Background: Telephone-based interviews can be used for screening and to obtain key study outcomes when participants in longitudinal studies die or cannot be seen in person, but must be validated among ethnically and educationally diverse people.

Objective: To determine the accuracy of a telephone interview in classifying (1) demented from non-demented participants, (2) cognitively impaired participants from cognitively normal participants, and (3) participants with mild cognitive impairment (MCI) from those with normal cognition or (4) MCI from dementia among an ethnically and educationally diverse community-based sample.

Method: The sample consisted of 377 (30.5% non-Hispanic white, 34.7% non-Hispanic black, and 33.7% Caribbean Hispanic) older adults. The validation standard was diagnosis of dementia and MCI based on in-person evaluation. The Telephone Interview for Cognitive Status (TICS) and the Dementia Questionnaire (DQ) were administered within the same assessment wave.

Results: The sample included 256 people (67.9%) with normal cognition, 68 (18.0%) with MCI, and 53 (14.1%) with dementia. Validity of the TICS was comparable among non-Hispanic whites, non-Hispanic blacks, and Hispanics. Among non-Hispanic whites, the DQ had better discrimination of those with dementia from those without dementia and from those with MCI than among other racial/ethnic groups. Telephone measures discriminated best when used to differentiate demented from non-demented participants (88% sensitivity and 87% specificity for the TICS; 66% sensitivity and 89% specificity for DQ) and when used to differentiate cognitively normal participants from those with cognitive impairment (ie, MCI and dementia combined; 73% sensitivity and 77% specificity for the TICS; 49% sensitivity and 82% specificity for DQ). When demographics and prior memory test performance were used to calculate pretest probability, consideration of the telephone measures significantly improved diagnostic validity.

Conclusions: The TICS has high diagnostic validity for identification of dementia among ethnically diverse older adults, especially when supported by the DQ and prior visit data. However, telephone interview data were unable to reliably distinguish MCI from normal cognition.

Arch Neurol. 2011;68(5):607-614
phone-based data even when participants are unable or unwilling to be seen in person at a follow-up visit. However, this would require the instruments to validly distinguish MCI from normal cognition and MCI from dementia. This has been a challenge in prior studies; although several research groups have documented high specificity for MCI using telephone-based measures, only 1 study showed high sensitivity for distinguishing MCI from normal cognition. Because participants in our study were seen in person at a prior assessment wave, we sought to determine whether the distinction of MCI from other classifications would improve if prior visit data were used along with data from the telephone interview.

**METHOD**

The Columbia University institutional review board approved this project. All individuals discussed the study with trained research staff and provided written informed consent.

**PARTICIPANTS**

The current sample comprised 377 English- and Spanish-speaking participants in a longitudinal study of aging, cognitive function, and dementia among Medicare-eligible older adults residing in neighborhoods in northern Manhattan, New York. The current sample was drawn from a cohort resulting from 2 recruitment efforts, the first in 1992 (n=2125) and the other in 1999 (n=2183). The sampling strategies and recruitment outcomes of these 2 cohorts are detailed in prior publications. Reevaluations occur during follow-up waves that are spaced approximately 18 to 30 months apart.

**RACIAL AND ETHNIC GROUP DEFINITIONS**

Ethnic group was determined by self-report using the format of the 2000 US Census. Participants were first asked to report their race (ie, American Indian/Alaska native, Asian, native Hawaiian or other Pacific Islander, black or African American, or white), then, in a second question, were asked whether they were Hispanic.

**LANGUAGE OF ADMINISTRATION**

Evaluations were conducted in either English or Spanish, on the basis of the participant’s opinion of which language would yield the best performance. Examiners were balanced bilinguals, who spoke both English and Spanish daily with friends, family, and colleagues.

**TELEPHONE INTERVIEW**

The validation study for the telephone interview was initiated during the 2005–2007 assessment wave of the cohort. The telephone interview was conducted by trained interviewers during the same assessment wave but independently from the in-person visit. On average, calls occurred 7.3 months after the in-person visit, with an SD of 10.9 months. One participant had only the TICS because the call was interrupted and the participant could not be recontacted. Of the participants for whom the Dementia Questionnaire (DQ) interview was conducted, 8 did not have the TICS because they were not well enough to come to the telephone (n=4), the participant died soon after the in-person visit (n=2), or the call was interrupted (n=2).

