Quality Payment
PROGRAM

Diabetes

Measure Justification Form

June 2021

Contents

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Diabetes measure. The form is intended to provide detailed information about the testing conducted on this measure, and accompanies the Measure Methodology[1] and measure Codes List[2] file, which together, comprise the specifications for this cost measure.

1.1 Project Title and Overview

The Centers for Medicare & Medicaid Services (CMS) has contracted with Acumen, LLC to develop care episode and patient condition groups for use in cost measures to meet the requirements of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). The contract name is "Physician Cost Measure and Patient Relationship Codes (PCMP)." The contract number is 75FCMC18D0015, Task Order 75FCMC19F0004.

1.2 Measure Name

Diabetes Episode-Based Cost Measure

1.3 Type of Measure

Cost/Resource Use

[1] CMS, "Diabetes Measure Methodology," MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.
[2] CMS, "Diabetes Measure Codes List" MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

2.0 Measure Testing: Importance

2.1 Evidence to Support the Measure Focus

2.1.1 Measure Description

The Diabetes cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients receiving medical care to manage diabetes. The measure score is a clinician's or clinician group's weighted average of risk-adjusted cost for each attributed episode, where each episode is weighted by the number of assigned days during the episode. This chronic measure includes services that are clinically related and under the reasonable influence of the attributed clinician or clinician group. Services are assigned during a Diabetes episode, which is a portion of the overall time period of a clinician's or clinician group's responsibility for managing a patient's diabetes. Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period are eligible for the measure.

2.1.2 Evidence for Measure Focus

The Diabetes measure was developed for use in the Merit-based Incentive Payment System (MIPS) to meet the requirements of the Social Security Act section 1848(r), added by the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). MIPS aims to reward high-value care by measuring clinician performance through four areas: quality, improvement activities, promoting interoperability, and cost. Each category assesses different aspects of care, and the categories are weighted such that they are combined into one composite score. CMS is introducing MIPS Value Pathways (MVPs) as a way to align and connect quality measures, cost measures, and improvement activities across performance categories of MIPS for different specialties or conditions. MVPs aim to provide a holistic assessment of clinician value for a specific type of care to achieve better healthcare outcomes and lower costs for patients. The use of cost measures is required by statute, and their purpose is to assess resource use. To be effective, they should capture costs related to a clinician's care decisions and account for factors outside of their influence.

This measure provides clinicians with information about their costs of care that they can use to understand the costs associated with their decision-making. Clinicians play an important role in variation in health care expenditures due to their ability to affect costs.[3] A cost measure offers opportunity for improvement if clinicians can exercise influence on the intensity or frequency of a significant share of costs during the episode, or if clinicians can achieve lower spending and better care quality through changes in clinical practice.

Diabetes mellitus is a group of metabolic disorders characterized by chronic hyperglycemia. The most common of these metabolic disorders in the Medicare population are type 1 and type 2 diabetes, both of which have their particular sets of causes, clinical manifestations, and management strategies, ranging from lifestyle changes to medication. Specifically, 7-12% of both the Medicare and broader United States diabetic population have type 1 diabetes, which is characterized by little to no insulin production by the insulin-producing beta cells of the

[3] David Cutler et al., "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," American Economic Journal: Economic Policy 11, no. 1 (February 1, 2019): 192–221, https://doi.org/10.1257/pol.20150421.

pancreatic islets.[4] Conversely, 87-91% of the Medicare and broader United States diabetic population have type 2 diabetes, which is characterized by insulin resistance.[5]

According to the literature and feedback received through stakeholder input activities to date, this measure's focus represents an area where there are opportunities for improvement. Primary opportunities for improvement include (i) promoting diabetes self-management education and support (DSME/S), (ii) increasing the use of appropriate medications, and (iii) encouraging adherence to correct preventive treatment guidelines. An increased focus on these types of preventative care can minimize downstream costs by mitigating the use of institutional post-acute care and inpatient stays, and reducing overutilization of other care for diabetes-related complications.

