

# 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 August 2005

Terence M. McMenamin<sup>1</sup>, Thomas W. Hale<sup>2</sup>, Douglas Kruse<sup>3</sup>, Haejin Kim<sup>4</sup>  
Division of Labor Force Statistics, U.S. Bureau of Labor Statistics<sup>1</sup>  
Office of Program Development and Research, Social Security Administration<sup>2</sup>  
Rutgers University<sup>3</sup>  
California State University – Dominguez Hills<sup>4</sup>

## ABSTRACT

In 1998, Executive Order 13078 mandated the development of an accurate and reliable measure of the employment rate of people with disabilities, to be published as frequently as possible. The Bureau of Labor Statistics, in cooperation with the Employment Rate Measurement Methodology interagency workgroup, identified the goal of placing a small set of questions within the Current Population Survey. A set of potential questions was drawn from existing surveys, cognitively tested, and placed in the National Comorbidity Survey (NCS) for a field test. The BLS analyzed the test data to determine the disability status of respondents, using a Dephi process for difficult cases. A variety of tests were used to identify which small set of questions best approximated the results of the initial analysis. The BLS then cognitively tested the set of questions to identify any comprehension problems respondents might have in a CPS context. Further field testing in a CPS environment is being planned to determine the compatibility between the question set and the survey.

## Measuring Employment of the Disabled

The effort to evaluate the employment status of the disabled population is clouded by a lack of consistent, reliable data.<sup>1</sup> Disability data are collected in many surveys, yet the questions used to collect these data are rarely tested to establish the accuracy and reliability of the data obtained. Due to varying constituencies and mandates, more than thirty definitions of disability have been documented in the analyses that are conducted in computing government support programs. It is clear that this is an unusual situation and, unfortunately, it does raise some problems. This wide range of definitions, combined with certain surveys' attempts to measure employment levels, yields disability employment rates that range from 20 to 50 percent. The lack of an official employment level

of disabled persons clearly exacerbates the already contentious issues in research on employment trends among disabled persons.

The U.S. Bureau of Labor Statistics, in cooperation with other federal agencies, has been involved in an effort to design a statistically reliable and accurate measure of the employment rate of adults with disabilities. This effort has spanned many years and has involved extensive testing. This paper will describe the mandate that originated this effort, the standards that were imposed upon the work, and the research process itself. Finally, there will be a discussion of the work that is left, along with the steps that have been taken to complete this effort.

## The Task

On March 13, 1998, President Clinton issued Executive Order 13078 establishing the Presidential Task Force on the Employment of Adults with Disabilities (PTFEAD). This task force was an over-arching organization that provided an operating base for several committees and work groups with mandates specified in the Order. With respect to disability statistics, the Executive Order states:

*The Bureau of Labor Statistics of the Department of Labor and the Census Bureau of the Department of Commerce, in cooperation with the Departments of Education and Health and Human Services, the National Council on Disability, and the President's Committee on the Employment of People with Disabilities shall design and implement a statistically reliable and accurate method to measure the employment rate of adults with disabilities as soon as possible, but no later than the date of termination of the Task Force [September 2002]. Data derived from this methodology shall be published on as frequent a basis as possible. (Clinton, 1998)*

Pursuant to this mandate, the Task Force established the Employment Rate Measurement Methodology (ERMM) Work Group. About 17

Federal Agencies are currently represented on the ERMM Work Group.

### **Identifying the Disability Population**

The problem of identifying the disability population exists in a complex social and technical milieu. From the social perspective, the definitional issues stem from the fact that there are many social constructs and views of disability, each requiring specific, narrowly focused data. The technical problems begin once one settles upon a definition; namely, how to operationalize the definition in a way that will yield accurate and reliable data.

The ADA uses a three-pronged approach to define disability, and the first prong directly relates to the work conducted by BLS. This prong defines a person as having a disability if he or she has a condition that significantly limits them in a major activity of daily living. This definition lacks precision, but one must remember that the ADA is a civil rights act, and its definition was crafted for the purposes of protecting a class of people, and not to serve as the basis for a quantitative measure of the population. While the second and third prongs of the ADA definition have important civil rights purposes, it is unclear what the benefit might be of collecting separate employment data for individuals fitting into these other categories<sup>ii</sup>.

Initial meetings of the ERMM Work Group centered on the potential uses of the data, the definitions of disability, and on the dimensions of disability that would be useful and practical to measure. Some agencies just needed a count of the total number of people with disabilities. Others thought a count of those with severe disabilities was important. Still others would have liked to see individuals with specific types of disabilities identified, such as those with mental disabilities.

