Reliability and stability of the 6-question disability measure in the Survey of Income and Program Participation*

Matthew W. Brault[1]
[1]U.S. Census Bureau, 4600 Silver Hill Rd, SEHSD 7H168E, Washington, DC 20233

Abstract
Researchers have long acknowledged that disability is a dynamic characteristic (Adler 1992, Verbrugge, Reoma and Gruber-Baldini 1994, Wolf and Gill 2008). Nonetheless, the concept is often treated as static over short periods in longitudinal studies. The disability status of a respondent is asked during one interview and assumed to remain constant over several interviews or for the life of the panel. I explore this assumption using reoccurring data on disability status from the Survey of Income and Program Participation (SIPP). In the 2008 panel, the six-question set of disability questions from the ACS were added to a reoccurring topical module. I employ structural models from Heise (1969) and Wiley and Wiley (1970) to separate reporting error from real change under two assumptions about the measures' reliability. Both methods assume that disability status follows a first-order Markov process. With these methods, I find that the disability measures in the SIPP had relatively moderate to low reliability with coefficients between 0.414 and 0.638. Conversely, an individual's true disability status is strongly correlated with the person's status one year later (r=0.937). Thus, the supposition that disability remains relatively consistent over short periods has some validity.

Key Words: Disability, longitudinal data, reliability, stability

1. Introduction

Researchers have long acknowledged that disability is a dynamic process (Adler 1992, Verbrugge, Reoma and Gruber-Baldini 1994, Wolf and Gill 2008). Over time, individuals may become disabled as aging and illness contribute to increased difficulty performing various basic life activities. Adverse events like injuries or medical errors may result in disablement. Alternatively, innovations in health treatments and technological accommodations may improve a person's functioning and reduce disability. Despite the acknowledgement of disability's dynamic nature, the concept is often treated as static over short periods in longitudinal studies. The disability status of a respondent is asked during one interview and assumed to remain constant over several interviews or for the life of the panel.

This paper attempts to explore the assumption that one's disability status remains relatively constant over a short period such as one year. Interest in this topic extends from other work with the Survey of Income and Program Participation (SIPP), where we intend to combine disability information from a topical module (supplemental questionnaire) in one wave (interview) with other data collected in an immediately preceding wave. In combining the two waves, we assume that a respondent's disability

*__Disclaimer:__ This paper is released to inform interested parties of ongoing research and to encourage discussion. The views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

status remained the same over the 4 months between waves. This analysis was conducted to assess how sound of an assumption it is.

This examination of the stability of disability has broader applications because other surveys make similar assumptions, often implicitly. The Bureau of Labor Statistics (BLS) collects disability information in the Current Population Survey (CPS) basic monthly sample during the first and fifth interview months and retains the respondent's disability status across months 2 through 4 and 6 through 8, respectively.[1] For example, estimates of poverty for people with disabilities in the CPS Annual Social and Economic Supplement (ASEC) – conducted around March of each year – uses disability statuses of respondents that were collected up to 3 months prior to the interview with the supplement questionnaire. Only 31 percent of households in the 2012 CPS ASEC received both the disability questions and the supplement information in the same interview. With a relative minority of observations with concurrent ASEC and disability information, the reliance on the implicit assumption about the dynamics of disability should be tested.

Dynamics in disability status – even short-term dynamics – have largely been studied in older populations and have focused on the determinants for changes with activities of daily living (ADLs) and instrumental activities of daily living (IADLs) (Mendes de Leon, Guralnik and Bandeen-Roche 2002, Hardy and Gill 2004, Cai, Schenker and Lubitz 2006). Analyses of disability spells in these populations have revealed shorter episodes with more frequent monitoring, implying that short-term dynamics do occur. Several transitions may occur between current-status measures, even at 12-month intervals (Wolf and Gill 2008). I do not speculate as to the length of spells but rather focus on the likelihood of having the same status over the short intervals, given that variability does occur.