One way that clinicians may be able to contain costs associated with the management of diabetes is the promotion of DSME/S. Given that diabetes is a chronic condition that requires patients to make several daily self-management decisions, DSME/S provides diabetes patients with a foundation to navigate these decisions and activities that are necessary to manage their condition (e.g., through medical nutrition therapy or other appropriate specialist referrals).[6] For clinicians, there are national standards for DSME/S, which include but are not limited to developing an individualized DSME/S plan with diabetes patients, making diabetes patients aware of options and resources available for ongoing support of their initial education, and monitoring and communicating whether diabetes patients are achieving their self-management goals and other outcomes.[7] Through promoting DSME/S, managing clinicians have an opportunity to reduce their patients' diabetes-related hospital admissions and readmissions, reduce their lifetime health care costs for diabetes-related complications, improve their glycated hemoglobin (HbA_{1C}), an indicator of patient blood glucose levels, by as much as 1%, and reduce the onset or advancement of their diabetes-related complications, among other benefits.[8]

Increasing the use of appropriate medications offers another way for clinicians to contain costs associated with the management of diabetes. These pharmacological options, which are often supplemented by lifestyle changes, may vary depending on the type of diabetes. For patients with type 1 diabetes or poorly-controlled type 2 diabetes, insulin therapy helps to maintain normal blood glucose levels. In patients with type 1 diabetes, early and chronic exogenous insulin coverage, either through multiple daily injections or through use of an infusion pump, can reduce diabetes-related microvascular and macrovascular complications.[9,10] In patients with type 2 diabetes, insulin therapy can reduce diabetes-related microvascular complications and in the long-term, can improve cardiovascular prognosis.[11] Other diabetes management medications, such as metformin, aim to further regulate blood glucose levels by decreasing

---

[4] Juan José Marín-Peñalver et al., "Update on the Treatment of Type 2 Diabetes Mellitus," World Journal of Diabetes 7, no. 17 (September 2016): 354-95, https://doi.org/10.4239/wjd.v7.i17.354.
[5] International Diabetes Federation, "IDF Diabetes Atlas - 8th Edition, "https://diabetesatlas.org/upload/resources/previous/files/8/IDF_DA_8e-EN-final.pdf."
[6] Powers et al., "Diabetes Self-management Education and Support in Type 2 Diabetes: A Joint Position Statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics," Diabetes Care 38, no. 7 (July 2015): 1372-1382, https://doi.org/10.2337/dc15-0730.
[7] Beck et al., "2017 National Standards for Diabetes Self-Management Education and Support," Diabetes Care 40, no. 10 (October 2017): 1409-1419, https://doi.org/10.2337/dci17-0025.
[8] Powers et al.
[9] Juan José Marín-Peñalver et al.
[10] Home et al., "Insulin Therapy in People with Type 2 Diabetes: Opportunities and Challenges?," Diabetes Care 37, no. 6 (June 2014): 1499-1508, https://doi.org/10.2337/dc13-2743.
[11] Ibid.

gluconeogenesis or increasing pancreatic insulin secretion.[12] For most patients with type 2 diabetes, metformin is recommended as the preferred initial glucose lowering medication. This is due, in part, to its effectiveness in lowering blood glucose levels, its minimal hypoglycemia risk when used as monotherapy, and its weight loss benefits in some patients with type 2 diabetes.[13] Through identifying these and other appropriate medication(s) and promoting patient adherence to their medication regimes, managing clinicians have an opportunity to prevent the onset or progression of costly diabetes-related complications in their patients.

Current literature also suggests that the managing clinician has an opportunity to contain diabetes-related costs by encouraging adherence to correct preventive treatment guidelines. It is well established that poor monitoring and control of blood glucose, lipid levels, and blood pressure can drastically increase the risk and severity of diabetes-related complications. This is especially salient for older adults whose diabetes treatment may be complicated by their clinical, cognitive, and functional heterogeneity.[14] For example, higher rates of cognitive impairment in older adults have been associated with an increased risk of hypoglycemia, which can lead to falls, seizures, and loss of consciousness.[15,16] One study showed that lower cognitive ability was associated with a twofold higher incidence of severe hypoglycemia.[17] This study demonstrates that by screening older adults with diabetes for cognitive impairment during clinical visits, clinicians can better assess their patients' potential risk for worsening of their glycemic control, allowing clinicians to modify a patient's treatment plan to accommodate these cognitive changes and to continue to effectively manage their patient's diabetes care.[18] Furthermore, diabetic patients also face an increased risk of cardiovascular disease and require close monitoring of lipid profiles and blood pressure to prevent stroke, coronary artery disease (CAD), and heart failure.[19] One study found that improved control of HbA1C, lipid levels, and blood pressure predicted a 28-49% reduction in the probability of diabetes-related complications and a 7-10% decrease in total cost of care.[20] To manage blood pressure, during each office visit, clinicians should measure their diabetic patients' blood pressure. If the readings on at least 2 of the visits are ≥ 130/80 mmHg, then clinicians should initiate medications (e.g., ACE inhibitors or angiotensin receptor blockers (ARB)) and lifestyle changes (e.g., diet and exercise) for these patients.[21] For lipid levels, it is recommended that clinicians screen patients with diabetes annually for their fasting serum lipid levels, and for those with dyslipidemia, clinicians should encourage lifestyle interventions (e.g., medical nutrition therapy or smoking cessation) and/or