An important use to which the agencies might put the data is to track their progress in improving the labor force status of people with disabilities, so it would be helpful if the definition of disability employed by the ERMM enabled the various agencies to measure their disability population of concern.

However, an issue with any approach to defining disability is that various definitions systematically deliver disability populations of

different sizes. Several different counts of the disability population have been published, based on the Survey of Income and Program Participation (SIPP) and the National Health Interview Survey on Disability (NHIS-D). Many consider having a disability count similar to that from another survey a form of validation. However, selecting a definition based upon how large one thinks the number of persons with disabilities ought to be is, of course, completely contrary to the scientific method.

The definition of disability given in the Executive Order states "An adult with a disability is a person with a physical or mental impairment that substantially limits at least one major life activity." As noted above, this definition is the first prong of the ADA definition. It embraces the view that disability is a function of the interaction between an individual with an impairment and his/her environment.

### **Choice of Survey Vehicle**

The primary vehicle for collecting labor force data for demographic groups is the Current Population Survey (CPS), a monthly survey of about 60,000 households conducted for BLS by the Census Bureau. This survey was chosen as the ideal destination for the disability questions for two main reasons. First, since the CPS is a monthly survey, it would satisfy the requirement in the Executive Order to present the data on "as frequent a basis as possible." Secondly, since the CPS is already the official source of labor force data for various demographic groups, it seemed logical that the disabled should be included among these.

The primary mission of the CPS, and the time and space constraints that this imposes on any new questions were (and remain) very important considerations. The key purpose of the CPS is to identify the employed and unemployed. This mission is legislatively mandated, and therefore cannot be compromised. (29 U.S.C., and other public laws specify the collection of these data, and for specific groups such as Hispanics and veterans.) Also, the CPS collects a range of demographic and labor force information, and adding a disability measure would increase the respondent burden. That could result in a decline in the response rate and reduction in the quality of CPS data as a whole. (Technical Paper 63, 2002) In recognition of both of these

restrictions, the ERMM Work Group decided that the question set should be designed to meet the requirements in the Executive Order using as few questions as possible.

### **The Standards**

The Executive Order required that accurate and reliable data be developed. The terms accurate and reliable require a bit of explanation. From a measurement perspective, validity (i.e., accuracy) and reliability are critical. An interpretation of the data requires an understanding of the correspondence between the concept being measured and the degree to which the questions measure the concept.

The term "face validity" means little more than "looks good." Perhaps identifying this concept as construct validity is more to the point. The question looks good because there appears to be a high correspondence between the objective of asking the question and the responses one would get when the question is answered. While this may be useful in the beginning of a question-selection process, there is no objective measure of this type of validity. Face validity may be important for the first pass, but concurrent and predictive validity are empirically verifiable.<sup>iii</sup> Since the Executive Order called for accurate and reliable statistics, research was required to ensure the data had these characteristics.

Although the CPS did not contain questions designed to identify persons with disabilities in a systematic manner, there were several other Federal agencies that did. (Hale, 2001) It was initially hoped that the disability questions contained in the Census Bureau's Survey of Income and Program Participation (SIPP) could be used as a "gold standard" against which potential questions could be measured.

At the first meeting of the ERMM Work Group, the first indication that SIPP might not be a "gold standard" for measuring disability appeared. Census Bureau's disability expert, John M. McNeil, suggested that the SIPP data might not be reliable. He soon provided data that demonstrated the point. Reliability problems cut across questions, even including those that focused on severe disabilities that one would not expect to disappear over the course of a year. It became clear that the SIPP could no longer be considered the "gold-standard." The

ERMM Work Group would have to embark on seminal research in this area.

The research plan evolved as more information became available. Originally, the hope was that BLS would be able to test some existing question sets that had proven to be accurate and reliable, and that identified the correct population. The ERMM had a consultant compile an annotated bibliography to determine if any of the existing disability survey questions had been pre-tested (e.g., cognitive interviews or field tests). This would enable further narrowing of the field of potential questions, and possibly build on previous research. As it turned out, there were virtually no data on the accuracy or reliability of existing question sets; little or no testing had been done on the question sets that were available, and, for those that had been tested, results were often not available from sponsoring organizations. (Furrie, 1999)

Upon considering the testing that would be necessary, it was decided to test individual questions rather than question sets. BLS was aware that some of the sets appeared not to work as well as hoped, and there was still a lot of uncertainty regarding the meaning of the results. Additionally, if questions are tested as sets, the properties of the individual questions would not necessarily be evaluated as well; the researcher would have less flexibility in customizing an instrument from the components of existing questions sets. In a test made to focus on individual questions, if the analysis showed that a question from one set complemented a question from another set, they could be combined. A further difficulty involved the limitations of field testing. If several question sets were to be tested in a field test so the results could be compared, it would require a split-panel test. With each additional question set chosen for testing, the number of respondents who respond to any given set lowers accordingly, affecting the accuracy of the data. Additionally, such a process would only serve to identify the best question set of those tested, rather than an ideal set of questions. Given the lack of data about the available question sets, this might only serve to identify an unreliable or inaccurate question set that happened to perform better than the other faulty question sets tested.

