

**DOCUMENTATION OF AFFECTIVE FUNCTIONING MEASURES
IN THE HEALTH AND RETIREMENT STUDY
(HRS/AHEAD)**

Working Paper Version

Diane E. Steffick¹

¹ University of Michigan

TABLE OF CONTENTS

I.	Rationale for Measuring Affective Functioning and Depression	2
II.	Inventory of Measures Including Variable Construction	4
	The HRS Depression Symptoms Measure: The Shortened CES-D Summary Scores	4 5
	The Short Form Composite International Diagnostic Interview (CIDI-SF) Used In HRS/AHEAD Summary Scores	6 7
	Self-Rated Emotional Health and Self-Reported Psychological Diagnosis	8
III.	Origin of the CES-D and CIDI-SF	9
	The CES-D Scale	9
	The CIDI-SF for Major Depressive Episodes	9
IV.	Special Methodological Issues	10
	Age-Eligibility	11
	HRS Wave 1 Imputations	12
	Proxy Respondents	12
	Determining Caseness	13
	The CES-D Measures	13
	The CIDI-SF Measure	15
V.	Comparability of the Shortened CES-D Measures Across HRS and AHEAD Waves	15
	Wording Changes Between Waves 1 and 2 of HRS	16

VI.	Evaluation of the Measures and Data Quality	19
A.	Univariate Distributions	20
	The CES-D Measure	20
	The CIDI-SF for Major Depressive Episodes	21
B.	Benchmarking Against Other Surveys and Prevalence Estimates	22
	Finding an Equivalent to “16+” for the HRS Wave 1 Measure: A More Rigorous Approach	22
	Determining a “16+” Equivalent for the 8-Item CES-D Scale	25
	The CIDI-SF for Major Depressive Episodes	26
C.	Internal Consistency and Measurement Properties	26
D.	Constructed Variables	28
E.	Item Non-Response / Missing Data	28
F.	Construct Validity	30
VII.	Conclusion	32
	REFERENCES	33
	Figures	
	Tables	
	Appendix Tables	
	Technical Appendix with SAS Code	

LIST OF FIGURES

		<u>Page</u>
Figure 1:	The CIDI-SF for Major Depressive Episodes in the HRS and AHEAD	6
Figure 2:	Comparison of HRS Wave 1 Women and NLS Mature Women Using Common CES-D Score From HRS Wave 1: Both Weighted, Ages 52-61 Only	23
Figure 3:	NLS Mature Women Distribution of HRS Wave 1 CES-D Measure By Whether Full Score Is Above or Below 16+ Cutoff: Weighted, Age 52-61 Only	23

LIST OF TABLES

Table 1:	Question Wording for CES-D	5
Table 2:	Affective Functioning Questions Asked in HRS & AHEAD: The CES-D	5
Table 3:	Affective Functioning Questions Asked in HRS & AHEAD: The Composite International Diagnostic Interview – Short Form (CIDI-SF) for Major Depressive Episodes	7
Table 4:	Affective Functioning Questions Asked in HRS & AHEAD: Self-Reported Measures	8
Table 5:	Age-Eligibility for HRS and AHEAD	11
Table 6:	Proxy Respondents in HRS/AHEAD	12
Table 7:	HRS Wave 2: Comparing CES-D Items Across Response Formats	17
Table 8:	Comparison of Frequency Response and Yes/No Response: Module Respondents, Both Formats Asked During Same Interview (1994)	17
Table 9:	HRS and AHEAD: Between-Wave Changes Using the CES-D Measure	18
Table 10:	HRS and AHEAD CES-D Summary Scores: Two-Year Change, Direction and Magnitude	19

	<u>Page</u>
Table 11: Summary CES-D Scores in the HRS: Frequency Response HRS Wave 1 (1992) Respondents	20
Table 12: Summary CES-D Scores in the HRS and AHEAD: Yes/No Response	21
Table 13: CIDI-SF Scale for Major Depressive Episodes: HRS Wave 3 (1996) and AHEAD Wave 2 (1995) — I	21
Table 14: CIDI-SF Scale for Major Depressive Episodes: HRS Wave 3 (1996) and AHEAD Wave 2 (1995) — II	22
Table 15: NLS Mature Women, Ages 52-61, 1989	23
Table 16: Comparison of HRS Wave 1 CES-D With Other Surveys	25
Table 17: Prevalence Estimates for HRS Waves 2 & 3 and AHEAD Waves 1 & 2	26
Table 18: Reliability and Factor Analysis for the 11 CES-D Items in Wave 1 of the HRS	26
Table 19: Psychometric Properties of the CES-D in the HRS and AHEAD: Results From Factor Analysis and Internal Consistency Investigations	27
Table 20: HRS and AHEAD Item Non-Response: CES-D Based Measure	28
Table 21: HRS and AHEAD Item Non-Response: CIDI-SF For Major Depressive Episodes	29
Table 22: CES-D Construct Validity: Bivariate Relationships	30
Table 23: OLS Regression Model of Depression Using HRS Wave 1 CES-D Score (11 Items, Frequency Response)	31

APPENDIX TABLES

Table A1:	HRS: Comparing CES-D Items Across Response Formats
Table A2:	Comparison of Frequency Response and Yes/No Response: Module Respondents, Both Formats Asked During Same Interview (1994)
Table A3:	CES-D Summary Scores: Comparison of HRS Wave 2 (1994) and HRS Wave 3 (1996) Scores
Table A4:	CES-D Summary Scores: Comparison of AHEAD Wave 1 (1993) and AHEAD Wave 2 (1995) Scores
Table A5:	Summary CES-D Scores in the HRS: Frequency Response For Module Respondents Only
Table A6:	Frequencies of Individual CES-D Items in the HRS: Frequency Response
Table A7:	Constructed Variables on AHEAD Wave 1: CESD8
Table A8:	CES-D Construct Validity: Cross-Tabulations With Significant Variables

TECHNICAL APPENDIX

	<u>Page</u>
HRS Wave 1	T-1
HRS Wave 2	T-2
HRS Wave 3	T-4
AHEAD Wave 1	T-8
AHEAD Wave 2	T-10

Documentation of Affective Functioning in the Health and Retirement Study (HRS/AHEAD)

This paper is one in a series of working papers that provides background information and documentation of several substantive areas of the Health and Retirement Study (HRS/AHEAD). The Health Working Group of the HRS project is in the process of providing similar papers on the assessment of cognitive functioning, physical functioning (including ADLs, IADLs, Nagi items), clinical conditions and their severity, and alcohol use.

This paper documents the HRS measures of mental health status, also called affective functioning. It is intended for a wide range of users, particularly those who would like to include mental health in their research but are not familiar with the measures available in the HRS. It also addresses topics of interest to mental health specialists who are concerned with how the abbreviated scales in the HRS correspond to the full scales used in the existing literature. This paper provides users with an overview of the measures included in the study, including their origins and pedigree. It details the specific steps used to create standard summary variables for the scales measuring depressive symptoms and major depressive episodes, and includes SAS code for this variable creation in a Technical Appendix¹.

Also included in this working paper is an examination of data quality. We provide univariate distributions of the HRS measures and compare these to other published studies, and also perform standard psychometric evaluations for reliability (the coefficient alpha) and exploratory factor analysis. The paper shows key bivariate relationships that support construct validity **C** showing that the HRS measures relate to several other respondent characteristics in the manner one would expect; for example, individuals with higher depressive symptoms are also less satisfied with their jobs and marriages than those with fewer symptoms of depression. It also addresses methodological issues specific to the mental health measures such as the change in the question wording of the depression symptoms scale between Wave 1 and Wave 2 of the HRS.

The paper is structured as follows. Section I provides a discussion of affective functioning in general and describes the decisions which led to the selection of specific measures for the HRS. Section II provides an inventory of the measures for the first three waves of HRS (the cohort aged 51 to 61 in 1992) and first two waves of AHEAD (the Asset and Health Dynamics Among the Oldest Old study, a cohort representative of those age 70 and older in 1993) and describes summary variable creation. Section III details the origins of the measures. Section IV outlines special methodological issues with the depression measures with subsections on age-eligibility, proxy respondents, HRS Wave 1 imputation, and determining caseness.

¹SAS is a statistical software package licensed by the SAS Institute.

Section V assesses the comparability of the measures across the HRS waves (1992, 1994, and 1996) and AHEAD waves (1993 and 1995). Section VI provides the results of investigations into the quality of the measures, using data from the first few waves of each survey (HRS/AHEAD). This section included subsections on univariate distributions, benchmarking against other surveys and prevalence measures, internal consistency and measurement properties, constructed variables, item nonresponse and missing data, and construct validity. The Appendix contains additional tables, which may be of interest to the reader, while the Technical Appendix contains the SAS code for constructed summary variables.

1. Rationale for Measuring Affective Functioning and Depression

Affective functioning, a person's mood and emotional health, is an important component of an individual's overall health and wellbeing. Research has shown that symptoms of depression and anxiety have a significant impact on a person's performance in the labor market. Individuals with poor affective functioning are less attached to the labor force, work fewer hours, have more work loss days, and earn lower wages (for example, see Ettner et al. (1997), Broadhead et al. (1990), and Steffick (1998)). Consequently, a person's emotional health is likely to affect his or her decision to retire.

Affective functioning is also related to physical health in many ways. A high level of psychological distress (symptoms of depression) is associated with poorer physical health. Research has shown that the direction of causation goes both ways. On the one hand, depression increases the risk of physical disease. For example, Barefoot and Schroll (1996) found that individuals with high levels of depression during their baseline observation were significantly more likely to experience a myocardial infarction (heart attack) than those with low levels of depression, even after controlling for baseline cardiovascular function. On the other hand, many physical illnesses and functional impairments have the effect of increasing depression and anxiety. For example, Palomäki et al. (1999) found that the prevalence of depression in a sample of ischemic stroke survivors increased with time from 6 percent at the initial stage, to 11 percent at one year, and 18 percent at eighteen months.

For these reasons, the Health and Retirement Study (HRS/AHEAD) includes affective functioning as one of its health domains. The remainder of this section describes the definition of affective functioning, introduces the measures, and provides the reasons behind their selection.

Affective functioning is not a well-defined concept that is easily quantified. It has many dimensions and not all of them can be measured well in a survey. A comprehensive inventory of psychological traits and disorders was prohibited by the nature of the survey and its time constraints. Therefore, one task facing the designers of the HRS/AHEAD was the selection of mental health concepts to comprise the affective functioning inventory from the wide range of alternatives C aspects of personality or self-esteem; general levels of stress or psychological distress; or recognized disorders such as major depression, schizophrenia, or generalized anxiety disorder (to name just a few).

The most direct method of characterizing affective functioning included in HRS/AHEAD is a set of questions asking the respondent to report about their own emotional health directly. This includes a global rating of emotional health as excellent, very good, good, fair, and poor. Respondents are also asked directly about receipt of a diagnosis from a doctor of an emotional or nervous problem and, whether he/she has sought treatment for this problem such as therapy or medication. This allows the respondent to acknowledge any mental health problem for which they have sought care from a health-care professional. However, the fact that these diagnosis-based questions are influenced by the decision to seek care, which is in turn influenced by economic considerations (insurance coverage, ability to pay) makes them undesirable for many research studies. To avoid this problem, the HRS/AHEAD sought to include more comprehensive measures of affective functioning, which are asked of all self-respondents.

It was decided that these measures would include one inventory of psychological distress and one measure of major depressive episodes. The measure of psychological distress, or symptoms of depression, consists of a subset of items from the Center for Epidemiologic Studies Depression Scale (CES-D)². The CES-D scale measures a continuum of psychological distress (symptoms of depression and anxiety), rather than determining the presence or absence of recognized psychiatric disorders. Fechner-Bates, Coyne, and Schwenk (1994) provide results of research on the relationship of the CES-D to both depressive disorders and anxiety disorders, as well as other psychiatric diagnoses.