[12] Ambady Ramachandran, Chamukuttan Snehalatha, and Arun Nanditha, "Classification and Diagnosis of Diabetes," in Textbook of Diabetes, 2016, 23-28, https://doi.org/10.1002/9781118924853.ch2.
[13] Davies et al., "Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes," Diabetes Care 41, no. 12 (December 2018): 2669-2701, https://doi.org/10.2337/dci18-0033.
[14] American Diabetes Association, "Older Adults: Standards of Medical Care in Diabetes – 2020," Diabetes Care 43 (January 2020): 152-162, https://doi.org/10.2337/dc20-S012.
[15] Ibid.
[16] Mousumi Sircar, Ashmeet Bhatia, and Medha Munshi, "Review of Hypoglycemia in the Older Adult: Clinical Implications and Management," Canadian Journal of Diabetes 40, no. 1 (February 2016): 66-72, https://doi.org/10.1016/j.jcjd.2015.10.004.
[17] Feinkohl et al., "Severe Hypoglycemia and Cognitive Decline in Older People with Type 2 Diabetes: The Edinburgh Type 2 Diabetes Study," Diabetes Care 37, no. 2 (February 2014): 507-515, https://doi.org/10.2337/dc13-1384.
[18] American Diabetes Association, "Older Adults: Standards of Medical Care in Diabetes – 2020."
[19] Iciar Martín-Timón et al., "Type 2 Diabetes and Cardiovascular Disease: Have all Risk Factors the Same Strength?," World Journal of Diabetes 5, no. 4 (August 2014): 444–470, https://doi.org/10.4239/wjd.v5.i4.444.
[20] Kathryn Fitch, Bruce S. Pyenson, and Kosuke Iwasaki, "Medical Claim Cost Impact of Improved Diabetes Control for Medicare and Commercially Insured Patients with Type 2 Diabetes," Journal of Managed Care & Specialty Pharmacy 19, no. 8 (October 2013): 609-20, https://doi.org/10.18553/jmcp.2013.19.8.609.
[21] Amanda H. Salanitro and Christianne L. Roumie, "Blood Pressure Management in Patients with Diabetes," Clinical Diabetes 28, no. 3 (July 2010): 107-114, https://doi.org/10.2337/diaclin.28.3.107.

pharmacological interventions (e.g., statins) to control lipid levels.[22] In following these and other preventive treatment guidelines, managing clinicians have another avenue to stem the onset or progression of diabetes-related complications in their patients.

Literature suggests that given the high impact of diabetes within the Medicare patient population and consequential effect on Medicare spending, the Diabetes episode group represents an area with significant opportunity for improvement with respect to cost containment.

2.2 Performance Gap

2.2.1 Rationale

The high prevalence and cost of diabetes mellitus and its associated complications to the United States health care system warrants the exploration of potential cost measures which aim to achieve more cost-effective care for a given condition. In the United States, there are approximately 13.5 million people ages 65 and older living with diabetes, and treatment of diabetes in the United States costs over $348 billion annually.[23] In 2012, 59% of healthcare costs related to diabetes were associated with patients over the age of 65.[24] In 2017, approximately 57% ($9,600 out of $16,750) of annual medical expenditures incurred for patients diagnosed with diabetes were related to their diabetes diagnosis.[25] Additionally, on average, patients with diabetes had medical expenditures 2.3 times higher than those for patients without a diabetes diagnosis.