**TELEPHONE INTERVIEW FOR COGNITIVE STATUS**

The TICS was administered and scored in accordance with published procedures. The TICS is modeled after the Mini-Mental State Examination, producing scores ranging from 0 to 41. High test-retest reliability has been demonstrated in several studies. The published Spanish-language adaptation of the TICS was used among Spanish-speaking participants. Total score was used in the analyses.

**DEMENTIA QUESTIONNAIRE**

The DQ is a semi-structured interview that includes yes-or-no questions assessing cognitive complaints in the domains of memory, confusion, and spatial orientation (8 questions) and language/verbal expression (3 items), as well as questions assessing problems with daily function (6 items). This questionnaire has established reliability and validity with high sensitivity and specificity for the detection of dementia and Alzheimer disease. Information about cognitive complaints and functional abilities could be provided by either the participant or an informant, as long as they were knowledgeable about the functional status and medical history of the participant. The 17 questions already mentioned were summed to create a score representing total burden of cognitive complaints and functional problems.

**IN-PERSON EVALUATION**

Medical history was recorded and neurologic and physical examinations were performed at the initial visit and each follow-up. A medical burden score was calculated as a sum of multiple nonpsychiatric medical conditions; it included hypertension, diabetes mellitus, heart disease, stroke, arthritis, chronic obstructive pulmonary disease or other pulmonary conditions, thyroid disease, liver disease, renal insufficiency, peptic ulcer disease, peripheral vascular disease, cancer, Parkinson disease, multiple sclerosis, and essential tremor. Current depressive symptoms were assessed using the Center for Epidemiologic Studies Depression Scale. The Disability and Functional Limitations Scale was used to assess instrumental activities of daily living via self and informant report, as well as perceived difficulty with memory.

Neuropsychological measures included the Buschke Selective Reminding Test (SRT), matching and delayed recognition conditions of a multiple-choice version of the Benton Visual Retention Test, the Rosen Drawing Test, a 15-item Boston Naming Test, the Controlled Oral Word Association Test, the Category Fluency Test, the Color Trails Test, and the Similarities subtest from the Wechsler Adult Intelligence Scale–Revised.

**CONSENSUS DIAGNOSIS**

After each clinical assessment, a group of physicians and neuropsychologists reviewed the functional, medical, neurologic, psychiatric, and neuropsychological data (but were blinded to TICS and DQ data) and reached a consensus regarding the presence or absence of dementia using criteria from the Diagnostic and Statistical Manual of Mental Disorders (Third Edition Revised). For follow-up evaluations, this group was shielded from the prior consensus diagnoses. If dementia was diagnosed, the etiology was determined using published research criteria for probable and possible Alzheimer disease, vascular dementia, Lewy body dementia, and other dementias. Mild cognitive impairment was not diagnosed in the consensus conference but was retrospec-
STATISTICAL ANALYSES

Characteristics of the 3 diagnostic groups were compared using χ² tests and analysis of variance, and correlations between measures and demographic variables were calculated. Receiver operating characteristic (ROC) curves were drawn for each of the telephone measures administered (TICS total score and DQ summary score), using 4 planned comparisons of interest: (1) nondemented vs demented, (2) cognitively normal vs cognitively impaired (ie, MCI and dementia), (3) normal vs MCI, and (4) MCI vs dementia. Areas under the curve were calculated and compared for all participants and separately for non-Hispanic whites, non-Hispanic blacks, and Hispanics. The cutoff yielding the best sensitivity and specificity for the overall sample was determined, and the sensitivity and specificity, negative and positive predictive values, and likelihood ratios were calculated for each telephone instrument for prediction of each of these diagnostic criteria in the entire sample and within the 3 primary racial/ethnic groups.

Binary logistic regression models were constructed to estimate the pretest probabilities of each of the 4 comparisons previously mentioned, with the in-person diagnostic classification as the dependent variable and the following independent variables: age, sex, race/ethnicity, years of education, and the delayed word-list recall score from the SRT taken from the prior assessment. Posttest probabilities for each participant were then estimated for 2 scenarios: (1) when both TICS and DQ were available, by adding the total scores from both measures to the model; and (2) when the participant was dead or too ill to come to the telephone and the TICS was not available, by adding to the model the total score from the DQ only. Predicted values from the models were then used to generate ROC curves. Areas under these curves were compared, and the differences were calculated with 95% confidence intervals.