Because the terms *stability* and *reliability* can and have been used in many contexts, I will define them here. For this paper, stability is the likelihood that an individual respondent's disability status in period t is the same as it was in period $t-1$. Stability is measured as a correlation between the statuses at the two periods. Reliability is the likelihood that an individual respondent's disability status would be the same had we asked them two times during the same period, also measured as a correlation. For many discrete variables, reliability is synonymous with terms like measurement error because the true value is fixed and variation in reporting is due to error from how the topic was measured. For disability, which can be quite subjective, the true value can be considered a distribution of functioning and even a good question will produce some variation. With this in mind, the reliability of disability deals may capture variability within a distribution of functioning whereas the stability deals with the shifting of the entire distribution. In this paper, reliability and reporting error are used interchangeably.

Determining the stability of responses requires separating within-time variation from between-times variation in disability status, like parsing noise from a signal. This may be done using *a priori* estimates of reliability to separate the two. In 2006, the Census Bureau conducted a content test of the American Community Survey (ACS) where the six disability questions – that were added to the SIPP and used here – were tested for reliability, among other criteria (Brault, Stern and Raglin 2007). Using test-retest methods with a follow-up interview two weeks after receiving the returned mail questionnaire or CAPI interview, the disability items demonstrated moderate reliability.[2]

---

[1] For more information on the CPS rotation groups and other aspects of its sample design, see *Current Population Survey Design and Methodology*. Technical Paper 66 (U.S. Census Bureau 2006).

[2] Reliability coefficients and their characterization as "moderate" were derived from the response cross-tabulations in the evaluation report's appendices (Brault, Stern and Raglin 2007).

Reliability, however, can be instrument specific, so differences between the methodologies in the ACS and SIPP preclude the use of the content test's estimates of reliability to parse out the real change. While I may not be able to use this estimate directly, the Content Test does provide one source for benchmarking these results.

Estimates of reliability from other disability questions in the SIPP may also help ballpark the levels we might expect to find here. In past studies, specific elements from the more comprehensive SIPP disability measure were shown to have moderate to poor reliability (McNeil 2000). Consequently, some researchers have questioned the validity of the comprehensive measure, using such evidence to justify developing new questions over adopting the SIPP questions (McMenamin, et al. 2005). Similar findings in the 6-question set might signal that the SIPP instrument or the disability topic in general, independent of how a question may be asked, could share some of the culpability for the measures' reliability.

Without usable prior estimates of reliability, this paper attempts first to determine the reporting error associated with the 6-question set of disability questions. Second, it uses this reliability to derive estimates of the stability of disability status over 1-year intervals and interpolates the results for the 4-month interval that correspond to the survey's waves.

2. Data and Methods

The 6-question set of disability questions were developed for the ACS, but have subsequently been added to a number of other federal household surveys including the CPS and SIPP. Researchers have long sought to have more consistent disability information about the population, particularly in the SIPP given its detailed collection of program enrollment (Adler 1992). This paper is inherently an analysis of the question set as it appears in the SIPP. These questions were added to the 2008 SIPP in the topical module on medical expenses and utilization of health care, asked during waves 4, 7, and 10. For this analysis, I use only cases from adult respondents age 15 and older who had valid interviews in each of the three waves and who provided valid responses to all of the disability questions. Survey weights were adjusted to account for the attrition and nonresponse to the disability items. All comparisons of weighted estimates use replicate weights to determine population variances and have been tested at the 90-percent confidence level.[3]

To assess the stability of disability over time I must start with a model for disablement; an individual's disability status can be expressed as a first-order Markov process. Broadly, a person's current disability status (x_t) is a function of the person's disability status in a prior period (x_{t-1}) and other observed or unobserved factors (u_t) that may result in the individual becoming limited or recovering from limitations in basic life activities, expressed in equation (1).

