Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 3707

Corresponding Measures:

Measure Title: Ratio of observed over predicted rates for diagnosis of mild cognitive impairment

Measure Steward: University of Southern California

sp.02. Brief Description of Measure:

Ratio of the number of patients 65 and older diagnosed with mild cognitive impairment attributed to a clinician or practice over the number predicted based on the demographic profile of that clinician or practice.

Once the clinician's or practice's O/E ratio (i.e., ratio of the observed and expected rates) is calculated, a computation of its associated standard error (SE) can be used to draw inference whether the O/E ratio is significantly different from 1 or not.

1b.01. Developer Rationale:

The recent announcement that the amyloid-targeting drug lecanemab met its primary and secondary endpoints in a phase 3 trial makes it likely that a disease-modifying Alzheimer's treatment will become available soon (Sep 28, 2022). This development lends renewed urgency to the problem of delayed and missed diagnosis of cognitive decline, as these drugs are only indicated in early disease stages, while today cognitive decline is usually detected in advanced stages. For example, Thoits et al.(Thoits, Dutkiewicz et al. 2018) found that about 79% of randomly selected patients, who were newly diagnosed at a memory clinic, had moderate or severe dementia, whereas the amyloid-targeting drugs are indicated only for mild cognitive impairment (MCI) and mild dementia due to Alzheimer's disease. These missed and delayed diagnoses have long taken away from patients and families the opportunity to adopt lifestyle changes to reduce the speed of decline(Ngandu, Lehtisalo et al. 2015), start symptomatic medication treatment, and consider measures to increase physical and financial safety and security(Dubois, Padovani et al. 2015). But soon failing to detect early-stage Alzheimer's disease will deprive patients of the prospect to alter the course of this devastating illness.

Unfortunately, limited data exist for the degree of missed diagnoses of MCI, the stage at which Alzheimer's disease would ideally be treated(Cummings and Salloway 2021). White et al. (White, Ingraham et al. 2021) used data from the Health and Retirement Study (HRS) to determine that 11.4% of subjects with incident MCI reported receiving a timely diagnosis, and Savva et al. (Savva and Arthur 2015) neuropsychiatric testing data from the Aging, Demographics and Memory Study ADAMS data to conclude that 15% of participants with a Clinical Dementia Rating of 0.5, a score reflective of MCI, were aware of a diagnosis of cognitive impairment.

More research has been conducted on dementia detection rates. One study linked Medicare claims data to information on 417 patients with a clinical diagnosis of Alzheimer's disease in the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) data and reported that only around 75% of patients had a corresponding diagnosis in claims data in the period from 1991 to 1995, (Taylor, Fillenbaum et al. 2002), a number similar to the 85% reported by Lee et al. for the 2007–2012 period of the same data. Zhu et al. published a dementia prevalence of 12.9% based on cognitive tests and of 12.4% based on diagnosis codes in the 20% Medicare sample in 2012(Zhu, Chen et al. 2021). Jutkowitz et al. found considerably lower dementia diagnosis rates of 5.6% and 6.5% in 2014 and 2016, respectively, in a convenience sample of three Medicare Advantage Plans (Jutkowitz, Bynum et al. 2020).

However, those studies commonly use older data and/or are limited to subsets of either the fee-for-service (FFS) Medicare population or members of Medicare Advantage Plans, which is important as the decision to enroll into a Medicare Advantage Plan is not random(Mirel, Wheatcroft et al. 2012, Nicholas 2013). More importantly, most prior studies are confined to identifying the prevalence of dementia diagnoses, and – to our knowledge – no analysis has looked into the gap at the stage of MCI in the full Medicare population.

In summary, there is substantial evidence that cognitive impairment remains underdiagnosed, in particular in early stages. Our analyses below confirm such prior evidence and also point to substantial disparities in diagnosis rates. Measuring and reporting diagnosis rates in primary care and comparing those to expected rates given the demographic composition of a clinician's or practice's panel can identify gaps in diagnosis and point primary care clinicians towards efforts to proactively inquire about cognitive concerns and follow up on subjective memory complaints, in particular in high-risk and disadvantaged populations.

References

(Sep 28, 2022). "Lecanemab confirmatory phase 3 clarity AD study met primary endpoint, showing highly statistically significant reduction of clinical decline in large global clinical study of 1,795 participants with early Alzheimer's disease. Eisai News Release." Retrieved Sep 29, 2022, from https://www.eisai.com/news/2022/news202271.html.

Thoits, T., A. Dutkiewicz, S. Raguckas, M. Lawrence, J. Parker, J. Keeley, N. Andersen, M. Vandyken and M. Hatfield-Eldred (2018). "Association Between Dementia Severity and Recommended Lifestyle Changes: A Retrospective Cohort Study." American Journal of Alzheimer's Disease & Other Dementias® **33**(4): 242-246.

Ngandu, T., J. Lehtisalo, A. Solomon, E. Levälahti, S. Ahtiluoto, R. Antikainen, L. Bäckman, T. Hänninen, A. Jula, T. Laatikainen, J. Lindström, F. Mangialasche, T. Paajanen, S. Pajala, M. Peltonen, R. Rauramaa, A. Stigsdotter-Neely, T. Strandberg, J. Tuomilehto, H. Soininen and M. Kivipelto (2015). "A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial." The Lancet **385**(9984): 2255-2263.

Dubois, B., A. Padovani, P. Scheltens, A. Rossi and G. Dell'Agnello (2015). "Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges." Journal of Alzheimer's Disease **49**(3): 617-631.

Cummings, J. and S. Salloway (2021). "Aducanumab: Appropriate use recommendations." Alzheimer's & Dementia.

White, L., B. Ingraham, E. Larson, P. Fishman, S. Park and N. B. Coe (2021). "Observational study of patient characteristics associated with a timely diagnosis of dementia and mild cognitive impairment without dementia." Journal of General Internal Medicine.

Savva, G. M. and A. Arthur (2015). "Who has undiagnosed dementia? A cross-sectional analysis of participants of the Aging, Demographics and Memory Study." Age and Ageing **44**(4): 642-647.

Taylor, D. H., Jr., G. G. Fillenbaum and M. E. Ezell (2002). "The accuracy of medicare claims data in identifying Alzheimer's disease." J Clin Epidemiol 55(9): 929-937

Lee, M., E. Whitsel, C. Avery, T. M. Hughes, M. E. Griswold, S. Sedaghat, R. F. Gottesman, T. H. Mosley, G. Heiss and P. L. Lutsey (2022). "Variation in Population Attributable Fraction of Dementia Associated With Potentially Modifiable Risk Factors by Race and Ethnicity in the US." JAMA Network Open **5**(7): e2219672.

Zhu, Y., Y. Chen, E. M. Crimmins and J. M. Zissimopoulos (2021). "Sex, Race, and Age Differences in Prevalence of Dementia in Medicare Claims and Survey Data." The Journals of Gerontology: Series B **76**(3): 596-606.

Jutkowitz, E., J. P. W. Bynum, S. L. Mitchell, N. M. Cocoros, O. Shapira, K. Haynes, V. P. Nair, C. N. McMahill-Walraven, R. Platt and E. P. McCarthy (2020). "Diagnosed prevalence of Alzheimer's disease and related dementias in Medicare Advantage plans." Alzheimers Dement (Amst) **12**(1): e12048.

Mirel, L. B., G. Wheatcroft, J. D. Parker and D. M. Makuc (2012). "Health characteristics of Medicare traditional fee-for-service and Medicare Advantage enrollees: 1999-2004 National Health and Nutrition Examination Survey linked to 2007 Medicare data." Natl Health Stat Report(53): 1-12.

Natalia Festa, M. P., Max Weiss, Lidia M. V. R. Moura, Nicole M. Benson, Sahar Zafar, Deborah Blacker, Sharon-Lise T. Normand, Joseph P. Newhouse, and John Hsu (2022). "Evaluating The Accuracy Of Medicare Risk Adjustment For Alzheimer's Disease And Related Dementias." Health Affairs **41**(9): 1324-1332.

Nicholas, L. H. (2013). "Better Quality of Care or Healthier Patients? Hospital Utilization by Medicare Advantage and Fee-for-Service Enrollees." Forum for Health Economics and Policy **16**(1): 137-161.

Petersen, R. C., O. Lopez, M. J. Armstrong, T. S. D. Getchius, M. Ganguli, D. Gloss, G. S. Gronseth, D. Marson, T. Pringsheim, G. S. Day, M. Sager, J. Stevens and A. Rae-Grant (2018). "Practice guideline update summary: Mild cognitive impairment." Neurology **90**(3): 126-135.

---

sp.12. Numerator Statement: Number of individuals age 65 and older under the care of a clinician or practice who are diagnosed with mild cognitive impairment

sp.14. Denominator Statement: Predicted number of individuals aged 65 and older under the care of a clinician or practice with mild cognitive impairment based on the demographic profile of the respective clinician or practice. Limit reporting to clinicians and practices with at least 25 attributed patients per CMS' guidance to ensure stability of measure results.

sp.16. Denominator Exclusions: The measures are not using any exclusions as they are based on the 100% samples for both Medicare fee-for-service and Medicare Advantage Plans. While we limit ***reporting*** of the measure to clinicians and practices with at least 25 attributed patients, this does not constitute an exclusion per NQF guidance, since those patients might be reported when reporting on higher levels of aggregation, such as a state. We merely follow CMS' recommendations for minimum sample size to report stable results.

Measure Type: Process

sp.28. Data Source:

Claims

sp.07. Level of Analysis:

Clinician: Group/Practice

Clinician: Individual

IF Endorsement Maintenance – Original Endorsement Date:

Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title: Observed over predicted rates for diagnosis of cognitive impairment

#3707 - Ratio of observed over predicted rates for diagnosis of mild cognitive impairment

#3729 - Ratio of observed over predicted rates for diagnosis of cognitive impairment of any stage

#3672 - Ratio of observed over predicted rates for diagnosis of dementia

sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?: The exact classification of the stage of cognitive impairment is difficult, as it requires not only neurocognitive testing but also a comprehensive evaluation of the patient's health status and functional abilities. Further, the stage may change over our three-year observation window because of natural disease progression on the one hand, and improvement on the other. For these reasons we recommend determining the overall diagnosis counts for any cognitive impairment (3729) first, then the counts for mild cognitive impairment (3707) and dementia (3672) as subsets.

All three should be compared to the predicted rates based on our model to form the respective ratios for observed versus predicted numbers. A ratio of close to 1.0 suggests that a clinician or practice has approximately as many cases diagnosed as expected. The ratio for 3729 shows the overall detection rate, and ratios for 3707 and 3672 provide information on how detection rates differ by stage. The three measures capture risk-adjusted rates of diagnoses of different stages of cognitive decline, i.e., reflect diagnostic quality along the continuity of the disease process. They should therefore be interpreted together.

1. Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria

---

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

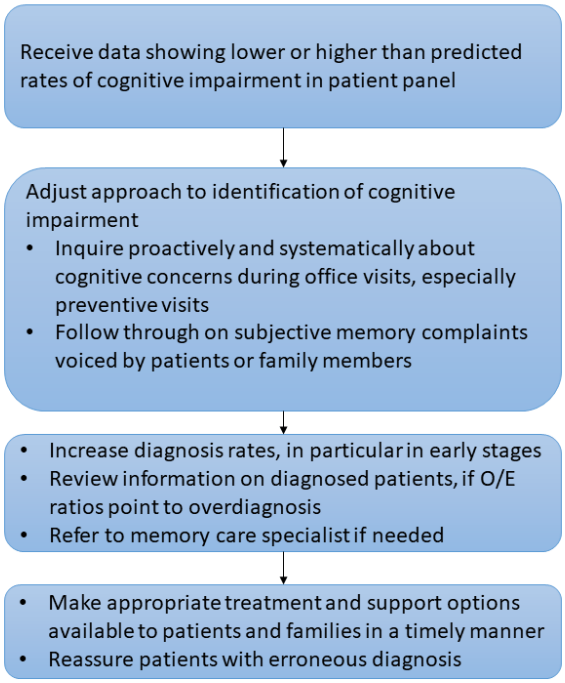
Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

[Response Begins]



Measure Logic Model

[Response Ends]

1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.

A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data.

[Response Begins]

Clinical Practice Guideline recommendation (with evidence review)

[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

Clinical Review Cited by the American Academy of Neurology (AAN) Summary of Practice Guideline for Clinicians (AAN, 2017)

Langa, Kenneth M, and Deborah A Levine. "The diagnosis and management of mild cognitive impairment: a clinical review." JAMA vol. 312,23 (2014): 2551-61. doi:10.1001/jama.2014.13806.

Cited in support of Recommendation A1 (see below) by:

American Academy of Neurology (AAN). (2017). AAN Summary of Practice Guideline for Clinicians. Minneapolis (MN): American Academy of Neurology; 2017 December. Available at https://www.aan.com/Guidelines/home/GuidelineDetail/881.

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

For patients for whom the patient or a close contact voices concern about memory or impaired cognition, clinicians should assess for MCI and not assume the concerns are related to normal aging.

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

The clinical review did not grade the evidence

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

The clinical review did not grade the evidence

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

The grade assigned to this recommendation was “Level B.”

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

Each recommendation is classified according to the AAN Levels of Recommendation as listed below:

A = Established as effective, ineffective, or harmful (or established as useful/predictive or not useful/predictive) for the given condition in the specified population

B = Probably effective, ineffective, or harmful (or probably useful/predictive or not useful/predictive) for the given condition in the specified population

C = Possibly effective, ineffective, or harmful (or possibly useful/predictive or not useful/predictive) for the given condition in the specified population

U = Data inadequate or conflicting; given current knowledge, treatment (test, predictor) is unproven

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

4,977 unique articles were found by the authors.

In terms of quality, randomized double-blind placebo-controlled trials (RCTs) with results reported as intention-to-treat analyses were considered as highest quality data by the authors. Large prospective cohort studies, meta-analyses, and systematic literature reviews were also included as appropriate for supplementing the RCT results.

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

Not provided

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

Not provided

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

N/A

[Response Ends]

Group 2 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

Report Cited by the Alzheimer Society of Canada (ASC) Brochure on "Alzheimer's disease: The importance of early diagnosis" (ASC, 2017)

Report:

Prince et al. (2011). World Alzheimer Report 2011: The benefits of early diagnosis and intervention.

https://www.alzint.org/resource/world-alzheimer-report-2011/

Cited in support of the benefits of early diagnosis of Alzheimer's disease by:

Alzheimer Society of Canada. (2017). Alzheimer's disease: The importance of early diagnosis [Brochure].

https://alzheimer.ca/sites/default/files/documents/the-importance-of-early-diagnosis_print-friendly.pdf

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

"Early diagnosis allows people with dementia and their families to receive timely practical information, advice and support. Only through receiving a diagnosis can they access available drug and non-drug therapies that may improve their cognition and enhance their quality of life."

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

The report did not grade the evidence.

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

The report did not grade the evidence.

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

The report did not grade the recommendation.

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

The report did not grade the recommendation.

