	CND_2_02
Numbness/Stiffness	1	CND_1A03	Hypertension	1	CND_2_03
Multiple sclerosis	1	CND_1A04	Stroke	1	CND_2_04
Cerebral palsy	1	CND_1A05	Circulation trouble	1	CND_2_05
Epilepsy	1	CND_1A06	Pneumonia	1	CND_2_06
Parkinson's disease	1	CND_1A07	Bronchitis	1	CND_2_07
Glaucoma	1	CND_1A08	Flu	1	CND_2_08
Diabetes	1	CND_1A09	Emphysema	1	CND_2_09
Cancer	1	CND_1A10	Asthma	1	CND_2_10
Constipation	1	CND_1A11	Broken hip	1	CND_2_11
Sleeping trouble	1	CND_1A12	Broken bones (Other)	1	CND_2_12
Headaches	1	CND_1A13	*Cognitive Impairment*		
Obesity	1	CND_1A14	Today's date	9	MNT_1
Arteriosclerosis	1	CND_1A15	Day of the week	9	MNT_2
Chronic pain	1	CND_1A16	Street address	9	MNT_3
Pressure sores	1	CND_1A17	Home state	9	MNT_4
Dementia (used for Cognitive Imp.)			Current age	9	MNT_5
Alzheimer's disease	1	CND_1B_1	Date of birth	9	MNT_6
Mental retardation	1	CND_1B_2	Current U.S. Pres.	9	MNT_7
Dementia	1	CND_1B_3	Previous U.S. Pres.	9	MNT_8
			Mom's maiden name	9	MNT_9
			Completing Math Seq.	9	MNT_10
			Reason SPMSQ Not Done	9	MNT_11TG_7

questionnaires. The ADL and IADL questions remained the same over all the survey waves. The tables are posted on the web:

1. ADL determination by the NLTCS screen interview in 2004: http://www.stat.Washington.edu/ ~ elena/NLTCS/ADL-Scr-Triggers.pdf
2. IADL determination by the NLTCS screen interview in 2004: http://www.stat.Washington.edu/ ~ elena/NLTCS/IADL-Scr-Triggers.pdf
3. ADL determination by the NLTCS detailed community interview in 2004: http://www.stat.Washington.edu/ ~ elena/NLTCS/ADL-Det-Triggers.pdf
4. IADL determination by the NLTCS detailed community interview in 2004: http://www.stat.Washington.edu/ ~ elena/NLTCS/IADL-Det-Triggers.pdf

B. Description of Variables

Tables 8 and 9 summarize variables used in this study, provide their locations in the NLTCS questionnaire and provide their identification names in the 2004 NLTCS SAS data files.

6. References

Agresti, A. (2003). Categorical Data Analysis, (Second Edition). New York: Wiley.

Bootsma-van der Wiel, A., Gussekloo, J., de Craen, A.J., van Exel, E., Knook, D.L., Lagaay, A.M., and Westendorp, R.G. (2001). Disability in the Oldest Old: "Can do" or "Do do?". Journal of the American Geriatrics Society, 49, 909–914.

Committee on National Statistics and Committee on Population. (2005). Summary from the Expert Meeting on the Future of the National Long-Term Care Survey: Data Needs for Disability Policy Making and Research for the Future. The National Academies, National Institute on Aging.

Committee on National Statistics and Committee on Population. (2006). Summary from the Expert Meeting on the Future of the National Long-Term Care Survey. The National Academies, National Institute on Aging.

Clayton, D., Spiegelhalter, D., Dunn, C., and Pickles, A. (1998). Analysis of Longitudinal Binary Data from Multi-phase Sampling (With Discussion). Journal of the Royal Statistical Society, Series B, 60, 71–102.

Cochran, W.G. (1977). Sampling Techniques. New York: Wiley.

Deming, W.E. (1977). An Essay of Screening, or on Two-phase Sampling, Applied to Surveys of a Community. International Statistical Review, 33, 29–37.

Freedman, V.A. (2000). The Implications of Asking Ambiguous Difficulty Questions: An Analysis of the Second Wave of the Asset and Health Dynamics of the Oldest Old Study. Journals of Gerontology: Social Sciences, 55, 288–297.

Freedman, V.A., Martin, L.G., and Schoeni, R.F. (2002). Recent Trends in Disability and Functioning Among Older Adults in the United States: A Systematic Review. Journal of the American Medical Association, 288, 3137–3146.

Freedman, V.A. and Soldo, B. (eds) (1994). Trends in Disability at Older Ages: Summary of a Workshop. Washington, DC: National Academy Press, Committee on National Statistics.

Gill, T.M. and Gahbauer, E.A. (2005). Overestimation of Chronic Disability Among Elderly Persons. Archives of Internal Medicine, 165, 2625–2630.

Gill, T.M., Hardy, S.E., and Williams, C.S. (2002). Underestimation of Disability in Community-living Older Persons. Journal of the American Geriatric Society, 50, 1492–1497.