$$x_t = \varphi x_{t-1} + u_t \tag{1}$$

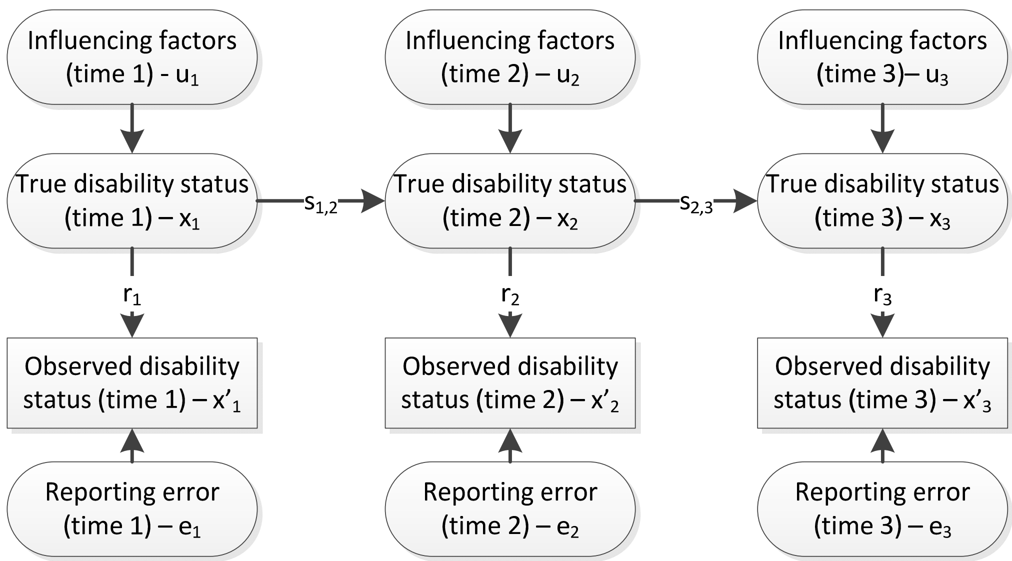
Furthermore, at any given interview at time *t*, the observed disability status ${x'}_t$ can be decomposed into the true disability status (x_t) and an error component (e_t) where $E(e_t) = 0$, shown in equation (2).

$${x'}_t = x_t + e_t \tag{2}$$

[3] Further information on the source and accuracy of estimates can be found online at http://www.census.gov/sipp/source.html

The combination of the two models shows that one cannot attribute an observed change in disability status entirely to reporting error or the factors that result in a true change in disability status. Heise (1969) and Wiley and Wiley (1970) demonstrate how three or more interviews of collection can be used to parse out the error and calculate estimates of stability using a structural equation model. From the structural model, Heise presents a formula for reliability and stability that is derived from certain assumptions: (1) the relationship between the true value and observed value is constant over time; (2) errors are not correlated with the true values; (3) errors at different times are not correlated; and (4) the factors that influence a change in the true value are not correlated with initial disability status. Wiley and Wiley suggest that assuming constant reliability may be doubtful. They present formulae for measuring reliability and stability that instead relies on an assumption that error variance remains constant across periods. These methods for determining the stability and reliability of a measure have been used to examine other topics, such as media exposure (Lee, Hornik and Hennessy 2008).

Figure 1. General Model of Disability Over Time



With these methods, the reliability of the disability measure can be determined from the correlations between statuses at 3 points in time following a general model of disability over time as shown in Figure 1. Per Heise's constant reliability assumption, $r_{XX} = r_1 = r_2 = r_3$, where r_{XX} is the reliability coefficient and r_1, r_2, and r_3 are the correlations between observed and true disability status, as shown in Figure 1. Equation (3) shows r_{XX} as a function of the correlations of x'_1 and x'_2 (as r_{12}), x'_2 and x'_3 (as r_{23}), and x'_1 and x'_3 (as r_{13}). If disability status is truly Markovian, then the correlation between times 1 and 3 should be the product of the correlations between times 1 and 2 and times 2 and 3. Any deviation from this relationship would be attributable to error, which is assumed constant.

$$r_{xx} = \frac{r_{12} r_{23}}{r_{13}} \tag{3}$$

Equations (4) show how the same correlations can be used to determine stability coefficients S_{12}, S_{23}, and S_{13}. These equations also show that if there is no reporting error ($r_{xx} = 1$), the correlations between periods would be equal to the stability coefficients.