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

N/A (not a systematic review)

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

Placeholder from NQF Maintenance Team

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

Not provided

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

none identified

[Response Ends]

Group 3 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

Literature Review on Timely Diagnosis of Alzheimer’s Disease (Dubois et al., 2016)

Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G. Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges. J Alzheimers Dis. 2016;49(3):617-31. doi: 10.3233/JAD-150692. PMID: 26484931; PMCID: PMC49278

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

"Diagnosis should generally occur earlier than is currently common practice, at a time when patients and their family first notice changes in cognitive function and can use the information to make sense of what is happening, make lifestyle changes, and plan for the future."

"Early intervention has the potential to improve the quality of life of patients and their informal family caregivers, both of whom are often relieved once the patient is diagnosed. A timely diagnosis at the prodromal stage may also improve patient access to support services or pathways of care and enable planning for the future."

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

The literature review did not grade the evidence

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

The literature review did not grade the evidence

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

The literature review did not grade the recommendation

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

The literature review did not grade the recommendation

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

451 records, including studies, reviews, editorials, letters, and commentaries, were reviewed and scored by at least two of the authors for inclusion in this literature review. 45 studies or surveys were further assessed for eligibility, of which 9 were included in the final results.

The studies were quantitative (e.g., cost studies) or qualitative (e.g., surveys, focus groups) with no geographical exclusions. Studies investigating the development of or cost/benefit of the tools used to make an early/timely diagnosis (e.g., biomarkers) were not included.

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

Placeholder from NQF Maintenance Team

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

Not provided

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

None identified

[Response Ends]

Group 4 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

Randomized Control Trial cited by "WHO Guidelines for Risk Reduction of Cognitive Decline and Dementia" (Stephen et al., 2021)

Randomized Control Trial:

Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, Bäckman L, Hänninen T, Jula A, Laatikainen T, Lindström J, Mangialasche F, Paajanen T, Pajala S, Peltonen M, Rauramaa R, Stigsdotter-Neely A, Strandberg T, Tuomilehto J, Soininen H, Kivipelto M. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet. 2015 Jun 6;385(9984):2255-63. doi: 10.1016/S0140-6736(15)60461-5. Epub 2015 Mar 12. PMID: 25771249.

Cited in support of the benefits of lifestyle intervention in preventing cognitive decline and dementia by the "WHO Guidelines for Risk Reduction of Cognitive Decline and Dementia":

Stephen R, Barbera M, Peters R, Ee N, Zheng L, Lehtisalo J, Kulmala J, Håkansson K, Chowdhary N, Dua T, Solomon A, Anstey KJ and Kivipelto M (2021) Development of the First WHO Guidelines for Risk Reduction of Cognitive Decline and Dementia: Lessons Learned and Future Directions. Front. Neurol. 12:763573. doi: 10.3389/fneur.2021.763573.

Risk Reduction of Cognitive Decline and Dementia: WHO Guidelines. Geneva: World Health Organization; 2019. PMID: 31219687.

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

"The comprehensive outcome measurements in FINGER suggested beneficial effects on both global cognition and cognitive domains highly relevant for everyday activities (eg, executive functioning, processing speed, and complex memory tasks)."

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

The guidelines did not grade the evidence.

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

The guidelines did not grade the evidence.

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

The guideline did not grade the recommendation

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

Not graded

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

N/A (not a systematic review)

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

Placeholder from NQF Maintenance Team

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

Not provided

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

none identified

[Response Ends]

1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

[Response Begins]

Narrative Evidence Overview

[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.

[Response Begins]

The recent emergence of disease-modifying treatments for Alzheimer's disease has brought a sense of urgency to detection of cognitive decline in early stages. Those drugs remove beta-amyloid deposits from the brain, which are assumed to be on the critical path to neuronal damage and subsequent cognitive and functional decline. However, treatment with those drugs must be initiated in early disease states (mild cognitive impairment (MCI) or mild dementia) in order to slow down disease progression. Clinicals trials in later stages of the disease have all failed to show an effect, which is biologically plausible because highly differentiated cells, such as neurons and cardiac myocytes have lost their ability to replicate. In other words, while the brain has some built-in redundancy, lost capacity is lost forever, rendering detection and treatment of neurodegenerative disorders, much like stroke, a race against time.

It is known that even manifest dementia is commonly diagnosed later if at all, with more than 60% of patients with dementia not detected in residential or community settings.[1] Limited data exist for underdiagnosis of MCI exist, but the gap in diagnosis is much larger as explained in the overview. Early symptoms develop slowly, allowing affected persons and their families to adjust and compensate, and they are commonly disregarded as signs of normal aging. Thus, efforts are needed to draw clinicians' attention to the problem of underdiagnosis and increase their vigilance to early signs of cognitive decline.[2] Clinicians should evaluate for MCI and/or dementia if the patient or caregiver(s) have explicit concerns regarding cognitive dysfunction, if they pick up observable signs of cognitive impairment during a patient visit, or if the patient tests abnormal on a brief cognitive assessment.[3] The initial evaluation should be followed by neuropsychological testing and determination of the underlying etiology to determine whether the patient is eligible for a disease-modifying treatment with either an approved drug or as a participant in a clinical trial.

Even in absence of disease-modifying treatments, early diagnosis of MCI and/or dementia can benefit patients and their caregivers in the following ways:

- Lifestyle risk reduction to decrease risk of disease progression[4]
- Symptomatic pharmacological treatment to improve cognition and reduce neuropsychiatric symptoms, as described by the American Psychiatric Association's "Practice Guideline for the Treatment of Patients with Alzheimer's Disease and Other Dementias"[5]
- Financial planning and prevention of scams[6],[7]
- Advance directives and end-of-life planning
- Driving safety
- Caregiver planning, education, and support

- Initiation of non-pharmacological treatment options, such as psychosocial or behavioral interventions

Thus, receiving a formal diagnosis clearly influences and improves the patient's as well as caregivers' quality of life through providing access to interventions, support systems, and planning.[8]

The clinical benefit of risk reduction interventions and symptomatic treatment has been studied in randomized clinical trials, as documented below, and a proper diagnosis is a precondition to receive disease-modifying treatment. In spite of their intuitive plausibility, the impact of social support and educational interventions has typically not been studied with such designs, partly because it would be unethical to randomly withhold a diagnosis and partly because of their commonsensical nature. The evidence in the following exhibits focuses on demonstrating the benefits of screening and early diagnosis of MCI and Alzheimer's disease, as well as highlighting what actions can be taken upon diagnosis.

[Response Ends]

1a.15. Detail the process used to identify the evidence.

[Response Begins]

Keyword-driven literature search with forward and backward searches

[Response Ends]

1a.16. Provide the citation(s) for the evidence.

[Response Begins]

[1] Lang L, Clifford A, Wei L, Zhang D, Leung D, Augustine G, Danat IM, Zhou W, Copeland JR, Anstey KJ, Chen R. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open. 2017 Feb 3;7(2):e011146. doi: 10.1136/bmjopen-2016-011146. PMID: 28159845; PMCID: PMC5293981.

[2] Ritchie CW, Russ TC, Banerjee S, Barber B, Boaden A, Fox NC, Holmes C, Isaacs JD, Leroi I, Lovestone S, Norton M, O'Brien J, Pearson J, Perry R, Pickett J, Waldman AD, Wong WL, Rossor MN, Burns A. The Edinburgh Consensus: preparing for the advent of disease-modifying therapies for Alzheimer's disease. Alzheimers Res Ther. 2017 Oct 26;9(1):85. doi: 10.1186/s13195-017-0312-4. Erratum in: Alzheimers Res Ther. 2018 Jul 30;10(1):73. PMID: 29070066; PMCID: PMC5657110.

[3] Ismail Z, Black SE, Camicioli R, Chertkow H, Herrmann N, Laforce R Jr, Montero-Odasso M, Rockwood K, Rosa-Neto P, Seitz D, Sivananthan S, Smith EE, Soucy JP, Vedel I, Gauthier S; CCCDTD5 participants. Recommendations of the 5th Canadian Consensus Conference on the diagnosis and treatment of dementia. Alzheimers Dement. 2020 Aug;16(8):1182-1195. doi: 10.1002/alz.12105. Epub 2020 Jul 29. PMID: 32725777; PMCID: PMC7984031.

[4] Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, Bäckman L, Hänninen T, Jula A, Laatikainen T, Lindström J, Mangialasche F, Paajanen T, Pajala S, Peltonen M, Rauramaa R, Stigsdotter-Neely A, Strandberg T, Tuomilehto J, Soininen H, Kivipelto M. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet. 2015 Jun 6;385(9984):2255-63. doi: 10.1016/S0140-6736(15)60461-5. Epub 2015 Mar 12. PMID: 25771249.

[5] Rabins PV, Rovner BW, Rummans T, Schneider LS, Tariot PN. Guideline Watch (October 2014): Practice Guideline for the Treatment of Patients With Alzheimer's Disease and Other Dementias. Focus (Am Psychiatr Publ). 2017 Jan;15(1):110-128. doi: 10.1176/appi.focus.15106. Epub 2017 Jan 11. PMID: 31997970; PMCID: PMC6519627.

[6] Nicholas LH, Langa KM, Bynum JPW, Hsu JW. Financial Presentation of Alzheimer Disease and Related Dementias. JAMA Intern Med. 2021 Feb 1;181(2):220-227. doi: 10.1001/jamainternmed.2020.6432. Erratum in: JAMA Intern Med. 2021 Feb 1;181(2):296. PMID: 33252621; PMCID: PMC7851732.

[7] Widera, E., Steenpass, V., Marson, D., & Sudore, R. (2011). Finances in the older patient with cognitive impairment: "He didn't want me to take over." Journal of the American Medical Association, 305(7), 698–706. doi: 10.1001/jama.2011.164.

[8] Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G. Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges. J Alzheimers Dis. 2016;49(3):617-31. doi: 10.3233/JAD-150692. PMID: 26484931; PMCID: PMC4927869.

[Response Ends]

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

The recent announcement that the amyloid-targeting drug lecanemab met its primary and secondary endpoints in a phase 3 trial makes it likely that a disease-modifying Alzheimer's treatment will become available soon (Sep 28, 2022). This development lends renewed urgency to the problem of delayed and missed diagnosis of cognitive decline, as these drugs are only indicated in early disease stages, while today cognitive decline is usually detected in advanced stages. For example, Thoits et al.(Thoits, Dutkiewicz et al. 2018) found that about 79% of randomly selected patients, who were newly diagnosed at a memory clinic, had moderate or severe dementia, whereas the amyloid-targeting drugs are indicated only for mild cognitive impairment (MCI) and mild dementia due to Alzheimer's disease. These missed and delayed diagnoses have long taken away from patients and families the opportunity to adopt lifestyle changes to reduce the speed of decline(Ngandu, Lehtisalo et al. 2015), start symptomatic medication treatment, and consider measures to increase physical and financial safety and security(Dubois, Padovani et al. 2015). But soon failing to detect early-stage Alzheimer's disease will deprive patients of the prospect to alter the course of this devastating illness.

Unfortunately, limited data exist for the degree of missed diagnoses of MCI, the stage at which Alzheimer's disease would ideally be treated(Cummings and Salloway 2021). White et al. (White, Ingraham et al. 2021) used data from the Health and Retirement Study (HRS) to determine that 11.4% of subjects with incident MCI reported receiving a timely diagnosis, and Savva et al. (Savva and Arthur 2015) neuropsychiatric testing data from the Aging, Demographics and Memory Study ADAMS data to conclude that 15% of participants with a Clinical Dementia Rating of 0.5, a score reflective of MCI, were aware of a diagnosis of cognitive impairment.

More research has been conducted on dementia detection rates. One study linked Medicare claims data to information on 417 patients with a clinical diagnosis of Alzheimer's disease in the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) data and reported that only around 75% of patients had a corresponding diagnosis in claims data in the period from 1991 to 1995, (Taylor, Fillenbaum et al. 2002), a number similar to the 85% reported by Lee et al. for the 2007–2012 period of the same data. Zhu et al. published a dementia prevalence of 12.9% based on cognitive tests and of 12.4% based on diagnosis codes in the 20% Medicare sample in 2012(Zhu, Chen et al. 2021). Jutkowitz et al. found considerably lower dementia diagnosis rates of 5.6% and 6.5% in 2014 and 2016, respectively, in a convenience sample of three Medicare Advantage Plans (Jutkowitz, Bynum et al. 2020).

However, those studies commonly use older data and/or are limited to subsets of either the fee-for-service (FFS) Medicare population or members of Medicare Advantage Plans, which is important as the decision to enroll into a Medicare Advantage Plan is not random(Mirel, Wheatcroft et al. 2012, Nicholas 2013). More importantly, most prior studies are confined to identifying the prevalence of dementia diagnoses, and – to our knowledge – no analysis has looked into the gap at the stage of MCI in the full Medicare population.

In summary, there is substantial evidence that cognitive impairment remains underdiagnosed, in particular in early stages. Our analyses below confirm such prior evidence and also point to substantial disparities in diagnosis rates. Measuring and reporting diagnosis rates in primary care and comparing those to expected rates given the

demographic composition of a clinician's or practice's panel can identify gaps in diagnosis and point primary care clinicians towards efforts to proactively inquire about cognitive concerns and follow up on subjective memory complaints, in particular in high-risk and disadvantaged populations.

References

(Sep 28, 2022). "Lecanemab confirmatory phase 3 clarity AD study met primary endpoint, showing highly statistically significant reduction of clinical decline in large global clinical study of 1,795 participants with early Alzheimer's disease. Eisai News Release." Retrieved Sep 29, 2022, from https://www.eisai.com/news/2022/news202271.html.

Thoits, T., A. Dutkiewicz, S. Raguckas, M. Lawrence, J. Parker, J. Keeley, N. Andersen, M. Vandyken and M. Hatfield-Eldred (2018). "Association Between Dementia Severity and Recommended Lifestyle Changes: A Retrospective Cohort Study." American Journal of Alzheimer's Disease & Other Dementias® **33**(4): 242-246.

Ngandu, T., J. Lehtisalo, A. Solomon, E. Levälahti, S. Ahtiluoto, R. Antikainen, L. Bäckman, T. Hänninen, A. Jula, T. Laatikainen, J. Lindström, F. Mangialasche, T. Paajanen, S. Pajala, M. Peltonen, R. Rauramaa, A. Stigsdotter-Neely, T. Strandberg, J. Tuomilehto, H. Soininen and M. Kivipelto (2015). "A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial." The Lancet **385**(9984): 2255-2263.

Dubois, B., A. Padovani, P. Scheltens, A. Rossi and G. Dell'Agnello (2015). "Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges." Journal of Alzheimer's Disease **49**(3): 617-631.

Cummings, J. and S. Salloway (2021). "Aducanumab: Appropriate use recommendations." Alzheimer's & Dementia.

White, L., B. Ingraham, E. Larson, P. Fishman, S. Park and N. B. Coe (2021). "Observational study of patient characteristics associated with a timely diagnosis of dementia and mild cognitive impairment without dementia." Journal of General Internal Medicine.

Savva, G. M. and A. Arthur (2015). "Who has undiagnosed dementia? A cross-sectional analysis of participants of the Aging, Demographics and Memory Study." Age and Ageing **44**(4): 642-647.