Gill, T.M. and Kurland, B. (2003). The Burden and Patterns of Disability in Activities of Daily Living Among Community-living Older Persons. The Journals of Gerontology, 58, 70–75.

Hardy, S.E., Allore, H.G., Gup, Z., Dubin, J.A., and Gill, T.M. (2006). The Effect of Prior Disability History on Subsequent Functional Transitions. Journal of Gerontology: Medical Sciences, 61A, 272–277.

Katz, S., Ford, A., Moskowitz, R., Jackson, B., and Jaffe, M. (1963). Studies of Illness in the Aged, the Index of ADL: A Standardized Measure of Biological and Psychosocial Function. Journal of the American Medical Association, 185, 914–919.

Kempen, G., Steverink, N., Ormel, J., and Deeg, D. (1996). The Assessment of ADL Among Frail Elderly in an Interview Survey: Self-report Versus Performance-based Tests and Determinants of Discrepancies. Journal of Gerontology: Psychological Sciences, 51B, P254–P260.

Lawton, M.P. and Brody, E.M. (1969). Assessment of Older People: Self-maintaining and Instrumental Activities of Daily Living. Gerontologist, 9, 179–186.

Lee, S., Mathiowetz, N.A., and Tourangeau, R. (2007). Measuring Disability in Surveys: Consistency Over Time and Across Respondents. Journal of Official Statistics, 23, 163–184.

Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis with Missing Data, (Second Edition). Hoboken: Wiley.

Lynch, S.M., Brown, S.J., and Harmsen, K.G. (2003). The Effect of Altering ADL Thresholds on Active Life Expectancy Estimates for Older Persons. The Journals of Gerontology: Social Sciences, 58, S171–S178.

Manton, K.G. (1988). A Longitudinal Study of Funtional Change and Mortality in the United States. The Journals of Gerontology: Social Sciences, 43, 153–161.

Manton, K.G. (1997). Analytic Data Files Memo. CD NLTCS public use data: 1982–1994.

Manton, K.G. (2007). Memorandum to Brenda Spillman: Beta-2 Public Use File. Unpublished.

Manton, K.G., Corder, L.S., and Stallard, E. (1993). Estimates of Change in Chronic Disability and Institutional Incidence and Prevalence Rates in the U.S. Elderly Population from the 1982, 1984, and 1989 National Long Term Care Survey. Journal of Gerontology: Social Sciences, 48, 153–166.

Manton, K.G., Corder, L.S., and Stallard, E. (1997). Chronic Disability Trends in Elderly United States Populations: 1982–1994. Proceedings of the National Academy of Sciences, 94, 2593–2598.

Manton, K.G. and Gu, X. (2001). Changes in The Prevalence of Chronic Disability in the United States Black and Nonblack Population Above Age 65 from 1982 to 1999. Proceedings of the National Academy of Sciences, 98, 6354–6359.

Manton, K.G., Gu, X., and Lamb, V.L. (2006). Change in Chronic Disability from 1982 to 2004/2005 as Measured by Long-term Changes in Function and Health in the U.S. Elderly Population. Proceedings of the National Academy of Sciences, 103(48), 18374–18379.

Mathiowetz, N.A. (2001). Methodological Issues in the Measurement of Persons with Disabilities. Research in Social Science and Disability, 2, 125–144.

Mathiowetz, N.A. and Lair, T.W. (1994). Getting Better? Change or Error in the Measurement of Functional Limitations. Journal of Economic and Social Measurement, 20, 237–262.

Neyman, J. (1938). Constitution to the Theory of Sampling from Human Populations. Journal of the American Statistical Association, 33, 101–106.

Pfeiffer, E. (1975). A Short Portable Mental Status Questionnaire for the Assessment of Organic Brain Deficit in Elderly Patients. Journal of American Geriatrics Society, 23, 433–441.

Rodgers, W. and Miller, B. (1997). A Comparative Analysis of ADL Questions in Surveys of Older People. The Journals of Gerontology: Series B, 52B, 21–36.

Särndal, C.-E. (1978). Design-based and Model-based Inference in Survey Sampling (With Discussion). Scandinavian Journal of Statistics, 5, 27–52.

Shrout, P.E. and Newman, S.C. (1989). Design of Two-phase Prevalence Surveys of Rare Disorders. Biometrics, 45, 549–555.

Spillman, B.C. (2004). Changes in Elderly Disability Rates and the Implications for Health Care Utilization and Cost. The Milbank Quarterly, 82, 157–194.

Todorov, A. (2000). Context Effects in National Health Surveys: Effects of Preceding Questions on Reporting Serious Difficulty Seeing and Legal Blindness. Public Opinion Quarterly, 64, 65–76.

Unicon Research Corporation. (2002). Documentation and Extraction Utilities for National Long Term Care Surveys.

Wolf, D.A., Hunt, K., and Knickman, J. (2005). Perspectives on the Recent Decline Indisability at Older Ages. The Milhank Quarterly, 83(3), 365–395.

Received September 2008
Revised November 2009