***Guidance from the MESA Cognitive Working Group****:*

**Using cognitive and dementia related data in MESA data** – updated April 4, 2019

The purpose of this guidance is help researchers best utilize MESA data resulting from efforts to address cognitive performance and dementia in MESA up to Exam 6, not to dictate that researchers adhere to this guidance.

Assessing the clinical diagnosis of dementia from ICD codes

The careful work of Fujiyoshi et al.1 improved the existing ascertainment of hospitalized dementia and dementia at death resulting in ICD9 codes with the following variable label ‘Possible ICD-based All Cause Dementia’ and variable name ‘ICDdementia’.

The working group recommends that MESA researchers start with this published derived variable. Details are described in Appendix A. In brief, the appendix lists candidate ICD 9 and ICD 10 codes, and links ICD 9 codes that are interpreted as corresponding to ICD 10 codes. It puts these codes into 7 groupings of similar diagnostic meaning (ICD code groups A-G).

**We propose to define “Possible ICD-based All Cause Dementia” as: a) meeting any 2 or more of criteria A-G or b) meeting only a single criterion A-E. The date of occurrence is the date of the hospitalization or death certificate of first occurrence of meeting any of the A-G ICD-based criteria.**

This ICD-based definition of ‘Possible ICD-based All Cause Dementia’ may be improved when combined with medication data (variables alzh1c – alzh5c) and available cognitive data. Details are described in Appendix B. The working group is in agreement that adding cases defined only by medication validly adds cases that are likely to be true positives. However, this definitional strategy was not recommended as the first-line definition because it is recognized that clinic information as a source of information is limited in scope by a) ability of people with cognitive impairment to attend clinic and b) frequency of clinic examinations. The working group recommends that individual users create this variable if they feel it is warranted, using the examination data at which the medication was first noted as the timing of diagnosis.

Addressing false negatives

There is probably a substantial group of false negatives, that is, people who suffer from cognitive impairment but have not been hospitalized or otherwise identified. Expansion of the definition to find these people could use clinic cognitive data or outpatient CMS codes.

Using MESA cognitive screening data (Exam 5 and Exam 6)

*CASI –* The Cognitive Abilities Screening Instrument (CASI) has a score range of 0 to 100 and provides ‘global’ quantitative assessment of attention, concentration, orientation, short-term memory, long-term memory, language abilities, visual construction, list-generating fluency, abstraction, and judgment.2 The raw CASI data include some individuals with invalid scores, missing data and erroneous low scores.

These guidelines recommend the following approach to CASI data quality control:

1. Limit CASI data to those participants with ‘valid’ data as deemed by test administrator.
2. If VALID5 or VALID6 is missing them include these data.
3. Calculate the number of missing CASI components. Only include data where the number of missing CASI components is 3 or less.
4. If the data include CASI scores < 20, then remove these as well. Scores <20 lack face validity.

Accordingly, flag variables (FlagCASI5, FlagCASI6, etc.) were created in order to exclude the total CASI scores that do not meet these criteria. The MESA Cognitive Working Group recommends the following approach to working with CASI data:

1. Analysts may limit analyses to “unflagged” CASI scores
2. If the analyst chooses to maximize sample size at the expense of including potentially invalid CASI score, then a sensitivity analysis should be conducted in order to limit analysis to “unflagged” CASI scores.

Consider that CASI scores significantly differ by age, race/ethnicity, education and other related factors.3 These factors should be considered as covariates.

**There are no well-supported CASI thresholds to define global cognitive impairment among the racial/ethnic groups represented in MESA.** Work by Hughes et al. addressed this issue by defining ‘low cognitive function’ based on the lowest 10% of each race/ethnicity-specific distribution.4 Subsequent statistical models should still be adjusted for age, education and other relevant potential confounders (see Fitzpatrick et al., 2015).3

*Digit Symbol Coding –* raw scores range from 0 to 133. It is a component of the Wechsler Adult Intelligence Scale-III (WAIS-III). Reference normative data can be found in the WAIS-III.5 DSC is a measure of processing speed and psychomotor function.

*Digit Span –* It is a component of the Wechsler Adult Intelligence Scale-III (WAIS-III). Reference normative data can be found in the WAIS-III.5 There are two parts to the test:  *Forwards* with raw scores that range from 0 to 16 and *Backwards* with raw scores that range from 0 to 14. WAIS-III normative data combine them into a single total score, so when comparing MESA participants’ performance to norms it is important to use the total score. DS-Forward and DS-Backward scores may be analyzed separately, however, if the research question specifically addresses attention/concentration (DS-F) or working memory (DS-B).

While the WAIS has established normative data for Digit Symbol Coding and Digit Span tests,5 these tests are primarily used as continuous variables in population based studies.