The lack of a gold standard further complicated our testing plans. We had some statistical tools to evaluate questions, but without a gold

standard there was no independent measure of validity. Some flexibility was required. First, BLS identified several surveys from which to draw questions, primarily because their disability questions had face validity, they were nationally known, or their data were widely reported. They included the NHIS-D, SIPP, WHO-DAS, the Census 2000 disability questions, the NOD/Harris poll, and the Behavioral Risk Factor Surveillance System Quality of Life Module (CDC).

While the process of further narrowing the list of candidate questions continued, BLS identified a test vehicle. The questions were selected for inclusion in the National Co-morbidity Survey (NCS). The NCS is a nationally representative survey. This survey is fielded by Harvard School of Health Care Policy with Professor Ronald Kessler as the Principal Investigator. It is primarily funded by the National Institute of Mental Health and, as one might expect, has extensive questions on mental health. It also has many questions on physical well being, and contains the General Assessment Schedule from the Diagnostic and Statistical Manual IV (commonly referred to as the General Assessment of Functioning, or GAF). (National Comorbidity Survey, 2005) The mental health and disability questions in particular were considered important because research conducted for the ERMM Work Group showed that individuals with these disabilities are among the most difficult to identify in a survey.

Because the NCS collected such extensive information, it would enable BLS to compare the respondents' answers to the test questions to the detailed information about their conditions. Using these data, BLS could determine a respondent's disability status using more complete data, and use this knowledge to judge whether a smaller set of questions had likewise correctly identified that respondent's status. The lack of a "gold standard" was no longer a major hindrance.

### **Testing**

Predictably, questions had to be modified to conform to the testing requirements. The questions came from different surveys, contexts, and modes, with different definitional objectives. The minimum length of time a disabling condition would have to have existed, or be expected to exist, needed to be standardized, and

a three-month reference period was selected. Several of the questions were modified to allow a measure of severity and duration.

The candidate questions were identified, and the questions were tested using cognitive interviews. The purpose of the cognitive testing was twofold. First, we wanted to ascertain if respondents had any difficulty understanding the questions or recalling the answers. Second, we wanted to learn what individuals heard and thought when certain questions were put to them. It was essential that the questions avoided difficulties such as:

- attempting to communicate too many concepts;
- covering subjects that were too complex or lengthy;
- probing into areas that were painful or difficult for the respondent.

The questions were subjected to three rounds of testing, each followed by a review to determine if any of the questions needed adjustment. The first round of cognitive testing was conducted at BLS headquarters in Washington DC and the final two rounds were held at the Westat Research Corporation offices in Rockville, Maryland. Even though all of the items included in the original instrument were taken from existing surveys, this cognitive process made several recommendations to change a number of the items prior to fielding the questionnaire in the NCS. (Cantor et al., 2000) The recommended changes were made to the questions, and then they were placed in the NCS.

Analysis of the NCS dataset began in July 2001.<sup>iv</sup> The results from the NCS interviews were analyzed by experts at Rutgers, Harvard, and Indiana University. Between July 2001 and August 2002, in consultation with BLS staff, a series of tables based on analysis of the NCS was created and updated. The final document contained presentations of (1) tabulations of the potential disability identifiers; (2) cross-tabulations of the potential disability identifiers with each other and with a variety of measures of medical and psychological conditions; (3) a summary of cases that were difficult to classify as having a disability; and (4) regressions predicting alternative measures of disability and severe disability.

In the summer of 2002, the above analysis was expanded into a disability classification system,

identifying 24 categories of respondents that could be broadly grouped into those who:

- A) definitely have a disability,
- B) probably have a disability,
- C) possibly have a disability,
- D) are very unlikely to have a disability, and

E) are not worth re-contacting in an attempt to gain further information. (See table 1.)