RESULTS

Demographic characteristics and scores on key study measures of participants with normal cognition, MCI, and dementia are described in Table 1. Most (87.3%) Hispanic participants in this study were immigrants from the Caribbean, including the Dominican Republic (59.5%), Cuba (17.5%), and Puerto Rico (10.3%). The MCI group comprised 68 participants of whom 44 were participants with MCI with memory impairment (64.7%) and 24 with MCI without memory impairment (35.3%). Of the 53 demented participants, most were diagnosed with probable Alzheimer disease, but the sample also included 2 people with Parkinson disease dementia, a participant with vascular dementia, and 1 with a diagnosis of Lewy body disease. The TICS total score was significantly correlated with age (r = −0.37; P < .001), years of education (r = 0.51; P < .001), prior visit SRT total recall score (r = 0.56; P < .001), prior visit SRT delayed recall score (r = 0.48; P < .001), and depressive symptoms (r = −0.28; P < .001). The mean (SD) score on the TICS for men (28.1 [9.1]) was slightly higher than that for women (25.6 [8.6]) (F₁,360 = 6.1, P = .01). Differences between mean (SD) scores of non-Hispanic whites (30.4 [7.5]), non-Hispanic blacks (27.4 [7.0]), and Hispanics (21.8 [10.2]) were all significant (omnibus F₁,360 = 32.3, P < .001; all pairwise comparisons P < .05). The DQ summary score was significantly related to age (r = 0.17; P < .001), years of education (r = −0.3; P < .001), prior visit SRT total recall score (r = −0.32; P < .001), prior visit SRT delayed recall score (r = −0.26; P < .001), depressive symptoms (r = 0.39; P < .001), and medical burden score (r = 0.21; P < .001).

There were no significant differences in mean (SD) DQ sum-

Table 1. Demographic Characteristics and Test Scores of the Validation Sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Normal Cognition</th>
<th>MCI</th>
<th>Demented</th>
<th>Overall F or χ²</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>256</td>
<td>68</td>
<td>53</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Age, y</td>
<td>80.3 (5.8)</td>
<td>82.2 (6.3)</td>
<td>85.9 (5.3)</td>
<td>20.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Educational level, y</td>
<td>11.2 (4.6)</td>
<td>10.4 (4)</td>
<td>6.4 (4.9)</td>
<td>24.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female sex, %</td>
<td>66.0</td>
<td>75.0</td>
<td>69.8</td>
<td>2.1</td>
<td>...</td>
</tr>
<tr>
<td>Race/ethnicity, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic white</td>
<td>33.2</td>
<td>32.4</td>
<td>15.1</td>
<td>7.7</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Hispanic</td>
<td>30.5</td>
<td>23.5</td>
<td>62.3</td>
<td>22.5</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Non-Hispanic black</td>
<td>35.2</td>
<td>44.1</td>
<td>20.8</td>
<td>7.5</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Other race</td>
<td>1.2</td>
<td>1.9</td>
<td>1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last SRT total recall score</td>
<td>39.6 (9.9)</td>
<td>31.9 (8.3)</td>
<td>22.1 (7.8)</td>
<td>74.6</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Last SRT delayed recall score</td>
<td>5.5 (2.7)</td>
<td>4.1 (2.2)</td>
<td>1.6 (1.9)</td>
<td>51.0</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CES-D score</td>
<td>1.3 (1.8)</td>
<td>1.2 (1.6)</td>
<td>2.3 (2.4)</td>
<td>6.1</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Medical burden score</td>
<td>3 (1.5)</td>
<td>3 (1.7)</td>
<td>3.3 (1.4)</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Time between telephone interview and in-person visit, mo</td>
<td>-6.4 (10.2)</td>
<td>-5.8 (7.5)</td>
<td>-13.3 (15.7)</td>
<td>9.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TICS total score</td>
<td>29.5 (6.2)</td>
<td>26.0 (5.0)</td>
<td>10.9 (10.2)</td>
<td>161.0</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DQ memory complaints score</td>
<td>1.3 (1.5)</td>
<td>2.1 (1.7)</td>
<td>4.3 (2.6)</td>
<td>66.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DQ language problems score</td>
<td>0.8 (0.9)</td>
<td>0.8 (1)</td>
<td>1.4 (1)</td>
<td>10.5</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DQ functional complaints score</td>
<td>0.9 (1.1)</td>
<td>0.8 (1.2)</td>
<td>3.2 (1.9)</td>
<td>78.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DQ summary score</td>
<td>2.9 (2.7)</td>
<td>3.7 (3.0)</td>
<td>8.8 (4.6)</td>
<td>78.6</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Abbreviations: CES-D, Center for Epidemiologic Studies Depression Scale; DQ, Dementia Questionnaire; MCI, mild cognitive impairment; SRT, Selective Reminding Test; TICS, Telephone Interview for Cognitive Status.

| aData are given as mean (SD) unless otherwise indicated. |

SAMPLE CHARACTERISTICS
mary score between men (3.6 [0.4]) and women (4.1 [0.2]) \( \left( F_{1,376} = 1.1, P = .29 \right) \). Although the mean DQ summary score (SD) of whites (3.1 [2.9]) and blacks (3.2 [2.9]) did not differ from each other, Hispanics (5.3 [4.5]) reported more problems on the DQ than each of the other 2 groups \( \left( F_{1,372} = 15.8, P < .001 \right) \). The TICS and DQ scores were significantly correlated with each other \( \left( r = -0.59; P < .001 \right) \).