Using CMS outpatient codes and other future research.

Additional guidance in consideration of CMS outpatient codes are discussed in Appendix C. Validity of cognitive testing for prediction of future dementia is discussed in Appendix D.

**Reference**

1. Fujiyoshi A, Jacobs DR, Jr., Alonso A, Luchsinger JA, Rapp SR, Duprez DA. Validity of Death Certificate and Hospital Discharge ICD Codes for Dementia Diagnosis: The Multi-Ethnic Study of Atherosclerosis. Alzheimer disease and associated disorders 2016.

2. Teng EL, Hasegawa K, Homma A, et al. The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. Int Psychogeriatr 1994;6:45-58; discussion 62.

3. Fitzpatrick AL, Rapp SR, Luchsinger J, et al. Sociodemographic Correlates of Cognition in the Multi-Ethnic Study of Atherosclerosis (MESA). The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry 2015.

4. Hughes TM, Craft S, Baker L, et al. Changes in Metabolic Risk Factors over 10 years and Their Associations with Late-Life Cognitive Performance: The Multi-Ethnic Study of Atherosclerosis (MESA). Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. *in press.*

5. Wechsler D. Wechsler Adult Intelligence Scale: Third Edition. San Antonio, TX, Pscyhological Corp; 1997.

**Appendix**

1. **Recommendation for a new hospital and death certificate ICD-based variable: Possible ICD-based All Cause Dementia**

Based on:

Akira Fujiyoshi, David R. Jacobs Jr, Alvaro Alonso, José A. Luchsinger, Stephen R. Rapp, Daniel A. Duprez. Validity of death certificate and hospital discharge ICD codes for dementia diagnosis: the Multi Ethnic Study of Atherosclerosis. Alzheimer Dis Assoc Disord, Published online, 2017.

**Definition of Possible ICD-based All Cause Dementia**

This paper reported on concordant/discordant/indeterminate status for a dementia diagnosis of dementia in textual medical record wording among a series of preselected ICD codes among hospitalizations and deaths in MESA. The supplement has the full list of codes and associated rubrics, matching ICD 9 and ICD 10. Akira Fujiyoshi acquired copies of all textual medical records that had been deposited with the MESA Coordinating Center (CoC) for the primary purpose of CVD event adjudication by the MESA Events Adjudication committee. These records were not primary information concerning dementia, so he looked for any wording consistent with or contradicting a dementia diagnosis: “We defined dementia as characterized by a significant decline in cognitive function compared to a previous level, not better accounted for by other mental disorders (such as major depressive disorder, schizophrenia) or secondary conditions (due to either delirium, infection, malignancy, trauma, or substance use). One clinician (AF) read medical records blinded to ICD codes, looking for phrases that would indicate, or contradict the conditions defined above. Dementia diagnosis textually documented in the medical records was considered the reference (i.e. true positive).”

We propose to call the resulting variable “Possible ICD-based All Cause Dementia” and to use as the date of occurrence the date of the hospitalization or death certificate of first occurrence of meeting the ICD-based criteria.

The list of candidate ICD codes considered is given in Table 1 and Table 2.

**Table 1. ICD-codes to identify candidate dementia cases**

|  |  |
| --- | --- |
| **ICD9 (used in hospital records)** | **Rubric** |
| 290.4 | Vascular Dementias |
| 290 except 290.4 | Several other dementias |
| 294 | Persistent mental disorders due to conditions classified elsewhere |
| 331.0 | Alzheimer's disease |
| 331.1 | Frontotemporal dementia |
| 331.2 | Senile degeneration of brain |
| 331.82 | Dementia with Lewy body |
| 331.83 | Mild cognitive impairment |
| 331.9 | Cerebral degeneration unspecified |
| 438.0 | Cognitive deficits, late effects of cerebrovascular disease |
| 780.93 | Memory loss |
|  |
| F00 | Dementia in Alzheimer |
| F01 | Vascular Dementia |
| F03 | Unspecified dementia |
| F04 | Organic amnesic syndrome not induced alcohol and other psychoactive substances |
| G30 | Alzheimer's disease |
| G31. except for G31.2 | Other degenerative disease of nervous system, not elsewhere classified |
| I69.91 | Cognitive deficits following unspecified cerebrovascular disease |
| R41 | Other symptoms and signs involving cognitive functions and awareness |

\*G31.2 (degeneration of nervous system due to alcohol)