Table 1. Summary of Disability Classification System

	Severe difficulties for 30 days	Impairment	GAF <=60	Other conditions
<b>A. DEFINITE DISABILITY</b>				
1	Yes	Yes	Yes	
2				Yes to vision or hearing question, GAF <=60
3				Yes to vision or hearing question, GAF over 60
4				Yes to vision or hearing question, no to activity questions or activities not limited for 3 months
<b>B. PROBABLE DISABILITY</b>				
5	Yes	Yes	No	GAF over 60
6	Yes	Yes	No	No to activity questions or activities not limited for 3 months
7	Yes	No	Yes	
8	No	Yes	Yes	No severe chronic condition
9	No	Yes	Yes	Severe chronic condition
<b>C. POSSIBLE DISABILITY</b>				
10	No	No	Yes	Severe chronic condition
11	No	No	Yes	No severe chronic condition
12	No	Yes	No	GAF over 60, no severe chronic condition
13	No	Yes	No	GAF over 60, and severe chronic condition
14	No	Yes	No	No to activity questions or activities not limited, no severe chronic condition
15	No	Yes	No	No to activity questions or activities not limited, severe chronic condition
16	Yes	No	No	GAF over 60
17	Yes	No	No	No to activity questions or activities not limited
18	No	No	No	Seen by others as having a disability
<b>D. VERY UNLIKELY TO HAVE DISABILITY</b>				
19	No	No	No	No to activity questions or activities not limited, no severe chronic condition
20	No	No	No	No to activity questions or activities not limited, severe chronic condition
21	No	No	No	GAF over 60, no severe chronic condition
22	No	No	No	GAF over 60, severe chronic condition
23				Duration less than 3 months
<b>E. NOT WORTH CONTACTING</b>				
24				Duration or GAF questions refused

The categories referred to in table 1 were constructed to separate respondents according to the likelihood that they have disabilities, and to identify those for whom additional information and review would be needed. The placement of respondents into categories was based on answers to the following groups of questions

(associated terminology used in table 1 are included in italics):

- 1.) seven types of activity limitations, building upon the disability questions used in the 2000 Census<sup>v</sup> (yes/no questions). [*activity questions*];
- 2.) seven specific impairments and disabilities<sup>vi</sup> (yes/no questions). [*impairment*];

- 3.) nineteen types of functional difficulties over the past 30 days<sup>vii</sup> (on a 5-point scale measuring degree of difficulty, with 1="none", 2="mild", 3="moderate", 4="severe", and 5="cannot do"). [*severe difficulties for 30 days*];
- 4.) the global assessment of functioning scale<sup>viii</sup> (on a 0-100 scale with 100= "excellent functioning in all areas of life" and 0="unconscious"). [*GAF <=60*];
- 5.) eighteen chronic conditions, of which one was selected randomly for questions about the degree of related interference with life activities on a 1-10 scale.<sup>ix</sup> [*severe chronic condition*] (Kruse, 2002)

In consultation with BLS staff, the information contained in the 24 categories was used to classify respondents according to the likelihood of disability. Based on the classification system, 540 respondents whose disability status was difficult to determine due to inconsistencies in their answers were selected for further review. Of these, 352 had their disability status determined by project staff, and 100 of the remaining respondents were successfully re-contacted to gain further information. (See table 2.) The data collected through the re-interviews were combined with the NCS data to create a more complete profile for each respondent.

Table 2. Summary of Disability Status Determination

	Total N	N pulled for further review	Status determined by staff	Re-interviewed and used in Delphi process	Percent identified as disabled
<b>A. DEFINITE DISABILITY</b>					
1	128	75	75	0	100.0
2	110	0	0	0	100.0
3	94	12	12	0	100.0
4	71	12	12	0	100.0
<b>B. PROBABLE DISABILITY</b>					
5	75	75	55	11	97.4
6	5	5	4	1	47.5
7	9	9	6	1	100.0
8	145	118	80	15	89.2
9	75	23	18	2	100.0
<b>C. POSSIBLE DISABILITY</b>					
10	22	7	4	3	63.3
11	72	22	6	7	61.5
12	252	71	47	16	71.4
13	98	27	21	5	86.9
14	205	55	3	28	19.0
15	40	11	1	5	40.7
16	5	5	3	1	50.0
17	14	3	3	0	50.0
18	10	10	2	5	28.6
<b>D. VERY UNLIKELY TO HAVE DISABILITY</b>					
19	1,694	0	0	0	0.0
20	161	0	0	0	0.0
21	187	0	0	0	0.0
22	42	0	0	0	0.0
23	268	0	0	0	0.0
<b>E. NOT WORTH CONTACTING</b>					
24	48	0	0	0	37.4
<b>TOTAL</b>					
	3830	540	352	100	

In the fall of 2003, the Delphi process was used to evaluate the revised profiles. The Delphi process is a technique that was designed to obtain the most reliable consensus amongst a group of experts by a series of questionnaires interspersed with controlled feedback. Essentially, the Delphi process involves submitting a group of cases to a selected team. Each team member evaluates the group of cases, and submits judgement and commentary on each case. In successive rounds, team members review the comments of fellow team members and use this input to decide whether or not they are persuaded to alter their original judgement. This process is repeated in successive rounds to achieve concordance.<sup>x</sup>