In the third wave of interviews for the HRS and the second wave for AHEAD, a short form of the World Health Organization's Composite International Diagnostic Interview (CIDI-SF) was administered. This scale determines a probable diagnosis of the psychiatric condition known as a major depressive episode, as defined by the *Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association*, third edition revised (DSM-III-R)³. The CIDI-SF elicits information relating to symptoms and duration as spelled out in the DSM-III-R and can be used to identify individuals whose depressive symptoms would be recognized as an illness by mental health practitioners.

The HRS/AHEAD focused on depression since it is the most prevalent psychiatric disorder among the elderly (Reiger et al., 1988) and may be quite disabling (see Wells et al., 1989, for example)⁴. The choice of these specific measures of affective functioning was subject to the constraints of the survey format. Structured interviews by mental health providers were not feasible, so the instruments had to be designed for administration by interviewers with no

²Wave 1 of the HRS included 11 items from the 20-item CES-D while all subsequent waves of the HRS and all waves of the AHEAD study contain 8 items.

³The HRS/AHEAD study uses a version of the CIDI-SF that implements DSM-III-R criteria. Later versions of the CIDI-SF have operationalized the DSM-IV criteria (Nelson et al., 1998).

⁴Reiger et al. (1988) found phobia and cognitive impairment to have the highest one-month prevalence rates among individuals age 65 or older, followed by the combination of depression and dysthymia. However, phobias tend to be very specific and do not often limit daily activities, so HRS/AHEAD chose not to ask about phobia. Cognitive impairment is evaluated in a separate section of the survey. See Ofstedal (1999) for details on HRS/AHEAD coverage of cognitive functioning.

advanced training in psychology or psychiatry. The instruments also had to be reliable when administered over the telephone, as well as in person. Both the CES-D and the CIDI-SF have been shown to be reliable in interviewing environments similar to the HRS/AHEAD (Ross and Mirowsky, 1984; Mroczek and Kessler, 1994).

An additional goal was to ensure comparability between the HRS/AHEAD and other national surveys. In addition to the CES-D, there are dozens of inventories of depressive symptoms used in psychiatric and psychological research. Many of these contain questions that are very similar to the CES-D. Ultimately, the CES-D was chosen because of its widespread use, which both attests to its reliability and validity for a variety of subpopulations, and allows for greater comparability of the HRS/AHEAD with existing research. The CES-D has been used in the National Health and Nutrition Examination Survey (NHANES), the Established Populations for Epidemiologic Study of the Elderly (EPESE), the National Longitudinal Surveys (NLS-Mature Women, NLS-Older Men, NLSY), and the Americans' Changing Lives study (ACL). The long form of the CIDI was used in the National Comorbidity Survey (NCS) and the Short Form was developed for use in the National Health Interview Survey (NHIS).

II. Inventory of Measures Including Variable Construction

This section of the working paper details the individual items that make up the mental health measures in the HRS and shows the user how the items were operationalized as questions in the respondent's interview. It notes changes between waves of the survey and includes tables listing the individual items for quick reference. It also details the most commonly used methods of combining the individual items into a summary score.

The HRS Depression Symptoms Measure: The Shortened CES-D

As mentioned previously, the HRS/AHEAD depression symptoms measure is a subset of the Center for Epidemiologic Studies Depression scale (CES-D). The original CES-D contains 20 items designed to assess the level of depressive symptomatology in epidemiologic studies of various populations. Due to interview time constraints, the first wave of the HRS included a shortened version of the CES-D which was developed for use in the Established Populations for Epidemiologic Study of the Elderly (EPESE) survey rather than including the full twenty-item scale.

In HRS Wave 1, respondents were asked to rate the frequency of eleven symptoms of psychological distress along the following categories: rarely/none of the time, some of the time, most of the time, or all of the time. (This four-level rating of frequency of symptoms will be referred to as "frequency response" throughout this document.) This retains the original response format of the CES-D but selects the subset of items used in the Iowa cohort of the EPESE. According to Kohout et al. (1993), these items were chosen based on the factor-analytic results presented by Radloff (1977).

The number of items and response format for the CES-D questions changed between the first and second waves of the HRS, in order to simplify telephone administration. In HRS Wave 2, three items were dropped from the subset, leaving 8 symptoms for respondents to evaluate. Additionally, the response format was changed from the frequency response to a question stem worded as “would you say yes or no?” (a Yes/No response). This eight-item scale was administered to the full sample. Table 1 below illustrates the question wording of the HRS/AHEAD depressive symptoms measure in HRS Wave 1 and all other waves.

Table 1 Question Wording for CES-D	
Wave 1 of HRS (Frequency Response)	Wave 2 and subsequent waves of HRS and all waves of AHEAD
<p>Please tell me how often you have experienced the following feelings during the past week C all or almost all of the time, most of the time, some of the time, or none or almost none of the time.</p> <p>I felt depressed.</p>	<p>Now think about the past week and the feelings you have experienced. Please tell me if each of the following was true for you much of the time this past week.</p> <p>Much of the time during the past week, you felt depressed. Would you say yes or no?</p>

In an experimental module (Module 1 of HRS Wave 2), the exact measure from HRS Wave 1 was administered to a random subset of 808 respondents, in addition to the new Yes/No response scale. This allows for comparison across the two response formats for a given individual. Analysis presented in Section V of this paper has shown that there is some disagreement between the two forms of the scale, leading to the conclusion that there is no simple way to accurately compare HRS Wave 1 and HRS Wave 2 depression symptomatology. Such comparisons are likely to be biased by measurement issues unless the analyst adjusts for this wording change with statistical models (refer to Section V of this paper for more information).

Subsequent waves of the HRS, and all of the AHEAD interviews, include the eight-item scale with the same Yes/No response format used in HRS Wave 2. Table 2 contains an inventory listing the CES-D items contained in the HRS and AHEAD studies.

[Table 2 about here]

Summary Scores. The traditional method of summarizing (also referred to as scoring) the CES-D is to assign each item a value from zero to three, with a response of “none of the time” counting as zero, “some of the time” counting as 1, “most of the time” counting as 2, and “all of the time” counting as 3. The items worded in the positive direction are “reverse-scored”; for

example, “I am happy” is valued at zero if answered “most of the time”. The items are then summed to yield a total score ranging from zero to thirty-three, in the case of the 11-item subset in HRS Wave 1. All other waves of the HRS and all waves of the AHEAD use a summary score ranging from zero to eight, created by summing the number of “yes” answers across the eight items (with the positive items reverse-scored).

In addition to the standard method of scoring, researchers have used the items that make up the CES-D scale in several alternate forms of summary scores. Many researchers have further transformed the CES-D summary score into a dichotomous variable indicating a high likelihood of clinical depression, following the example of Comstock and Helsing (1976). The subsection on determining caseness in Section IV contains more information on this. Pimley (1990) suggests using differential weights on symptoms (instead of the equal weight given when the items are simply summed), giving a higher weight to the rarer symptoms, which presumably discriminate between clinical and non-clinical depression better than the more common symptoms. Gottlib, Lewinsohn, and Seeley (1995) create two different summary scores, one measuring the number of symptoms endorsed and one representing the frequency. This strategy could be employed when analyzing the HRS Wave 1 data, however, only the number of symptoms is available in the other waves of the study. Obviously, the researcher is free to use the individual items in any manner he or she deems acceptable, but the summary score described in the previous paragraph is most commonly used in the psychology and epidemiology literatures.

The Short Form Composite International Diagnostic Interview (CIDI-SF) Used In HRS/AHEAD

The short form of the CIDI for major depressive episodes (MDE) is only administered once to each respondent in the surveys. The HRS cohort answered the CIDI-SF for MDE in Wave 3 (1996) and the AHEAD cohort in Wave 2 (1995). New cohorts added to the combined survey (starting with HRS-1998) will be administered the CIDI-SF for MDE during their baseline interview. As of this writing, there are no plans for a follow-up administration of the CIDI-SF for any respondents.

The short form for major depressive episodes in its entirety consists of thirty-three questions. However, respondents do not answer all thirty-three questions. Respondents are first asked a screen question regarding depressed mood (dysphoria): whether they ever felt sad, blue or depressed for more than two weeks in a row in the past 12 months (E1006 in HRS), as shown in Figure 1. If they answer “no” to this question, they are then skipped past the detailed questions (E1007-E1024 in the HRS) and are asked the second screen question about anhedonia (the inability to experience pleasure from normally pleasurable activities): “during the past 12 months was there ever a time lasting two weeks or more when you lost interest in most things like hobbies, work, or activities that usually give you pleasure?” (E1028 in the HRS). If they also answer “no” to this question, they are finished with the CIDI-SF (this is called being screened out of the CIDI-SF) and continue with the remainder of the HRS/AHEAD interview.

[Figure 1 about here]

Similarly, if a respondent volunteers that they did not experience a dysphoric period of 2 weeks or more because they were taking antidepressant medications, they are coded as such and skipped to the second screen question. If they answer “no” to that question also, and indicate that this was due to the medication, they are coded as such and are screened out of the CIDI-SF (skipped to the next section of the HRS interview).

If a respondent answers “yes” to one of the screen questions, they are then asked about the intensity and duration of the depressed mood or anhedonia (E1007 and E1008 for the first screen question, E1029 and E1030 for the second in the HRS interview). If the intensity was less than “all” or “most” of the day or it occurred less often than “every day” or “almost every day”, the respondent is screened out of the CIDI-SF. If the respondent indicates that the depressed mood or anhedonia was for most/all of the day *and* for almost every/every day, he or she is then probed for specific symptoms. The symptom questions asked after the depressed screen question (E1009 through E1017 in HRS) and the symptom questions asked after the anhedonia screen question (E1031 through E1038 in HRS) parallel each other.

Respondents that endorse the first screen question (depressed mood) for the necessary duration and intensity are then asked about seven symptoms. The first symptom is anhedonia, probed by E1009 in the HRS. Anhedonia is only counted as a separate symptom for those respondents who endorse the depressed mood screen question. Respondents endorsing the second screen question (E1028-anhedonia) are asked about the same six remaining symptoms that the depressed-mood respondents are asked.

Both groups are asked about feeling tired or having low energy (E1010, E1031). The third symptom is appetite change, either decreased appetite (E1011, E1032) or increased appetite (E1012, E1033)⁵. The fourth symptom is trouble sleeping (E1012, E1034) and a “yes” response is probed about frequency **C** every night, nearly every night, or less often (E1013, E1035). To qualify as a depressive symptom for the CIDI-SF diagnosis, the respondent must have trouble sleeping nearly every night or every night. The fifth symptom is trouble concentrating (E1015, E1036), the sixth is feeling worthless (E1016, E1037), and the seventh symptom is thinking about death (E1017, E1038). Note that the respondent is asked whether they think about death in general, their own or someone else’s, rather than asking about suicidal thoughts directly. The remaining questions in the CIDI-SF consist of review questions, asking about the length of the spell (from 2 to 52 weeks) and the month of onset of the most recent spell. Table 3 lists the items included in the CIDI-SF for major depressive episodes.

[Table 3 about here]

Summary scores. The summary variable for the CIDI-SF ranges from zero to seven. It is zero for those who answered “no” to the depressed and anhedonia screen questions, those who

⁵Note that this is a change from the original CIDI and the CIDI-SF v1.0 questions, which ask about weight gain and loss rather than appetite.

did not meet the frequency and duration criteria for those questions, and those that did not endorse any MDE symptoms beyond one of the screen questions. For those who *did* endorse one of these screen questions with the appropriate frequency/duration, the summary variable is the count of the number of symptoms endorsed out of the seven. Note that the possible score for respondents endorsing the second screen question only ranges from zero to six, because anhedonia is counted as an additional symptom for those with depressed mood.