Significant cost drivers in the care of diabetes are the occurrence of acute complications such as acute hyperglycemic crises (diabetic ketoacidosis and hyperglycemic hyperosmolar nonketotic syndrome) and longer-term complications of diabetes such as retinopathy, neuropathy, diabetic foot ulcers, cardiovascular events, and amputations.[26] For example, over $2.4 billion in costs from hospital treatment were attributed to acute hyperglycemic crises, and over $1.84 billion for acute hypoglycemia and related injuries.[27,28] Overall, patients with multiple diabetes complications had a higher risk of readmissions for severe dysglycemia (hyperglycemia or hypoglycemia) as well as causes that are unrelated to diabetes. It was also estimated that the prevalence of diabetic retinopathy among diabetic patients 65 years and older was 29.5%.[29] Similarly, in 2007, 8.1% of Medicare diabetic beneficiaries enrolled in Medicare Parts A and B had diabetic foot ulcers, incurring spending that was significantly higher than that for beneficiaries without chronic wounds ($31,363 vs. $11,692, respectively).[30] Given

[22] Jaiswal et al., “Lipids and Lipid Management in Diabetes,” Best Practice & Research Clinical Endocrinology & Metabolism 28 (2014): 325-338, http://dx.doi.org/10.1016/j.beem.2013.12.001.
[23] International Diabetes Federation, "IDF Diabetes Atlas - 8th Edition."
[24] Mousumi Sircar, Ashmeet Bhatia, and Medha Munshi.
[25] American Diabetes Association, “Economic Costs of Diabetes in the U.S. in 2017,” Diabetes Care 41, no. 5 (May 2018): 917–928, https://doi.org/10.2337/dci18-0007.
[26] Baxter et al., "Estimating the Impact of Better Management of Glycaemic Control in Adults with Type 1 and Type 2 Diabetes on the Number of Clinical Complications and the Associated Financial Benefit," Diabetic Medicine 33, no. 11 (January 2016): 1575-1581, https://doi.org/10.1111/dme.13062.
[27] Guillermo Umpierrez and Mary Korytkowski, "Diabetic Emergencies — Ketoacidosis, Hyperglycaemic Hyperosmolar State and Hypoglycaemia," Nature Reviews Endocrinology 12 (February 2016): 222-232, https://doi.org/10.1038/nrendo.2016.15.
[28] Zhao et al., "Economic Burden of Hypoglycemia: Utilization of Emergency Department and Outpatient Services in the United States (2005–2009)," Journal of Medical Economics 19, no. 9 (April 2016): 852-857, https://doi.org/10.1080/13696998.2016.1178126.
[29] Zhang et al., “Prevalence of Diabetic Retinopathy in the United States, 2005-2008,” JAMA 304, no. 6 (August 2010): 649–656, https://doi.org/10.1001/jama.2010.1111.
[30] Michael Sargen, Ole Hoffstad, and David Margolis, “Geographic Variation in Medicare Spending and Mortality for Diabetic Patients with Foot Ulcers and Amputations.” Journal of Diabetes and its Complications 27, no. 2 (March-April 2013):128-133, https://doi.org/10.1016/j.jdiacomp.2012.09.003.

the prevalence of diabetes in the Medicare population, and the high costs associated with the management of the disease and its complications, the Diabetes cost measure represents an opportunity for improvement on overall cost performance.

The Diabetes episode-based cost measure was recommended for development by an expert clinician committee—the Chronic Condition and Disease Management Clinical Subcommittee. Based on the initial recommendations from the Clinical Subcommittee, the subsequent measure-specific Clinician Expert Workgroup provided extensive, detailed input on this measure.

2.2.2 Performance Scores

To demonstrate the performance gap captured in the measure, Table 1 below presents a distribution of performance scores for 38,996 clinician group practices (identified by Taxpayer Identification Number, or TIN) and 81,786 practitioners (identified by a unique TIN and National Provider Identifier pair, or TIN-NPI) attributed at least episodes in 2019. These counts represent attributed clinicians and clinician groups billing Part B Physician/Supplier claims under a MIPS eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced Alternative Payment Model participation).

Table 1. Distribution of Observed over Expected (O/E) Ratio

Metric	TIN	TIN-NPI
Mean O/E ratio	1.00	0.96
O/E ratio Interquartile Range (IQR)	0.34	0.38
O/E ratio Percentile		
10th	0.68	0.61
25th	0.81	0.75
50th	0.98	0.92
75th	1.16	1.13
90th	1.35	1.34

3.0 Scientific Acceptability

3.1 Data Sample Description

3.1.1 Type of Data Used for Testing

Medicare administrative claims, Long-Term Minimum Data Set (MDS), Medicare Enrollment Database (EDB), Common Medicare Environment (CME), and United States Census Bureau's American Community Survey (ACS).

