

Additional Testing for Non-Pressure Ulcers

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 (https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/reliability).</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. http://www.rand.org/pubs/technical_reports/TR653.html).</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: 9,174 At the 20-episode volume threshold: 4,060 At the 30-episode volume threshold: 2,111</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: 0.783 At the 20-episode volume threshold: 0.841 At the 30-episode volume threshold: 0.875</p>

<p>Measure Score Level (Accountable Entity Level) Testing</p>	<p>037</p>	<p>*Signal-to-Noise: Interpretation of results</p>	<p>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>	<p>Reliability testing of the Non-Pressure Ulcers measure is conducted for clinicians (TIN-NPIs) and constructed using episodes ending between January 1, 2023, and December 31, 2023. 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 Non-Pressure Ulcers measure is high, specifically 0.804 at the TIN-NPI level, and the majority of TIN-NPIs (96.65%) meet or exceed the moderate reliability threshold of 0.4 at the 20-episode testing volume threshold.</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 (https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/validity)</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>4,060 TIN-NPIs who meet the 20-episode volume threshold</p>

<p>Measure Score Level (Accountability Entity Level) Testing</p>	<p>045</p>	<p>*Empiric Validity: Methods and findings</p>	<p>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>	<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>Model 1: Mean O/E = Mean Cost of Treatment Choices + Mean Cost of Adverse Events</p> <ul style="list-style-type: none"> - Adverse events: 0.06 [0.06,0.07] (p < 0.01) - Outpatient Evaluation & Management Services: -0.03 [-0.08,0.01] (p = 0.13) - Major Procedures: 0.24 [0.04,0.44] (p = 0.01) - Ambulatory/Minor Procedures: 0.06 [0.05,0.07] (p < 0.01) - Laboratory, Pathology, and Other Tests: 0.90 [0.72,1.08] (p < 0.01) - Imaging Services: 0.34 [0.21,0.46] (p < 0.01) - Durable Medical Equipment and Supplies: 0.02 [0.02,0.02] (p < 0.01)
--	------------	--	--	---

				<ul style="list-style-type: none"> - Chemotherapy and Other Part B-Covered Drugs: 0.00 [0.00,0.00] (p = 0.95) - Part-D Drugs: 0.05 [0.00,0.11] (p = 0.05) <p>Model 2: Mean Cost of Adverse Events = Mean Cost of Treatment Choices</p> <ul style="list-style-type: none"> - Outpatient Evaluation & Management Services: 1.93 [1.74,2.12] (p < 0.01) - Major Procedures: 0.83 [-0.06,1.71] (p = 0.07) - Ambulatory/Minor Procedures: 0.28 [0.24,0.32] (p < 0.01) - Laboratory, Pathology, and Other Tests: -0.24 [-1.04,0.56] (p = 0.56) - Imaging Services: 1.53 [0.98,2.09] (p < 0.01) - Durable Medical Equipment and Supplies: 0.03 [0.02,0.04] (p < 0.01) - Chemotherapy and Other Part B-Covered Drugs: 0.03 [0.01,0.05] (p < 0.01) - Part-D Drugs: 0.99 [0.74,1.23] (p < 0.01) <p>Overall, testing results demonstrated that the cost measure reflects both the cost directly related to treatment choices and the cost of related adverse outcomes. Therefore, there is evidence that the measure captures what it purports to measure.</p> <p>Model 1 demonstrates that adverse events are associated with worse clinician performance at the group and individual reporting levels. Ambulatory/minor procedures, imaging services, and durable medical equipment are also associated with a worse measure score. Moreover, these services are associated with a</p>
--	--	--	--	---

				<p>higher cost of adverse events in Model 2, suggesting that the opportunities to reduce these costs are linked to the reduction of adverse events.</p> <p>Laboratory, pathology, and other test services, and major procedures are associated with worse clinician performance in Model 1, but not associated with the cost of adverse events in Model 2. This suggests that there is a potential for overuse of these services.</p> <p>Lastly, the cost of outpatient evaluation and management services, Part B drugs, and Part D drugs is shown to not be a significant driver of the measure score.</p>
Measure Score Level (Accountable 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	<p>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."</p> <p>Please enter a single value and do not enter a range.</p> <p>If unknown or not available, enter 9999.</p>	There are 4,060 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 4,060 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>	The 10th percentile score is \$3,702.85.
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>	The 50th percentile median score is \$7,757.83.