Taylor, D. H., Jr., G. G. Fillenbaum and M. E. Ezell (2002). "The accuracy of medicare claims data in identifying Alzheimer's disease." J Clin Epidemiol 55(9): 929-937

Lee, M., E. Whitsel, C. Avery, T. M. Hughes, M. E. Griswold, S. Sedaghat, R. F. Gottesman, T. H. Mosley, G. Heiss and P. L. Lutsey (2022). "Variation in Population Attributable Fraction of Dementia Associated With Potentially Modifiable Risk Factors by Race and Ethnicity in the US." JAMA Network Open **5**(7): e2219672.

Zhu, Y., Y. Chen, E. M. Crimmins and J. M. Zissimopoulos (2021). "Sex, Race, and Age Differences in Prevalence of Dementia in Medicare Claims and Survey Data." The Journals of Gerontology: Series B **76**(3): 596-606.

Jutkowitz, E., J. P. W. Bynum, S. L. Mitchell, N. M. Cocoros, O. Shapira, K. Haynes, V. P. Nair, C. N. McMahill-Walraven, R. Platt and E. P. McCarthy (2020). "Diagnosed prevalence of Alzheimer's disease and related dementias in Medicare Advantage plans." Alzheimers Dement (Amst) **12**(1): e12048.

Mirel, L. B., G. Wheatcroft, J. D. Parker and D. M. Makuc (2012). "Health characteristics of Medicare traditional fee-for-service and Medicare Advantage enrollees: 1999-2004 National Health and Nutrition Examination Survey linked to 2007 Medicare data." Natl Health Stat Report(53): 1-12.

Natalia Festa, M. P., Max Weiss, Lidia M. V. R. Moura, Nicole M. Benson, Sahar Zafar, Deborah Blacker, Sharon-Lise T. Normand, Joseph P. Newhouse, and John Hsu (2022). "Evaluating The Accuracy Of Medicare Risk Adjustment For Alzheimer's Disease And Related Dementias." Health Affairs **41**(9): 1324-1332.

Nicholas, L. H. (2013). "Better Quality of Care or Healthier Patients? Hospital Utilization by Medicare Advantage and Fee-for-Service Enrollees." Forum for Health Economics and Policy **16**(1): 137-161.

Petersen, R. C., O. Lopez, M. J. Armstrong, T. S. D. Getchius, M. Ganguli, D. Gloss, G. S. Gronseth, D. Marson, T. Pringsheim, G. S. Day, M. Sager, J. Stevens and A. Rae-Grant (2018). "Practice guideline update summary: Mild cognitive impairment." Neurology **90**(3): 126-135.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

The first two tables show the ratings for individual clinicians and practices, respectively. Those ratings are based on a calculation of 95% confidence intervals around each unit of measurement's O/E ratio, which allows us to determine whether the O/E ratio is within range, too low or too high.

For MCI, almost all clinicians and practices score too low and around a tenth of a percent within range. Dementia diagnosis rates are within range for 72% of clinicians and 60% of practices and 7% and 11% too high. As expected, O/E ratio are close to the weighted average of the two components.

Those findings underscore the large gap in diagnosis of MCI with better, albeit variable as shown later, diagnosis rates for dementia. Any stage cognitive impairment O/E ratios are too low for around 90% of clinicians and practices.

*	Result for O/E ratio	Frequency	Percent
MCI	Too low	226,420	99.91%
*	Within range	159	0.07%
*	Too high	40	0.02%
Dementia	Too low	46,751	20.63%
*	Within range	163,238	72.03%
*	Too high	16,630	7.34%
Either stage cognitive impairment	Too low	203,003	89.58%
*	Within range	22,166	9.78%
*	Too high	1,450	0.64%

Frequency and percentage of diagnosis rates that are too low, within range or too high: Clinician level ratings (Showing that MCI and either stage cognitive impairment are sufficiently diagnosed only by a small proportion of clinicians, whereas around three-quarters diagnosed dementia at the expected rate.)

*Cell intentionally left empty.

*	Result for O/E ratio	Frequency	Percent
MCI	Too low	54,461	99.85 %
*	Within range	63	0.12 %
*	Too high	17	0.03%
Dementia	Too low	15,771	28.92%
*	Within range	32,575	59.73%
*	Too high	6,195	11.36%

*	Result for O/E ratio	Frequency	Percent
Either stage cognitive impairment	Too low	50,407	92.42%
*	Within range	3,812	6.99%
*	Too high	322	0.59%

Frequency and percentage of diagnosis rates that are too low, within range or too high: Practice level ratings (Showing that MCI and either stage cognitive impairment are sufficiently diagnosed only by a small proportion of practices, whereas over half diagnosed dementia at the expected rate.)

*Cell intentionally left empty.

The next table shows the population-wide results over time to illustrate the limited change.

The subsequent tables show the full distribution of the clinician and practice-level results. We also provide the full histograms for the O/E ratios of the three measures.

In the interest of parsimony, we show only detailed distribution data for the most recent 2017-2019 window but the results for the other windows are similar.

*		Observation period	2015-2017	2016-2018	2017-2019
*		Sample size	38,739,387	39,965,446	41,205,474
MCI		Predicted rate	0.2237	0.2207	0.2175
		O/E ratio	0.059	0.067	0.074
Dementia		Predicted rate	0.128	0.1245	0.121
		O/E ratio	0.84	0.86	0.88
Any cognitive impairment		Predicted rate	0.2794	0.2752	0.2708
		O/E ratio	0.43	0.44	0.45

Ratio of observed to expected diagnosis rates in Medicare claims (Suggesting limited changes in detection rates over time)

*Cell intentionally left empty.

Measure	Statistic	MCI	Dementia	Any stage cognitive impairment
Observed rate	Mean	0.0153	0.0975	0.1128
	Standard deviation	0.0190	0.0573	0.0624
	95th percentile	0.0495	0.1977	0.2212
	75th percentile	0.0227	0.1222	0.1400
	Median	0.0106	0.0898	0.1053

Measure	Statistic	MCI	Dementia	Any stage cognitive impairment
	25th percentile	0.0000	0.0619	0.0741
	5th percentile	0.0000	0.0233	0.0303
Expected rate	Mean	0.2147	0.1195	0.2674
	Standard deviation	0.0718	0.0651	0.0874
	95th percentile	0.3765	0.2631	0.4624
	75th percentile	0.2371	0.1380	0.2969
	Median	0.1926	0.1004	0.2427
	25th percentile	0.1682	0.0785	0.2104
	5th percentile	0.1398	0.0533	0.1715
O/E ratio	Mean	0.0760	0.8905	0.4291
	Standard deviation	0.0961	0.4812	0.2106
	95th percentile	0.2503	1.7078	0.7874
	75th percentile	0.1127	1.1443	0.5385
	Median	0.0506	0.8478	0.4130
	25th percentile	0.0000	0.5692	0.2933
	5th percentile	0.0000	0.2227	0.1287

Ratio of observed to expected diagnosis rates in Medicare claims for individual clinicians (Suggesting that over a quarter of clinicians do not diagnose any of their expected MCI cases, slightly less than half diagnose more dementia cases than expected and the average detection rate of any stage cognitive impairment is 43%.)

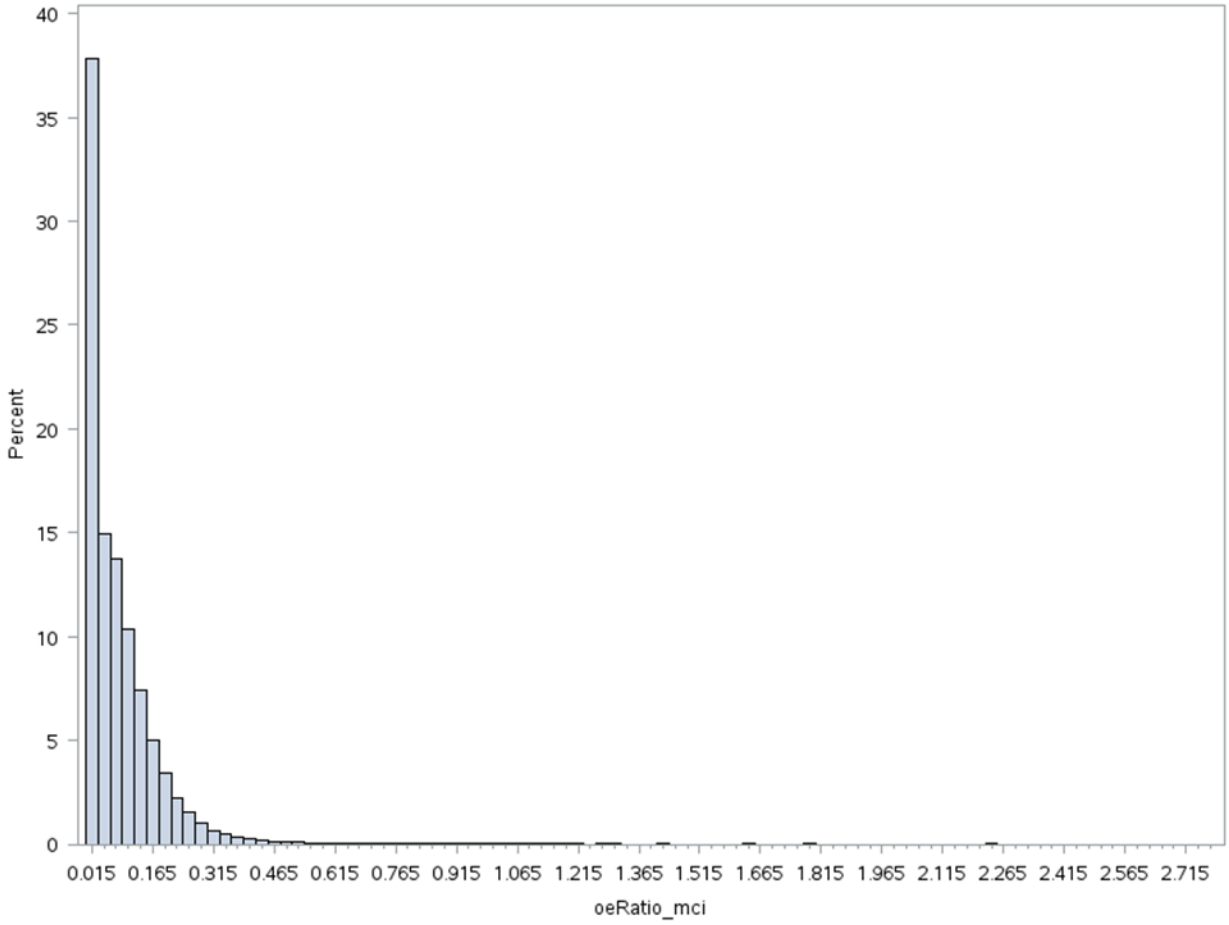
*Cell intentionally left empty.

Measure	Statistic	MCI	Dementia	Any stage cognitive impairment
Observed rate	Mean	0.0152	0.1081	0.1233
	Standard deviation	0.0181	0.0631	0.0674
	95th percentile	0.0800	0.2283	0.2500
	75th percentile	0.0209	0.1323	0.1496
	Median	0.0113	0.0974	0.1120
	25th percentile	0.0026	0.0692	0.0818
	5th percentile	0.0000	0.0286	0.0366
Expected rate	Mean	0.2291	0.1323	0.2849
	Standard deviation	0.0809	0.0751	0.0985
	95th percentile	0.4027	0.2964	0.4950
	75th percentile	0.2640	0.1594	0.3300
	Median	0.2016	0.1076	0.2540
	25th percentile	0.1734	0.0821	0.2170

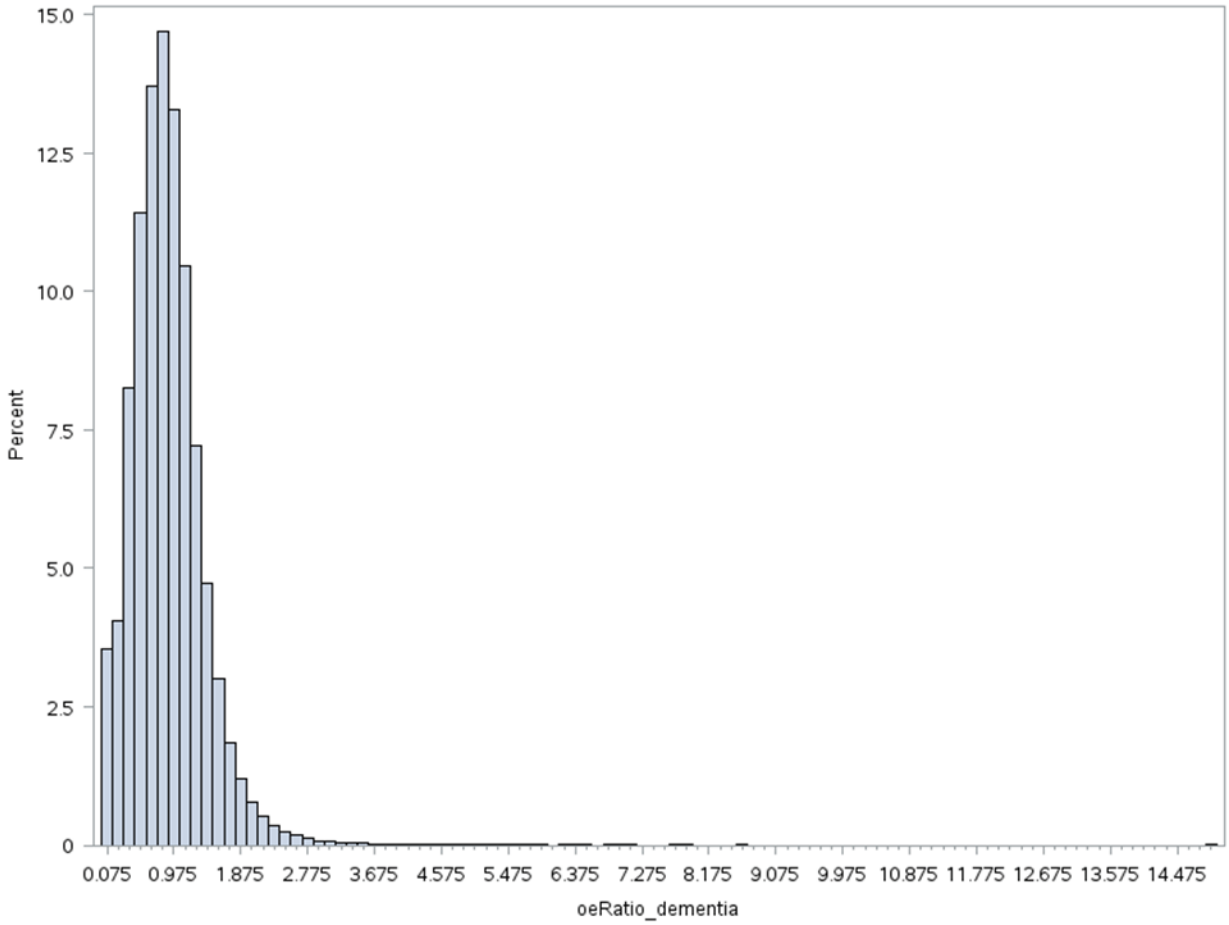
Measure	Statistic	MCI	Dementia	Any stage cognitive impairment
	5th percentile	0.1413	0.0540	0.1730
O/E ratio	Mean	0.0665	0.8167	0.4328
	Standard deviation	0.2242	0.8405	0.6844
	95th percentile	0.1987	0.7702	0.5051
	75th percentile	0.0793	0.8300	0.4533
	Median	0.0560	0.9052	0.4409
	25th percentile	0.0150	0.8426	0.3771
	5th percentile	0.0000	0.5288	0.2117

Ratio of observed to expected diagnosis rates in Medicare claims for practices (Suggesting that over five percent of practices do not diagnose any of their expected MCI cases, most practices diagnose approximately as many dementia cases as expected and the average detection rate of any stage cognitive impairment is 44%.)