According to Nelson et al. (1998), a score of three or more is indicative of a diagnosis of depression. Researchers may implement a more stringent cutoff of five or more symptoms to correspond to the DSM III-R requirements for major depression, if desired. Also, Nelson et al. (1998) detail a strategy for using the CIDI-SF as a continuous measure, by associating each symptom level (zero to seven) with a probability of “caseness” derived from the National Comorbidity Survey. This probability of caseness estimate is actually the probability that the *full* CIDI (rather than the short form) would designate the individual as having the disorder, and not the probability that a clinician would diagnose the individual with a disorder. Researchers should be wary of the applicability of this approach to the HRS/AHEAD data, since the NCS has a different age range, 15 to 54 years. One must believe that the presentation of major depressive episodes is the same for those over age 50 as those under age 50 to use the probabilities of caseness from the NCS with the HRS/AHEAD.

Self-Rated Emotional Health and Self-Reported Psychological Diagnosis

A final set of measures that addresses psychological health in general is the self-report questions. In each wave, respondents are asked to rate their own emotional health on a scale ranging from excellent, very good, good, fair, to poor. They are also asked about a doctor’s diagnosis of psychological problems, the receipt of psychological treatment, and the use of psychotropic medications. In the baseline interview (Wave 1 for the HRS and AHEAD cohorts), respondents are asked if a doctor ever told them that they had psychiatric, emotional, or nervous problems. If they answered affirmatively, they are asked about treatment and medication. In subsequent waves, respondents that answered “no” in previous waves are asked if they have received such a diagnosis since the last interview. All respondents who reported a diagnosis in the current wave or any previous wave are then asked about treatment and medication at each interview. Table 4 shows these questions across waves of the HRS and AHEAD.

[Table 4 about here]

Reliance on physician diagnoses of psychological problems is believed to underestimate the true prevalence of psychiatric disorder. A majority of individuals with mental disorders do not receive treatment from a general care doctor or a mental health professional (Kessler et al., 1994). The use of medical services for treatment of mental disorders is influenced by a person’s economic situation, as well as other factors. As a result, the use of self-reported physician diagnosis as an indication of psychiatric illness when studying the relationship between mental health and economic behaviors is problematic.

III. Origin of the CES-D and CIDI-SF

This section of the paper gives background information on the development of the original scales from which the HRS/AHEAD derives its affective functioning measures. These are the Center for Epidemiologic Studies Depression Scale (CES-D) and CIDI-SF for major depressive episodes.

The CES-D Scale

The CES-D was developed to measure the frequency of depressive symptoms in the general population and was designed for inclusion in surveys (Radloff, 1977). The twenty items that comprise the full scale were taken from existing measures of depression including Zung (1965); Beck, Ward, and Mendelson (1961); Raskin, Schulterbrandt, and Reating (1967); and the depression subscale of the Minnesota Multiphasic Personality Inventory (Dahlstrom and Welsh, 1960). Individuals evaluate how frequently they experienced each of the twenty items during the past week, from none or almost none of the time, some of the time, most of the time, to all/almost all of the time. With exploratory factor analytic techniques, Radloff (1977) found that the full (20-item) CES-D has four factors C depressed affect, positive affect, somatic complaints, and interpersonal problems. National norms for the CES-D came from its inclusion in the 1974-1975 National Health and Nutrition Examination Survey.

Hundreds of studies have used the CES-D scale to measure depressive symptoms in a wide range of both clinical and non-clinical populations. It has been used with many ethnic minority groups and has been translated for use in many non-English speaking countries C examples include Korean immigrants, Chinese-Americans, American Indians, Guatemalans, African-Americans, Hispanics, and Japanese. It has also been used with specific subpopulations suffering from physical health complaints: chronic pain patients, hospitalized physically-ill patients, persons with rheumatoid arthritis, and stroke survivors (Geisser et al., 1997).

The CIDI-SF for Major Depressive Episodes

The CIDI-SF is the most recent version of the highly-structured research interview which implements the diagnostic criteria of the DSM, can be administered by lay interviewers, and is scored by computer⁶. The initial instrument of this type was the Diagnostic Interview Schedule (DIS) developed at the National Institute of Mental Health and used in the Epidemiologic Catchment Area studies during the early 1980's (see Robins et al. (1981) for the history of the DIS). The World Health Organization (WHO) and the National Institutes of Health then modified the DIS to form the Composite International Diagnostic Interview (WHO-CIDI) by adding questions to generate ICD-10 (International Classification of Diseases, version 10) diagnoses in addition to DSM-III-R. The ICD-10 diagnoses allow cross-cultural and cross-national comparisons.

⁶See Kessler et al. (1998a), Wittchen and Kessler (1994), and Wittchen (1994).

Additional modifications to the WHO-CIDI led to the UM-CIDI, which was used in the National Comorbidity Survey (NCS). To form the UM-CIDI, the NCS researchers dropped diagnoses from the WHO-CIDI that were not of interest to their study, retaining only the questions needed to ascertain DSM-III-R diagnoses. They re-organized questions to minimize the “no” response set **C** when respondents realize that a “yes” answer to a stem question results in several more questions being asked, they have an incentive to answer “no” in order to minimize interview burden, even if their true answer is “yes”. The UM-CIDI moves all stem questions to the beginning of the interview, then begins the detailed questions once all of the general probes have been answered.

Researchers from the NCS and WHO developed “short forms” of the CIDI for use in the National Health Interview Survey⁷. One difference between the full CIDI and the CIDI-SF has to do with the time frame; the short form asks about the past twelve months while the full CIDI asks about lifetime experience. Respondents are asked about fewer symptoms than are in the full CIDI for major depression. In the short form, respondents are asked to identify their worst episode during the past twelve months and then are probed about symptoms. Questions were added to assess the DSM-III-R criteria regarding the persistence of symptoms⁸. The CIDI-SF does not ask questions regarding the organic nature of the symptoms; that is, whether the symptoms are a result of physical illness, medication, substance usage, or alcohol usage. Additionally, as a requirement for inclusion in the National Health Interview Survey, the short form scales were tested in a cognitive research laboratory. The wording of some questions was changed to clarify meaning, based on the results of this laboratory research.

IV. Special Methodological Issues

This section documents two general methodological issues of the HRS/AHEAD study and two special methodological issues relating to the depression measures. The first general issue concerns the determination of age-eligibility and why that impacts analyses. The second general issue concerns the imputed data that was released as part of the standard HRS Wave 1 data.

The last two issues are of special concern for the affective functioning measures. The first of these issues concerns the presence of proxy respondents, which appear as “don’t know” answers to the CES-D and CIDI-SF measures. The last issue discussed in this section concerns the use of these survey measures for detecting mental illness, which is often called determining “caseness.”

⁷As of this writing, these measures were planned as part the 1999 NHIS as indicated in the draft questionnaire.

⁸The CIDI-SF implements criterion A through C of the DSM-III-R but not criterion D that the depression is “not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.” It also does not impose the hierarchy of diagnoses in the DSM (Kessler et al., 1998).

Age-Eligibility

The desire of the HRS surveys to collect information on both spouses of a married couple has led to two mutually exclusive groups **C** those age-eligible for their survey and those who are spouses of an age-eligible respondent but are not age-eligible themselves. The HRS sampled households with at least one individual born between 1931 and 1941 and also interviewed the spouse when that individual was married, regardless of the spouse's birth year⁹. Therefore, some households contain two respondents who are age-eligible while some contain only one age-eligible respondent.

It is important to note that these not-age-eligible respondents are not a random sample representative of their age cohort. Rather than being chosen at random, they were sampled because they were married to an age-eligible person. In the HRS, not-age-eligible respondents born before 1931 (therefore older than the HRS cohort) tend to be the male husbands of age-eligible female respondents. Those not-age-eligible respondents born in 1942 or later (therefore younger than the HRS cohort) tend to be female **C** the wives of HRS-age-eligible men. Including these not-age-eligible respondents in an analysis can lead to biases. Therefore, the individual-level sampling weights on the HRS and AHEAD files are zero for not-age-eligible respondents. However, in unweighted analyses, the researcher must be careful to account for this conditionally selected group. Excluding these individuals is recommended. Hence, not-age-eligible respondents are excluded from the tabulations that appear in all of the tables in this paper.

Table 5 shows the number of age-eligible and not-age-eligible respondents in the HRS and AHEAD samples.

Age-Eligibility	HRS Sample	AHEAD Sample
Total Respondents	12,652	8,222
Age-Eligible	9,772	7,447
Not Age-Eligible	2,880	775

⁹Likewise, the AHEAD study sampled households with at least one individual born in 1923 or earlier and also interviewed their spouse, if married, who would not be age-eligible if born in 1924 or later.

HRS Wave 1 Imputations

This subsection describes the imputation procedures that are unique to Wave 1 of the HRS survey. Wave 1 of the HRS was released to the public with all missing items replaced with imputations. This is to allow analysts to quickly use the data without having to dwell on the issue of missing data. Imputation flags are provided for users who wish to “unimpute” these missing data and analyze the pattern of missing responses.

Of all the tabulations and analyses presented in this working paper for HRS Wave 1 use data that have had imputed responses replaced with missing values. For most of the tables in this paper, respondents with missing CES-D or CIDI-SF items are excluded from the sample that is presented in the table. All tables are clearly marked when they exclude respondents with missing data. For analysts trying to replicate tables in this paper, it is important to use the imputation flags to derive the appropriate sample.

Proxy Respondents

The methodology of the HRS/AHEAD study includes the use of “proxy” respondents when the original (i.e., sampled) respondent could not or would not complete the interview. In the introduction to the Special Issue (regarding the AHEAD study) of the Journals of Gerontology Series B (Vol. 52B), Myers, Juster, and Suzman (1997) explain proxy interviews as follows:

“Finally, a substantial number of the AHEAD interviews were completed by a proxy respondent rather than by the designated respondent. This typically occurred because the designated respondent was ill, cognitively impaired, or unable to participate in a relatively lengthy interview. The incidence of proxy responses varied with the age of the designated respondent, with almost a third of the interviews in the oldest old age group being conducted with proxy respondents rather than with the designated respondent.” (Myers et al., 1997: p. vii)

The following table (Table 6) shows the number of proxy respondents in each wave of the surveys to date.

Proxy Respondents	HRS			AHEAD	
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2
Total Age-Eligible Respondents	9,772	8,933	8,364	7,447	6,299
Proxy Used	475	651	483	791	856
Not a Proxy	9,297	8,282	7,881	6,656	5,443

Since the need for a proxy respondent increases with the likelihood of illness and/or cognitive impairment and these conditions increase with age, the use of a proxy respondent is not independent of age. The use of proxy respondents interacts with the age-eligibility question discussed earlier in this section, particularly for the HRS cohort. In the HRS, most of the respondents older than 61 in the base year (1993) were male, the husbands of age-eligible women. These older men were much more likely to need a proxy respondent than the members of the HRS cohort or the female respondents who were not-age-eligible, since most of these wives were younger than the HRS cohort. Therefore, including not-age-eligible respondents will show a gender bias in the use of proxy respondents for the HRS. Respondents who are not age-eligible in AHEAD are, by definition, younger than the AHEAD cohort and therefore are less likely to need a proxy than the age-eligible respondents.

The CES-D and the CIDI-SF are self-report instruments calling on the respondent to evaluate his or her state of mind. For that reason, neither the CES-D nor the CIDI-SF is asked of proxy respondents. The items for the CES-D and CIDI-SF measures will have the code “8” meaning “don’t know” for individuals with proxy respondents.

In HRS Wave 1, however, the CES-D items *were* mistakenly asked of proxy respondents. Since this was not intentional, we can not document whether the proxies reported on their own state of mind or gave their opinion of the designated respondent’s state of mind. We suggest that the researcher use the variable `AAPROXY@` (from the tracker file, which is available on the HRS web site: <http://www.umich.edu/~hrswww>) to remove proxy respondents from any analysis of the CES-D in Wave 1 of the HRS. Proxy respondents were not asked the depression questions in any other wave of HRS or AHEAD.

Determining Caseness

This subsection discusses the use of the HRS/AHEAD affective functioning measures to diagnose mental disorders; which is referred to as determining caseness.