There were five Delphi teams, each with ten original members. Team members participated upon invitation by BLS staff, and most participants had a professional or personal interest in the study of disability. Each team received 20 cases to review in the first round. In the second round, all 20 profiles were sent back to each of the participants. Those who had not responded to round 1 were asked to review all of the cases, while those who had responded were only required to provide feedback for the cases where concordance had not been achieved. (Concordance was defined as agreement among 75 percent or more members of a group.) Any comments provided in the first round were sent to all participants in the second round for consideration. In the third round, only those cases which had not reached concordance were sent to the participants for review. After three Delphi rounds, concordance had been achieved in all but 10 cases.<sup>xi</sup>

### **Analysis of NCS Data**

After establishing the disability status of the respondents, statistical analysis was conducted to determine which potential disability identifiers most appropriately and efficiently reflect the sample of people with disabilities. Thirty-three questions were used as potential disability identifiers: seven activity limitation questions, seven questions about specific impairments and disabilities, and nineteen questions about functional difficulties in the past 30 days.

The analysis used three basic techniques to ascertain the best predictors of disability status:

- 1) stepwise regression;

- 2) highest R-squared regressions using all 5-, 6-, and 7-question sets; and

- 3) a combinatorial approach, comparing the classification accuracy of all 5-, 6-, and 7-question sets.

The stepwise regressions included both forward selection and backward selection models, with both 95% and 99% levels of statistical significance, in order to test for consistency in identifying significant predictors of disability status. The highest R-squared regressions compared the collective predictive power of different combinations of predictors, examining which 5-, 6-, and 7-question sets collectively explained the greatest variance in disability status. Finally, the combinatorial approach examined how different combinations of question sets performed when measurement is based on an affirmative response to one or more questions in the set.

In order to more accurately identify people with disabilities and minimize the overcount, the question sets were tested to determine the optimal number of positive responses that would be needed to indicate a high probability of a disability. Receiver operator characteristic (ROC) analysis was used for this, which is a system designed to quantify the accuracy of diagnostic tests in distinguishing between two states or conditions. After removing those with vision or hearing impairments (who are automatically assigned to disability status), the results showed that nearly all question sets require two positive answers for the most accurate measure of disability. Since the optimal prediction of disability status will often include a more complex configuration of positive answers, the results were probed using classification tree analysis (with the CART program). This analysis showed that the optimal configurations of questions were almost identical to the above algorithm (a “yes” to hearing or vision impairment, or at least two of the other questions in the set), and the performance of the question sets was very similar, so the simpler algorithm was employed.

Once the most promising question sets were identified, the third stage of analysis involved examining how the overcount and undercount varied by demographic group (age-sex cohorts and race) and employment status. This was examined both with simple calculations of overcount and undercount in each of the subgroups, and with logistic and OLS

regressions of overcount and undercount on the demographic and employment variables.

The performance of the best sets was evaluated in the following areas:

- a quality index defined as  $100 - \text{undercount} - (2 * \text{overcount})$ ;
- overall accuracy;
- percentage identified as disabled;
- overcount<sup>xii</sup>;
- undercount;
- overcount by demographic characteristics;
- undercount by demographic characteristics.

There was no question set that outperformed all the others in all of these areas of consideration. However, it was possible to select question sets that performed consistently well in most, or all areas. The question set that was eventually identified as the best performed well in all of the areas of consideration. Additionally, any question set that outperformed it in any specific area was invariably found to be inferior in a number of other focus areas.

The best question set contains seven questions that correctly identify the disability status of 94.4 percent of the population. Disabled persons account for 17.6 percent of the population with this question set, as compared with 19.8 percent with the full data set. The overcount is 9.6 percent of those identified as disabled by the question set. The undercount is 19.8 percent of persons identified by the full dataset as having a disability.

### **The Question Set**

Does anyone in this household who is 15 years old or over have:

A hearing problem that prevents them from hearing what is said in normal conversation even with a hearing aid?

A vision problem that prevents them from reading a newspaper even when wearing glasses or contacts?

Any condition that substantially limits one or more basic physical activities, such as walking, climbing stairs, reaching, lifting, or carrying?

Any other physical disability?

Any emotional or mental disability?

Because of a physical, mental or emotional condition lasting three months or longer, does anyone in the household who is 15 years old or over have difficulty doing any of the following:

Learning, remembering or concentrating?