**ROC CURVE ANALYSES**

**Figure 1** shows the ROC curves for separation of demented vs nondemented (i.e., normal cognition and MCI combined) participants. The area under the curve (AUC), diagnostic characteristics, and optimal cutoffs derived from the ROC analyses for the TICS (Table 2) and the DQ (Table 3) are shown for each of the comparisons among all participants and then separately for non-Hispanic whites, non-Hispanic blacks, and Hispanics. The ability of the TICS to discriminate between demented and nondemented participants was comparable across racial/ethnic groups, but the DQ's discrimination was higher among non-Hispanic whites than among racial/ethnic minorities. Figure 1B depicts the ROC curves when MCI participants were combined with the demented participants to form a cognitively impaired group and then compared with participants with normal cognition. The AUCs for the TICS and the DQ were comparable across ethnic groups for this comparison (Tables 2 and 3).

We sought to determine the diagnostic accuracy of the telephone-based measures when making more subtle distinctions between participants with normal cognition and MCI and between MCI and dementia. Figure 1C depicts the ROC curves when demented participants were eliminated from the analysis and participants with MCI and normal cognition were compared. The AUC for both measures for this comparison was lower (0.71 for the TICS and 0.58 for the DQ), but discriminability was similar across ethnic groups (Tables 2 and 3). We then determined the ability of the instruments to distinguish people with MCI from those with dementia, when participants with normal cognition were omitted from the analysis (Figure 1D). For this comparison, the AUC was 0.91 for the TICS, and discriminability was comparable across ethnic groups. The AUC was 0.81 for the DQ in the whole sample, but for this comparison, the DQ had better discrimination among non-Hispanic whites than...
among ethnic minorities. Examination of the odds ratios in Tables 2 and 3 reveals that both the TICS and DQ performed best in distinguishing nondemented (normal cognition and MCI combined) from demented participants, and in distinguishing people with MCI from people with dementia.

**DERIVING AN OPTIMAL CLASSIFICATION ALGORITHM**

*Figure 2* depicts ROC curves for pretest and posttest probabilities as calculated in the logistic regression models. As shown in Table 4, adding information gathered from the TICS and DQ to the pretest model significantly improved the diagnostic performance for all key clinical outcomes in the study. For example, the addition of the TICS and DQ to the pretest prediction of dementia vs no dementia improved the AUC by 6.5%. The best diagnostic performance was in distinguishing nondemented from demented participants (AUC, 0.96) and MCI from demented participants (AUC, 0.95) using demographic information, prior SRT delayed memory score, and both TICS and DQ. Accurate identification of MCI among nondemented participants was poor overall, even when both TICS and DQ were available (AUC, 0.73). Predicted classification as normal cognition, MCI, and dementia, using the optimal cutoffs for the predicted values from the models separating demented from nondemented participants and normal cognition from cognitive impairment, was compared with the observed diagnoses. The cutoffs correctly identified 66.5% of participants with normal cognition, 55.4% of those with MCI, and 92.7% of those with dementia.
Models using only the DQ summary score showed improved classification over pretest probabilities when the goal was to distinguish demented from nondemented people, and cognitively impaired from cognitively normal people. However, addition of the DQ summary score did not improve diagnostic accuracy over pretest probabilities when the goal was to identify MCI among nondemented participants or to identify dementia among cognitively impaired participants (Table 4).

The sensitivity and specificity of the TICS and DQ was variable and depended on the diagnostic groups serving as the standard for comparison. There were no consistent racial/ethnic differences in the ability of the TICS to discriminate diagnostic classifications. However, in distinguishing demented people from nondemented (normal cognition and MCI combined), and people with dementia from people with MCI, the DQ performed better among non-Hispanic whites than among non-Hispanic blacks and Hispanics.