3.1.2 Specific Dataset Used for Testing

The Diabetes measure uses Medicare Part A, Part B, and Part D claims data maintained by CMS. Part A, B, and D claims data are used to build episodes of care, calculate episode costs, and construct risk adjustors. To ensure that the measure accurately reflects Medicare costs, Part D branded drug costs were adjusted to account for drug rebates. More detailed information on the Part D payment standardization methodology and the Part D rebate adjustment methodology is available from the CMS Research Data Assistance Center.[31]

Episode costs are payment standardized and risk adjusted to ensure accurate comparison of cost across clinicians. Payment standardization adjusts the allowed amount for a Medicare service to limit observed differences in costs to those that may result from health care delivery choices. Data from the EDB are used to determine beneficiary-level (or patient-level) exclusions and secondary risk adjustors, specifically Medicare Parts A, B, and C enrollment, primary payer, disability status, end-stage renal disease (ESRD), patient birth dates, and patient death dates. The risk adjustment model also accounts for expected differences in payment for services provided to patients in long-term care based on data from the MDS. Specifically, the MDS is used to create the long-term care indicator variable in risk adjustment.

For measure testing, data from the ACS and CME are used in analyses evaluating social risk factors in risk adjustment.

3.1.3 Dates of the Data Used in Testing

Diabetes episodes ending from January 1, 2019 through December 31, 2019.

3.1.4 Levels of Analysis Tested

Individual clinician (identified by combination of TIN and NPI) and clinician group/practice (identified by TIN).

3.1.5 Entities Included in the Testing and Analysis

After applying exclusions and the case minimum, the final population for testing and analyses included 38,996 clinician group practices and 81,786 practitioners who were attributed 20 or more Diabetes episodes across all 50 states and the District of Columbia during the measurement period. The most frequent settings in which a Diabetes episode was triggered included:

- Ambulatory/office-based care
- Skilled nursing facility (SNF)
- Hospital outpatient department (HOD)

[31] CMS, Research Data Assistance Center, https://resdac.org/articles/cms-price-payment-standardization-overview.

3.1.6 Patient Cohort Included in the Testing and Analysis

4,527,680 Medicare patients, with a mean age of 72.80, (from 6,215,678 episodes) were included in measure testing and analyses.

The patient population for the Diabetes measure calculation consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) who receive medical care to manage diabetes that triggers a Diabetes episode. A Diabetes episode is identified by a "trigger event", which is the occurrence of 2 Part B Physician/Supplier (Carrier) claims billed by the same clinician group practice within 180 days of one another. These claims include:

- A trigger claim that is a "primary care" Evaluation & Management (E&M) code with a relevant diabetes diagnosis, and
- A confirming claim that is either another "primary care" E&M code with a relevant diabetes diagnosis, or a chronic condition-related Current Procedural Terminology/Healthcare Common Procedure Coding System (CPT/HCPCS) code for related services with a relevant diabetes diagnosis.

Patients and their episodes were excluded from the sample if they met a set of exclusion criteria (listed below) meant to ensure completeness of data and to focus the measure on a clinically homogeneous cohort of patients receiving medical care to manage diabetes.

The exclusion criteria are:

- The patient does not have Medicare as their primary payer for the entire episode window, as well as the 120-day lookback period prior to the episode window.
- The patient was not continuously enrolled in Medicare Parts A and B, and not enrolled in Part C, for the entirety of the episode window and the 120-day lookback period.
- The patient resided outside of the United States or its territories during the episode window.
- The patient was not found in the Medicare EDB.
- The patient has an episode window shorter than one year.
- The episode is an outlier case in the regression.
- The episode has no attributed clinician (only applied at the TIN-NPI level).
- The episode does not fall in any defined measure sub-groups (Type 1 Diabetes or Type 2 Diabetes).[32]
- The patient received hospice care.

To determine whether the Diabetes measure's exclusion criteria distort patient characteristics on episodes, we produced and analyzed distributions of patient characteristics (age, race, sex, dual eligibility status, income, unemployment, hierarchical condition categories [HCCs]) for (i) episodes with exclusion criteria, (ii) episodes without exclusion criteria, (iii) patients with exclusion criteria, and (iv) patients without exclusion criteria.

This analysis shows that the Diabetes measure's exclusion criteria have a minimal effect on the percentage of patients in any particular demographic category. The difference between patients being excluded and included in the measure is 3.81 or less percentage points across each of the characteristics in the analysis at TIN level testing, and 5.41 or less percentage points at TIN-NPI level testing. To illustrate, the percentage of patients aged 65 to 69 is 22.73% without applying the exclusion criteria, compared to 21.47% after applying the exclusion criteria at the TIN level. Furthermore, the difference in the percentage of patients across race categories with

[32] Sub-groups represent more granular, mutually exclusive and exhaustive patient populations defined by clinical criteria collected from claims found during a year of the patient's data.

and without the exclusion criteria is 3.08 or less percentage points at both TIN and TIN-NPI level testing. When it comes to sex, there is a difference of 0.38 or less percentage points between the included and excluded populations with regards to the share of male and female patients (for both TIN and TIN-NPI level testing). These results indicate that there is minimal shift in patient characteristics as a result of using the exclusion criteria listed above at both TIN and TIN-NPI level testing.