$$S_{12} = r_{12}/r_{xx} = r_{13}/r_{23}$$
$$S_{23} = r_{23}/r_{xx} = r_{13}/r_{12} \quad (4)$$
$$S_{13} = r_{13}/r_{xx} = {r_{13}}^2/r_{12}r_{23} = S_{12}S_{23}$$

For the Wiley and Wiley method, the relationships between observed statuses from the three interviews are restated in equations (5).

$$x'_1 = u_1 + e_1$$
$$x'_2 = \alpha_{12}u_1 + u_2 + e_2 \quad (5)$$
$$x'_3 = \alpha_{23}(\alpha_{12}u_1 + u_2) + u_3 + e_3$$

From the covariance matrix between values of x', relevant parameters α_{12}, α_{23}, V(e), V(u_1), V(u_2), and V(u_3) are estimated, based on the assumption that V(e_1) = V(e_2) = V(e_3) = V(e). These parameters then fit equations (6), which produce estimates of reliability at different periods.

$$\rho_1 = \frac{V(u_1)}{[V(u_1)+V(e)]}$$
$$\rho_2 = \frac{[\alpha_{12}^2 V(u_1)+V(u_2)]}{[\alpha_{12}^2 V(u_1)+V(u_2)+V(e)]} \quad (6)$$
$$\rho_3 = \frac{\{\alpha_{23}^2[\alpha_{12}^2 V(u_1)+V(u_2)]+V(u_3)\}}{\{\alpha_{23}^2[\alpha_{12}^2 V(u_1)+V(u_2)]+V(u_3)+V(e)\}}$$

The stability coefficients can likewise be calculated, as shown in equations (7).

$$\gamma_{12} = \alpha_{12}\frac{\sqrt{V(u_1)}}{\sqrt{\alpha_{12}^2 V(u_1)+V(u_2)}}$$
$$\gamma_{23} = \alpha_{23}\frac{\sqrt{\alpha_{12}^2 V(u_1)+V(u_2)}}{\sqrt{\alpha_{23}^2[\alpha_{12}^2 V(u_1)+V(u_2)]+V(u_3)}} \quad (7)$$
$$\gamma_{13} = \alpha_{12}\alpha_{23}\frac{\sqrt{V(u_1)}}{\sqrt{\alpha_{23}^2[\alpha_{12}^2 V(u_1)+V(u_2)]+V(u_3)}} = \gamma_{12}\gamma_{23}$$

These methods were employed to calculate reliability and stability estimates for the overall disability measure (‘yes’ to any of the six disability types), for each type, and for the work disability measure that is part of the SIPP core instrument, for comparison purposes.

3. Results

If one looks at the prevalence of disability at each of the three waves, it appears that there is little change.[4] As shown in Table 1, about 15.1 percent of the adult population had a disability in 2009 (wave 4) and that level was not statistically different from the prevalence in 2010 (wave 7) and 2011 (wave 10) at 15.0 percent and 15.1 percent,

---

[4] The prevalence of vision difficulty in wave 4 was statistically different from in wave 7, the prevalence of self-care difficulty in wave 4 was statistically different from in wave 7, and the prevalence of independent living difficulty in wave 4 was statistically different from in wave 10. All other differences across waves were not statistically significant.

respectively. For the individual disability types, the ranges in prevalence across the three years never exceed 0.2 percentage points. Hearing difficulty was between 4.3 and 4.5 percent, vision difficulty between 2.6 and 2.8, cognitive difficulty between 5.2 and 5.4, ambulatory difficulty between 9.0 and 9.2, self-care difficulty between 2.6 and 2.8, and independent living difficulty between 5.1 and 5.3. The net difference in disability prevalence between any two interviews (4 to 7, 4 to 10, or 7 to 10) would likely be small and disability status would appear to be stable.

