

## Additional Testing for Parkinson's Syndromes, Multiple Sclerosis (MS), and Amyotrophic Lateral Sclerosis (ALS)

This document includes testing results at the TIN-NPI level and a summary of empirical data that demonstrates support for the measure concept.

### TIN-NPI Level Testing Results

#### Reliability and Validity

Subsection	Row	Field Label	Guidance	ADD YOUR CONTENT HERE
Measure Score Level (Accountable Entity Level) Testing	032	*Reliability	<p>Indicate whether reliability testing was conducted for the accountable entity-level measure scores. Acceptable reliability tests include signal-to-noise (or inter-unit reliability) or random split-half correlation. For more information on accountable entity-level reliability testing, refer to the Blueprint content on the CMS Measures Management System (MMS) Hub (<a href="https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/reliability">https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/reliability</a>).</p> <p>Select "Yes" if acceptable accountable entity-level reliability testing has been completed as of submission of this form.</p> <p>Select "No" if you are not able to provide the results of acceptable accountable entity-level reliability testing in this submission. If testing results are incomplete, or if you are submitting a different type of reliability testing, provide as an attachment.</p> <p>Note: This section refers to the reliability of the accountable entity-level measure scores in the final performance measure. For testing of surveys or patient reported tools, refer to the Patient-Reported Data section.</p> <p>Note: for MIPS-Quality submissions, please provide individual clinician-level results. If the measure was also tested at the clinician group level, you may include those results in an attachment.</p>	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

Measure Score Level (Accountable Entity Level) Testing	033	*Reliability: Type of analysis	<p>Select all that apply.</p> <p>Signal-to-noise (or inter-unit reliability) is the precision attributed to an actual construct versus random variation (e.g., ratio of between unit variance to total variance) (Adams J. The reliability of provider profiling: a tutorial. Santa Monica, CA: RAND; 2009. <a href="http://www.rand.org/pubs/technical_reports/TR653.html">http://www.rand.org/pubs/technical_reports/TR653.html</a>).</p> <p>Random split-half correlation is the agreement between two measures of the same concept, using data derived from split samples drawn from the same entity at a single point in time.</p>	<input checked="" type="checkbox"/> Signal-to-Noise <input type="checkbox"/> Random Split-Half Correlation
Measure Score Level (Accountable Entity Level) Testing	034	*Signal-to-Noise: Level of Analysis	<p>Select the level of analysis at which the signal-to-noise analysis was conducted. If the measure is specified and intended for use at more than one level, ensure the results in this section are at the same level of analysis selected in the Measure Information section of this form.</p> <p>For MIPS-Quality submissions, you must report the results of individual clinician-level testing. If group-level testing is available, you may submit those results as an attachment.</p>	<input type="checkbox"/> Accountable Care Organization <input checked="" type="checkbox"/> Clinician – Individual only <input type="checkbox"/> Clinician – Group only <input type="checkbox"/> Facility <input type="checkbox"/> Health plan <input type="checkbox"/> Integrated Delivery System <input type="checkbox"/> Population: Community, County or City <input type="checkbox"/> Population: Regional and State
Measure Score Level (Accountable Entity Level) Testing	035	*Signal-to-Noise: Sample size	<p>Indicate the number of accountable entities sampled to test the final performance measure. Note that this field is intended to capture the number of measured entities and not the number of individual patients or cases included in the sample.</p>	<p>At the 10-episode volume threshold: 6,283            At the 20-episode volume threshold: 3,019            At the 30-episode volume threshold: 1,859</p>
Measure Score Level (Accountable Entity Level) Testing	036	*Signal-to-Noise: Median Statistical result	<p>Indicate the median result for the signal-to-noise analysis used to assess accountable entity level reliability. Results should range from 0.00 to 1.00. Calculate reliability as the measure is intended to be implemented (e.g., after applying minimum denominator requirements, appropriate type of setting, provider, etc.).</p>	<p>At the 10-episode volume threshold: 0.454            At the 20-episode volume threshold: 0.537            At the 30-episode volume threshold: 0.596</p>