The CES-D Measures

Originally, the CES-D scale was not designed to measure the prevalence or incidence of depressive disorders in the general population, although it has been used this way in the literature. There are many ways in which the CES-D questions do not match the DSM (Diagnostic and Statistical Manual) criteria for depressive disorders. They do not address duration and intensity, which are important components for a diagnosis of disorder, nor do they probe respondents as to whether the depressive symptoms were the result of bereavement, medication side effects, drugs and alcohol, or physical illness. They also ignore the possibility of other psychological disorders that have symptoms similar to depression, most notably anxiety disorders. This information is important in making a diagnosis of depressive disorder that would be comparable to a clinician’s diagnosis.

However, even with these drawbacks, the CES-D has been used to indicate the prevalence of depression in the literature. Consequently, a number of studies have assessed the ability of the CES-D to identify clinical depression by comparing CES-D scores to the results of structured interviews by mental health practitioners. To explore this question, researchers have evaluated several dimensions of the scale as listed below:

- ! Specificity: the ability to accurately identify noncases or “true negatives.”
- ! Sensitivity: the ability to accurately identify cases, the “true positives.”
- ! Positive Predictive Value: the proportion of true cases among those *exceeding* the cutoff.
- ! Negative Predictive Value: the proportion of true noncases among those *below* the cutoff.

In each of the measures described above, the “truth” refers to the diagnosis made by a clinician; these measures compare the CES-D above and below a cutoff level with a psychological diagnosis for the same individual. Estimates of sensitivity for the full CES-D scale range from 70 percent to 99 percent, and estimates of specificity range from 56 percent to 94 percent (Gotlib et al., 1995). Positive predictive value is generally low, between 20 and 30 percent, while the negative predictive value is high, above 90 percent (Roberts et al., 1991). This is because the cutoff is generally chosen to be conservative.

Although the HRS and AHEAD samples have not been evaluated by mental health professionals, the surveys do include the CIDI-SF for major depressive episodes, which can be used for comparison, as a substitute for clinician diagnosis. Turvey et al. (1997) do this comparison using the AHEAD Wave 2 data. At a cutoff point of 3 or more on the 8 item CES-D, they found a sensitivity of 71 percent and a specificity of 79 percent using the CIDI-SF diagnosis as the “true caseness.” This means that the abbreviated CES-D in AHEAD classified 71 percent of individuals as “not depressed” that the CIDI-SF also determined were “not depressed.” Conversely, the CES-D identified 79 percent of individuals as depressed that the CIDI-SF also classified as depressed.

Although the sensitivity and specificity of the eight-item CES-D measure were similar in the AHEAD sample when compared against the CIDI-SF, the published evaluations of the full CES-D scale (compared with clinician diagnosis) imply that the CES-D is able to identify individuals who are currently depressed pretty well; that is, it does not miss many cases of true depression. However, it tends to have many false positives; its ability to discriminate between those who have many symptoms of depression but do not qualify as clinically depressed, and those who qualify for a diagnosis of depression, is poor. Of course, these evaluations vary depending on what level of CES-D score is used as the cutoff to indicate depression.

Comstock and Helsing (1976) originated the traditional cutoff point of 16 or higher on the full CES-D scale (which ranges from zero to sixty) to indicate the likelihood of clinical depression. This point corresponded to the 80th percentile of the CES-D score distribution of

their epidemiologic survey. Although a score of 16 or higher is often used as a cutoff, many researchers have found different optimal cutoff points for various subpopulations. Of course, a score of 16 on the full scale (with range 0-60) does not correspond to a score of 16 on the shortened scale used in the HRS and AHEAD. For more discussion of cutoff points for the HRS/AHEAD measures, see Section VI, Evaluation of Measures and Data Quality.

More recent research using receiver operating curve (ROC) analysis has identified different optimal cutoff points. Additionally, these optimal cutoff points vary by subpopulation under study. For example, a cutoff of 28 or higher maximized sensitivity and specificity among a Native American village population; 24 or higher was optimal for American high school students; 17 or higher for Cuban Americans; 26 or higher for Puerto Ricans; 27 or higher for patients at primary care clinics; and 13 or higher for nursing home residents (Furukawa et al., 1997).

The CIDI-SF Measure

The full WHO-CIDI measure is currently the most widely accepted method for determining the prevalence of psychiatric disorders in the United States (and elsewhere) through large surveys with lay interviewers (Kessler et al., 1994). The CIDI-SF included in the HRS and AHEAD surveys approximates the diagnosis given by the full CIDI. The authors of the short form recommend a cutoff point of 3 or more symptoms to indicate a diagnosis of clinical depression, based on their analysis comparing the CIDI-SF to the full CIDI. Other researchers have advocated the use of 5 or more symptoms as the cutoff for clinical relevance, to better correspond with the guidelines in the DSM-III-R (Turvey et al., 1997).

The use of this measure will provide the researcher with an estimate of the 12-month prevalence of major depressive episodes among the HRS and AHEAD cohorts. It should be noted that the CIDI-SF does not distinguish between major depressive episodes that occur with major depressive disorder, bipolar disorder (manic-depression), or psychotic disorders. Because major depressive episodes occur within several different psychiatric disorders, the prevalence estimates of major depressive episodes from the CIDI-SF will not necessarily correspond to the prevalence of actual major depressive disorder. Additionally, since it is not used as a longitudinal measure in HRS/AHEAD, the CIDI-SF will not be able to estimate the incidence, or onset, of depression.

V. Comparability of the Shortened CES-D Measures Across HRS and AHEAD Waves

This section provides information about using the measures of affective functioning for longitudinal analysis. Only the CES-D measure of depressive symptoms is repeated in each wave, so any longitudinal analysis must make use of the shortened CES-D scale.

This section documents the changes in the shortened CES-D scale between waves of the surveys. There was a significant change in the wording of the CES-D response options between Wave 1 and Wave 2 of the HRS (AHEAD has consistent wording throughout all waves of the

survey). Since Wave 2 of the HRS, the CES-D scale format and items have not changed. Only Wave 1 of the HRS differs from the rest of the survey.

Making use of an experimental module included in Wave 2 of the HRS, the effect of wording changes in the CES-D between Wave 1 and Wave 2 of the HRS is evaluated. This section also provides basic statistics on between-wave changes in the level of depression for both the HRS (Wave 2 and Wave 3) and the AHEAD (Wave 1 and Wave 2) surveys.

Wording Changes Between Waves 1 and 2 of the HRS

To evaluate the effect of changing the CES-D response categories from the frequency response (4 choices **C** rarely/none of the time, some of the time, most of the time, all/almost all of the time) to the Yes/No response, a subset of the Wave 2 HRS respondents was given the CES-D scale with the frequency response format (i.e., the exact wording from Wave 1) in an experimental module. This subset also had answered the new CES-D format (Yes/No) in the main HRS interview. The frequency response CES-D questions are contained in Module 1 of the HRS Wave 2 data set. They are labeled “Module 1” in the tables that follow.

The dilemma of CES-D scales with different response formats was also encountered in the Established Populations for Epidemiologic Studies of the Elderly (EPESE). Kohout et al. (1993) details their experiences. The Yale/New Haven site administered the full, twenty-item CES-D with the traditional four-level frequency response format. The East Boston and Iowa sites only asked a subset of the 20 items (10 in East Boston, 11 in Iowa) and changed the response format¹⁰. East Boston offered the Yes/No format, while the Iowa site offered three choices **C** hardly ever or never; some of the time; and much or most of the time. To compare the measurement properties of these shortened CES-D scales the authors used the full scale in the Yale data to simulate the Iowa and East Boston scales. To replicate the Yes/No response format of the East Boston scale, they considered “rarely/none of the time” and “some of the time” to be “No” answers and “much of the time” and “most or all of the time” to be “Yes” answers.

To resolve the problem of different response formats for the HRS Wave 1 and HRS Wave 2 shortened CES-D scale, the survey staff intended to follow this same strategy used to compare the difference response formats between EPESE sites. We assumed collapsing “most” and “all” of the time into “yes” and “none” and “some” of the time into “no” would be equivalent to the format where respondents answered with yes or no.

The results from following this strategy with the HRS data are presented in Table 7. Note that for all of the tables presented in this paper, only age-eligible respondents who did not use a proxy response (called self-respondents) with no missing or imputed CES-D items were included in the sample. Counts and frequencies are unweighted. In Table 7, the sample consists of only age-eligible respondents who answered all of the CES-D items in the experimental module (the frequency response) *and* the main HRS interview (the Yes/No response) in both Waves 1 and 2. This leaves 594 respondents out of the 808 originally administered Module 1.

¹⁰The Duke/Piedmont site of the EPESE administered the full, twenty-item CES-D scale, but used the Yes/No response format (Blazer et al., 1991). The Duke site is not discussed in Kohout et al. (1993).

[Table 7 about here]

The first column of Table 7 contains the tabulations of Yes and No responses to the 8 CES-D items given in the main section of the HRS interview, for the sample of 594 Module Respondents. The second column reports tabulations of the frequency response CES-D items (the Module questions) which have been converted to Yes/No responses by coding “most of the time” and “all of the time” as “Yes” and “some of the time” and “none of the time” as “No” C as was done in the EPESE study.

Comparing these two columns highlights the problem with this conversion strategy. For example, take the “Felt Depressed” symptom. When given a choice between “Yes” and “No”, 14.8 percent of the Module respondents said, “Yes, I felt depressed.” However, when the four choices of the frequency response format are collapsed into two categories (Yes and No), only 4.7 percent are coded as “Yes.” This pattern is repeated for all of the symptoms, with the converted frequency response yielding much lower estimates of the prevalence of each symptom than the actual Yes/No response. This problem was not apparent in the EPESE because they did not have a group of the *same* individuals answering both forms of the CES-D during the same interview. With the inclusion of the experimental module in HRS Wave 2, this problem was discovered. Table A1 in the Appendix contains additional tabulations.

Further exploration with the Module respondents uncovered the reason why the proposed conversion strategy did not work. Table 8 cross-classifies the Module respondents on the two formats C the frequency response (from the Module questions) and the Yes/No response (from the main HRS interview). Each of the 8 symptoms is listed in bold type. In the rows beneath each symptom title are the four frequency-response options (all of the time, most of the time, some of the time, and rarely/none of the time). For each frequency response level group (as reported in the Module questions), the columns show the percentage of that group that responded “yes” or “no” to the corresponding question from the main HRS interview.

[Table 8 about here]

For example, 90.9 percent of the respondents who answered “I felt depressed almost all of the time” in the Module also answered “yes” when asked if they felt depressed in the main HRS interview. However, 9.1 percent (one respondent) that answered “I felt depressed all of the time” in the Module said “no” they did not feel depressed in the main HRS interview. This is an obvious contradiction, since one would expect feeling depressed all of the time to translate to “yes, I felt depressed.” Contradictions like these, which do not correspond to our conversion plan (most and all of the time **Y** yes; some and none of the time **N** no), are outlined in the table.

Examining these contradictions reveals the major source of discrepancy from collapsing the four-level response categories into yes/no responses. It is the respondents that report “some of the time” on the frequency response. For example, on the “I felt depressed” question, 45.1 percent (55 respondents) responding “some of the time” to the frequency response choice

answered “yes”, while 54.9 percent (67 respondents) answered “no”¹¹. This pattern is repeated for the rest of the questions with 40 to 69 percent of the “some of the time” respondents endorsing “yes” in the main questionnaire. There is no clear strategy to resolve this issue; designating “some of the time” respondents as “yes” will overstate the endorsement of the item, while designating them as “no” understates the prevalence.

There are many possible statistical methods to get around this problem. One could randomly assign the “some of the time” respondents to Yes and No based on the proportions derived from the Module respondents. Or, one might want to fully impute Yes and No answers from the Module for all combinations of frequency response and Yes/No response, rather than just the “some of the time” group. Depending on the purpose of the analysis, the researcher may want to include covariates in the imputations. Additionally, one circumvents this problem by modeling depression as an underlying latent variable with each CES-D question being a separate indicator of the latent variable.