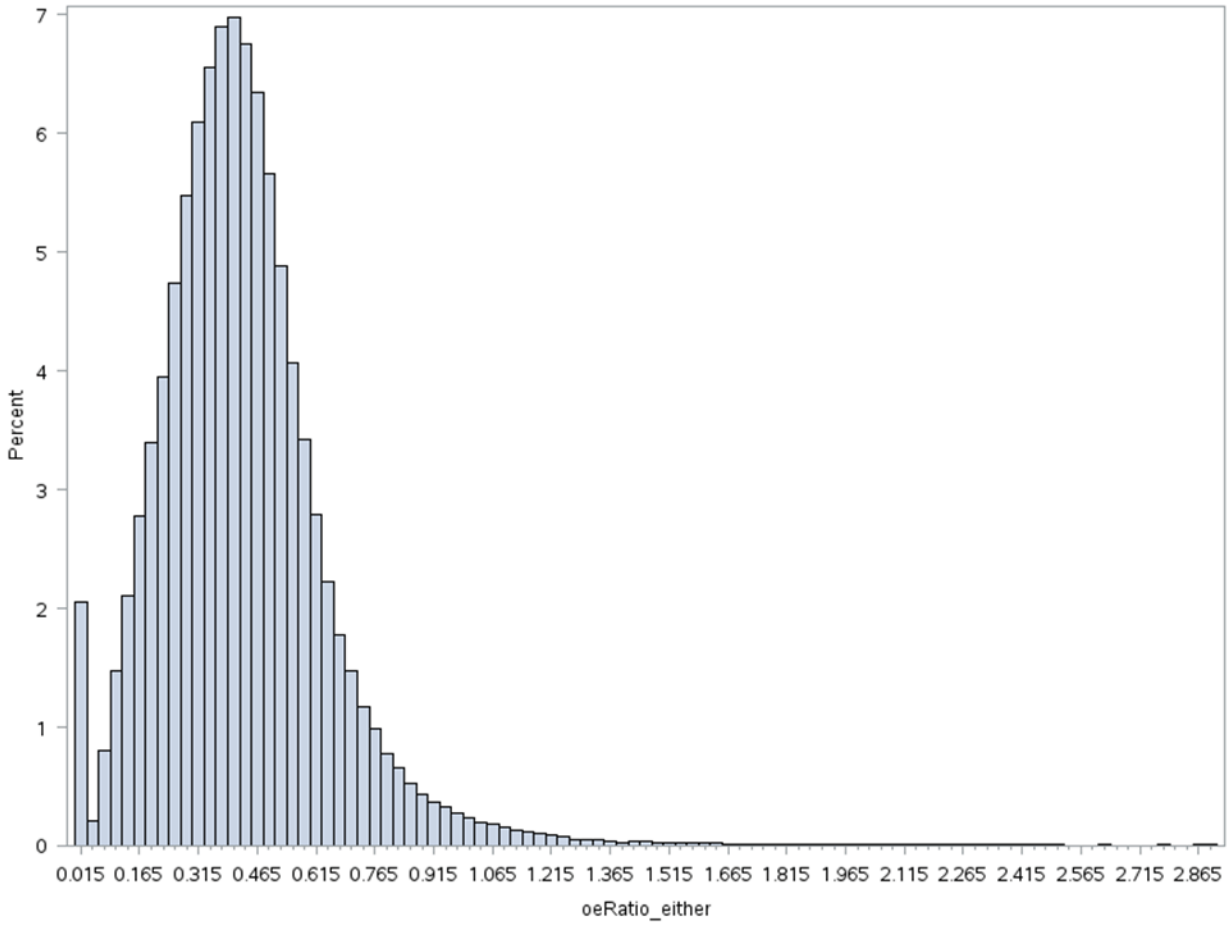
*Cell intentionally left empty.



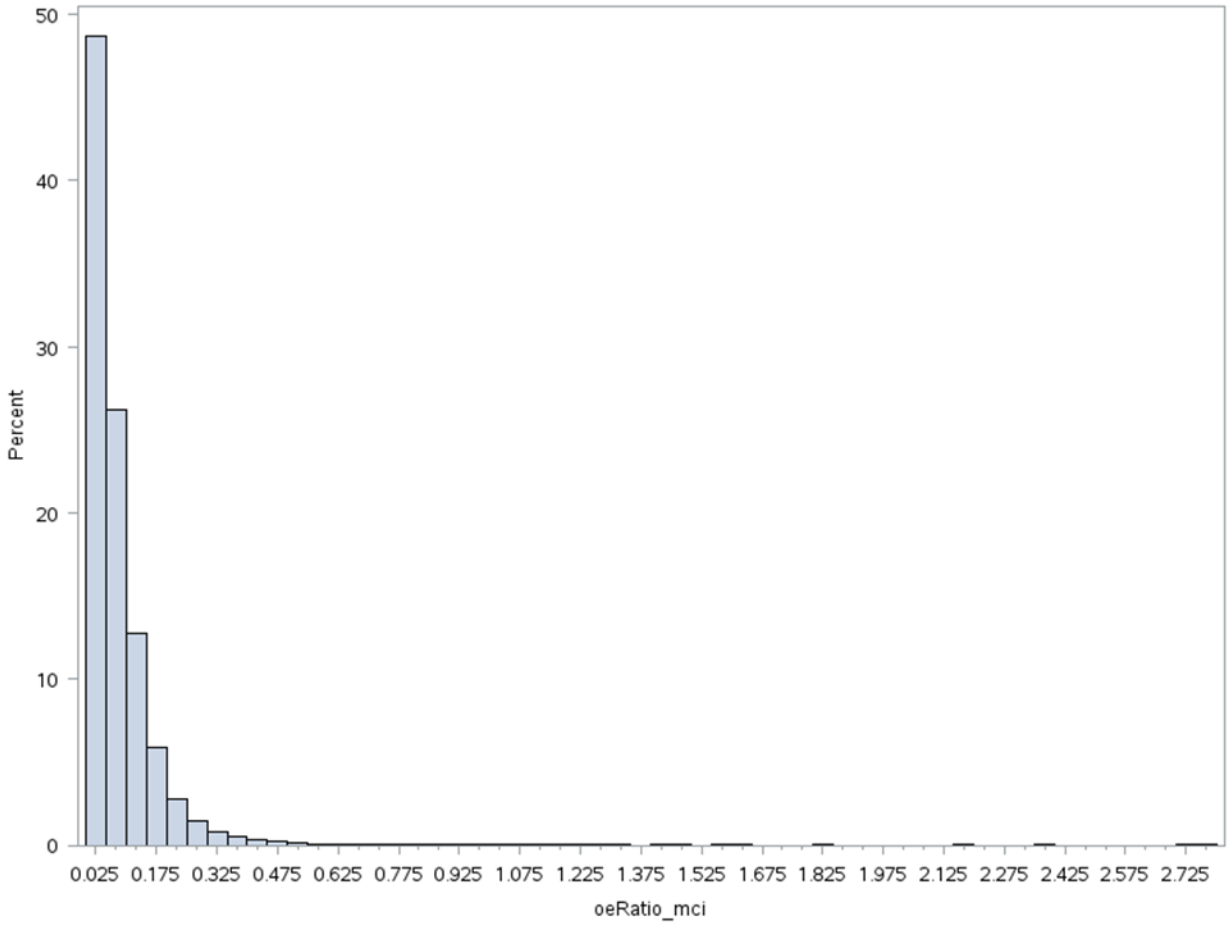
Clinician level distribution of O/E ratios: MCI



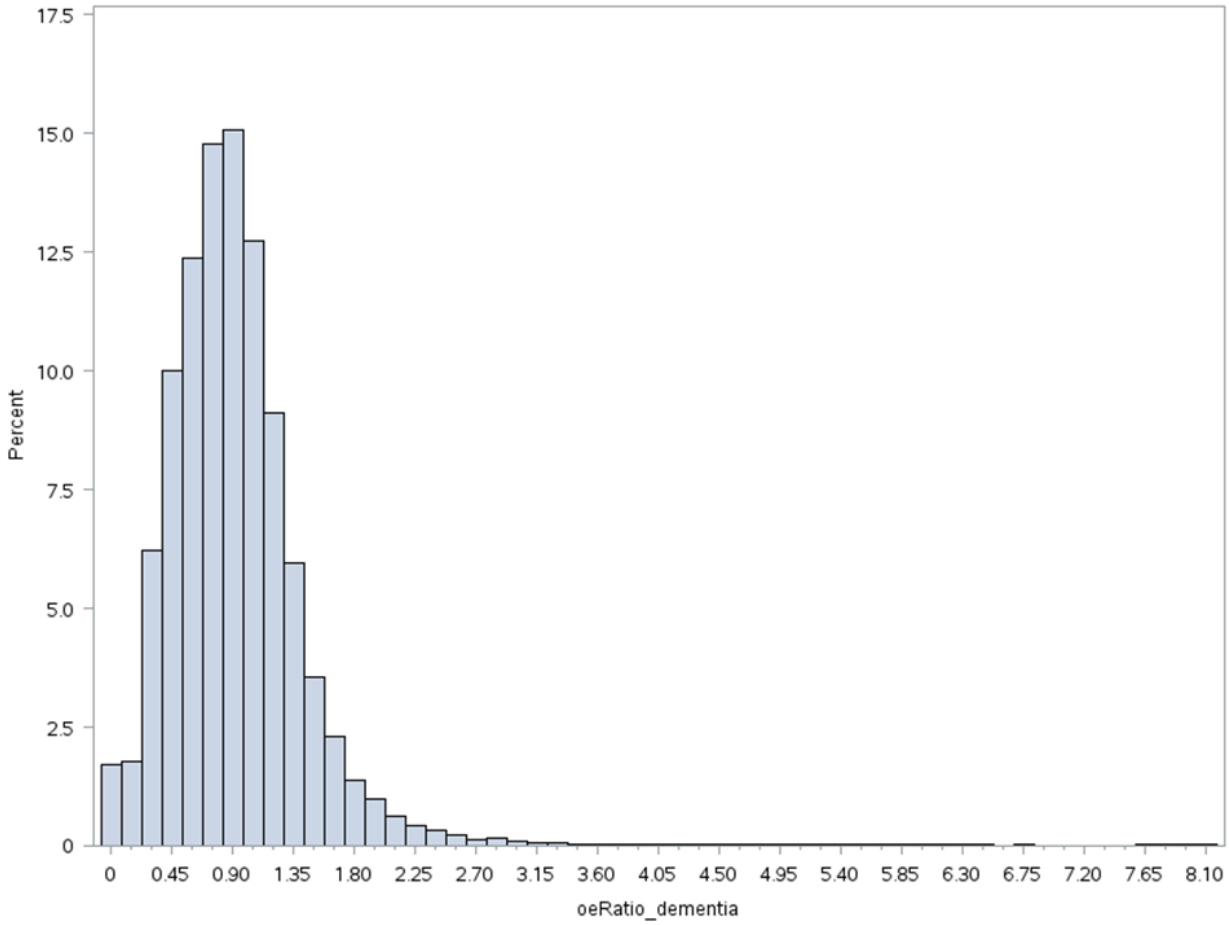
Clinician level distribution of O/E ratios: Dementia



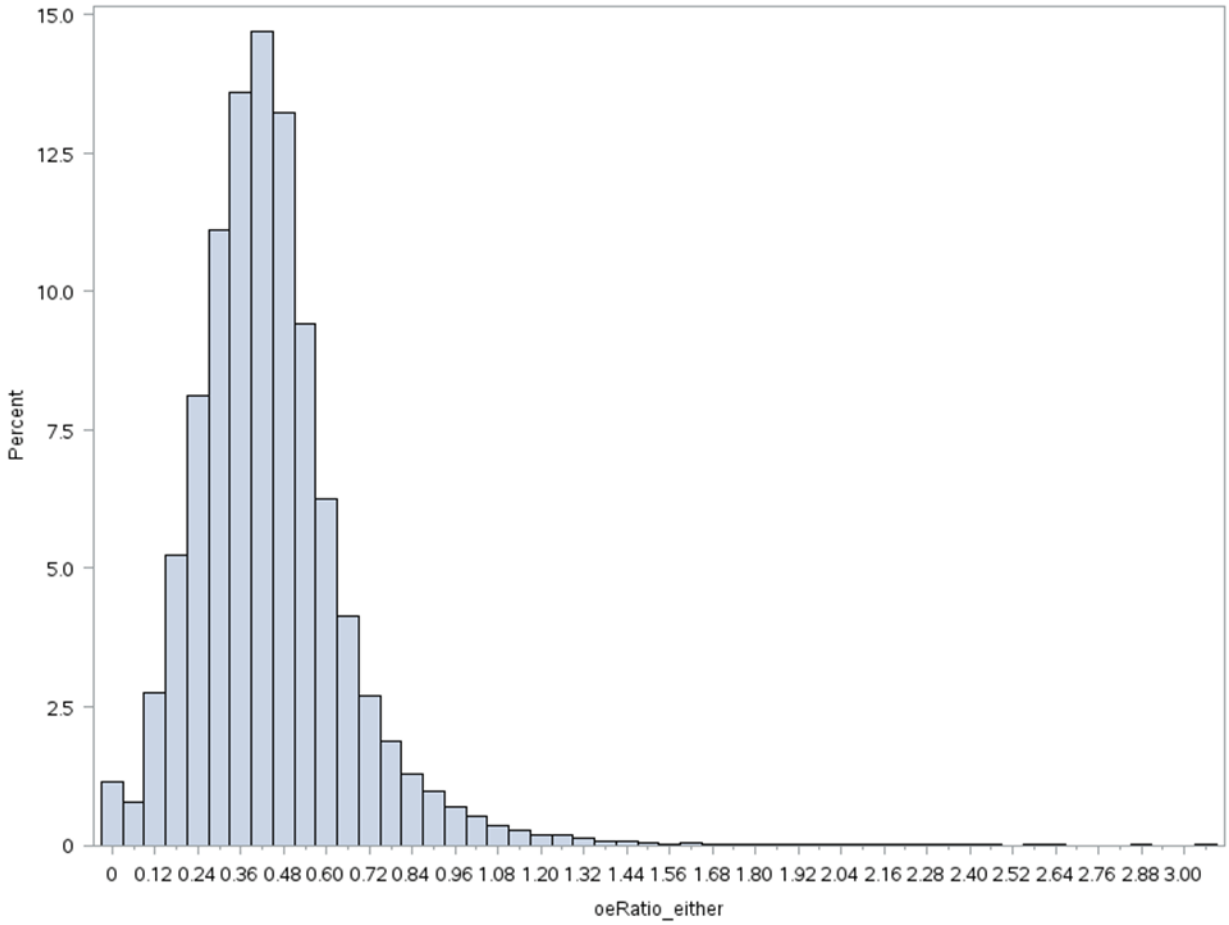
Clinician level distribution of O/E ratios: Either stage cognitive impairment



Practice level distribution of O/E ratios: MCI



Practice level distribution of O/E ratios: dementia



Practice level distribution of O/E ratios: Either stage cognitive impairment

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]

As stated under "rationale, our findings confirm gaps in diagnosis rates, in particular for early-stage cognitive impairment.