Measure Score Level (Accountable Entity Level) Testing	037	*Signal-to-Noise: Interpretation of results	Describe the type of statistic and interpretation of the results (e.g., low, moderate, high). Provide the distribution of signal-to-noise results across measured entities (e.g., min, max, percentiles). List accepted thresholds referenced and provide a citation. If applicable, include the precision of the statistical result (e.g., 95% confidence interval) and/or an assessment of statistical significance (e.g., p-value).	<p>Reliability testing of the Parkinson's Syndromes, Multiple Sclerosis (MS), and Amyotrophic Lateral Sclerosis (ALS) measure is conducted for clinicians (TIN-NPIs) and constructed using episodes ending between January 1, 2022, and December 31, 2022. Reliability evaluates a measure's ability to differentiate one clinician's performance from another consistently. The reliability metric captures how much of the variance in a measure is due to systematic differences in episode spending between clinicians ("signal") rather than differences in episode spending within a clinician's set of episodes ("noise"). A measure with high reliability suggests that performance comparisons across clinicians reflects systematic differences in actual performance better. Based on existing scientific evidence on the different interpretations and methods of estimating reliability, CMS finalized in the CY 2022 Physician Fee Schedule (86 FR 64996) rule that the 0.4 threshold for mean reliability continues to be appropriate for indicating moderate reliability for performance measures in the Cost category in the MIPS program. Mean reliability levels above 0.7 continue to demonstrate high reliability for cost measures, as previously established in the CY 2017 Quality Payment Program final rule (81 FR 77169 through 77171).</p> <p>At the 20-episode volume threshold, testing results indicated that the mean reliability for the measure is 0.537 at the TIN-NPI level, and approximately 73.17% of clinicians meet or exceed the moderate reliability threshold of 0.4.</p>
--	-----	---	--	---

Measure Score Level (Accountability Entity Level) Testing	042	*Empiric Validity	<p>Indicate whether empiric validity testing was conducted for the accountable entity-level measure scores. For more information on accountable entity level empiric validity testing, refer to the CMS Measures Management System Blueprint (<a href="https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/validity">https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/validity</a>)</p> <p>Note: This section refers to the empiric validity of the accountable entity level measure scores in the final performance measure. Refer to the Patient-Reported Data section for testing of surveys or patient reported tools.</p> <p>Note: for MIPS-Quality submissions, please provide individual clinician-level results. If the measure was also tested at the clinician group level, you may include those results in an attachment.</p>	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Measure Score Level (Accountable Entity Level) Testing	043	*Empiric Validity: Level of Analysis	<p>Select the level of analysis at which the empiric validity analysis was conducted. If the measure is specified and intended for use at more than one level, ensure the results in this section are at the same level of analysis selected in the Measure Information section of this form.</p> <p>For MIPS-Quality submissions, you must report the results of individual clinician-level testing. If group-level testing is available, you may submit those results as an attachment.</p>	<input type="checkbox"/> Accountable Care Organization <input checked="" type="checkbox"/> Clinician – Individual only <input type="checkbox"/> Clinician – Group only <input type="checkbox"/> Facility <input type="checkbox"/> Health plan <input type="checkbox"/> Integrated Delivery System <input type="checkbox"/> Population: Community, County or City <input type="checkbox"/> Population: Regional and State
Measure Score Level (Accountability Entity Level) Testing	044	* Empiric Validity: Sample size	<p>Indicate the number of accountable entities sampled to test the final performance measure. Note that this field is intended to capture the number of measured entities and not the number of individual patients or cases included in the sample.</p>	<p>3,019 TIN-NPIs who meet the 20-episode volume threshold</p>

Measure Score Level (Accountability Entity Level) Testing	045	*Empiric Validity: Methods and findings	Describe the methods used to assess accountable entity level validity. Describe the comparison groups or constructs used to verify the validity of the measure scores, including hypothesized relationships (e.g., expected to be positively or negatively correlated). Describe your findings for each analysis conducted, including the statistical results and the strongest and weakest results across analyses. If applicable, include the precision of the statistical result(s) (e.g., 95% confidence interval) and/or an assessment of statistical significance (e.g., p-value). If methods and results require more space, include as an attachment.	<p>Validity is a criterion used to assess whether the cost measure can quantify the construct it aims to measure, which is the cost directly related to treatment choices and the cost of adverse outcomes resulting from care. Validity is evaluated empirically by estimating the effect of relevant treatment choices on the measure score. This analysis first estimates the correlation between treatment choices and the measure score while controlling for adverse outcomes. Then the correlation between treatment choices and related adverse outcomes is calculated to demonstrate the indirect effect. Generally, adverse outcomes are non-trigger inpatient hospitalizations, non-trigger emergency room visits, and post-acute care. The remaining service categories are typically considered treatment.</p> <p>At the individual clinician reporting level, below are the estimated coefficients [95% CI] (p-value), scaled to thousands of dollars:</p> <p><b>Model 1:</b> Mean O/E = Mean Cost of Treatment Choices + Mean Cost of Adverse Events</p> <ul style="list-style-type: none"> <li>- Adverse Events: 0.06 [0.06,0.07](p &lt; 0.01)</li> <li>- Major Procedures: 0.17 [0.09,0.24](p &lt; 0.01)</li> <li>- Ambulatory/Minor Procedures: 0.05 [0.00,0.10] (p = 0.03)</li> <li>- Imaging Services: -0.07 [-0.18,0.03](p = 0.15)</li> <li>- Outpatient Physical, Occupational, or Speech and Language Pathology Therapy: 0.10 [0.08,0.12](p &lt; 0.01)</li> <li>- Laboratory, Pathology, and Other Tests: 0.25 [0.11,0.39](p &lt; 0.01)</li> </ul>
---	-----	--	---	---