3.1.7 Social Risk Factors Included in Analysis

The social risk factors analyzed were variables from the ACS, EDB, and CME. ACS variables are either at the Census Block Group or Zone Improvement Plan (ZIP) Code level. Social risk variables analyzed include the following:

- Race (EDB)
  - Asian, Black, Hispanic, North American Native, White, and Other
- Sex (EDB)
  - Female, male
- Dual status (CME)
  - Full dual, partial dual, non-dual to indicate whether a patient is dually enrolled in Medicare and Medicaid
- Income (ACS)
  - Low Income: median income < 33rd percentile nationally
  - Medium Income: median income in the interval spanning the 33rd percentile to the 66th percentile nationally
  - High Income: median income > 66th percentile
- Education (ACS)
  - Education < High School: when % with < high school education is the highest for a given Census Block Group
  - Education = High School: when % with only high school is the highest
  - Education > High School: when % with > high school is the highest
- Employment (ACS)
  - Unemployment Rate > 10%
  - Unemployment Rate <= 10%
- Agency for Healthcare Research and Quality (AHRQ) Socioeconomic Status (SES) Index (ACS)
  - Continuous variable (composite score of multiple community-level metrics, such as property values, density of living spaces, and poverty level) that can theoretically range from 0 to 100[33]

3.2 Reliability Testing

3.2.1 Level of Reliability Testing

The following levels of reliability were tested: critical data elements used in the measure and performance measure score (e.g., signal-to-noise analysis).

3.2.2 Method of Reliability Testing

Data Element Reliability

The Diabetes measure is constructed using CMS claims data, as described in Section 3.1.2. CMS has implemented several auditing programs to assess overall claims code accuracy,

[33] Refer to Section 3, page 42 of this AHRQ publication for the scoring algorithm used to calculate the AHRQ SES index variable.

ensure appropriate billing, and recoup any overpayments. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in this measure, including diagnosis and procedure codes and other elements that are consequential to payment. Specifically, CMS works with Zone Program Integrity Contractors, and formerly Program Safeguard Contractors, to ensure program integrity; the agency also uses Recovery Audit Contractors to identify and correct for underpayments and overpayments.

CMS also uses the Comprehensive Error Rate Testing (CERT) Program to ensure that Medicare payments are correct in accordance with coverage, coding, and billing rules. Between 2005 and 2019, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3% to 96.4% of total payments each year.[34] The fiscal year 2020 Medicare fee-for-service program proper payment rate was 93.7%.[35] CMS continues to perform successful corrective actions and give providers additional education to ensure accurate billing.

To ensure claims completeness and inclusion of any corrections, the measure was developed and tested using data with a three month claims run-out from the end of the measurement period.

Measure Reliability

Measure reliability is the degree to which repeated measurements of the same entity agree with each other. For measures of clinician performance, the measured entity is the TIN or TIN-NPI, and reliability is the extent to which repeated measurements of the TIN or TIN-NPI give similar results. To estimate measure reliability, we used a signal-to-noise analysis.

This approach seeks to determine the extent to which variation in the measure is due to true, underlying clinician performance, rather than random variation (i.e., statistical noise) within clinicians due to the sample of cases observed. To achieve this, we calculate reliability scores as:

$$R_j = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{w_j}^2}$$

Where:

$\sigma_{w_j}^2$ is the within-group variance of the mean measure score of clinician *j*

σ_b^2 is the between-group variance of clinicians within the episode group

That is, reliability is calculated as the ratio of between-group variance to the sum of between-group variance and within-group variance. Reliability closer to a value of one indicates that the between-group variance is relatively large compared to the within-group variance, which suggests that the measure is effectively capturing the systematic differences between the clinician and their peer cohort.

[34] Comprehensive Error Rate Testing (CERT) Program. “Appendices Medicare Fee-for-Service 2020 Improper Payments Report”. Table A6. https://www.cms.gov/files/document/2020-medicare-fee-service-supplemental-improper-payment-data.pdf-1.
[35] Ibid.