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

*Cell intentionally left empty.Our data show the analysis of disparities in measure rates for the most recent 2017-2019 window. Patterns for the other windows are similar.

We observe increasing diagnosis rates relative to expected rates by age, except for the oldest cohort that still do not increase in line with the increasing burden of this ageing related disorder. Women are slightly more likely to be diagnosed than men. There is a striking difference in observed over expected diagnosis rates for all three measures by race and ethnicity in that the O/E rates are about twice as high in white compared to black and Hispanic individuals. Similar disparities exist for the highly vulnerable group of dually eligible beneficiaries, who have around half of the O/E ratios of Medicare only beneficiaries for MCI and dementia. O/E ratios are similar in Medicare fee-for-service and Medicare Advantage.

*	Age group	Age 65-69	Age 70-74	Age 75-79	Age 80-84	Age 85+
*	Sample size	8,913,079	12,030,038	8,682,044	5,750,802	5,829,511
MCI	Predicted rate	0.1235	0.1554	0.2145	0.2924	0.4203
	O/E ratio	0.065	0.0737	0.0844	0.0824	0.0658
Dementia	Predicted rate	0.0378	0.0546	0.0992	0.1787	0.3607
	O/E ratio	0.7763	0.8314	0.9018	0.9155	0.876
Any cognitive impairment	Predicted rate	0.1447	0.1847	0.2641	0.3719	0.5517
	O/E ratio	0.2583	0.3076	0.4073	0.5046	0.6228

Ratio of observed to expected diagnosis rates in Medicare claims, 2017-2019, by age group (Suggesting that detection rates of cognitive decline increase with age, except for the oldest cohort of age 85 and older

*Cell intentionally left empty.

*	Sex	Female	Male
*	Sample size	17,796,969	23,408,505
MCI	Predicted rate	0.2151	0.2207
	O/E ratio	0.0768	0.0713
Dementia	Predicted rate	0.1317	0.1068
	O/E ratio	0.8986	0.8394
Any cognitive impairment	Predicted rate	0.273	0.268
	O/E ratio	0.494	0.3934

Ratio of observed to expected diagnosis rates in Medicare claims, 2017-2019, by sex (Suggesting that detection rates are slightly higher in women than in men)

*Cell intentionally left empty.

*	Race/ethnicity	White	Black	Hispanic	Other
*	Sample size	31,701,890	3,553,686	3,489,177	2,460,721
MCI	Predicted rate	0.1733	0.4165	0.3952	0.2484
	O/E ratio	0.0968	0.0328	0.0425	0.0455
Dementia	Predicted rate	0.0876	0.2784	0.2392	0.1559

*	Race/ethnicity	White	Black	Hispanic	Other
	O/E ratio	1.181	0.48	0.5027	0.5019
Any cognitive impairment	Predicted rate	0.2204	0.4992	0.4664	0.3136
	O/E ratio	0.5455	0.2951	0.2937	0.2855

Ratio of observed to expected diagnosis rates in Medicare claims, 2017-2019, by race/ethnicity (Showing that detection rates for cognitive decline are much lower in Black and Hispanic than in White individuals)

*Cell intentionally left empty.

*	Medicaid eligibility	Dually eligible	Medicare only
*	Sample size	5,102,819	36,102,655
MCI	Predicted rate	0.4333	0.187
	O/E ratio	0.0425	0.0848
Dementia	Predicted rate	0.3807	0.0842
	O/E ratio	0.602	1.051
Any cognitive impairment	Predicted rate	0.5598	0.23
	O/E ratio	0.4424	0.4539

Ratio of observed to expected diagnosis rates in Medicare claims, 2017-2019, by Medicaid eligibility (Showing that detection rates for cognitive decline are lower in dually eligible than in Medicare-only beneficiaries)

*Cell intentionally left empty.

*	Type of Medicare coverage	FFS	MAP
*	Sample size	22,957,446	18,248,028
MCI	Predicted rate	0.2058	0.2323
	O/E ratio	0.0802	0.0679
Dementia	Predicted rate	0.1128	0.1312
	O/E ratio	0.9510	0.7949
Any cognitive impairment	Predicted rate	0.2576	0.2875
	O/E ratio	0.4806	0.4177

Ratio of observed to expected diagnosis rates in Medicare claims, 2017-2019, by type of Medicare coverage (Showing similar detection rates in traditional Medicare and Medicare Advantage Plans.)

*Cell intentionally left empty.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

n/A

[Response Ends]

2. Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

---

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

Ratio of observed over predicted rates for diagnosis of mild cognitive impairment

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

Ratio of the number of patients 65 and older diagnosed with mild cognitive impairment attributed to a clinician or practice over the number predicted based on the demographic profile of that clinician or practice.

Once the clinician's or practice's O/E ratio (i.e., ratio of the observed and expected rates) is calculated, a computation of its associated standard error (SE) can be used to draw inference whether the O/E ratio is significantly different from 1 or not.

[Response Ends]

sp.03. Provide a rationale for why this measure must be reported with other measures to appropriately interpret results.

[Response Begins]

The exact classification of the stage of cognitive impairment is difficult, as it requires not only neurocognitive testing but also a comprehensive evaluation of the patient's health status and functional abilities. Further, the stage may change over our three-year observation window because of natural disease progression on the one hand, and improvement on the other. For these reasons we recommend determining the overall diagnosis counts for any cognitive impairment (3729) first, then the counts for mild cognitive impairment (3707) and dementia (3672) as subsets.

All three should be compared to the predicted rates based on our model to form the respective ratios for observed versus predicted numbers. A ratio of close to 1.0 suggests that a clinician or practice has approximately as many cases diagnosed as expected. The ratio for 3729 shows the overall detection rate, and ratios for 3707 and 3672 provide information on how detection rates differ by stage. The three measures capture risk-adjusted rates of diagnoses of different stages of cognitive decline, i.e., reflect diagnostic quality along the continuity of the disease process. They should therefore be interpreted together.

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Neurology: Alzheimer's Disease

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Disparities Sensitive

Health and Functional Status

Screening

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Elderly (Age >= 65)

Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid

Populations at Risk: Individuals with multiple chronic conditions

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Clinician: Group/Practice

Clinician: Individual

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Ambulatory Care

[Response Ends]

sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".

[Response Begins]

N/A

[Response Ends]

sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

No data dictionary/code table – all information provided in the submission form

[Response Ends]

Attachment: 3707 _3707 _3707 _predictive model coefficients-508.xlsx

sp.13. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]

Number of individuals age 65 and older under the care of a clinician or practice who are diagnosed with mild cognitive impairment

[Response Ends]

sp.14. Provide details needed to calculate the numerator.

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

Follow the algorithms of the Chronic Condition Warehouse that is used to identify persons diagnosed with dementia, which uses a three-year observation window, i.e., the numerator is reported over rolling three-year periods. Diagnosis based on ICD-10-CM code G31.84 in any position on one inpatient or SNF claim or two claims on separate days on any other claim.

Only individuals who meet the denominator inclusion criteria are included in the numerator.

[Response Ends]

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

Predicted number of individuals aged 65 and older under the care of a clinician or practice with mild cognitive impairment based on the demographic profile of the respective clinician or practice. Limit reporting to clinicians and practices with at least 25 attributed patients per CMS' guidance to ensure stability of measure results.

[Response Ends]

sp.16. Provide details needed to calculate the denominator.

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

Identify individuals with near-continuous Medicare coverage for three years. Following the coverage definition used by the Chronic Condition Warehouse (CCW), our definition of nearly continuous enrollment requires an average of 11 months of both Part A and B or Part C coverage each year (at least 33 out of a possible 36 months) or, if the beneficiary died during the third year of the surveillance period, fully continuous Part A and B or Part C coverage with no interruption up until the month of death. The continuous enrollment requirement means that individuals will have to reach age 67 at a minimum to be included in the denominator.

Apply estimated weights from predictive model to enrollment data. Choose mid-point year of the rolling three-year window for calculation for time-variant variables. The weights are multipliers for each individual-level categorical variable for age group, sex, race/ethnicity and dual eligibility as well as the linear trend. Add the estimated constant.

The linear combination of those weights is then the individual's predicted risk of having mild cognitive impairment. The sum of the predicted probabilities for the individuals attributed to a clinician or practice is their expected number of cases with mildcognitive impairment.

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

The measures are not using any exclusions as they are based on the 100% samples for both Medicare fee-for-service and Medicare Advantage Plans. While we limit ***reporting*** of the measure to clinicians and practices with at least 25 attributed patients, this does not constitute an exclusion per NQF guidance, since those patients might be reported when reporting on higher levels of aggregation, such as a state. We merely follow CMS' recommendations for minimum sample size to report stable results.

[Response Ends]

sp.18. Provide details needed to calculate the denominator exclusions.

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

Set reporting threshold to 25 or more patients attributed to clinician or practice

[Response Ends]

sp.19. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]

N/A

[Response Ends]

sp.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

Statistical risk model

[Response Ends]

sp.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Ratio

[Response Ends]

sp.23. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Score within a defined interval

[Response Ends]

sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

1. Define three-year observation period
    - Rolling window for longitudinal comparison
2. Attribute Medicare beneficiaries to individual primary care clinicians based on NPI
    - Attribution based on plurality of office visits over the three-year observation period
    - In case of ties, use multiple attribution for shared responsibility assumption
3. Identify number of patients diagnosed with cognitive impairment of any stage, mild cognitive impairment and dementia for each clinician
    - Based on ICD-10 diagnosis codes provided
4. Calculate expected number of patients with cognitive impairment of any stage, mild cognitive impairment and dementia for each clinician
    - Based on weights from predictive model
5. Calculate ratio of observed (step 3) over expected (step 4) number, which is the O/E ratio
6. Roll up clinicians and their observed and expected numbers to practices
    - Based on the Tax ID under which they report the majority of their claims
7. Calculate practice-level O/E ratios
8. Limit further analysis and reporting to clinicians/practices with at least 25 attributed patients
9. Compute standard error (SE) associated with O/E ratio to draw inference whether the O/E ratio is significantly different from 1 or not.

Note on Step 2: Identification of primary care clinicians is described under Reliability Testing

To clarify Step 8: The measure should be calculated for all eligible persons, but the reporting needs to follow guidelines for minimum denominator sizes, which could change. Also, a patient could be attributed to a physician, who does not meet the minimum denominator threshold but is part of a practice that does. Also, while the

measure is intended to be used at the clinician/practice level, it might get reported at higher levels of aggregation, like health plan or state. Therefore, restrictions are imposed on reporting and not on calculation of the measure.

To clarify Step 9: Once the clinician's or practice's O/E ratio (i.e., ratio of the observed and expected rates) is calculated, a computation of its associated standard error (SE) can be used to draw inference whether the O/E ratio is significantly different from 1 or not.

As in our proposed calculation for the O/E ratio, we assume the denominator "E" is error free or ignorable, and most of the sampling error comes from the numerator "O", which reflects what the clinician does. To compute the variance of "O", we follow Adams' (2010) approach by assuming that: (1) each observation's (in this case, observed diagnosis on a patient attributed to this physician) sampling error is a function of the patient's demographic information; (2) different observations are independent to one another (i.e., how a physician diagnoses a patient has nothing to do with how the same physician diagnoses other patients); and (c) the total sampling error of this physician's observed rate is a function of the number of patients he/she has and the demographic composition of his/her patient pool.

Therefore, the SE of the O/E ratio, and thus the 95% confidence interval of the O/E ratio, can be computed using the following steps:

1. Compute each observation's sampling variance as p*(1-p), where p is the population-level observed rate for the demographic combination that the patient has (e.g., female, age 65-69, black, non-dual). These rates can be found in the "Additional" section.
2. Sum the quantities from step 1 across all observations attributed to this specific NPI or practice, and divide the sum by squared sample size (i.e., square of the total number of patients attributed to this physician or practice). This quantity is essentially the variance of the O rate, or var(O).
3. Dividing the quantity from step 2 by square of the E rate gives the variance of the O/E rate, i.e., var(O/E) = var(O)/(E^2).
4. Taking the square root from the quantity from step 3 gives the SE of the O/E ratio.
5. Multiplying the quantity from step 4 by 1.96 gives the margin of error assuming 95% confidence level.
6. Thus, the 95% confidence interval can be constructed as the O/E ratio of the physician plus or minus the quantity derived in step 5. If this range covers the value of 1, we consider this physician's observed diagnosis rate is similar to what's expected; when this interval is entirely above 1, it suggests the physician is significantly over diagnosing; if the interval is entirely below 1, the physician is underperforming.

If the measures were endorsed, we would develop a tool into which one would only have to enter an enrollment table for the clinician's or practice's panel to automate the calculation.

Reference

Adams JL, McGlynn EA, Thomas JW, Mehrotra A. Incorporating statistical uncertainty in the use of physician cost profiles. *BMC Health Services Research*. 2010;10(1):57. doi:10.1186/1472-6963-10-57

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

• Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.

• The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited

generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.

• The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.

• When possible, units of measurement and patients within units should be randomly selected.

[Response Begins]

The measure is based on the 100% Medicare data, no sampling

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

[Response Ends]

sp.31. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

Medicare RIF files for traditional Medicare and Medicare Advantage Plans, 100% files, carrier, inpatient, outpatient, SNF

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.

- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Claims

[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]

Medicare fee-for-service and Advantage Plan claims and enrollment data

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

01-01-2017 - 12-31-2019

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Clinician: Group/Practice

Clinician: Individual

[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]

The total number of clinicians and practices that have at least 25 attributed patients are 319,522 clinicians and 88,532 practices, respectively.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

The demographic composition of the Medicare beneficiaries in the sample is shown in the table below. The numbers reflect only those who were attributed to a clinician or practice with at least 25 patients.

*	*	Clinician level	Practice level
*	Total sample	26,870,191	36,167,092
Sex	Percent Female	57.79%	57.89%
Race/Ethnicity	Percent white	78.30%	78.28%
Race/Ethnicity	Percent black	8.19%	8.24%
Race/Ethnicity	Percent Hispanic	7.65%	7.61%
Age group	<70	21.33%	21.43%
Age group	70-74	29.19%	29.19%
Age group	75-79	21.47%	21.42%
Age group	80-84	14.20%	14.16%
Age group	>85	13.80%	13.79%
*	Percent dually eligible	11.49%	11.61%

Demographic composition of the Medicare beneficiaries with respect to sex, race/ethnicity, age group and dual eligibility for Medicare and Medicaid included in the measure; numbers reflect only those who were attributed to a clinician or practice with at least 25 patients.