Since there is no easily agreed upon method to handle these problems with the wording change, the HRS staff recommend starting any longitudinal analysis of depression with Wave 2 for the HRS. Table 9 presents a comparison of Wave 2 and Wave 3 CES-D scores for the HRS respondents, and the same statistics for AHEAD respondents between Wave 1 and Wave 2. Each survey (HRS and AHEAD) compares the same sample of individuals in both waves; those with no missing values for any CES-D item in either wave (also age-eligible and not proxy, as in all of the tables).

[Table 9 about here]

The left-hand side of Table 9 shows the prevalence of the individual CES-D items. The first two columns are for the HRS cohort (Wave 2 and Wave 3), while the second two columns are for the AHEAD cohort (Wave 1 and Wave 2). For each cohort as a whole, the distribution of CES-D scores and endorsement of individual items does not change much between waves. For example, in Wave 2 of the HRS, 15.9 percent of respondents said “yes, I felt depressed” while 14.0 percent did in Wave 3. Similarly, in AHEAD Wave 1, 18.7 percent of respondents endorsed the depressed mood symptom while 18.0 percent did in Wave 2.

The right-hand side of Table 9 shows the summary scores for each cohort. The summary score creation is described in Section II of this working paper. In Wave 2, 52.1 percent of HRS respondents had zero symptoms while 51.6 percent had zero in Wave 3. In AHEAD, there was slightly more movement in the lowest end of the scale, with 40.8 percent having zero symptoms in Wave 1 and 44.2 percent scoring zero in Wave 2. However, for scores above two, the distribution is very similar in Wave 1 and Wave 2.

This aggregate analysis, however, masks significant individual change, as shown in Table 10. As the left-hand side of Table 10 shows, more than half of the HRS respondents experienced a change in their total CES-D score between Wave 2 and Wave 3. Nearly the same fraction increased their score as decreased their score, accounting for the aggregate stability. Most of

¹¹Table A2 in the Appendix contains the sample counts corresponding to Table 8.

those whose score did change between waves only changed by a small amount C one or two points (85.7 percent had a change of two points or less). Very few respondents went from highly symptomatic to asymptomatic and vice versa. The right-hand side of Table 10 shows a similar pattern for the AHEAD respondents. A somewhat higher fraction experienced a change in score, roughly 60 percent, with more having a decrease in their score (32.4 percent) than an increase (28.0 percent). As seen with the HRS cohort, the size of the change was small — 85.6 percent with a change of 2 points or less. Tables A3 and A4 in the Appendix show more detail on these between wave changes.

[Table 10 about here]

VI. Evaluation of the Measures and Data Quality

This section of the working paper examines the HRS and AHEAD data on affective functioning in great detail. These tabulations and analyses are to allow the user to evaluate the quality of the HRS and AHEAD measures and to use as background to their own research.

The first subsection (Section A) contains the univariate distributions of the CES-D-based measure and the CIDI-SF for major depressive episodes within each wave of HRS and AHEAD, through 1997.

The next subsection, entitled “Benchmarking Against Other Surveys” (Section B), describes the use of the National Longitudinal Survey of Mature Women (NLS-MW) dataset to calibrate the shortened form of the CES-D scale used in HRS Wave 1 with the full CES-D. The full 20-item CES-D scale with the four-level frequency response (rarely/none, some, most, all/almost all of the time) was administered to the NLS-MW cohort in 1989. This section also compares the estimated prevalence of depression in the HRS and AHEAD with previously-published figures.

The section on “Internal Consistency and Measurement Properties” (Section C) evaluates the scales’ reliability and validity using basic psychometric techniques. Coefficient alpha scores are presented, as well as the results of exploratory factor analysis; both of which replicate the previously reported reliability and factor structure of the CES-D.

Constructed variables are discussed in the following subsection, Section D. At the time of this writing, only the first wave of the AHEAD study (AHEAD Wave 1) has been released with the inclusion of constructed variables for affective functioning. The method of constructing these variables is documented and the reader is referred to SAS coding in the technical appendix for all other waves of HRS/AHEAD.

Section E describes the amount of missing data, or item-nonresponse, associated with the depression measures, the CES-D and the CIDI-SF. Section F concludes this investigation with a look at the construct validity of the CES-D-based measure. That is, it describes bivariate and multivariate relationships between the CES-D summary score and many relevant items.

A. Univariate Distributions

This section presents the univariate distributions of the depression measures in the HRS and AHEAD for the use of the reader. The univariate distributions show all of the characteristics of the distribution of CES-D scores from sources outside of the HRS. For each wave, the distribution of scores is highly skewed towards the low end of the scale with significant fractions of the sample receiving a total score of zero. To date, there is no outside data to contrast the CIDI-SF against so the distribution is presented for the user's information.

The CES-D Measure

Tables 11 and 12 show the univariate distributions of the CES-D. Table 11 shows the distribution of the 11-item CES-D summary score in Wave 1 of the HRS, while Table 12 shows the distributions of the eight-item CES-D for Waves 2 and 3 of the HRS and Waves 1 and 2 of AHEAD.

[Table 11 about here]

The summary score shown in Table 11 was created following the traditional method of construction C adding up the eleven items with each frequency level receiving a different score from 0 to 3. The sample included all age-eligible respondents in Wave 1 of the HRS, and excluded those who were proxy respondents (see Section IV for rationale) and those with an imputed value for any of the CES-D items, leaving a sample size of 9,137. The mean CES-D score is a low 4.97 and the median is 4. As is the case with other samples that use the CES-D, the distribution is highly skewed towards the low end, with 23.1 percent of respondents scoring zero or one.

The right-hand side of Table 11 contains several cutoffs for indicating clinically-relevant symptoms of depression, or Acaseness@. Three different Aback of the envelope@ strategies were employed to translate the 16 or higher cutoff point of the full CES-D (which ranges from zero to 60) to a cutoff point on the shortened CES-D scale included in Wave 1 of the HRS. A fourth method is described in Subsection B. First, a proportional strategy was employed. We set up the following equation: $16/60=x/33$. The solution for "x" is 8.8 which is rounded this up to 9. A cutoff of 9 and above yields a prevalence of 18.2 percent.

The second strategy uses the fact that 16 can be interpreted as reporting slightly more than 5 symptoms out of 20 with the frequency of "all or almost all of the time." Five symptoms out of 20 is equivalent to 2.75 symptoms out of 11. When valued at the "all or almost all of the time" frequency, 2.75 symptoms yields a total CES-D score of 8.25, which rounds to 8. This cutoff results in a prevalence of 22.6 percent.

The third strategy for finding a cutoff point, uses the interpretation of a "16 plus" score as 8 symptoms at "most of the time". Using the same approach described for the second method

gives a cutoff on the HRS scale of 8.8 which rounds to 9, a prevalence of 18.2 percent. The fourth method listed on Table 11 is described in detail in the subsection on Benchmarking below.

[Table 12 about here]

Table 12 shows the distribution of the eight-item CES-D with yes/no responses for Waves 2 and 3 of HRS and Waves 1 and 2 of AHEAD. In both waves of the HRS, more than half of the respondents had zero symptoms of depression, while in AHEAD, 38.4 percent scored zero in Wave 1 and 44 percent had zero in Wave 2.

In HRS Wave 2, 14.3 percent of the HRS sample scored 4 or higher, while 12.8 percent scored 4 or higher in Wave 3. In the AHEAD cohort, 17.3 scored 4 or higher in Wave 1 and 15.1 percent scored 4 or higher in Wave 2. Tables A5 and A6 in the Appendix provide additional tabulations. Table A5 shows the distributions for the Module respondents, and Table A6 shows the individual CES-D items for Wave 1 of the HRS. See Table 9 in Section V for the individual items for HRS Waves 2 and 3 and AHEAD Waves 1 and 2.

The CIDI-SF for Major Depressive Episodes

Tables 13 and 14 present the results of the CIDI-SF administration for the HRS (in Wave 3) and AHEAD (in Wave 2) cohorts. Throughout the CIDI-SF, the AHEAD cohort appears to exhibit fewer indications of depression than the HRS cohort. As shown in the top panel of Table 13, while 16.2 percent of the HRS cohort said “yes” to the first screen question regarding depression (i.e., has there been a two week period when you felt depressed, blue, etc.), only 10.9 percent of the AHEAD cohort did. The persistence questions (i.e., were the depressed feelings most or all of the day during the two-week period; and, did they occur every day or almost every day during the spell) screened out more than half of those initially endorsing the depression screen question for both HRS and AHEAD. A smaller fraction of both samples endorsed the anhedonia stem question (losing interest in things), 6.7 percent in both cohorts. Over three-quarters of these respondents did not meet the persistence requirements. As a result, 9.4 percent of the HRS cohort and 5.6 percent of the AHEAD cohort “screened in” to the CIDI-SF and were asked the symptom questions.

[Table 13 about here]

The bottom panel of Table 13, labeled Individual Symptoms, shows the prevalence of each symptom asked in the CIDI-SF for major depression. Note that respondents who did not pass the initial screening questions were not asked the symptom questions and were automatically assigned a “no” response. Feeling tired and/or having low energy was the most frequent symptom for the HRS cohort (7.9 percent), followed by trouble concentrating (7.5 percent). For the AHEAD cohort, feeling tired and/or having low energy was the most frequent

symptom (4.6 percent) followed by change in appetite (3.7 percent), thoughts of death (3.7 percent), and trouble concentrating (3.6 percent).

The top panel of Table 14, labeled Total Score on CIDI-SF for MDE Scale, shows the range of the total number of symptoms for the full HRS and AHEAD samples. The majority of the sample for both cohorts had zero symptoms on the CIDI-SF scale. For those with at least one symptom, scores of 5 or 6 were the most common in the HRS cohort, and scores of 4, 5, or 6 were most common in the AHEAD cohort. The next panel, CIDI-SF Diagnosis of Depression (MDE), shows the fraction of the samples scoring above the recommended three or more cutoff and the more stringent five or more cutoff. 8.2 percent of the HRS cohort and 5.2 percent of the AHEAD cohort are depressed by the three-or-more cutoff; while only 5.6 percent of HRS and 3.7 percent of AHEAD pass the five-or-more cutoff.

[Table 14 about here]

The final panel of Table 14 shows the percentage endorsing each individual symptom, out of the group that was asked the symptom questions (those meeting the persistence criteria of either the first or second screen questions). These frequencies show the types of symptoms experienced by the subgroups of HRS and AHEAD that passed the initial screen and potentially have a major depressive episode. Again, we see that feeling tired and/or having low energy and having trouble concentrating were the most frequently endorsed symptoms among the HRS subgroup. Feeling tired and/or having low energy was also the most frequent symptom among the AHEAD subgroup, followed by thoughts of death and change in appetite.

B. Benchmarking Against Other Surveys and Prevalence Estimates

One method of evaluating the quality of the depression measures is to compare them against results found in other surveys using the same or similar measures. Because of the shortened versions of the CES-D scale used in the HRS/AHEAD, it is impossible to compare directly to data collected with an identical measure. Using the microdata from the National Longitudinal Survey of Mature Women (NLS-MW), we were able to approximate the full CES-D score for Wave 1 of the HRS. Going one step further and using data from the individuals in HRS that answered the questions from Module 1 in HRS Wave 2, we were able to approximate the traditional cutoff point of 16 or higher for the eight-item Yes/No response scale. This section describes the process of estimating these cutoff points and compares the resulting prevalence rates to other published figures.