				<ul style="list-style-type: none"> <li>- Durable Medical Equipment and Supplies: -0.02 [-0.03,-0.01](p &lt; 0.01)</li> <li>- Part-D Drugs: 0.02 [0.02,0.02](p &lt; 0.01)</li> <li>- Chemotherapy and Other Part B-Covered Drugs: 0.02 [0.02,0.03](p &lt; 0.01)</li> </ul> <p><b>Model 2:</b> Mean Cost of Adverse Events = Mean Cost of Treatment Choices</p> <ul style="list-style-type: none"> <li>- Major Procedures: -1.03 [-1.59,-0.46](p &lt; 0.01)</li> <li>- Ambulatory/Minor Procedures: 0.58 [0.24,0.93] (p &lt; 0.01)</li> <li>- Imaging Services: -1.49 [-2.24,-0.74](p &lt; 0.01)</li> <li>- Outpatient Physical, Occupational, or Speech and Language Pathology Therapy: -0.22 [-0.35,-0.08](p &lt; 0.01)</li> <li>- Laboratory, Pathology, and Other Tests: -2.93 [-3.96,-1.91](p &lt; 0.01)</li> <li>- Durable Medical Equipment and Supplies: 0.38 [0.29,0.46](p &lt; 0.01)</li> <li>- Part-D Drugs: 0.01 [-0.01,0.03](p = 0.44)</li> <li>- Chemotherapy and Other Part B-Covered Drugs: 0.01 [-0.01,0.02](p = 0.49)</li> </ul> <p>Model 1 shows that the cost of adverse events is associated with a worse measure score. Outpatient evaluation and management and durable medical equipment are associated with a better measure score. Major and minor procedures, physical/occupational/speech pathology therapy, laboratory testing, part B and D drugs are associated with a worse measure score. Among these services, model 2 shows that minor procedures and durable medical equipment are associated with higher cost of</p>
--	--	--	--	---

				adverse events, which suggests that the opportunities to reduce these costs are linked to the reduction of adverse events. On the other hand, major procedures, and physical/occupational/speech pathology therapy, laboratory testing are associated with lower cost of adverse events, which suggests that they are important in avoiding adverse events but also prone to overuse because the reduction in cost of adverse events do not fully offset the costs of these services.
Measure Score Level (Accountability Entity Level) Testing	046	*Empiric Validity: Interpretation of results	Indicate whether the statistical result affirmed the hypothesized relationship for the analysis conducted.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Measure Score Level (Accountable Entity Level) Testing	047	*Face validity	<p>Indicate if a vote was conducted among experts and patients/caregivers on whether the final performance measure scores can be used to differentiate good from poor quality of care.</p> <p>Select “No” if experts and patients/caregivers did not provide feedback on the final performance measure at the specified level of analysis or if the feedback was related to a property of the measure unrelated to its ability to differentiate performance among measured entities.</p> <p>This item is intended to assess whether face validity testing was conducted on the final performance measure and is not intended to assess whether patient-reported surveys or tools have face validity. Survey item testing results can be provided in an attachment and described in the Patient-Reported Data Section.</p>	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