Table 1. Disability prevalence in 2008 SIPP

Disability Type	2009 (wave 4)		2010 (wave 7)		2011 (wave 10)	
	Estimate	SE	Estimate	SE	Estimate	SE
With a disability	15.1	0.18	15.0	0.19	15.1	0.19
hearing difficulty	4.5	0.10	4.3	0.10	4.3	0.09
vision difficulty	2.8	0.08	2.6	0.08	2.6	0.09
cognitive difficulty	5.4	0.11	5.2	0.12	5.2	0.11
ambulatory difficulty	9.0	0.13	9.2	0.14	9.1	0.14
self-care difficulty	2.8	0.07	2.6	0.07	2.7	0.08
independent living difficulty	5.1	0.09	5.2	0.10	5.3	0.11

Source: U.S. Census Bureau, Survey of Income and Program Participation, 2008 Panel

However, if one cross-tabulates disability from one wave with the individuals' responses in another wave, as shown in Table 2, the stability of disability becomes questionable. Of the 41,328 respondents with valid interviews in all three waves, 4,859 had a change in disability status (10.6 percent weighted). 2,321 people reported a disability in wave 4 but reported no disability in wave 7 and 2,538 reported no disability in wave 4 but reported a disability in wave 7. Response combinations between waves 4 and 10 and waves 7 and 10 displayed similar patterns.

Furthermore, inconsistent responses occur for each of the disability types, including those that may be associated with permanent health conditions. Serious difficulty hearing or seeing are not often considered temporary, as most conditions associated with hearing or vision loss are degenerative. Despite this, more than half of respondents who reported a hearing difficulty in wave 4 did not report difficulty in wave 7; the same was true for those who reported vision difficulty in wave 4.

Based on these cross-tabulations, Table 3 shows the variance-covariance matrices and correlations between disability statuses at different interviews. For the overall disability measure, the status in waves 4 and 7 had a correlation of 0.578. For disability between wave 7 and wave 10, the correlation was 0.607; and between wave 4 and wave 10, the correlation was 0.555. The correlations between wave 4 and 7 for the individual disability types ranged from 0.339 for vision difficulty to 0.567 for ambulatory difficulty. Work disability between the two waves had a correlation of 0.788. For most items, the correlations between waves 7 and 10 were higher than the correlations between waves 4 and 7.[5]

[5] The correlation between waves 4 and 7 for hearing difficulty was not statistically different from the correlation between waves 7 and 10.

Table 2. Crosstabulations of Disability in Waves 4, 7, and 10

	With a disability		Hearing difficulty		Vision difficulty		Cognitive difficulty		Ambulatory difficulty		Self-care difficulty		Independent living difficulty	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total	41,328	100.0	41,328	100.0	41,328	100.0	41,328	100.0	41,328	100.0	41,328	100.0	41,328	100.0
With a disability in wave 4														
With a disability in wave 7	4,529	9.4	986	2.0	436	0.9	1,130	2.5	2,618	5.2	454	0.9	1,193	2.5
No disability in wave 7	2,321	5.1	1,053	2.2	809	1.7	1,105	2.4	1,507	3.1	636	1.3	909	1.9
With a disability in wave 10	4,505	9.3	937	1.9	429	0.9	1,103	2.4	2,575	5.1	474	1.0	1,195	2.5
No disability in wave 10	2,345	5.2	1,102	2.3	816	1.7	1,132	2.5	1,550	3.2	616	1.3	907	1.9
No disability in wave 4														
With a disability in wave 7	2,538	5.5	1,095	2.4	750	1.6	1,083	2.4	1,815	3.8	653	1.4	1,143	2.3
No disability in wave 7	31,940	80.0	38,194	93.4	39,333	95.8	38,010	92.7	35,388	87.9	39,585	96.4	38,083	93.2
With a disability in wave 10	2,896	6.2	1,229	2.6	855	1.8	1,249	2.6	2,074	4.3	873	1.8	1,417	2.9
No disability in wave 10	31,582	79.3	38,060	93.1	39,228	95.6	37,844	92.4	35,129	87.4	39,365	96.0	37,809	92.6
With a disability in wave 7														
With a disability in wave 10	4,896	10.2	1,045	2.1	498	1.0	1,235	2.7	2,945	5.9	586	1.2	1,430	3.0
No disability in wave 10	2,171	4.8	1,036	2.2	688	1.5	978	2.2	1,488	3.1	521	1.1	906	1.9
No disability in wave 7														
With a disability in wave 10	2,505	5.4	1,121	2.4	786	1.7	1,117	2.3	1,704	3.5	761	1.5	1,182	2.5
No disability in wave 10	31,756	79.7	38,126	93.2	39,356	95.8	37,998	92.8	35,191	87.5	39,460	96.2	37,810	92.7