Finding an Equivalent to A16+@ for the HRS Wave 1 Measure: A More Rigorous Approach

The NLS-MW survey administered the full twenty-item CES-D to its respondents in 1989. In 1989, the NLS-MW respondents were age 52 to 66, which corresponds well to the HRS Wave 1 age range of 51 to 61. We created a sample of women from the NLS-MW who were age

52 to 61 and a sample of women from HRS Wave 1 who were age 52 to 61. Because the CES-D measure in Wave 1 of HRS uses the same response categories as the original CES-D, it was possible to create identical CES-D measures for the NLS-MW and HRS Women by selecting from the NLS-MW the items common to both studies.

Figure 2 shows a graph of the distributions of the identical 11-item CES-D measure for both the NLS-MW (the solid line) and the HRS Women (the dashed line). Both distributions are weighted. All possible CES-D scores are listed along the X-axis (from 0 to 33) and the fraction of the weighted sample that scored zero is plotted on the Y-axis, then the fraction that scored one, etc. These points are then joined with a line. The figure shows that the HRS Women and NLS-MW distributions track each other rather well, with some deviation around the lowest end of the scale. The NLS-MW sample has a much larger fraction scoring zero than the HRS Women. Consequently, the sample of HRS Women has a larger fraction scoring one through 10, with the distributions coming together at the score of 11.

[Figure 2 about here]

Figure 3 shows the distribution of the 11-item CES-D in the NLS-MW by whether the respondent scored above or below the traditional 16+ cutoff on the full CES-D. The dashed line shows the scores on the 11-item CES-D (the measure used in Wave 1 of the HRS) for the group scoring less than 16 on the full CES-D. The solid line shows the distribution of scores on the 11-item CES-D for those scoring 16 or higher. No one scoring less than 16 on the full CES-D scored higher than 13 on the 11-item measure. No one scoring above the 16 or higher cutoff had a score lower than 6 on the 11-item CES-D. However, scores on the 11-item CES-D between 6 and 13 represent a mix of those meeting the traditional cutoff for depression and those who do not.

[Figure 3 about here]

Table 15 below shows the sensitivity and specificity of cutoffs on the 11-item CES-D measure where the “truth standard” is scoring above or below 16 on the full CES-D scale.

Proposed Cut-off for the 11-Item CES-D	Specificity (identify true non-cases)	Sensitivity (identify true cases)
A Score of:		
6 and above	85.6	99.7
7 and above	94.4	96.8
8 and above	95.4	94.6
9 and above	97.4	90.2
10 and above	98.6	84.1
11 and above	99.3	74.5
12 and above	99.4	64.1
13 and above	99.9	55.6

Table 16					
Comparison of HRS Wave 1 CES-D With Other Surveys					
Measure	HRS Wave 1 1992 Age 51-61		NLS-MW 1989 Age 52-61	NHANES I 1974-75 Age 65-74	EPESE 1982 Age 65+ New Haven
	Blacks & Whites	All Races	Blacks & Whites	Blacks & Whites	All Races
Total Percent Above Cutoff	12.5%	13.3%	---	14.8%	16.4%
Men	10.5%	11.1%	---	---	11.3%
Women	14.3%	15.1%	15.9%	---	19.2%

Source: Author's calculations for HRS Wave 1 and NLS-MW; NHANES I from Eaton and Kessler (1981); New Haven EPESE from Berkman et al. (1986).

Determining a 16+ Equivalent for the 8-Item CES-D Scale

Taking the estimation strategy one step further allows us to estimate an equivalent to the 16+ cutoff for the 8-item scale. Using data from the module respondents (N=594), we estimated the following equation, weighted by the Wave 1 HRS sampling weights, using White's heteroskedasticity-consistent (robust) estimator of the standard errors:

$$(Eq. 2) \quad [\text{Module 11-item CES-D}] = 1.9784 * [\text{Main HRS 8-item CES-D}] + 1.5633$$

(Standard errors) (0.0851) (0.1140)

Setting the Module 11-item CES-D equal to 9.41 and solving for the Main HRS 8-item CES-D yields 3.96, which rounds to a cutoff point of 4 or more symptoms.

Table 17 below shows the percentage of each sample scoring 4 or above for each of the HRS and AHEAD waves using the 8-item CES-D measure. The prevalence of 14.3 percent in HRS Wave 2 is slightly higher than the 13.3 percent estimated with the 11-item CES-D in Wave 1. In HRS Wave 3, the prevalence rate is slightly lower, at 12.8 percent. The AHEAD surveys, the cohort age 70 and above in 1993, show higher prevalence rates which also decrease over time, with the Wave 1 rate being 17.3 percent which drops to 15.1 in Wave 2.

	HRS Wave 2	HRS Wave 3	AHEAD Wave 1	AHEAD Wave 2
Sample Size	8,377	7,861	6,566	5,396
Percent scoring 4 or higher:	14.3%	12.8%	17.3%	15.1%

Notes: Uses 8-item CES-D with Yes/No response format. Age-eligible, self-respondents with no missing responses to any CES-D item.

The CIDI-SF for Major Depressive Episodes

As of this writing, the CIDI-SF for Major Depressive Episodes (MDE) has not been used in any surveys of older Americans. There is no appropriate data to benchmark the CIDI-SF against. Reiger et al. (1998) present 12-month prevalence rates for MDE for a sample of ages 18 to 54 from three samples: the first wave of the Epidemiologic Catchment Area (ECA), which used the DIS, had a prevalence of 4.2 percent; the two-wave ECA had a prevalence of 6.4 percent; and the National Comorbidity Survey, which used the full CIDI, had a prevalence of 10.1 percent. The prevalence of MD in the HRS cohort was 8.2 percent and in the AHEAD cohort was 5.2 percent.

C. Internal Consistency and Measurement Properties

This section explores the reliability and factor structure of the shortened CES-D included in the HRS and AHEAD. Being a well-established scale, the full, twenty-item CES-D scale has been shown to have high internal consistency and to exhibit a four-factor structure when analyzed using exploratory factor analysis techniques (Radloff, 1977). Below, we analyze the reliability and factor structure of the 11-item and 8-item version used in the HRS/AHEAD.

[Table 18 about here]

Table 18 shows a psychometric analysis of the eleven-item CES-D scale in Wave 1 of the HRS. This includes the standard statistic on scale reliability, Cronbach's alpha, in the top panel of Table 17. In HRS Wave 1, the alpha coefficient was 0.843 using the standardized version of the CES-D items and 0.838 using the raw items. These values are quite high and indicate that the eleven-item CES-D is reliable in the Wave 1 HRS data.

Focusing on the bottom of Table 18 (the last two lines) shows that a principal components analysis yielded three components (estimated linear approximations of the factors; hereafter, they will be referred to as factors) with eigenvalues greater than one, which are identifiable as depressed mood, somatic complaints, and interpersonal relationship problems. The depressed mood factor explained 40.2 percent of the variance, the somatic factor explained an additional 10.4 percent, and the interpersonal factor explained an additional 9.5 percent.

The center of Table 18 shows two types of rotation. Rotating the factors identified by the principal components analysis results in a more interpretable factor loading matrix. Varimax rotation (which was used by Radloff (1977)) maintains the orthogonality of the factors after the rotation (that is, the factors remain uncorrelated), while promax oblique rotation allows the rotated factors to be correlated with each other. In both cases, five symptoms loaded most strongly on the depressed mood factor **C** depressed, happy, lonely, enjoyed life, and felt sad. “Everything was an effort”, “restless sleep”, “could not get going”, and “poor appetite” loaded most strongly on the somatic complaints factor. “People were unfriendly” and “people dislike me” loaded on the interpersonal relationships factor, while “feeling lonely” cross-loaded on both the depressed mood and the interpersonal factor.

Table 19 presents the psychometric evaluation of the eight-item CES-D scale used in Waves 2 and 3 of HRS, and all of AHEAD. The top panel shows the Cronbach alphas, measuring reliability, which are lower for the two waves of AHEAD (0.77 to 0.79) than the HRS (0.81 to 0.83); however, they still show good reliability for the shortened scale with the Yes/No format.

[Table 19 about here]

The principal components analysis with subsequent rotation showed a similar factor structure for all four waves considered. The “eigenvalue greater than one” criterion led to the identification of two factors **C** depressed mood and somatic complaints¹². This is reasonable since the items loading most strongly on the interpersonal factor for the 11-item scale are not included in the 8-item scale. The same five symptoms loaded on the depressed mood factor as did in the 11-item scale: felt depressed, was happy, felt lonely, enjoyed life, and felt sad. The remaining three symptoms loaded on the somatic complaints factor: everything was an effort, restless sleep, and could not get going. Overall, the abbreviated CES-D scales used in the HRS and AHEAD show good internal consistency.

¹²In Wave 2 of the HRS, the somatic factor had an eigenvalue of 0.99, technically not meeting the traditional criterion of being larger than one. Since the value of one as a cutoff for a salient factor is somewhat arbitrary, the second factor was retained in the factor analysis of the HRS Wave 2 data.

D. Constructed Variables

This subsection explains the constructed variables included in Wave 1 of AHEAD. Code for the summary variables used throughout this paper is included in the Technical Appendix. Currently constructed variables exist for the depression measures on the Wave 1 data file of AHEAD only. The main constructed variable included with the AHEAD Wave 1 data is CESD8. This is a sum of the eight CES-D items, with yes counting as “one” and no counting as “zero” except for the two positively worded items. For “I felt happy” and “I enjoyed life”, an answer of yes counted as “zero” and an answer of “no” counted as “one.”

This variable was created for everyone with at least one non-missing CES-D item. This is a difference from all of the results presented in this working paper which exclude observations with any missing CES-D items. For the CESD8 variable, if a respondent refused to answer one of the CES-D items, the variable CESD8 contains the sum of the other seven items, and so on, for those missing up to five items (see Appendix Table A7 for details). The researcher should be careful to understand this variable before using it in analysis.

There are two additional constructed variables regarding affective functioning on AHEAD Wave 1. They are AFF9 and AFF10. The variable AFF9 adds the response to “much of the time last week I felt interested in things” to the CESD8 score, with an answer of “No” receiving one point and “Yes” receiving zero. The AFF10 adds the response to “much of the time last week I had a lot of energy” to the AFF9 score, with an answer of “No” receiving one point and “Yes” receiving zero. As was done with CESD8, these variables contain values for all respondents with at least one non-missing item.

Constructed variables for inclusion on the other data files (HRS Waves 1, 2, and 3; AHEAD Wave 2) have not yet been released. The Technical Appendix to this working paper contains the SAS programs used to construct the summary measures presented in this document for both the CES-D scale and the CIDI-SF scale for use by researchers¹³.

E. Item Non-Response / Missing Data

This subsection documents the amount of item non-response missing data in the affective functioning measures in the HRS and AHEAD. Missing data refers to respondent answers of “don’t know” and “refused.” Table 20 shows counts of missing data for the shortened CES-D scale in the first three waves of the HRS and the first two waves of AHEAD. As noted in Section IV, CES-D items on the HRS Wave 1 data file will not contain missing data; this is because all items with missing data in Wave 1 of HRS were replaced by imputed responses. Therefore, for Wave 1 of the HRS, missing data also refers to responses that have been imputed. These responses are identified by the imputation flags that are also contained in the HRS Wave 1 file.

[Table 20 about here]

¹³SAS is a statistical software package produced by the SAS Institute in Cary, North Carolina.

As Table 20 shows, there were 161 age-eligible, self-respondents who were missing at least one CES-D item in HRS Wave 1. The majority of this group missed only one item (139 respondents, which is 85.8 percent of the total with missing items). Ten respondents refused or could not answer the entire set of eleven CES-D questions in HRS Wave 1.

The amount of missing data was much lower in the remaining waves of the HRS and in the AHEAD. This reduction in missing data is most likely due to the simpler response categories (the yes or no choice) implemented in these waves. In HRS Wave 2, there were only 44 respondents with any missing data, with only seven missing more than one item. Three of these seven respondents refused to answer all eight of the CES-D items. In Wave 3 of the HRS, there were even fewer respondents with missing data; 19 had at least one missing item, five of whom were missing more than one item, and two of those five were missing the entire set of CES-D questions.