*cell intentionally left blank

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]

all analyses are restricted to clinicians and practices with at least 25 attributed patients

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]

N/A

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

We tested reliability of the set of three measures at the individual clinician or clinician-group/practice. As we restrict reporting to clinicians and practices with at least 25 attributed patients, following CMS' guidance for minimum sample size to achieve stable results, reliability testing and subsequent validity testing uses that sample of clinicians and practices as well. Initially, we attributed to any clinician but saw that this approach leads to implausibly high O/E ratios, particularly for dementia. As the below table shows, those ratios with a maximum of 17.60 are usually observed in neurologists and psychiatrists. Those clinicians have a low expected but very high observed rate, leading to the outlier ratios, and are most likely memory care specialists (Table 1). We also discovered multiple implausible attributions to other specialists, for example to dermatologists because a patient received monthly injection treatment for psoriasis, or to podiatrists for regular foot care (Table 2). We therefore decided to attribute to primary care clinicians only, with the exception of general gynecologists, whom many women use as primary source of care, in keeping with the philosophy of the measure to attribute patients to the main overall source of care.

Number of attributed patients	Observed dementia rate	Expected dementia rate	O/E ratio dementia	Specialty
28	0.89	0.05	17.60	Neurology
29	0.72	0.05	16.03	Neurology
35	0.83	0.05	15.33	Neurology
27	0.41	0.03	13.95	Neurology
59	0.85	0.06	13.05	Neurology
56	0.86	0.07	12.06	Neurology
38	0.68	0.06	11.23	Physical medicine and rehabilitation
28	0.61	0.05	11.15	Neurology
26	0.88	0.08	11.12	Neurology
74	0.97	0.09	10.73	Nurse practitioner
82	0.85	0.08	10.71	Neurology
35	0.89	0.08	10.59	Neurology
64	0.77	0.07	10.35	Physician assistant
25	0.48	0.05	10.33	Neurology
37	0.35	0.04	9.91	Neurology

Number of attributed patients	Observed dementia rate	Expected dementia rate	O/E ratio dementia	Specialty
46	0.98	0.10	9.83	Geriatric medicine
31	0.39	0.04	9.74	Nurse practitioner
27	0.52	0.05	9.71	Neurology
25	0.72	0.07	9.69	Nurse practitioner
43	0.77	0.08	9.67	Psychiatry
469	0.74	0.08	9.61	Internal medicine
61	0.75	0.08	9.60	Neurology
27	0.33	0.03	9.57	Psychiatry
41	0.44	0.05	9.54	Psychiatry

Table 1: Specialty of physicians with highest O/E ratio (Illustrating that most clinicians with very high O/E rates are either brain specialists or mid-level clinicians that likely work in memory clinics)

Specialty	Frequency	Percent
Family practice	68434	21
Internal medicine	52963	16
Nurse practitioner	30551	9
Physician assistant	17599	5
Cardiovascular disease (cardiology)	15042	5
Orthopedic surgery	14483	4
Podiatry	10487	3
Dermatology	10060	3
Urology	7934	2
Ophthalmology	6918	2
Hematology/Oncology	6667	2
Neurology	6617	2
Otolaryngology	6428	2
Optometry	5094	2
Pulmonary disease	4939	2
Nephrology	4898	2
Gastroenterology	4537	1
Psychiatry	4496	1
Endocrinology	4168	1
Rheumatology	3727	1
General practice	3187	1
Interventional cardiology	3048	1

Specialty	Frequency	Percent
Physical medicine and rehabilitation	3046	1
General surgery	3000	1
Medical oncology	2652	1
Obstetrics/Gynecology	2332	1
Emergency medicine	1889	1
Vascular surgery	1670	1

Table 2: Specialty of physicians with at least 1% of all patients attributed (Explaining that patients are frequently attributed to clinicians are not likely to diagnose cognitive impairment, just because they happen to see them frequently.)

Lastly, we discovered that nurse practitioners, physician assistants and internists were among those outliers and learned in conversations with memory clinics that there is commonly a within-clinic specialization on care for patients with cognitive impairment in larger practices. We therefore excluded from attribution individual providers in whom the share of patients with cognitive impairment was greater than 50%. This sample restriction eliminated the 1% of clinicians with the highest O/E ratios for dementia. Table 3 summarizes the changes in the sample. While the number of included clinicians and patients decreased, the average O/E ratio remained largely unchanged.

	Attribution to all clinicians	Attribution to primary care only
Number of attributed patients	41,798,052	26,870,191
Number of included clinicians	331,179	175,968
Mean O/E rate MCI	0.08	0.08
Mean O/E rate dementia	0.94	0.89
Mean O/E rate MCI or dementia	0.46	0.43

Table 3: Comparison of attribution to any clinician to primary care only (Explaining that restriction of attribution to primary care clinicians only did not change O/E ratios meaningfully.)

1. Approach

We followed the standard method developed by RAND [1] for reliability testing. This method conceptualizes the reliability index as a function of the signal-to-noise ratio. Specifically, it assumed that each individual clinician or practice has a true state of performance (the signal), in other words, the extent to which the provider's true diagnosing capability is deviating from the expectation based on their patient pool. However, the observed performance by individual provider, the ratio of observed over expected diagnosed cases (O/E ratio), may vary due to factors irrelevant to their true quality of care (the noise). The amount of such noise can be assessed through intra-provider variability, with respect to variability across providers.

The reliability index is therefore expressed as a ratio between variability across individual clinician-level or practice-level's performances (i.e., the strength of the signal) and the total variability in the data (i.e., the strength of signal + noise). Reliability closer to 1 suggests that diagnostic performance of a clinician is consistent. Thus, noise is limited and the assessment of individual clinician-level quality is accurate. On the other hand, reliability closer to 0 implies that the patients attributed to the same individual clinician behave very differently, and thus the noise is dominant and the assessment of individual clinician-level quality is poor. Adams [2] suggested a reliability of 0.7 or higher to be considered acceptable when the observed clinician-profiling measure is used to infer an individual clinician's true performance. To compute the reliability, we adapt from the derivation by Adams [3] which specifically examines reliability testing in O/E ratio-type measures, leveraging a two-level Hierarchical Linear Model. While Adams' original proposal is based O/E ratio of two continuous variables, we adapt his formulas to the scenario when both O and E are rates bounded between 0 and 1. We require an individual clinician or practice to have at least 25 attributed patients to be included to qualify for the measure, consistent with Medicare reporting rules. Thus, this cutoff also applies to the reliability analysis.

1. Adams JL. *The Reliability of Provider Profiling: A Tutorial*. RAND Corporation; 2009.

2. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician Cost Profiling — Reliability and Risk of Misclassification. *New England Journal of Medicine*. 2010;362(11):1014-1021. doi:10.1056/nejmsa0906323

3. Adams JL, Mehrotra A, McGlynn EA. *Estimating Reliability and Misclassification in Physician Profiling*. RAND Corporation; 2010.Method of testing reliability

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

A. Individual Clinician level results

Table 1 shows the results of the reliability testing at the clinician level. The first column contains the average reliability, which is greater than 0.7 in all cases, suggesting adequate reliability of all three measures.

Measure	Mean	Standard deviation	5th percentile	25th percentile	Median	75th percentile	95% percentile
MCI	0.9870	0.0131	0.9589	0.9820	0.9921	0.9963	0.9986
Dementia	0.9865	0.0184	0.9516	0.9834	0.9932	0.9970	0.9990
Any stage cognitive impairment	0.9881	0.0114	0.9637	0.9833	0.9924	0.9964	0.9985

Table 1: Results of reliability analysis: individual clinicians (suggesting excellent reliability of all three measures)

A. Clinician-group/practice level results

Table 2 shows the results of the reliability testing at the clinician level. The first column contains the average reliability, which is close to 1.0 for all measures with very limited variability around the average estimate, suggesting high reliability of all three measures.

Measure	Mean	Standard deviation	5th percentile	25th percentile	Median	75th percentile	95% percentile
MCI	0.9911	0.0116	0.9662	0.9891	0.9956	0.9982	0.9996
Dementia	0.9905	0.0168	0.9616	0.9901	0.9963	0.9986	0.9997
Any stage cognitive impairment	0.9919	0.0099	0.9703	0.9898	0.9957	0.9982	0.9996

Table 2: Results of reliability analysis: practices (suggesting excellent reliability of all three measures)

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

The analysis suggests that all three measures can be reliably reported at the individual clinician and practice levels. We find that the measures well exceed established criteria for reliability, as expressed by the signal-to-noise ratio, based on the commonly accepted threshold of 0.7 and a method that follows the assessment method developed by Adams for O/E ratio type measures.

[Response Ends]

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

The measures capture the ratio of the number of patients attributed to a clinician or practice, who are diagnosed with cognitive impairment, over the number that would be expected based on the demographic composition of the clinician's or practice's panel. NQF#3729 is capturing any stage of cognitive impairment, and NQF#3707 and NQF#3672 mild cognitive impairment (MCI) and dementia, respectively, to give a better understanding of whether performance is better in early or later stages of cognitive decline. Better performance is a ratio of close to 1.0, indicating that approximately as many patients were diagnosed as expected.

Face validity testing

1. Technical approach

We identified nine clinical experts who are familiar with the diagnosis of cognitive impairment in primary care to rate the measures of the detection of cognitive impairment in primary care on their face validity, understandability, and usability using a web-based questionnaire (developed using SurveyMonkey®). One of the identified experts did not complete the rating by the time of submission. The names and affiliations of the eight clinical experts, who participated in the rating are listed in Table 1. These experts represent a range of years in practice, setting, and training.

Name	Affiliation
Samantha Cotler, DO, MBA	Medical Director, Family Medicine, AdventHealth Medical Group Core Faculty, Family Medicine Residency Program

Name	Affiliation
Fred Kobylarz, MD	Professor, Department of Family Medicine and Community Health Co-Director, Geriatric Fellowship Program Rutgers Robert Wood Johnson Medical School Medical Director, Parker at Monroe Nursing Care Residence
Michelle Panlilio, DNP, GNP-BC	National Lead Dementia Care Specialist, UCLA Alzheimer's Dementia Care Program
Tatiana Sadak, PhD, PMHNP, RN, FAAN, FGSA	Interim Associate Dean, Academic Affairs Director, Dementia Palliative Education Network University of Washington School of Nursing
Magdalena Stepien, MD	Neurologist, AdventHealth Medical Group
Diana Summanwar, MD	Assistant Professor of Family Medicine, Indiana School of Medicine
Po-Heng Tsai, MD	Behavioral Neurologist, Banner Alzheimer's Institute
Amy Walsh, MSc	Project Manager, Institute for Healthcare Improvement

Table 1: Expert panel members who conducted face validity and usability evaluation

The clinical experts were asked to review three measures: "ratio of observed over predicted rates for diagnosis of cognitive impairment of any stage," "ratio of observed over predicted rates for diagnosis of dementia," and "ratio of observed over predicted rates for diagnosis of mild cognitive impairment." After reviewing the background material, they were instructed to rate the face validity of the three measures at the clinician and practice level by indicating their degree of agreement on a 5-point scale (see Table 2).

Value	Description
1	Strongly Disagree
2	Disagree
3	Neither Agree nor Disagree
4	Agree
5	Strongly Agree

Table 2: Five-point Scale

A. Empirical validity testing

1. Technical approach

Our empirical assessment of the measures' validity had three components. We looked at the association with utilization of the Medicare Annual Wellness Visit, with the share of attributed patients aged 80 and older and with the attributed clinician being a geriatrician. Validity testing only includes clinicians and practices with at least 25 attributed patients, as explained in the reliability section.

a. Rationale for using Annual Wellness Visit

The Medicare Annual Wellness Visit (AWV) was introduced in 2011 to provide a broad range of preventive services without cost to beneficiaries, as the first comprehensive prevention benefit of the Medicare program. Uptake of the benefit, however, has been slow and AWV utilization increased from 8.1% to just 23.0% of all beneficiaries between 2011 and 2016. Uptake among minority beneficiaries, who have higher risk of cognitive impairment, was substantially lower(Lind, Hildreth et al. 2019). An assessment of cognitive function is a required component of the AWV. We therefore hypothesized that clinicians and practices, who conduct the AWV in a larger share of their attributed patients, have **higher O/E ratios for the three measures,** i.e., a positive correlation, since they would investigate potential cognitive decline more frequently. We tested this hypothesis by correlating the clinician- and practice-level AWV utilization rates [SM2] with their O/E ratios with a Spearman correlation.

b.Rationale for using share of patients 80 years and older

Physicians [SM3] and practices with a higher share of very old Medicare beneficiaries, defined as 80 years and older, can be expected to be more attuned to cognitive decline, given the strong age-related increase of incidence prevalence. Thus, we hypothesized that the share of this subgroup among the attributed patients is **positively correlated with the O/E ratio** and test the strength of the association with a Spearman correlation.

Of note, while age is part of our predictive model to calculate expected diagnosis rates, using share of very old patients in a clinician's or practice's panel remains a valid test of empirical validity. A higher share of very old patients will increase the expected number of cases, if detection does not increase accordingly, the O/E ratios will be lower. In other words, we are using the share of very old patients to understand the focus of a clinician or practice, and a geriatrics-focused provider is expected to have higher detection rates relative to casemix.

c. Rationale for comparing geriatricians to other clinicians

Lastly, we hypothesized that geriatricians, identified based on PECOS code=38, have a higher O/E ratio than other primary care clinicians. While training in assessment of cognitive function plays a minor role in general primary care residency programs, geriatricians are specifically trained to diagnose and treat ageing-related conditions, and we would expect them to have **higher O/E ratios on average**. We use a t-test to evaluate this hypothesis. This validity test can only be applied to individual clinicians.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

There were two rating questions on face validity, one at the individual clinician level and one at the practice level (See Tables 1 and 2). The experts were asked to rate the statement: "This measure possesses face validity at the clinician/practice level."

*	**strongly disagree**	*	**disagree**	*	**neither agree nor disagree**	*	**agree**	*	**strongly agree**	*
*	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number
mild cognitive impairment.	0.00%	0	0.00%	0	12.50%	1	12.50%	1	75.00%	6
dementia.	0.00%	0	12.50%	1	12.50%	1	25.00%	2	50.00%	4
cognitive impairment of any stage	0.00%	0	0.00%	0	12.50%	1	37.50%	3	50.00%	4

Table 1: Expert rating of the face validity of the three measures of the detection of cognitive impairment in primary care at the **clinician level** (question 1) (Showing that 87.5% of the experts

strongly agreed or agreed that all three measures of the detection of cognitive impairment in primary care at the clinician level have face validity.)

*Cell intentionally left empty.

Comment:

- Basing MCI and dementia diagnosis on ICD-10-CM codes is potentially problematic. For example, some of the dementia codes are for conditions not necessarily related to dementia such as F05: Delirium due to known physiological condition; F06.1: Catatonic disorder due to known physiological condition; R54: Age-related physical debility/Frail elderly. Therefore, the observed rate could be an overestimate. On the other hand, only one code G31.84 was used for MCI, which is a syndromic and not disease-specific diagnosis. It is possible that clinicians used one of the listed dementia codes for a patient with MCI to be more specific about the underlying disease process.

Developer's comment: We are aware of the limitations of using diagnoses in administrative data. However, we followed the validated algorithms for a claims-based dementia diagnosis from the CMS' Chronic Conditions Warehouse. There is no CCW algorithm for MCI, but after reviewing several coding instructions, we found G31.84 to be the only code for it. Also, our low MCI detection rates are in line with other published estimates. Acknowledging the potential misclassification of MCI as dementia in claims data, we added a measure for cognitive impairment of any stage.