3.2.3 Statistical Results from Reliability Testing

Measure Reliability

At the proposed case minimum of 20 episodes, the mean reliability for TINs is 0.57 and for TIN-NPIs is 0.61. The majority of TINs and TIN-NPIs meet or exceed 0.4 reliability at the 20 episode case minimum, with 82.51% of TINs and 81.66% of TIN-NPIs meeting or exceeding the 0.4 threshold.

3.2.4 Interpretation

Measure Reliability

The mean reliability of the Diabetes measure exceeds 0.4 at the proposed case minimum of 20 episodes or more for both TINs and TIN-NPIs partly due to the large number of episodes attributed to clinicians. CMS generally considers 0.4 as the threshold indicating 'moderate' reliability, which is supported by previous work into reliability and the threshold was finalized in the CY 2017 Quality Payment Program final rule.[36,37] See the CY2021 Physician Fee Schedule (PFS) proposed rule for further discussion of cost measure reliability thresholds.

Though higher volume thresholds typically yield even higher reliability results, it is at the cost of further reducing the number of clinicians and clinician groups that are eligible to receive a measure score. The proposed Diabetes measure case minimum of 20 episodes was selected in part to strike a balance between these considerations.

3.3 Validity Testing

3.3.1 Level of Validity Testing

Our performance measure score validity testing included systematic assessment of both face validity and empirical validity testing.

3.3.2 Method of Validity Testing

Face Validity

The Diabetes measure was developed through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. Experts in this clinical area evaluated specifications to ensure that each aspect of the measure (e.g., assigned services) was intentionally capturing only the costs of care within the reasonable influence of the attributed clinician for a defined patient population (i.e., the ability of the measure score to differentiate good from poor performance).

In developing this measure, Acumen incorporated input from:

(i) a Chronic Condition and Disease Management Clinical Subcommittee;
(ii) a Diabetes Clinician Expert Workgroup;
(iii) a Technical Expert Panel (TEP); and
(iv) person and family partners.

This process is detailed in the Episode-Based Cost Measures Development Process document posted on the MACRA Feedback Page.[38]

---

[36] Mathematica, Inc., "Memorandum: Reporting Period and Reliability of AHRQ, CMS 30-Day and HAC Quality Measures – Revised," http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf.
[37] CMS, "CY 2017 Quality Payment Program final rule," 81 FR 77169-77170.
[38] CMS, "2020 Episode-Based Cost Measure Field Testing Wave 3 Measure Development Process," MACRA Feedback Page, https://www.cms.gov/files/document/macra-cmft-ebcm-process-2020.pdf.

One of the key roles of the measure-specific Clinician Expert Workgroup was to develop service assignment rules for the cost measure. These service assignment rules are intended to ensure clinicians are evaluated on services and costs that are clinically related to the attributed clinician's role in managing a patient's diabetes care, thus limiting cost variation unrelated to clinician care this measure. Services performed in the following service categories are considered for assignment to the episode: outpatient (OP) facility and clinician services, emergency department (ED), acute inpatient (IP) – medical, acute IP – surgical, inpatient rehabilitation facility (IRF), long term care hospital (LTCH), durable medical equipment, prosthetics, orthotics, and supplies (DME), home health (HH), SNF, and Part D prescription drugs.

Empirical Validity Testing

We evaluated the empirical validity of the Diabetes measure by examining correlation with known indicators of resource or service utilization based on a literature review, specifically complications related to diabetes. For this analysis, we compared the ratio of observed over expected (O/E) spending at the provider level for Diabetes episodes with and without complications. The analysis sought to confirm the expectation that the Diabetes measure captures variation in service utilization as an indicator of clinician cost performance. We expect episodes with downstream acute readmissions or post-acute care (IRF, LTCH, HH, and SNF) would have higher O/E cost ratios, since complications like these should yield higher cost, even after accounting for patient clinical characteristics via risk adjustment. Conversely, episodes without these downstream costs should have lower O/E cost ratios, demonstrating that the measure can differentiate good from poor cost performance.

3.3.3 Statistical Results from Validity Testing

Table 2 below presents the results from the validity analysis. The mean O/E cost ratio for all episodes is 0.98. The mean O/E cost ratio for episodes with downstream acute readmission is 2.98 compared with 0.70 for episodes without downstream acute readmission. Similarly, the mean O/E cost ratio for episodes with post-acute care is 2.60 compared with 0.83 for episodes without post-acute care. Additionally, there is greater variation in the O/E cost ratio among episodes with downstream acute readmission and post-acute care.