The initial wave of AHEAD, similar to the initial wave of the HRS, had more missing data than the second wave. 90 respondents had at least one missing item with 15 respondents missing more than one item, two of whom were missing the entire CES-D scale. In Wave 2 of AHEAD the number of missing responses went down, with only 47 respondents missing any CES-D items and only four missing more than one item. One person in AHEAD Wave 2 refused to answer every item in the CES-D scale.

The bottom of Table 20 shows the individual CES-D items and their non-response frequencies. Four items stand out as the most frequently missing: the two positively-worded items C I was happy and I enjoyed life; and two of the somatic complaints items C everything was an effort and could not get going. This pattern was evident in all waves of both surveys.

Table 21 shows the item non-response for the CIDI-SF for Major Depressive Episodes, which was administered in HRS Wave 3 and AHEAD Wave 2. Overall, there was very little missing data for the CIDI-SF. In the HRS only 5 respondents had missing data for both screener questions, which is equivalent to having missing data for the whole CIDI-SF instrument. Only one respondent in AHEAD was missing both screener questions.

[Table 21 about here]

Since only those who answer “yes” to one of the screener questions are asked the subsequent questions, respondents that answered “no” to both screens are appropriately coded as missing for the remaining questions. Of those asked the subsequent questions in the HRS, thirteen were missing the portion of the day duration question and two were missing the frequency during the spell question. For each of the individual symptom questions, the number of missing responses ranged from zero to three. Similar response patterns are evident in AHEAD.

These tabulations lead us to conclude that item non-response is not a large problem with the affective functioning measures in HRS and AHEAD. The researcher should use whatever

method they choose to deal with item non-response — excluding those respondents from the sample or imputing a response using the questions that were answered or other variables from the file.

F. Construct Validity

This subsection of the paper evaluates the construct validity of the shortened CES-D scale used in HRS/AHEAD. Construct validity refers to how well a scale reflects the underlying concept it is trying to measure. It asks the question: Is the relationship between the scale score and other characteristics of the respondent what theory would predict for the relationship between depression and the other characteristics?

Multivariate and bivariate relationships between the abbreviated CES-D scale used in Wave 1 of the HRS and many other respondent characteristics are presented to allow the reader to evaluate construct validity. In our opinion, these results provide strong evidence that the HRS depression measure does tap the underlying level of psychological distress and depression.

Table 22 shows bivariate relationships between the HRS Wave 1 CES-D measure and various demographic, health, and attitudinal characteristics of the respondents. Using the coefficients from the NLS-MW, a full CES-D scale score (ranging from 0 to 60) was predicted for all HRS Wave 1 respondents who were age-eligible, not a proxy interview, and did not have missing values for any HRS CES-D items. Sample sizes for each characteristic are shown in the column labeled *AN*, while the remaining three columns show the percentage of Wave 1 respondents with the given characteristic who scored under 10, 10 to 15, or 16 and over using the predicted full CES-D score¹⁴. Summing across these three columns yields 100 percent.

[Table 22 about here]

As mentioned in the subsection on prevalence estimates, the fraction of HRS Wave 1 respondents scoring 16 or higher on the imputed full CES-D scale is 13.3 percent. As the literature on depression predicts, there is a higher fraction of women scoring 16 and above than men; 15.1 percent of female respondents versus 11.1 percent of male respondents. The bivariate relationship between race and depression is also as expected from prior literature; more non-whites score above 15 than whites. While 11.4 percent of whites score 16 or higher, 21.6 percent of blacks and 24.2 percent of Hispanics do.

The relationship between depression and marital status is similar to that seen in other studies (see Steffick (1998) for example). Currently married people have the lowest prevalence of depression symptoms with only 9.5 percent scoring 16 and above. Separated individuals show the highest prevalence of depression with 30.2 percent scoring 16 or higher. The next highest

¹⁴For more details regarding this imputation procedure, refer to subsection B of this section, Benchmarking Against Other Surveys and Prevalence Estimates.

prevalence of depression is seen in widowed respondents (27 percent), followed by divorced respondents (20.9 percent), never-married individuals (19.8 percent), and those living with a partner (16.8 percent).

We see the expected relationship between physical health and depression, using the HRS Wave 1 abbreviated CES-D scale. Very few respondents reporting excellent physical health experienced high levels of depression, only 3.7 percent scored 16 or higher. Also, the majority of respondents reporting poor physical health experienced high levels of depression with 53.7 percent scoring 16 or more.

Self-rated emotional health also shows a strong negative relationship with depressive symptoms, as expected. Over 90 percent of those claiming excellent emotional health scored below 10 on the full (imputed) CES-D scale. Seventy-three percent of those reporting poor emotional health scored above 15 on the CES-D. Likewise, self-reported diagnosis of emotional or psychological illness correlates positively with depressive symptoms, 42.6 of those saying they had ever been diagnosed scored 16 or above.

The final set of variables presented in Table 22 are the respondents rating of their overall satisfaction with various aspects of their life: house or apartment, neighborhood, physical health, financial situation, friendships, marriage (if married), job (if employed), family life, the way he or she handles problems in life, life as a whole, and rating of time spent with their spouse (if married). In each case, those reporting that they are satisfied or very satisfied with a specific aspect of their life show the lowest prevalence of scores of 16 and higher. Those saying they are unsatisfied or very unsatisfied with various aspects of their life score much higher, with a substantial fraction scoring 16 or higher. Table A8 in the Appendix shows additional bivariate relationships between the CES-D and respondent characteristics such as economic situation, job characteristics, and social support.

To further support the construct validity of the HRS Wave 1 CES-D, we examined the multivariate relationships between the shortened CES-D measure and several known predictors of depression. Table 23 shows the results of ordinary least squares regression analysis with the eleven-item, frequency response CES-D from Wave 1 of the HRS as the dependent variable. This multivariate regression model confirms various relationships found in the literature (Blazer et al., 1991; Hays et al., 1998; Roberts et al., 1997; for example). Being female and also being non-white increase the level of depressive symptoms; depressive symptoms increase as physical and cognitive functioning decrease. Respondents rating their satisfaction with various aspects of their lives as poor or fair is strongly predictive of depressive symptoms, as one would expect. Being unemployed or disabled is also associated with higher depressive symptoms. The results of both of these explorations indicate that the abbreviated CES-D shows good construct validity.

[Table 23 about here]

VII. Conclusion

This working paper has described the affective functioning measures used in the HRS and AHEAD surveys. Detailed analysis has shown that the quality of these measures in HRS/AHEAD is very good. There are no structural changes planned for the affective functioning measures in future waves of the surveys at this date. This means that the information in this working paper is applicable to HRS-1998 and HRS-2000, as well as future waves. Any proposed changes to the measures will be publicized on the HRS web site, and updated versions of this working paper will be provided.

REFERENCES

- American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised*. Washington, DC: American Psychiatric Association.
- Barefoot, John C. and Marianne Schroll. 1996. ASymptoms of Depression, Acute Myocardial Infarction, and Total Mortality in a Community Sample.@ *Circulation*, Vol. 93, pp. 1976-1980.
- Blazer, D., B. Burchett, C. Service, and L. George. 1991. AThe Association of Age and Depression Among the Elderly: An Epidemiologic Exploration.@ *Journals of Gerontology: Medical Sciences*, Vol. 46, No. 6, pp. M210-M215.
- Beck, A.T, C.H. Ward, M. Mendelson, J. Mock, and J. Erbaugh. 1961. AAn Inventory for Measuring Depression.@ *Archives of General Psychiatry*, Vol. 4, pp. 561-571.
- Broadhead, W. Eugene, Dan G. Blazer, Linda K. George, and Chiu Kit Tse. 1990. ADepression, Disability Days, and Days Lost From Work in a Prospective Epidemiologic Survey.@ *Journal of the American Medical Association*, Vol. 264, No. 19, Nov. 21, pp. 2524-2528.
- Comstock, G.W. and K. Helsing. 1976. ASymptoms of Depression in Two Communities.@ *Psychological Medicine*, Vol. 6, pp. 551-563.
- Dahlstrom, W.G. and G.S. Welsh. 1960. *An MMPI Handbook*. Minneapolis: University of Minnesota Press.
- Dufouil, Carole, Rebecca Fuhrer, Jean-Francois Dartigues, and Annick Alperovitch. 1996. ALongitudinal Analysis of the Association Between Depressive Symptomatology and Cognitive Deterioration.@ *American Journal of Epidemiology*, Vol. 144, No. 7, pp. 634-641.
- Ettner, Susan L, Richard G. Frank, and Ronald C. Kessler. 1997. AThe Impact of Psychiatric Disorders on Labor Market Outcomes.@*Industrial and Labor Relations Review*, Vol. 51, No. 1, pp. 64-81.
- Fechner-Bates, Suzanne, James C. Coyne, and Thomas Schwenk. 1994. AThe Relationship of Self-Reported Distress to Depressive Disorders and Other Psychopathology.@*Journal of Consulting and Clinical Psychology*, Vol. 62, No. 3, pp. 550-559.
- Furukawa, T., T. Hirai, T. Kitamura, and K. Takahashi. 1997. AApplication of the Center for Epidemiologic Studies Depression Scale Among First-Visit Psychiatric Patients: a New Approach to Improve Its Performance.@ *Journal of Affective Disorders*, Vol. 46, pp. 1-13.