*	**strongly disagree**	**	**disagree**	**	**neither agree nor disagree**	**	**agree**	**	**strongly agree**	**
*	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number
mild cognitive impairment	0.00%	0	0.00%	0	12.50%	1	12.50%	1	75.00%	6
dementia	0.00%	0	12.50%	1	0.00%	0	25.00%	2	62.50%	5
cognitive impairment of any stage	0.00%	0	0.00%	0	12.50%	1	25.00%	2	62.50%	5

Table 2: Expert rating of the face validity of the three measures of the detection of cognitive impairment in primary care at the **practice level** (question 2) (Showing that 87.5% of the experts strongly agreed or agreed that all three measures of the detection of cognitive impairment in primary care at the practice level have face validity.)

*Cell intentionally left empty.

Comment:

- Similar concern as mentioned in [question] 1.

Interpretation

In summary of the face validity ratings results, 87.5% of the experts strongly agreed or agreed that all three measures of the detection of cognitive impairment in primary care at both the clinician and practice levels have face validity. Notably, at both levels, one respondent disagreed that the measure, "ratio of observed over predicted rates for diagnosis of dementia," has face validity (See comment under Table 1) because of concerns

about the ICD 10 codes. in sum, however, these results indicate strong support for the face validity of the measures.

Empirical validity testing

1. Individual Clinician level results

The results displayed in Table 1 show that all three tests confirmed the empirical validity of the set of three measures. As hypothesized, the O/E ratios of all three measures are positively and significantly correlated with the proportion of attributed patients receiving the AWV and of patients aged 80 years and older. While the correlation coefficient on the AMV measure is mathematically low, probably because of the low and inconsistent uptake of this benefit, the correlation is highly significant. Further, it is plausible that the association is strongest for the MCI measure, because the AMV is meant to proactively explore cognitive status even in patients without a subjective memory complaint. Geriatricians have significantly higher O/E ratios than other primary care physicians, confirming our hypothesis that they are better trained to identify cognitive decline.

*	*	**O/E ratio for**	**O/E ratio for**		**O/E ratio for**
*	*	**any stage of cognitive impairment**		**MCI**	**Dementia**
Proportion of attributed patients receiving annual wellness visit	Spearman r coefficient	0.11		0.14	0.10
*	p value	<0.001		<0.001	<0.001
Proportion of attributed patients aged 80 years and older	Spearman r coefficient	0.34		0.15	0.20
*	p value	<0.001		<0.001	<0.001
O/E ratio for geriatricians compared to other primary care physicians	mean - geriatrician	0.761		0.161	1.26
	mean - other PCP	0.426		0.080	0.875
*	F value	35.01		18.78	22.37
	p value	<0.001		<0.001	<0.001

Table 1: Empirical validity testing: individual clinicians (The O/E ratios of all three measures are positively and significantly correlated with the proportion of attributed patients receiving the AWV and of patients aged 80 years and older. Geriatricians have significantly higher O/E ratios than other primary care physicians)

*cell intentionally left empty

1. Practice level results

The results displayed in Table 2 show that both tests confirmed the empirical validity of the set of three measures. As hypothesized, the O/E ratios of all three measures are positively and significantly correlated with the proportion of attributed patients receiving the AWV and of patients aged 80 years and older. While the correlation coefficient

on the AMV measure is mathematically low, probably because of the low and inconsistent uptake of this benefit, the correlation is highly significant. Further, it is plausible that the association is strongest for the MCI measure, because the AMV is meant to proactively explore cognitive status even in patients without a subjective memory complaint.

*	*	**O/E ratio for**	**O/E ratio for**	**O/E ratio for**
*	*	**any stage of cognitive impairment**	**MCI**	**Dementia**
Proportion of attributed patients receiving annual wellness visit	Spearman r coefficient	0.06	0.12	0.03
*	p value	<0.001	<0.001	<0.001
Proportion of attributed patients aged 80 years and older	Spearman r coefficient	0.32	0.11	0.17
*	p value	<0.001	<0.001	<0.001

Table 2: Empirical validity testing: practices (the O/E ratios of all three measures are positively and significantly correlated with the proportion of attributed patients receiving the AWV and of patients aged 80 years and older)

*cell intentionally left empty

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]

In summary of the face validity ratings results, 87.5% of the experts strongly agreed or agreed that all three measures of the detection of cognitive impairment in primary care at both the clinician and practice levels have face validity. Notably, at both levels, one respondent disagreed that the measure, "ratio of observed over predicted rates for diagnosis of dementia," has face validity (See comment under Figure 1) because of concerns about the ICD 10 codes. in sum, however, these results indicate strong support for the face validity of the measures.

The analysis of the empirical validity suggests that the set of three measures possess empirical validity at the individual clinician and practice levels, as they show the expected associations with indicators for better expected performance of the measures. Our export panel ratings mirrored these findings. We therefore conclude that the three measures have acceptable validity. The analyses suggests that the set of three measures possess empirical validity at the individual clinician and practice levels, as they show the expected associations with indicators for better expected performance of the measures. We are in the process of conducting face validity testing with an expert panel and will add those results with the final submission.

The measures have adequate discriminatory ability to distinguish top and bottom performing providers, and the underlying prediction model was tested and shown to perform adequately.

[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

We are testing whether the measure results are significantly different between the top and bottom performing quintiles of providers.

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

Results are shown in the table below. We find statistically significant differences in the O/E ratios of all three measures and both for the true O/E ratio and the capped ratios. The differences are of large magnitude with an increase factor of three to four times. Our sensitivity analysis of removing the effect of outliers by capping O/E rates at 2.0 lowers the mean of the top quintile, as expected, but does not reduce our ability to distinguish high and low-performing practices.

*	O/E ratio; any stage of cognitive impairment	*	*	O/E ratio; MCI	*	*	O/E ratio; Dementia	*	*
*	1st quintile mean	5th quintile mean	p-value	1st quintile mean	5th quintile mean	p-value	1st quintile mean	5th quintile mean	p-value
Clinician, O/E ratio	0.1673	0.7297	<.0001	0	0.2166	<.0001	0.3089	1.5843	<.0001
Practice, O/E ratio	0.1933	0.7497	<.0001	0	0.2027	<0.001	0.3473	1.5833	<.0001

Differences in measure performance for top and bottom quintiles of physicians and practices (Showing large and statistically significant differences between clinicians and practices in the top and bottom performing quintiles for all three measures.)

*cell deliberately left blank

[Response Ends]

2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

There are statistically significant differences between top and bottom performing clinicians and practices. These differences and their large magnitude imply that the measures have adequate discriminatory ability.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

N/A, we are using fully adjudicated Medicare data

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

N/A. We are using fully adjudicated Medicare data.

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

n/A

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

N/A or no exclusions

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]

The measures are not using any exclusions as they are based on the 100% samples for both Medicare fee-for-service and Medicare Advantage Plans. While we limit ***reporting*** of the measure to clinicians and practices with at least 25 attributed patients, this does not constitute an exclusion per NQF guidance, since those patients might be reported when reporting on higher levels of aggregation, such as a state. We merely follow CMS' recommendations for minimum sample size to report stable results.

[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]

N/A

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]

N/A

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

Predictive model to estimate number of cases

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

To estimate the underlying prevalence of MCI and dementia based on cognitive assessments, we used 2000 to 2016 data of the Health and Retirement Study (HRS), a nationally representative longitudinal survey of older adults aged 50 and over in the U.S., which includes formal cognitive assessments. We applied the Langa-Weir classification (Crimmins et al. 2011) to identify respondents with dementia and cognitive impairment but no dementia (CIND), as representing MCI, based on cognitive assessments of self-respondents. For individuals with a proxy respondent, a score was created using definitions and cutoffs from Zhu et al. by summing the following: number of limitations with instrumental activities of daily living (0-5); proxy's rating of the respondent's difficulty finishing the interview (0-2 with a higher score indicating poorer cognition); and proxy's rating of the respondent's memory (0-4 with a higher score indicating poorer cognition). Proxy scores were then used to classify respondents into groups: cognitively normal (score 0-2), CIND (3-5), and dementia (6-11).

We used a probit model to separately predict CIND (versus cognitively normal) and dementia (versus cognitively normal) as well as any cognitive impairment (CIND or dementia versus cognitively normal). Predictors were sex, age groups (50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84 and 85+), race/ethnicity (White, Black, Hispanic, and Other), dual eligibility status (individuals covered by both Medicare and Medicaid), and a continuous linear trend to account for the secular decline in dementia incidence (Wolters, Chibnik et al. 2020). Our sample consists of

individuals aged 65 and older and person-level weights were applied to make it representative of the national population, as the HRS oversamples black and Hispanic populations.

The estimated regression weights using the 2000 to 2014 data were applied to each individual in the 2016 wave to calculate the predicted outcome, i.e., probability of having CIND or dementia and the weighted average of the predicted probabilities represents the national estimates of persons with CIND and dementia. We used receiver operating characteristic (ROC) curves to examine model performance.

As all predictors in the HRS-based model are also available in Medicare enrollment data, the estimated weights can be applied to additional years of Medicare data to generate expected diagnosis rates for MCI and dementia, given the changing demographic composition and secular trend. The ratio of the rates observed based on diagnoses documented in the claims data of the included beneficiaries to those expected rates provides a measure for potential gaps in diagnoses. Such observed to expected (O/E) ratios are frequently used in quality measurement (Cho et al, 2022) and can be interpreted as the proportion of expected cases, who were diagnosed, and can also be interpreted as the proportion of expected cases who were actually diagnosed.

The estimated regression weights derived from the 2000 to 2014 HRS data are listed in Table 3 as are the predicted and observed rates of MCI, dementia and any cognitive impairment for 2016 based on those estimated weights. The predicted probability for MCI in 2016 was 18.67% compared to an observed probability of 19.29%, for dementia 9.06% compared to 9.99% and for any cognitive impairment 0.25 compared to 0.26. The area under the ROC curve (AUC) was 0.7128 for the MCI model, 0.8156 for the dementia model and 0.7449 for the combined impairment model (Table 3, and Figures).

Table 3: Probit model estimates using HRS respondents age 65 and over (2000-2014) and reported and predicted rates for 2016

Y	**MCI (vs normal)**	*	*	*	**Dementia (vs normal)**	*	*	*	**MCI or Dementia (vs normal)**	*	*	*
Y (2016)	19.29%	*	*	*	9.99%	*	*	*	25.93%	*	*	*
Y^ (2016)	18.67%	*	*	*	9.06%	*	*	*	24.79%	*	*	*
*	Coefficient	Robust SE	95% CI	95% CI	Coefficient	Robust SE	95% CI	95% CI	Coefficient	Robust SE	95% CI	95% CI
Constant	22.08	2.95	16.30	27.86	34.36	4.26	26.01	42.71	27.55	2.80	22.06	33.04
Sex: male	0.12	0.02	0.08	0.15	0.03	0.02	-0.02	0.08	0.10	0.02	0.07	0.14
Age groups	*	*	*	*	*	*	*	*	*	*	*	*
70-74	0.20	0.02	0.16	0.23	0.28	0.03	0.23	0.34	0.23	0.02	0.20	0.26
75-79	0.45	0.02	0.41	0.48	0.66	0.03	0.60	0.73	0.53	0.02	0.50	0.57
80-84	0.70	0.02	0.66	0.75	1.07	0.03	1.00	1.13	0.85	0.02	0.81	0.90
85+	1.08	0.02	1.03	1.13	1.70	0.03	1.63	1.77	1.35	0.02	1.30	1.40
Race & Ethnicity	*	*	*	*	*	*	*	*	*	*	*	*
Black	0.74	0.03	0.69	0.79	0.82	0.03	0.76	0.89	0.80	0.02	0.75	0.85

Y	MCI (vs normal)	*	*	*	Dementia (vs normal)	*	*	*	MCI or Dementia (vs normal)	*	*	*
Hispanic	0.62	0.03	0.56	0.69	0.56	0.05	0.47	0.65	0.63	0.03	0.57	0.69
Other	0.26	0.06	0.13	0.38	0.31	0.08	0.15	0.47	0.29	0.06	0.17	0.41
Dual eligible	0.56	0.03	0.51	0.62	1.02	0.03	0.95	1.09	0.76	0.03	0.71	0.81
Year	-0.01	0.00	-0.01	-0.01	-0.02	0.00	-0.02	-0.01	-0.01	0.00	-0.02	-0.01
R^2	0.0962	*	*	*	0.2447	*	*	*	0.1466	*	*	*
AUC	0.7128	*	*	*	0.8156	*	*	*	0.7449	*	*	*
N (2016)	8,946	*	*	*	7,904	*	*	*	9,808	*	*	*
N (2000-2014)	77,206	*	*	*	68,612	*	*	*	86,559	*	*	*

A table comparing Probit model estimates using HRS respondents age 65 and over between 2000 and 2014 as well as reported and predicted rates for 2016 for MCI (vs normal), Dementia (vs normal), and MCI or dementia (vs normal) by sex, age category breakouts from 70-85+, race and ethnicity, dual eligible, year, R^2, and AUC

*cell intentionally left blank

References

Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study. *J Gerontol B Psychol Sci Soc Sci*. Jul 2011;66 Suppl 1:i162-71. doi:10.1093/geronb/gbr048

Zhu Y, Chen Y, Crimmins EM, Zissimopoulos JM. Sex, Race, and Age Differences in Prevalence of Dementia in Medicare Claims and Survey Data. *The Journals of Gerontology: Series B*. 2021;76(3):596-606. doi:10.1093/geronb/gbaa083

Cho SK, Mattke S, Sheridan M, Ennis W. Association of wound healing with quality and continuity of care and sociodemographic characteristics. *Am J Manag Care*. Apr 1 2022;28(4):e146-e152. doi:10.37765/ajmc.2022.88868

[Response Ends]

Attachment: 3707 _3707 _3707_AUC curves-508_(1).docx

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

Published literature

Internal data analysis

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

We included as predictors sex, as women have a higher risk of cognitive impairment than men, age groups (50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84 and 85+), as incidence increases with age, race/ethnicity (White, Black, Hispanic, and Other), dual eligibility status (individuals covered by both Medicare and Medicaid), and a continuous linear trend to account for the secular decline in dementia incidence (Wolters, Chibnik et al. 2020).

The limited set of predictors is a consequence of us using the HRS data to develop and validate the model to have access to cognitive testing data for a nationally representative panel of individuals. However, we applied the coefficients of the model to Medicare administrative data to estimate expected diagnosis rates in the Medicare population. Thus, we could only include variables that are present and identically defined in the HRS and the Medicare data, confining to demographic variables, because presence and control of comorbid conditions, which are important predictors for cognitive impairment is based on self-report in HRS but on care utilization in the Medicare data.