## Measure Performance Scores and Performance Gap Analysis

Subsection	Row	Field Label	Guidance	ADD YOUR CONTENT HERE
Measure Performance	060	*Measure performance - type of score	Select one. Measure performance score type should be at the level of accountable entity.	<input type="checkbox"/> Categorical (e.g., measured entity scores yes/no, pass/fail, or rating scale/score) <input type="checkbox"/> Composite scale/non-weighted score <input type="checkbox"/> Composite scale/weighted score <input type="checkbox"/> Continuous variable (e.g., average) <input type="checkbox"/> Count <input type="checkbox"/> Frequency Distribution <input type="checkbox"/> Proportion <input type="checkbox"/> Rate <input checked="" type="checkbox"/> Ratio
Measure Performance	061	*Measure performance score interpretation	Select one	<input type="checkbox"/> Better quality = Higher score <input checked="" type="checkbox"/> Better quality = Lower score <input type="checkbox"/> Better quality = Score within a defined interval <input type="checkbox"/> Passing score above a specified threshold defines better quality <input type="checkbox"/> Passing score below a specified threshold defines better quality
Measure Performance	062	*Number of accountable entities included in analysis	Provide the number of accountable entities included in the analysis of the distribution of performance scores described in "Overall mean performance score" -"Overall standard deviation of performance scores."  Please enter a single value and do not enter a range.  If unknown or not available, enter 9999.	There are 3,019 TIN-NPIs included in the analysis
Measure Performance	063	*Number of accountable entities: unit	Provide the unit of accountable entities included in the analysis of the distribution of performance scores described in "Overall mean performance score" -"Overall standard deviation of performance scores."	TIN-NPIs with at least 20 attributed episodes



Subsection	Row	Field Label	Guidance	ADD YOUR CONTENT HERE
Measure Performance	064	*Number of persons	Provide the number of persons included in the analysis of the distribution of performance scores	There are 3,019 TIN-NPIs included in the analysis
Measure Performance	065	*10th percentile	<p>Provide the performance score at the 10th percentile for the testing sample that is relevant to the intended use of the measure.</p> <p>If this is a proportion measure, provide the 10th percentile score in percentage form, without the symbol. For example, if the 10th percentile performance score is 21.2%, enter 21.2 and not 0.212.</p> <p>If a 10th percentile performance score is not available, enter 9999.</p>	Performance score at the 10th percentile is \$9,184.30
Measure Performance	066	*50th percentile (median)	<p>Provide the median performance score (50th percentile) for the testing sample that is relevant to the intended use of the measure.</p> <p>Please enter only one value in the response field and do not enter a range of values.</p> <p>If this is a proportion measure, provide the median performance score in percentage form, without the symbol. For example, if the median performance score is 85.6%, enter 85.6 and not 0.856.</p> <p>If a median performance score is not available, enter 9999.</p>	Performance score at the 50th percentile is \$13,877.00

Subsection	Row	Field Label	Guidance	ADD YOUR CONTENT HERE
Measure Performance	067	*90th percentile	<p>Provide the performance score at the 90th percentile for the testing sample that is relevant to the intended use of the measure.</p> <p>If this is a proportion measure, provide the 90th percentile score in percentage form, without the symbol. For example, if the 90th percentile performance score is 85.6%, enter 85.6 and not 0.856.</p> <p>If a 90th percentile performance score is not available, enter 9999.</p>	Performance score at the 90th percentile is \$20,532.19
Measure Performance	068	*Additional measure performance information	<p>Provide the following additional measure performance information, <u>as applicable</u>:</p> <ul style="list-style-type: none"> <li>- Mean performance score across accountable entities in the test sample that is relevant to the intended use of the measure.</li> <li>- Minimum and maximum performance score for the testing sample that is relevant to the intended use of the measure.</li> <li>- Standard deviation of performance scores for the testing sample that is relevant to the intended use of the measure.</li> <li>- Passing score for the performance measure.</li> <li>- Performance score's defined interval, including upper and lower limit of the performance score.</li> </ul>	<p>Analysis of all clinicians (TIN-NPIs) with at least 20 attributed episodes for the 2022 performance period shows a wide range of provider scores of the Parkinson's Syndromes, MS, and ALS measure. The measure score has the following distributional characteristics:</p> <ul style="list-style-type: none"> <li>• Mean (SD): \$14,553.99 (\$4,874.39)</li> <li>• Min: \$1,453.99</li> <li>• Max: \$47,794.81</li> </ul> <p>There is no passing score for the performance measure and there are no lower or upper limits for the performance score.</p>

## Summary of Empirical Data Supporting the Measure Concept

Neurological disorders influencing movement affect almost 40 million Americans across different neurological conditions affecting movement.<sup>1</sup> Parkinson's disease, other degenerative diseases of basal ganglia, Multiple Sclerosis (MS), and Amyotrophic Lateral Sclerosis (ALS) affect nearly half a million of Medicare beneficiaries, and patients with these disorders have higher utilization of

<sup>1</sup> University of Michigan Health, "Movement Disorders," <https://www.uofmhealth.org/conditions-treatments/brain-neurological-conditions/movement-disorders>

healthcare services. For example, patients with Parkinson's present 31% higher emergency department (ED) admissions and double the number of Skilled Nursing Facilities (SNF) stays.<sup>2</sup> Patients with MS present double the number of ED admissions and 3.5 times the number of inpatient stays,<sup>3</sup> and patients with ALS alternatively have high rates of home health service utilization as well as the highest national economic burden amongst movement disorders.<sup>4</sup>