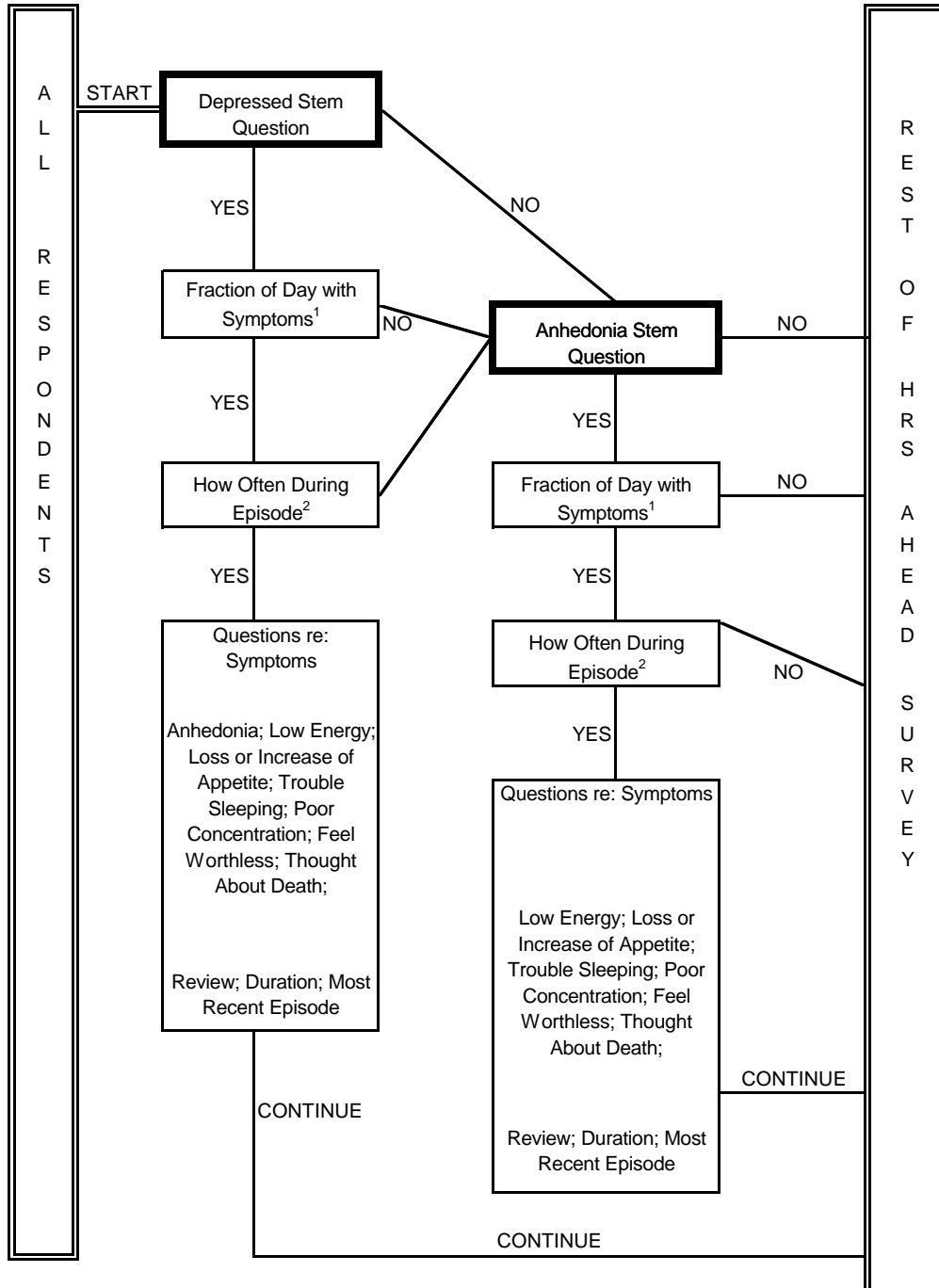
- Geisser, Michael E., Randy S. Roth, and Michael E. Robinson. 1997. Assessing Depression among Persons with Chronic Pain Using the Center for Epidemiologic Studies-Depression Scale and the Beck Depression Inventory: A Comparative Analysis. *The Clinical Journal of Pain*. Vol. 13, pp. 163-170.
- Gottlib, Ian H., Peter M. Lewinsohn, and John R. Seeley. 1995. Symptoms Versus a Diagnosis of Depression: Differences in Psychosocial Functioning. *Journal of Consulting and Clinical Psychology*, Vol. 63, No. 1, pp.90-100.
- Hays, J.C., L.R. Landerman, L.K. George, E.P. Flint, H.G. Koenig, K.C. Land, and D.G. Blazer. 1998. Social Correlates of the Dimensions of Depression in the Elderly. *Journals of Gerontology: Psychological Sciences*, Vol. 53B, No. 1, pp. P31-P39.
- Kessler, Ronald C., G. Andrews, D. Mroczek, B. Ustun, H-U Wittchen. 1998a. The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). *International Journal of Methods in Psychiatric Research*, Vol. 7, No. 4, pp. 171-185.
- Kessler, Ronald C., H-U. Wittchen, J.M. Abelson, K. McGonagle, N. Schwarz, K.S. Kendler, B. Knäuper, S. Zhao. 1998b. Methodological Studies of the Composite International Diagnostic Interview (CIDI) in the U.S. National Comorbidity Survey (NCS). *International Journal of Methods in Psychiatric Research*, Vol. 7, No. 1, pp. 33-55.
- Kessler, Ronald C., K.A. McGonagle, S. Zhao, C.B. Nelson, M. Hughes, S. Eshleman, H-U. Wittchen, and K.S. Kendler. 1994. Lifetime and 12-Month Prevalence of DSM-III-R Psychiatric Disorders in the United States. *Archives of General Psychiatry*, Vol. 51, No. 8, pp. 8-19.
- Kohout, Frank J., Lisa F. Berkman, Denis A. Evans, and Joan Cornoni-Huntley. 1993. Two Shorter Forms of the CES-D Depression Symptoms Index. *Journal of Aging and Health*, Vol. 5, No. 2, pp. 179-193.
- Mroczek, Dan and Ron Kessler. 1994. Final CIDI Screening Scales With Timing Estimates. University of Michigan memorandum.
- Nelson, Christopher B., R.C. Kessler, and D. Mroczek. 1998. Scoring the World Health Organization's Composite International Diagnostic Interview Short Form (CIDI-SF; v1.0 NOV98). Available on the World Health Organization web site at: www.who.int/msa/cidi/cidi-sf.
- Ofstedal, Mary Beth. 1999. Documentation for Cognitive Functioning Measures in AHEAD. University of Michigan Memorandum.

- Palomäki, H, M Kaste, A Berg, et al. 1999. APrevention of Post-Stroke Depression: 1 Year Randomized Placebo Controlled Double-Blind Trial of Mianserin with 6 Months Follow Up After Therapy.@ *Journal of Neurology and Neurosurgical Psychiatry*, Vol. 66, pp. 490-494. As cited in Hachinski, Vladimir. 1999. APost Stroke Depression, Not To Be Underestimated.@ *The Lancet*, Vol. 353, May 22, p. 1728.
- Pimley, Scott. 1990. AProblems of Meaning in Psychiatric Screening Scales: A Critique and Possible Solutions.@ *International Journal of Social Psychiatry*, Vol. 36, No. 2, pp. 143-151.
- Radloff, Lenore S. 1977. AThe CES-D Scale: A Self-Report Depression Scale for Research in the General Population.@ *Applied Psychological Measurement*, Vol. 1, No. 3, pp.385-401.
- Raskin, A., J. Schulterbrandt, and N. Reating. 1967. AFactors of Psychopathology in Interview, Ward Behavior and Self-Report Ratings of Hospitalized Depressives.@ *Journal of Consulting Psychology*, Vol. 31, pp.270-278.
- Roberts, R.E., G.A. Kaplan, S.J. Shema, and W.J. Strawbridge. 1997. APrevalence and Correlates of Depression in an Aging Cohort: The Alameda County Study.@ *Journals of Gerontology: Social Sciences*, Vol. 52B, No. 5, pp. S252-S258.
- Roberts, Robert E., Peter M. Lewinsohn, and John R. Seely. 1991. AScreening for Adolescent Depression: A Comparison of Depression Scales.@ *Journal of the American Academy of Child and Adolescent Psychiatry*, Vol. 30, No. 1, pp. 58-66.
- Robins, L.N., J.E. Helzer, J. Croughan, and K.S. Ratcliff. 1981. ANational Institute of Mental Health Diagnostic Interview Schedule: Its History, Characteristics, and Validity.@ *Archives of General Psychiatry*, Vol. 38, pp. 381-389.
- Silberg, J.L., A.C. Heath, R. Kessler, M.C. Neale, J.M. Meyer, L.J. Eaves, and K.S. Kendler. 1990. AGenetic and Environmental Effects on Self-Reported Depressive Symptoms in a General Population Twin Sample.@ *Journal of Psychiatric Research*, Vol. 24, No. 3, pp. 197-212.
- Steffick, Diane E. 1998. ALabor Market Consequences of Psychological Depression: An Exploratory Analysis.@ Manuscript in progress, University of Michigan.
- Turvey, Carolyn L., Robert B. Wallace, and Regula Herzog. 1997. AThe Relation Between a Revised CES-D Measure of Depressive Symptoms and a DSM-Based Measure of Major Depressive Episodes in the Elderly.@ Manuscript, University of Iowa.
- Wittchen, Hans-Ulrich. 1994. AReliability and Validity Studies of the WHO-Composite International Diagnostic Interview (CIDI): A Critical Review.@ *Journal of Psychiatric Research*, Vol. 28, No. 1, pp.57-84.

Wittchen, Hans-Ulrich and Ronald C. Kessler. 1994. A Modifications of the CIDI in the National Comorbidity Survey: The Development of the UM-CIDI.® National Comorbidity Survey Working Paper No. 2.

Zung, W.W.K. 1965. AA Self-Rating Depression Scale.® *Archives of General Psychiatry*, Vol. 12, pp. 63-70.

Figure 1
The CIDI-SF for Major Depressive Episodes In The HRS And AHEAD



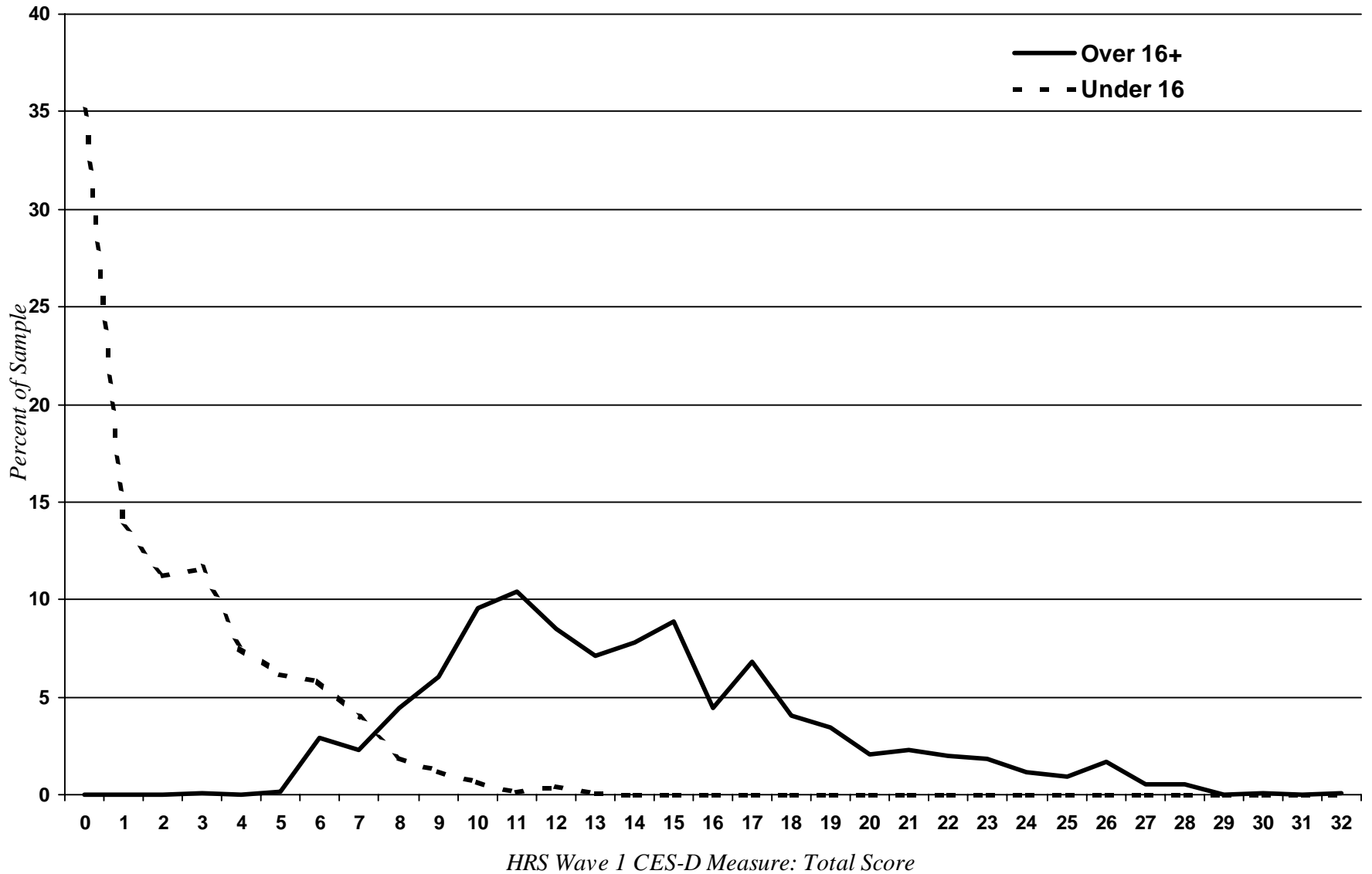
NOTES:

¹ For "Fraction of Day with Symptoms", a "YES" answer is "all day long" or "most of the day". "NO" is "less often".

² For "How Often During Episode", a "YES" answer is "every day" or "almost every day". "NO" is "less often".

Figure 3

NLS Mature Women Distribution of HRS Wave 1 CES-D Measure By Whether Full Score Is Above or Below 16+ Cutoff: Weighted, Age 52-61 Only



APPENDIX TABLES

TECHNICAL APPENDIX