Participating fully in school, housework, or other daily activities?

### **Final Cognitive Interviews**

The BLS designed a cognitive test for the question set that had been identified. The test was designed to address three areas of concern:

1.) It was important to test the set of questions in the context of the CPS instrument. For example, how well do the questions work within CPS; are there any problems of transition or flow from the original CPS questions to the new set of disability questions?

2.) The second concern was how the set of questions work together. Since the questions were selected from a larger set of questions, they had not yet been tested as a set to see how well they function. The questions were tested for clarity, comprehension, length, and any repetitiveness among questions.

3.) Lastly, it was necessary to determine the best way to administer the questions, i.e., whether each question should be asked at the individual person level or at the household level.

There were 17 participants in the test, and people with and without disabilities were included, as well as proxy respondents. The results showed that there were some minor conceptual issues with certain questions, but on the whole, the question set was clearly understood by most participants. No problems due to the inclusion of the questions in a labor force survey were identified, although the sentence providing the transition to the disability questions needed some adjustment. The cognitive test did not reveal any issues that were unique to either the household or individual version of the question set, so in the interest of brevity the household version was selected. (Rho, 2004)



## Split-Panel Test

When BLS felt that a question set had been identified through research that satisfied the requirements of the Executive Order, these questions and the research were presented for assessment to the ERMM interagency workgroup and other interested parties. One concern that was raised was that there might be an element of sample bias that had not been discovered through testing. In an effort to address this concern, the National Comorbidity Survey (NCS) sample was split in half randomly, then all possible question sets containing 5 to 7 questions were analyzed using one half of the data. Two possible classification paradigms were considered: 1) the assignment of disability status based on a “yes” to any question in the set, and 2) the assignment of disability status if there is a “yes” to the vision or hearing question, or to two of the remaining questions. The results were compared to an analysis of the other half of the data. There were two specific hypotheses that this test was designed to address:

### *Hypothesis 1*

It can be shown by this method that there is a better question set than the one which has already been identified.

### *Hypothesis 2*

The chosen question set displays inconsistent results when the data from each half of the NCS are compared.

One half of the respondents were randomly selected, and the combinatorial approach was used to identify the 50 most promising question sets from this half of the dataset (using percent correctly classified, overcount, undercount, and the index with a double penalty for overcount). A complete analysis of overcount and undercount was done for these 50 sets, and the BLS staff chose the best 20 of these question sets to be analyzed with the other half of the dataset. The results for both halves of the dataset were then compared to determine quality and consistency of the question sets in identifying disability status.

Insufficient evidence was found to support acceptance of either of the hypotheses. Based on these results, BLS decided to proceed with testing of the question set within the actual CPS instrument. (McMenamin, 2005)

## Future Testing

Whenever a change to the CPS is considered, one of the main concerns of the agencies involved is to ensure that there are as few adverse affects to the response rate as possible. In order to determine what effect the addition of a set of disability questions might have, a joint effort between BLS, the Census Bureau, the Office of Disability Employment Policy (ODEP), and the National Institute on Disability and Rehabilitation Research (NIDRR) is underway to test the questions in a supplement to the CPS in February 2006. This is the final test planned for these questions. The results of this study will determine whether the question set will be placed in the CPS. This supplement will contain the disability question set, which will be asked directly after the end of the regular CPS questions. It will be a split-panel test, administered to the majority of CPS rotation groups. The test will enable BLS to measure the effects the addition of the disability question set will have on CPS response rates, as well as to determine a disability rate which will be compared with results from prior testing in the NCS.

---

<sup>i</sup> Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics or the Social Security Administration. Thomas W. Hale's contributions to this paper were completed while he was employed at the Bureau of Labor Statistics.

<sup>ii</sup> Under the ADA, an individual with a disability is a person who:

1. has a physical or mental impairment that substantially limits one or more major life activities;
2. has a record of such an impairment; or
3. is regarded as having such an impairment.

<sup>iii</sup> Predictive validity evaluates a concept's ability to measure what it is designed to measure. Concurrent validity evaluates a concept's ability to distinguish the difference between groups it has been designed to identify.

<sup>iv</sup> The NCS was underway at the time the analysis began, so the initial analysis was based on the 3,458 observations that were available. A second wave of interviews became available in January 2002, increasing the sample size to 6,318, and a third wave became available in March 2002, increasing the total sample size to 7,403.

---

<sup>v</sup> The NCS has a total of 42 sections devoted to mental or physical disabilities. The questions used for this analysis were included in the first section ( Screener—SC) and the twentieth section (30-Day Functioning and Disability--FD). The first set of potential disability identifiers used in this effort asks about several types of activity limitations, building upon the disability questions used in the 2000 Census. These are in the first (“SC”) section of the NCS:

10.1 Because of a physical, mental, or emotional condition lasting three months or longer, do you have any difficulty in doing any of the following activities:

10.1a. Learning, remembering, or concentrating?