Note: Percentages are weighted.
Source: U.S. Census Bureau, Survey of Income and Program Participation, 2008 Panel

Table 3. Variance-Covariance Matrices and Correlation Coefficients between Waves 4, 7, and 10

With a disability

	Wave 4	Wave 7	Wave 10
Wave 4	0.720 [1.000]		
Wave 7	0.421 [0.578]	0.740 [1.000]	
Wave 10	0.411 [0.555]	0.456 [0.607]	0.763 [1.000]

With a hearing difficulty

	Wave 4	Wave 7	Wave 10
Wave 4	0.235 [1.000]		
Wave 7	0.105 [0.440]	0.241 [1.000]	
Wave 10	0.099 [0.407]	0.111 [0.449]	0.251 [1.000]

With a vision difficulty

	Wave 4	Wave 7	Wave 10
Wave 4	0.148 [1.000]		
Wave 7	0.049 [0.339]	0.144 [1.000]	
Wave 10	0.047 [0.312]	0.057 [0.381]	0.153 [1.000]

With a cognitive difficulty

	Wave 4	Wave 7	Wave 10
Wave 4	0.275 [1.000]		
Wave 7	0.134 [0.489]	0.271 [1.000]	
Wave 10	0.128 [0.460]	0.145 [0.526]	0.280 [1.000]

With an ambulatory difficulty

	Wave 4	Wave 7	Wave 10
Wave 4	0.445 [1.000]		
Wave 7	0.262 [0.567]	0.478 [1.000]	
Wave 10	0.252 [0.537]	0.294 [0.604]	0.495 [1.000]

With a self-care difficulty

	Wave 4	Wave 7	Wave 10
Wave 4	0.129 [1.000]		
Wave 7	0.051 [0.393]	0.130 [1.000]	
Wave 10	0.052 [0.371]	0.066 [0.463]	0.155 [1.000]

With an independent living difficulty

	Wave 4	Wave 7	Wave 10
Wave 4	0.247 [1.000]		
Wave 7	0.134 [0.521]	0.270 [1.000]	
Wave 10	0.133 [0.490]	0.159 [0.558]	0.300 [1.000]

Work disability (from core instrument)

	Wave 4	Wave 7	Wave 10
Wave 4	0.638 [1.000]		
Wave 7	0.516 [0.788]	0.672 [1.000]	
Wave 10	0.490 [0.739]	0.555 [0.815]	0.690 [1.000]

Source: U.S. Census Bureau, Survey of Income and Program Participation, 2008 Panel

Using these values in the method proposed by Heise, the reliability of the disability measure was calculated at 0.632, shown in Table 4. From the variance-covariance matrix between waves shown in Table 3, per the Wiley and Wiley method, the reliabilities in wave 4, 7, and 10 were 0.621, 0.632, and 0.643, respectively. For the individual disability types, Heise reliability coefficients ranged from 0.414 (low reliability) for vision difficulty to 0.638 (moderate reliability) for ambulatory difficulty. Comparatively, the reliability coefficient for work disability was 0.870.