While this restriction limits the precision of the model, our testing results show that we achieve acceptable accuracy given that the intent of the model is not to **identify cases with high likelihood of cognitive impairment** but to estimate the **prevalence of cognitive impairment in a larger sample.** Further, as minority populations have a higher burden of chronic conditions, such as diabetes, hypertension and lipid disorders, minority status can serve as a proxy for cardiovascular and thus dementia risk..(CDC 2022) Similarly, dually eligibles tend to be poorer and have lower education levels than Medicare-only beneficiaries, which is also associated with higher risk of cognitive impairment. (MACPAC & MedPAC 2022)

References

Wolters, F. J., L. B. Chibnik, R. Waziry, R. Anderson, C. Berr, A. Beiser, J. C. Bis, D. Blacker, D. Bos, C. Brayne, J.-F. Dartigues, S. K. L. Darweesh, K. L. Davis-Plourde, F. De Wolf, S. Debette, C. Dufouil, M. Fornage, J. Goudsmit, L. Grasset, V. Gudnason, C. Hadjichrysanthou, C. Helmer, M. A. Ikram, M. K. Ikram, E. Joas, S. Kern, L. H. Kuller, L. Launer, O. L. Lopez, F. E. Matthews, K. McRae-Mckee, O. Meirelles, T. H. Mosley, M. P. Pase, B. M. Psaty, C. L. Satizabal, S. Seshadri, I. Skoog, B. C. M. Stephan, H. Wetterberg, M. M. Wong, A. Zettergren and A. Hofman (2020). "Twenty-seven-year time trends in dementia incidence in Europe and the United States." Neurology **95**(5): e519-e531.

CDC. Racial and Ethnic Approaches to Community Health. https://www.cdc.gov/nccdphp/dnpao/state-local-programs/reach/

MACPAC & MedPAC (2022). Beneficiaries Dually Eligible For Medicare and Medicaid.

https://www.macpac.gov/wp-content/uploads/2022/02/Beneficiaries-Dually-Eligible-for-Medicare-and-Medicaid-February-2022.pdf

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

In light of being limited to demographic variables that are present in both Medicare and HRS data, we used the full set of variables and did not conduct stepwise testing of their contribution. Of note, the objective is not to determine whether a particular variable is significantly associated with the outcomes, but to estimate a patient's predicted probability of cognitive impairment. The sum of the predicted probabilities by clinician or practice is then the expected number of cases.

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

As Medicare claims and enrollment data do not include information on social determinants of health, we were unable to explore those. However, dual eligibility can be seen as a proxy for low income, as the elderly duals typically become Medicaid edibility because of poverty.

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

The estimated regression weights using the 2000 to 2014 HRS data (development sample) were applied to each individual in the 2016 wave (validation sample) to calculate the predicted outcome, i.e., probability of having CIND or dementia. The adequacy of the model was determined based on its ability to predict the outcome in the validation sample. We used receiver operating characteristic (ROC) curves to examine model performance. An Area Under the Curve (AUC) of >0.70 is an accepted standard for acceptable model performance.

We also compared the weighted averages of the predicted probabilities, which represent the national estimates of prevalence of CIND and dementia, for 2016 (Y^) to the measured prevalence in the 2016 data (Y) as further evidence for predictive accuracy.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

As stated above, the AUCs are the most appropriate measures of model accuracy for a binary outcome. As they are >0.7 for all three models, they perform adequately. We report the R^2 statistics, which are in the range expected for a well-specific probity model.

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

Not applicable for a predictive model

[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

Not applicable for a predictive model

[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

N/A

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

Strictly speaking, we are not risk-adjusting the measure but we are using a predictive model to define a benchmark number of diagnosed cases for a clinician or practice, given the demographic profile of their patient pool, against which their actual number of diagnosed cases will be compared. The model was calibrated in external data and performed well with AUCs >0.7. As we are not predicting for the purposes of identifying individual cases with cognitive impairment but to estimate the overall number of cases in a clinician or practice, this performance is adequate.

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

We apologize for the usual presentation of the data. It is a set of three measures, and it is important to see results for all three together.

Further, we are not conducting risk adjustment but use a predictive model to estimate an expected number of cases. Thus, the presentation of our results does not perfectly fit into the categories provided for a classical risk adjustment model.

[Response Ends]

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

ALL data elements are in defined fields in electronic claims

[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]

N/A

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

N/A

[Response Ends]

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]

While claims data for fee-for-service Medicare are available within less than a year, there is a delay of two years or longer in obtaining the Medicare Advantage data.

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),

Attach the fee schedule here, if applicable.

[Response Begins]

Measure will be available free of charge for noncommercial use

[Response Ends]

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

4a.01. Check all current uses. For each current use checked, please provide:

- **Name of program and sponsor**
- **URL**
- **Purpose**
- **Geographic area and number and percentage of accountable entities and patients included**
- **Level of measurement and setting**

[Response Begins]

Not in use

[Not in use Please Explain]

new measure

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

Payment Program

Quality Improvement (internal to the specific organization)

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

N/A, new measure

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]

Assuming a supportive vote from the Standing Committee, we plan to submit the measure to the Measures Under Consideration call for the Spring 2023 cycle and to the Measures Application Partnership – Clinicians – for the fall 2023 cycle. Both processes will get the measure on the 2024 agenda for CMS to assess whether it would be suitable for the Quality Payment Program. Depending on the timing of the CMS decision process, the measure could be added to the rulemaking process for CY 2024 or 2025, and use in an accountability application as well as reporting in 2025 or 2026.

In parallel, we will communicate the measure to health systems that might be interested in improving early detection of cognitive impairment. For example, Advent Health in Central Florida and Indiana University Health are current sites for a pilot program on improving early detection supported by the Davos Alzheimer's Collaborative (https://www.davosalzheimerscollaborative.org/healthcare-system-preparedness) and interested in performance measurement for their clinicians. Dr. Mattke is a scientific advisor to this Collaborative and in regular contact with both sites. Advocate Aurora Health in Illinois, University of Washington, and Cleveland Clinic are joining the program and might be interested as well. We will contact them regarding those measures and potential for use.

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]

N/A, new measure

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

N/A, new measure

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

N/A, new measure

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

N/A, new measure

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

N/A, new measure

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

N/A, new measure

[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]

This is a new measure. We expect that the information contained in the measure will draw clinicians' and practices' attention to potential underdiagnosis of cognitive impairment and trigger efforts to improve detection rates. We conducted an expert panel rating exercise to assess usability.

We identified nine clinical experts in the treatment of Alzheimer's disease and related dementias (ADRD) to rate the measures of the detection of cognitive impairment in primary care on their face validity, understandability, and usability using a web-based questionnaire (developed using SurveyMonkey®). One of the identified experts did not complete the rating by the time of submission. The names and affiliations of the clinical experts who participated in the rating are listed under face validity testing.

The clinical experts were asked to review three measures: "ratio of observed over predicted rates for diagnosis of **cognitive impairment of any stage,**" "ratio of observed over predicted rates for diagnosis of dementia," and "ratio of observed over predicted rates for diagnosis of **mild cognitive impairment**." After reviewing the background material, they were instructed to rate the usability of the three measures at the clinician and practice level by indicating their degree of agreement on a 5-point scale as described under validity testing.

There were two rating questions on understandability and usability, one at the clinician level and one at the practice level (See Figures 3 and 4). The experts were asked to rate the statement: "Clinicians/practices can understand the results of the measure and are likely to find them useful for decision making."

Figure 3: Expert rating of the understandability and usability of the three measures of the detection of cognitive impairment in primary care at the **clinician level** (question 3)

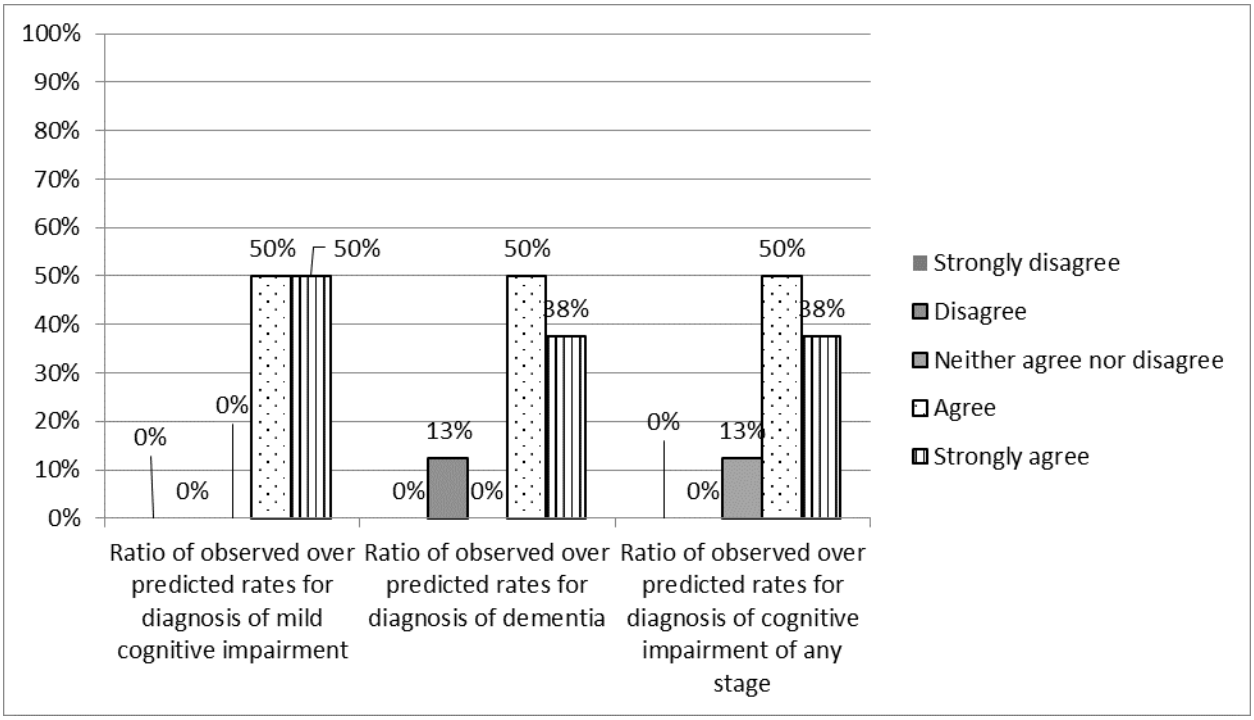


Figure 3: Expert rating of the understandability and usability of the three measures of the detection of cognitive impairment in primary care at the **clinician level** (question 3)

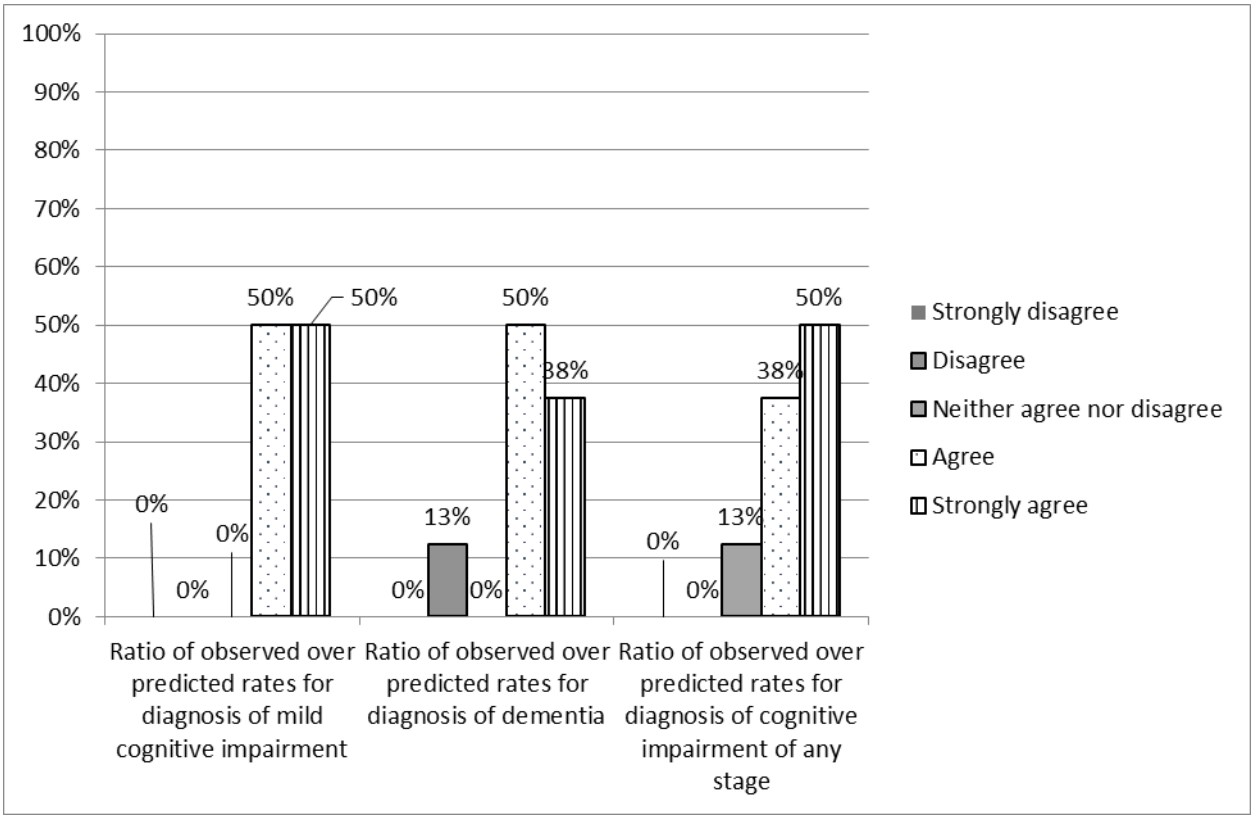


Figure 4: Expert rating of the understandability and usability of the three measures of the detection of cognitive impairment in primary care at the **practice level** (question 4)

As for the understandability and usability of the measures, 100% of the expert panel rated the measure, "ratio of observed over predicted rates for diagnosis of mild cognitive impairment," and 87.5% of the expert panel rated the other two measures as strongly agree or agree. Again, one respondent disagreed that the "ratio of observed over predicted rates for diagnosis of dementia" is understandable and usable at both the clinician and practice levels.

Regarding the single expert who disagreed on the usability of the dementia measure, we acknowledge the expert's concern about accurate coding of dementia in administrative data. However, our definitions follow the validated algorithm of CMS' Chronic Condition Warehouse and we decided not to deviate from those established standards. We would update accordingly if the algorithm changed.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

N/A, new measure

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

N/A, new measure

[Response Ends]

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

2872e: Dementia: Cognitive Assessment

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

Not to our knowledge

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

Yes

[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]

N/A

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

No competing measures. 2872 measures annual cognitive assessment for patients with an established dementia diagnosis, whereas ours looks at undiagnosed cognitive impairment.

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.:

No appendix

Contact Information

Measure Steward (Intellectual Property Owner): University of Southern California

Measure Steward Point of Contact: Mattke, Soeren, mattke@usc.edu

Measure Developer if different from Measure Steward: University of Southern California

Measure Developer Point(s) of Contact: Mattke, Soeren, mattke@usc.edu

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

No appendix

[Response Ends]

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]

N/A

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

2022

[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]

11/2022

[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]

annual updates for coding changes, maintenance review per NQF requirements

[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]

2023

[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]

(C) University of Southern California. 2022

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]

N/A

[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]

Tables for calculations of the standard errors around a clinician's or practice's O/E ratio

[Response Ends]

Attachment: 3707 _3707 _Observed Prevalence and Expected Prevalence -by- demographic groupings, for benes in physician-level measure-508.xlsx

Attachment: 3707 _3707 _Observed Prevalence and Expected Prevalence -by- demographic groupings, for benes in practice-level measure-508.xlsx