10.1b. Getting along with people?

10.1c. Dressing, bathing, or getting around inside your home?

10.1d. Going outside your home alone to shop or visit a doctor’s office?

10.1e. Working at a job or business?

10.1f. Participating fully in school, housework, or other daily activities?

Respondents who answered “no” to all of the above were then asked:

10.1h. Have you been limited in any way for the past three months because of any impairment or health problem?

<sup>vi</sup> The second set of potential disability identifiers, also in the SC section of the NCS, contains questions about specific impairments and disabilities. All of those who answered “yes” to any of the above activity limitation questions, and a random sample of those who answered “no” to all of those questions, were then asked:

10.4. Do you have any of the following conditions:

10.4a. Blindness, deafness, or a serious speech problem?

10.4b. A hearing problem that prevents you from hearing what is said in normal conversation even with a hearing aid?

10.4c. A vision problem that prevents you from reading a newspaper even when wearing glasses or contacts?

10.4d. A learning disability of any kind?

10.4e. Any other physical handicap or disability?

10.4f. Any other condition that substantially limits one or more basic physical activities, such as walking, climbing stairs, reaching, lifting, or carrying?

10.4g. Any emotional or mental disability?

---

<sup>vii</sup> The third set of questions containing potential disability identifiers is in the FD section of the NCS, asking about any functional difficulties in the past 30 days. The questions identified 19 types of potential functional difficulties, arranged into five categories. The first four categories had an initial screener question asking whether the respondent had any of several types of health-related difficulties in the past 30 days (FD10, FD12, FD14, and FD16). Those who answered “yes” were asked how many days they experienced those difficulties, and how much difficulty they had for each of the several types of activities, measured on a 5-point scale with 1=“none”, 2=“mild”, 3=“moderate”, 4=“severe”, and 5=“cannot do”. The first 15 types of functional difficulty are:

11a. Concentrating on doing something for ten minutes?

11b. Understanding what was going on around you?

11c. Remembering to do important things?

11d. Learning a new task--for example, learning how to get to a new place?

13a. Standing for long periods, such as 30 minutes?

13b. Moving around inside your home?

13c. Walking a long distance such as (a kilometer/half a mile)

15a. Washing your whole body?

15b. Getting dressed?

15c. Staying by yourself for a few days?

17a. Starting and maintaining a conversation?

17b. Dealing with people you did not know well?

17c. Maintaining friendships?

17d. Making new friends?

17e. Controlling your emotions when you were around people?

Those who gave answers indicating serious limitations were then asked the following four questions, with the same 5-point scale for answer options:

18a. During the past 30 days, how much health-related difficulty did you have in taking care of your household responsibilities?

18b. How much were you emotionally affected by your health during the past 30 days?

18c. How much was your health a drain on the financial resources of you or your family in the past 30 days?

18d. During the past 30 days, how much of a problem did you have in joining community activities, for example, festivities, religious or

---

other activities in the same way as anyone else because of your health?

Respondents were asked 18a-18d if they (a) gave two or more responses of "moderate, severe, or cannot do" regarding the above types of functional difficulties, (b) reported three or more days in the past thirty of staying in bed or being unable to work or carry out normal activities, or (c) reported having severe physical discomfort for all or most of the past thirty days.

<sup>viii</sup> Those who answered "yes" to one of the SC10.1 questions about activity limitations were asked about the nature of the condition causing such an activity limitation and the length of time their activities have been limited. Those who said their activities have been limited for three months or more were then asked question SC10.14:

The 0-to-100 scale on page 2 in your booklet describes amounts of activity limitation. Using this scale, what number describes how much your activities have been limited because of your health problems during the past 3 months?

- 100 Excellent functioning in all areas of life
- 90 Good functioning in all areas of life
- 80 Slight difficulty
- 70 Some difficulty
- 60 Moderate difficulty
- 50 Serious impairment in one area
- 40 Serious impairment in more than one area
- 30 Unable to function in most areas
- 20 Difficulty with basic needs
- 10 Unable to meet basic needs
- 0 Unconscious

Those who gave scores of 60 or below on this scale were judged to have high likelihoods of having disabilities.