**Table 2. Seven groupings of these codes were formed, also drawing equivalence over ICD-9 and ICD-10.**

|  |
| --- |
| **Criterion definition** |
|  | **ICD9 code** | **ICD10 code** |
| A. Vascular dementia | 290.4 | F01 |
| B. Alzheimer's dementia | 331.0 | F00, G30 |
| C. Other dementias and chronic organic psychotic conditions | 290, 294 (not 290.4, 294.8, 294.9) | F03 |
| D. Other persistent mental disorders  | 294.8 or 294.9 | F04 |
| E. Other specified dementias and non-specific conditions | 331.1, 331.2, 331.8, 331.9 | G31 not G31.2 |
| F. Cognitive deficits: late effects of cerebrovascular disease | 438.0 | I69.91 |
| G. General signs and symptoms: memory loss | 780.93 | R41 |

The paper shows several aspects of positive predictive value for these codes. In the paper, 306 people were found to have any of these codes, of whom medical record review found evidence for any dementia in 224, found evidence for a non-dementia condition in 18, and information was insufficient in 64. Among the insufficient information cases, 31 did not have a medical record at the CoC while 33 did not contain any remark one way or the other about presence of dementia or a competing condition. In a check of the indeterminate cases in the Minnesota MESA clinic (n=11), additional records were found in 9 people, with 4 cases found to be concordant and 5 cases to be discordant, so there is more information in the clinics than at the CoC, but we are not likely ever to access that; the authors did not have access to the medical records which had been procured at other clinics but were not deemed relevant to CVD diagnosis and therefore not sent to the CoC. Therefore we assume that the 64 indeterminate cases are not particularly different from the 242 cases in which a judgment could be made. It is reasonable to say that the positive predictive value for a rule that takes any qualifying ICD code as evidence of possible dementia is then 224/242 = 92.6%.

Among the 306 cases, the least specific categories for all cause dementia are groupings F and G, constituting 4 ICD codes which have low positive predictive value. Removing category F and G ICD codes when they are not supported by any of the other code leaves a base of 292 cases, among whom 223 were concordant with dementia, 13 were discordant, and 56 were indeterminate. Under the assumption that the indeterminate cases are distributed similarly to the determinate cases, positive predictive value becomes 223/236 = 94.5%.

Also, we think it is likely that CMS inpatient ICD codes will a) agree with the ICD codes for hospitalizations found by MESA and b) expand the above with hospitalizations and ICD codes which MESA did not find. This possibility has not been investigated, however.

**We therefore propose to define “Possible ICD-based All Cause Dementia” as a) meeting any 2 or more of criteria A-G or b) meeting only a single criterion A-E. The date of occurrence is the date of the hospitalization or death certificate of first occurrence of meeting any of the A-G ICD-based criteria.** This definition provides a fairly pure case pool going forward for MESA or from any study which collected ICD codes for all hospitalizations and deaths. We recognize that the people who meet the criteria only in death may be a less complete set of all dementia cases (because the death certificate ICD record may be sparse in a person who died with multiple conditions) and that the occurrence date is a date of the study learning about dementia, not of the beginning of dementia.

1. **Recommendation for the combination of ICD dementia codes with medication data**

MESA data include a derived variable for medications used in the treatment of Alzheimer’s disease and related dementias. These derived variables alzh1c--alzh5c represent the use of acetylcholinesterase inhibitor and NMDA receptor blocking medications. The combination of the medication data with the ‘Possible ICD-based All Cause Dementia’ adds additional 50 MESA participants with only medication and no ICD codes for dementia. The combination of the medication and ‘Possible ICD-based All Cause Dementia’ would be beneficial.

1. **False negative dementia and mild cognitive impairment cases**

It is recognized that the algorithm for “Possible ICD-based All Cause Dementia” excludes people with dementia or mild cognitive impairment who were never hospitalized and were not known to be dead. These people are considered false negatives. They may be classifiable using two sources: CMS outpatient ICD codes and in-clinic cognitive testing. The positive predictive value of CMS outpatient ICD codes is an open question which should be studied. Criteria to be investigated may be a) agreement with the existing assessment of medical records for cases included in Fujiyoshi, et al.; b) concordance with in-clinic cognitive testing, and c) concordance with full cognitive assessment being collected in the Wake Forest MESA sub-cohort by Timothy Hughes and the fairly complete cognitive assessment being collected in the Johns Hopkins clinic by Jingzhong Ding. It is possible that the outpatient codes, especially in the most specific categories A-D have similar positive predictive value as those found in hospital, because the hospitalizations were not generally primarily for dementia. The less specific category E may have lower positive predictive value, while the non-specific categories F and G, with no other supporting evidence for dementia, may be inaccurate.

**D. Prediction of future dementia in people who had exam 5 cognitive testing**

Another topic of great interest is to follow the people who had cognitive testing at exam 5 for occurrence of dementia. This can be done with occurrence after exam 5 of “Possible ICD-based All Cause Dementia”, with agreement with the diagnoses made by ancillary studies from the Wake Forest and Johns Hopkins sites, and by a decline in cognitive score between exams 5 and 6.