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>	The 90th percentile score is \$13,956.80.
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"> - Mean performance score across accountable entities in the test sample that is relevant to the intended use of the measure. - Minimum and maximum performance score for the testing sample that is relevant to the intended use of the measure. - Standard deviation of performance scores for the testing sample that is relevant to the intended use of the measure. - Passing score for the performance measure. - Performance score's defined interval, including upper and lower limit of the performance score. 	<p>Analysis of all clinicians (TIN-NPIs) with at least 20 attributed episodes for the 2023 performance period shows a wide range of provider scores of the Non-Pressure Ulcers measure. The measure score has the following distributional characteristics:</p> <ul style="list-style-type: none"> • Mean (SD): \$8,480.14 (\$4,669.90) • Min: \$594.88 • Max: \$87,295.15 <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

Chronic non-pressure ulcers are highly prevalent in the US Medicare population. In 2019, 16.3% of Medicare beneficiaries were affected by chronic ulcers, up from 14.5% in 2014.¹ Chronic ulcers can last over a year, are recurring in up to 70% of patients, and

¹ Sen CK. Human Wound and Its Burden: Updated 2022 Compendium of Estimates. *Adv Wound Care (New Rochelle)*. 2023;12(12):657-670. doi:10.1089/wound.2023.0150.

can lead to loss of function, decreased quality of life (QOL), and poor health outcomes.² Ulcers can heavily impact QOL for patients, as more than 85% of lower limb amputations are preceded by foot or ankle ulcers.³ Chronic non-pressure ulcers are also costly to the U.S. healthcare system. Total Medicare spending for all wound types is \$28.1 billion annually. Including noninfected and infected wound costs, the estimated cost of care for diabetic foot ulcers ranges from \$6.2 billion to \$18.7 billion, and \$0.7 billion to \$1.5 billion for venous leg ulcers,⁴ with the total cost for wounds ranging from \$31.7 to \$96.8 billion when they are included as a secondary diagnosis.⁵

The literature scan identified two critical areas of opportunities to improve care outcomes for patients with non-pressure ulcers and reduce costs associated with managing the condition. These include reducing recurring ulcers as well as lower limb amputations caused by non-healing wounds and creating a care management plan to coordinate appropriate treatment technologies. It is estimated that more than 85% of lower limb amputations are preceded by foot or ankle ulcers.⁶ Methods to correctly identify ulcer types and severity, such as color-flow duplex ultrasounds and plain radiographs,^{7,8} as well as continuous care of already identified wounds are vital components to preventing amputations. Additionally, the wide variety of existing technologies to treat ulcers raises the need to create care management plans tailored to specific patient needs. For instance, unless a diabetic wound has not healed by at least 50% in four weeks, clinicians should not consider skin grafts, as they have shown to slow healing time for neuropathic and arterial ulcers.^{9,10}

Existing literature and Acumen's testing indicate a high cost to Medicare for treating and managing non-pressure ulcers, 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 over 314,800 beneficiaries and over 8,000 clinician groups and individual clinicians combined (using 2023 as the study year). Table 1 shows the distribution of the measure score for TIN and TIN-NPI levels. There is variation observed in the measure

² *Optimal Care of Chronic, Non-Healing, Lower Extremity Wounds: A Review of Clinical Evidence and Guidelines*. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; December 17, 2013.

³ Suthar M, Gupta S, Bukhari S, Ponemone V. Treatment of chronic non-healing ulcers using autologous platelet rich plasma: a case series. *J Biomed Sci*. 2017 Feb 27;24(1):16. doi: 10.1186/s12929-017-0324-1. PMID: 28241824; PMCID: PMC5327512. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5327512/>

⁴ Agency for Healthcare Research and Quality (2018). Skin Substitutes for Treating Chronic Wounds. <https://effectivehealthcare.ahrq.gov/products/skin-substitutes/protocol>

⁵ Nussbaum SR, Carter MJ, Fife CE, et al. An Economic Evaluation of the Impact, Cost, and Medicare Policy Implications of Chronic Nonhealing Wounds. *Value Health*. 2018;21(1):27-32. doi:10.1016/j.jval.2017.07.007

⁶ Suthar M, Gupta S, Bukhari S, Ponemone V. Treatment of chronic non-healing ulcers using autologous platelet rich plasma: a case series. *J Biomed Sci*. 2017 Feb 27;24(1):16. doi: 10.1186/s12929-017-0324-1. PMID: 28241824; PMCID: PMC5327512. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5327512/>

⁷ Schneider C, Stratman S, Kirsner RS. Lower Extremity Ulcers. *Med Clin North Am*. 2021;105(4):663-679. doi:10.1016/j.mcna.2021.04.006

⁸ Eastman DM, Dreyer MA. Neuropathic Ulcer. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; September 28, 2022.

⁹ Agency for Healthcare Research and Quality (2018). Skin Substitutes for Treating Chronic Wounds. <https://effectivehealthcare.ahrq.gov/products/skin-substitutes/protocol>

¹⁰ Eastman DM, Dreyer MA. Neuropathic Ulcer. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; September 28, 2022.

score at both TIN and TIN-NPI levels, indicated by the interquartile ranges, standard deviations, and coefficients of variation. The 90th percentile of score is over triple the 10th percentile at both the 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	\$9,108.90	\$8,480.14
Score Interquartile Range (IQR)	\$4,615.62	\$4,925.52
Standard Deviation	\$4,398.51	\$4,669.90
Coefficient of Variation	0.48	0.55
Score Percentile		
10 th	\$4,484.67	\$3,702.85
25 th	\$6,405.43	\$5,517.01
50 th	\$8,504.64	\$7,757.83
75 th	\$11,021.05	\$10,442.53
90 th	\$14,156.62	\$13,956.80