<sup>ix</sup> Respondents were asked if they had ever experienced any of 18 chronic conditions. For those who had, one of the conditions they reported was selected randomly by the interview program, and the respondent was then asked to think of the month or longer in the past twelve months when the condition or its consequences were most severe. The respondent was asked to report, on a 0 to 10 scale, how much interference this condition caused during that time with a) home management, b) ability to work, c) ability to form and maintain close relationships with other people, and d) social life. In line with the wording represented on the scale, those who gave answers of 7 or more on any of those four activities were designated as having a severe chronic condition.

---

The eighteen conditions, in questions CC1a-t, are: arthritis or rheumatism, chronic back or neck problems, frequent or severe headaches, any other chronic pain, seasonal allergies like hay fever, stroke, heart attack, heart diseases, high blood pressure, asthma, tuberculosis, other chronic lung disease like COPD or emphysema, diabetes or high blood sugar, ulcer in stomach or intestine, HIV infection or AIDS, epilepsy or seizures, or cancer.

<sup>x</sup> The Delphi process has four necessary features:

1) *Anonymity* is achieved through the use of questionnaires. By allowing group members to consider and answer their replies privately, undue social pressures should be avoided. In some instances it may be appropriate for the members of the Delphi group to be identified. However their answers will be anonymous, i.e. the individuals' answers are anonymous even if the participants themselves are not.

2) *Iteration* occurs through the submission of a questionnaire over a series of rounds, allowing members to change their opinions.

3) *Controlled feedback* occurs between rounds. The results of each round are analysed by a central researcher and the responses for each given statement are fed back to all members of the Delphi group. This allows members of the group to assess their views in the light of the group's responses.

4) *Statistical aggregation of group response* is obtained at the end of the procedure. This is an expression of the degree of consensus of the group on a particular issue. It is commonly expressed as a mean value and spread of opinion, which can be combined to indicate the "strength" of opinion.

<sup>xi</sup> To measure disability status for the 10 split decisions in the Delphi cases, respondents were assigned the percentage of Delphi panel members who voted in favor of disability. To measure disability status for the 48 respondents who refused to give necessary information, the likelihood of a disability was imputed based on probit regressions using other activity limitation, impairment, and chronic condition variables. Tests were conducted both with and without these imputations. There were 88 respondents whose status could not be determined by BLS staff but who were not part of the Delphi process. Corresponding weights were assigned to the Delphi cases in those categories to account for these 88 respondents.

---

<sup>xiii</sup> In the interest of clarity, the terms overcount and undercount used in this paper refer to the following concepts: Overcount—the portion of persons who were identified as *not disabled* using the full set of information collected via the NCS, but were identified as *disabled* using the short question set. Undercount-- the portion of persons who were identified as *disabled* using the full set of information collected via the NCS, but were identified as *not disabled* using the short question set.

---

Rho, C. (2004), “CPS Disability Questions: Summary of Findings from Cognitive Interviews,” unpublished technical report, U.S. Bureau of Labor Statistics.

*National Comorbidity Survey* (2005), retrieved May 25, 2005, from the Harvard School of Medicine Web site: <http://www.hcp.med.harvard.edu/ncs/>

*Technical Paper 63: Current Population Survey-Design and Methodology* (2002), U.S. Bureau of Labor Statistics, U.S. Census Bureau.

## References

Cantor, D., Kudela, M. S., Robbins, C., Strickland, D. (2000), “Pre-Testing Disability Questions for the Current Population Survey,” unpublished technical report, U.S. Bureau of Labor Statistics.

Cherry Engineering Support Services, Inc. (CESSI), (2003), “Federal Statutory Definitions of Disability,” retrieved May 25, 2005, from the Interagency Committee on Disability Research’s Web site: <http://www.icdr.us/documents/definitions.htm>

Clinton, W. J. (1998), “Executive Order 13078: Increasing Employment of Adults With Disabilities,” retrieved May 25, 2005, from the U.S. Department of Labor’s Web site: [http://permanent.access.gpo.gov/lps9586/www.dol.gov/dol/\\_sec/public/programs/ptfead/2000rpt/execorder.htm](http://permanent.access.gpo.gov/lps9586/www.dol.gov/dol/_sec/public/programs/ptfead/2000rpt/execorder.htm)

Furrie, A. (1999), “Summary Review of Disability Survey Questions,” unpublished technical report, U.S. Bureau of Labor Statistics.

Hale, T. W. (2001), “The Lack of a Disability Measure in Today’s Current Population Survey,” *Monthly Labor Review*, Vol. 124 No. 6, 38-40.

Kruse, D. (2002), “NCS Groups to Reinterview,” unpublished technical report, U.S. Bureau of Labor Statistics.

McMenamin, T. M. (2005), “BLS’ Split-Panel Test of National Comorbidity Survey Data,” unpublished technical report, U.S. Bureau of Labor Statistics.