Table 4. Reliability of disability items by wave

Disability Type	Heise (1969)	Wiley & Wiley (1970)		
		Wave 4	Wave 7	Wave 10
With a disability	0.632	0.621	0.632	0.643
hearing difficulty	0.485	0.472	0.485	0.506
vision difficulty	0.414	0.432	0.414	0.450
cognitive difficulty	0.560	0.567	0.560	0.575
ambulatory difficulty	0.638	0.612	0.638	0.651
self-care difficulty	0.490	0.487	0.490	0.574
independent living difficulty	0.594	0.556	0.594	0.635
Work disability	0.870	0.863	0.870	0.874

Source: U.S. Census Bureau, Survey of Income and Program Participation, 2008 Panel

Having determined the reliability of the measures, the true change in status can be parsed out. Tables 5 and 6 show stability coefficients across the waves. Assuming constant reliability (the Heise method), disability status in waves 4 and 7 were correlated at 0.914, which improved to 0.960 between waves 7 and 10. The average annual stability coefficient, determined by geometric average, was 0.937. Interpolating the estimates for per-wave stability (based on the compounding nature of stability coefficients), I estimate the coefficient to be 0.979. The differences between Wiley and Wiley-based estimates of stability and those derived with the Heise method – especially for the average annual coefficients – were within the bounds of sampling error.

Table 5. Stability of disability items across waves (Heise Method)

Disability Type	W4-W7	SE	W7-W10	SE	Avg/yr	SE	Avg/wave	SE
With a disability	0.914	0.011	0.960	0.010	0.937	0.009	0.979	0.003
hearing difficulty	0.906	0.022	0.925	0.025	0.915	0.020	0.971	0.007
vision difficulty	0.818	0.035	0.920	0.028	0.868	0.028	0.954	0.010
cognitive difficulty	0.874	0.020	0.939	0.017	0.906	0.017	0.968	0.006
ambulatory difficulty	0.889	0.013	0.947	0.011	0.918	0.011	0.972	0.004
self-care difficulty	0.801	0.031	0.945	0.029	0.870	0.029	0.955	0.011
independent living difficulty	0.877	0.019	0.940	0.017	0.908	0.017	0.968	0.006
Work disability	0.901	0.007	0.931	0.006	0.915	0.005	0.971	0.002

Source: U.S. Census Bureau, Survey of Income and Program Participation, 2008 Panel

Table 6. Stability of disability items across waves (Wiley and Wiley Method)

Disability Type	W4-W7	SE	W7-W10	SE	Avg/yr	SE	Avg/wave	SE
With a disability	0.922	0.011	0.952	0.010	0.937	0.009	0.978	0.003
hearing difficulty	0.918	0.027	0.907	0.028	0.913	0.024	0.970	0.008
vision difficulty	0.801	0.039	0.882	0.039	0.841	0.032	0.944	0.012
cognitive difficulty	0.868	0.020	0.927	0.020	0.897	0.017	0.964	0.006
ambulatory difficulty	0.908	0.014	0.938	0.014	0.923	0.012	0.974	0.004
self-care difficulty	0.803	0.033	0.873	0.032	0.838	0.030	0.943	0.011
independent living difficulty	0.907	0.021	0.909	0.018	0.908	0.017	0.968	0.006
Work disability	0.905	0.007	0.928	0.006	0.916	0.005	0.971	0.002

Source: U.S. Census Bureau, Survey of Income and Program Participation, 2008 Panel

For the individual disability types, annual stability coefficients ranged from 0.868 (vision difficulty) to 0.918 (ambulatory difficulty). The average per-wave coefficients ranged from 0.954 to 0.972. Work disability had an annual stability coefficient of 0.921 and a per-wave coefficient of 0.973.

4. Conclusions

While the methods used in this paper depend on the assumption that disability follows a first-order autoregressive Markov process – which likely would not hold up to scrutiny – these results provide a significant benchmark in determining the reliability and stability of the 6-question disability measure in the SIPP. Some of the disability items in this survey do demonstrate relatively poor reliability, compared to conventional levels; however, this may not reflect quality of the survey questions directly. Some aspects of the survey in general may contribute to the low reliability.