The literature scan identified three critical areas of opportunities to improve care outcomes for patients with Parkinson's, MS, and ALS and reduce costs associated with managing the disease. These include improving fall-related education and treatment, screening patients for additional comorbidities not related to physical complications, and mitigating drug interactions or use of inappropriate medications. Increased activity improves both physical health and mental acuity in both Parkinson's and MS patients<sup>5</sup> and, in fact, significantly improves fall-related outcomes in Parkinson's patients and costs of subsequent hospitalizations.<sup>6</sup> Beyond physical constraints, a significant portion of patients also report cognitive impairment and mental health issues, which both require regular assessments to detect and follow up on.<sup>7,8</sup> Lastly, Parkinson's, MS, and ALS patients are also heavily dependent on medications to manage symptoms and additional comorbidities, which increases the risk of potential drug-drug interactions, raising the need for effective management of multiple medications.<sup>9</sup>

Existing literature and Acumen's testing indicate a high cost to Medicare for treating and managing Parkinson's, MS, and ALS, opportunities for improvement through best practices, and a substantial empirical performance gap. Our testing indicates that the measure would have a significant impact on Medicare, whether through measuring beneficiaries, clinicians, or cost. The measure would capture almost 300,000 beneficiaries and almost 6,000 clinician groups and individual clinicians combined (using 2022 as the study year). Table 1 shows the distribution of the measure score for TIN and TIN-NPI levels. There is substantial variation observed in the measure score at both TIN and TIN-NPI levels, indicated by the interquartile ranges, standard deviations, and coefficients of

---

<sup>2</sup> Gandhi, Aakash Bipin et al. "Health Care Resource Utilization Associated With Parkinson Disease Among Medicare Beneficiaries." *Neurology* vol. 97,6 (2021): e597-e607. doi:10.1212/WNL.00000000000012290

<sup>3</sup> Asche, Carl V et al. "All-cause health care utilization and costs associated with newly diagnosed multiple sclerosis in the United States." *Journal of managed care pharmacy : JMCP* vol. 16,9 (2010): 703-12. doi:10.18553/jmcp.2010.16.9.703

<sup>4</sup> Winston Wong, PharmD "Managed Care Considerations to Improve Health Care Utilization for Patients With ALS." *Am J Manag Care*. 2023;29(suppl 7):S120-S126. <https://doi.org/10.37765/ajmc.2023.89388>

<sup>5</sup> Döring, Andrea et al. "Exercise in multiple sclerosis -- an integral component of disease management." *The EPMA journal* vol. 3,1 2. 24 Dec. 2011, doi:10.1007/s13167-011-0136-4

<sup>6</sup> Shen, Xia et al. "Effects of Exercise on Falls, Balance, and Gait Ability in Parkinson's Disease: A Meta-analysis." *Neurorehabilitation and neural repair* vol. 30,6 (2016): 512-27. doi:10.1177/1545968315613447

<sup>7</sup> Langdon, D W et al. "Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS)." *Multiple sclerosis (Houndmills, Basingstoke, England)* vol. 18,6 (2012): 891-8. doi:10.1177/1352458511431076

<sup>8</sup> Baek, William S et al. "Quality care assessment of Parkinson's disease at a tertiary medical center." *The International journal of neuroscience* vol. 123,4 (2013): 221-5. doi:10.3109/00207454.2012.751024

<sup>9</sup> Bachmann, Paula et al. 2022. "Prevalence and Severity of Potential Drug-Drug Interactions in Patients with Multiple Sclerosis with and without Polypharmacy" *Pharmaceutics* 14, no. 3: 592. <https://doi.org/10.3390/pharmaceutics14030592>

variation. The 90th percentile of score is approximately double the 10th percentile at both TIN and TIN-NPI levels. The results highlight an opportunity for improvement by closing the gap between the most and least efficient providers.

Table 1. Distribution of the Measure Score

Metric	TIN	TIN-NPI
Mean Score	\$14,704.36	\$14,553.99
Score Interquartile Range (IQR)	\$4,646.72	\$6,025.30
Standard Deviation	\$4,057.16	\$4,874.39
Coefficient of Variation	0.28	0.33
Score Percentile		
10 <sup>th</sup>	\$9,985.94	\$9,184.30
25 <sup>th</sup>	\$12,168.09	\$11,185.35
50 <sup>th</sup>	\$14,315.24	\$13,877.00
75 <sup>th</sup>	\$16,814.81	\$17,210.64
90 <sup>th</sup>	\$19,640.65	\$20,532.19