Used alone, the TICS had high sensitivity for distinguishing demented from nondemented participants (normal cognition and MCI combined), and excellent specificity when distinguishing people with dementia from people with MCI. The DQ had lower sensitivity and higher specificity than the TICS for all comparisons but was most valid when distinguishing demented people from those with MCI. The superior specificity of the DQ to the TICS was expected, given the original purpose of developing the instruments: the DQ was designed to pick up on changes in memory and function that are specific to dementia and are not seen in normal aging or MCI.

Comparing likelihood ratios with those of the recent Mayo clinic study by Knopman et al,10 our use of the TICS overall and within each racial/ethnic group yielded superior performance to the Modified TICS when distinguishing demented from nondemented participants (MCI and normal cognition combined), and MCI from dementia. Identification of cognitive impairment (MCI and dementia) from normal cognition was comparable with the Mayo Modified TICS among non-Hispanic whites, non-Hispanic blacks, and Hispanics.

The sensitivity and specificity of the TICS and DQ was variable and depended on the diagnostic groups serving as the standard for comparison. There were no consistent racial/ethnic differences in the ability of the TICS to discriminate diagnostic classifications. However, in distinguishing demented people from nondemented (normal cognition and MCI combined), and people with dementia from people with MCI, the DQ performed better among non-Hispanic whites than among non-Hispanic blacks and Hispanics.

Used alone, the TICS had high sensitivity for distinguishing demented from nondemented participants (normal cognition and MCI combined), and excellent specificity when distinguishing people with dementia from people with MCI. The DQ had lower sensitivity and higher specificity than the TICS for all comparisons but was most valid when distinguishing demented people from those with MCI. The superior specificity of the DQ to the TICS was expected, given the original purpose of developing the instruments: the DQ was designed to pick up on changes in memory and function that are specific to dementia and are not seen in normal aging or MCI.

Comparing likelihood ratios with those of the recent Mayo clinic study by Knopman et al,10 our use of the TICS overall and within each racial/ethnic group yielded superior performance to the Modified TICS when distinguishing demented from nondemented participants (MCI and normal cognition combined), and MCI from dementia. Identification of cognitive impairment (MCI and dementia) from normal cognition was comparable with the Mayo Modified TICS among non-Hispanic whites, non-Hispanic blacks, and Hispanics.
Table 4. Pretest and Posttest Probabilities for Key Diagnostic Outcomes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Non-demented vs Demented</th>
<th>Cognitively Normal vs Cognitively Impaired</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC vs Pretest Model</td>
<td>99% CI</td>
</tr>
<tr>
<td>Pretest model</td>
<td>0.899</td>
<td>. . .</td>
</tr>
<tr>
<td>Posttest model: TICS and DQ</td>
<td>0.964</td>
<td>0.065</td>
</tr>
<tr>
<td>Posttest model: DQ only</td>
<td>0.940</td>
<td>0.042</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cognitively Normal vs MCI</th>
<th>MCI vs Demented</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC vs Pretest Model</td>
<td>99% CI</td>
</tr>
<tr>
<td>Pretest model</td>
<td>0.692</td>
<td>. . .</td>
</tr>
<tr>
<td>Posttest model: TICS and DQ</td>
<td>0.750</td>
<td>0.058</td>
</tr>
<tr>
<td>Posttest model: DQ only</td>
<td>0.718</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under the curve; CI, confidence interval; DQ, Dementia Questionnaire; MCI, mild cognitive impairment; TICS, Telephone Interview for Cognitive Status; ellipses, not applicable.

Accepted for Publication: September 1, 2010.

Correspondence: Jennifer J. Manly, PhD, Taub Institute for Research on Alzheimer’s Disease and the Aging Brain, Columbia University Medical Center, 630 W 168th St, P&S Box 16, New York, NY 10032 (jjm71@columbia.edu).

Author Contributions: All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Manly, Schupf, Stern, Brickman, and Mayeux. Acquisition of data: Manly and Schupf. Analysis and interpretation of data: Manly, Tang, and Mayeux. Drafting of the manuscript: Manly, Brickman, and Mayeux. Critical revision of the manuscript for important intellectual content: Manly, Schupf, Stern, and Tang. Statistical analysis: Manly, Schupf, and Tang. Obtained funding: Manly and Mayeux. Administrative, technical, and material support: Manly, Stern, and Mayeux.

Financial Disclosure: Dr Schupf serves as a consultant to Al Janssen.

Funding/Support: This study was supported by grants P01-AG07232 (Dr Mayeux) and R01-AG16206 (Dr Manly) from the National Institute on Aging.

REFERENCES