For many respondents, the SIPP takes a long time to complete and these disability questions appear in a supplemental module that is asked towards the end of the questionnaire.[6] Respondent fatigue may play a large role in reducing the reliability or accuracy of responses. The SIPP also allows for proxy responses; one family member may respond for another who is absent for the interview. Aside from direct proxy misreporting, a second time through the instrument could cause further fatigue. Further analysis could attempt to tease out how much of the reporting error is due to issues associated with proxy response.

Second, topical module questions tend not to use feedback to help inform respondents of past responses. In this analysis, estimates were compared to the work disability question that appears in the core questionnaire and which exhibited higher reliability than the “ACS” style questions. The better reliability may be a function of the fact that the question appears earlier in the instrument and that respondents’ answer are fed back in subsequent interviews. Respondents may not be re-asked the disability question if their health and employment situation were reported as unchanged.

---

[6] Internal Census Bureau memoranda suggest that an adult respondent may take 20 to 35 minutes to complete the core questions in the SIPP before they proceed to the topical module questions.

Despite the low to moderate reliability of the questions, the data also show that over short periods of time – annual measures – disability remains relatively stable. Assumptions made by survey designers to collect disability status at one time and hold the respondents' status constant over several months (such as in the CPS) has validity. Using disability from either wave 4 or wave 7 with other characteristics in, say, wave 6 would not appear to cause egregious harm as respondents' true disability statuses were likely the same. But because disability is not completely stable, researchers should try to minimize the length of time that they hold this assumption. Given the choice to use either wave 4 or 7 disability with wave 6 characteristics, using the disability status from wave 7 would be safer.

5. References

Adler, Michele. "The Future of SIPP for Analyzing Disability and Health." *Journal of Economic and Social Measurement* 18 (1992): 91-124.

Brault, Matthew, Sharon Stern, and David Raglin. *Evaluation Report Covering Disability*. 2006 American Community Survey Content Test Report P.4, Washington, DC: U.S. Census Bureau, 2007.

Cai, Liming, Nathaniel Schenker, and James Lubitz. "Analysis of Functional Status Transitions by Using a Semi-Markov Process Model in the Presence of Left-Censored Spells." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 55, no. 4 (2006): 477-491.

Hardy, Susan E, and Thomas M Gill. "Recovery from Disability Among Community-Dwelling Older Persons." *JAMA* 291, no. 13 (2004): 1596-1602.

Heise, David R. "Separating Reliability and Stability in Test-Retest Correlation." *American Sociological Review* 34, no. 1 (1969): 93-101.

Lee, Chul-joo, Robert Hornik, and Michael Hennessy. "The Reliability and Stability of General Media Exposure Measures." *Communication Methods and Measures* 2, no. 1 (2008): 6-22.

McMenamin, Terence M, Thomas W Hale, Douglas Kruse, and Haejin Kim. "Designing Questions to Identify People with Disabilities in Labor Force Surveys: The Effort to Measure the Employment Level of Adults with Disabilities in the CPS." *JSM Proceedings.* ASA Social Statistics Section, 2005. 1986-1997.

McNeil, John M. *Employment, Earnings, and Disability*. Conference Paper, Washington, DC: Western Economic Association, 2000.

Mendes de Leon, Carlos F, Jack M Guralnik, and Karen Bandeen-Roche. "Short-term Change in Physical Function and Disability: The Women's Health and Aging Study." *Journal of Gerontology: SOCIAL SCIENCES* 57B, no. 6 (2002): S355-S365.

U.S. Census Bureau. *Current Population Survey Design and Methodology*. Technical Paper 66, Washington, DC: U.S. Government Printing Office, 2006.

Verbrugge, Lois M, Joseto M Reoma, and Ann L Gruber-Baldini. "Short-term Dynamics of Disability and Well-being." *Journal of Health and Social Behavior* 35, no. June (1994): 97-117.

Wiley, David E, and James A Wiley. "The Estimation of Measurement Error in Panel Data." *American Sociological Review* 35, no. 1 (1970): 112-117.

Wolf, Douglas A, and Thomas M Gill. *Fitting Event-History Models to Uneventful Data*. TRENDS Working Paper 08-07, Ann Arbor, MI: University of Michigan, 2008.