Table 2: Distribution of Observed to Expected Ratios

Episode Type	Observed / Expected Ratio										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
All Final Episodes	0.98	1.37	0.04	0.08	0.12	0.22	0.51	1.19	2.32	3.35	6.67
Episodes with Downstream Acute (Re)admission	2.98	2.37	0.55	0.81	1.01	1.49	2.30	3.60	5.64	7.57	12.70
Episodes without Downstream Acute (Re)admission	0.70	0.85	0.04	0.08	0.11	0.20	0.42	0.89	1.61	2.22	4.07
Episodes with Post-Acute Care (IRF LTCH HH SN)	2.60	2.30	0.27	0.49	0.67	1.15	1.98	3.23	5.08	6.96	12.20
Episodes without Post-Acute Care (IRF LTCH HH SN)	0.83	1.13	0.04	0.08	0.12	0.21	0.45	1.01	1.90	2.75	5.46

3.3.4 Interpretation

As expected, the average O/E cost ratios for episodes with complications (i.e., downstream acute readmissions and post-acute care) are higher than for episodes without downstream complications. These results demonstrate that the Diabetes measure is able to accurately capture higher resource use, and suggests that episodes with complications (the frequency or severity of which could be reasonably expected to be influenced by the treatment of the attributed clinician or clinician group) will yield higher costs, even after risk adjustment.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the Diabetes measure to ensure a comparable patient population within the scope of the measure's focus on the management of diabetes and that episodes provide meaningful information to attributed clinicians. Exclusions are also used as part of data processing so that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode. For the exclusions analysis discussed in this section, we focused on exclusions added to ensure a homogenous patient population. These exclusions, along with their rationales, are listed below:

- Episodes where the patient's episode window length is less than one year.
    - These episodes were excluded because the methodology for the chronic measures requires at least one year of claims data to measure clinician cost performance during an open attribution window for a performance period. Additionally, this exclusion may capture episodes during which a patient died, given that there may be insufficient data for these episodes. However, episodes with a death event are still included as long as the episode window is at least one year long.
- Episodes where there is not an attributed clinician.
    - These episodes were excluded because the episode does not have any TIN-NPIs that billed at least 30% of 'primary care' E&M codes with a relevant diabetes diagnosis and/or chronic condition-related CPT/HCPCS codes for related services with a relevant diabetes diagnosis on Part B Physician/Supplier (Carrier) claim lines during the episode within the attributed TIN. This exclusion only applies to episodes at the TIN-NPI level, while attributed TIN would continue to be attributed these episodes.
- Episodes where the patient is not in a defined measure sub-group.
    - These episodes were excluded because the patient's diabetes type could not be determined based on their available claims data. Episodes are sub-grouped as being either Type 1 Diabetes or Type 2 Diabetes to ensure clinical comparability so that the measure fairly compares clinicians with a similar patient case-mix.
- Episodes where the patient received hospice care.
    - These episodes were excluded because patients receiving hospice care are more ill and clinically complex than the overall patient cohort. The variance in costs for this high-risk patient cohort is also expected to be higher and would likely not be adequately accounted for by risk adjustment.
- Episodes classified as outlier cases.
    - To account for limitations of risk adjustment, episodes predicted to have expected costs that are substantially different from observed costs are excluded as outliers. Specifically, episodes with residuals from the risk adjustment model below the 1st percentile and above the 99th percentile are considered outliers and removed from measure calculation.

Given the rationales for these exclusions, we would expect these excluded episodes to have a different risk profile than the included episodes, such as a higher mean cost, or a different distribution of costs (e.g., a long tail of high-cost episodes). For the exclusions, we examined the number of episodes and patients affected, as well as the distributions of observed cost and ratio of observed over expected spending (calculated by applying existing risk factor coefficients to the excluded episodes) for excluded episodes. We then compared the cost characteristics of the excluded episodes to those of final episodes included in measure calculation to assess the distinctness between the 2 patient cohorts. A full list of the exclusions used for the Diabetes measure is provided in the Measure Codes List available on the MACRA Feedback Page.[39]

3.4.2 Statistical Results from Testing Exclusions

Table 3 below presents observed cost statistics and O/E cost ratios for the Diabetes measure exclusions. Cost statistics are also provided for the set of final episodes included in the Diabetes measure for comparison, with a testing volume threshold of 20 episodes at the TIN and TIN-NPI levels. For the standard exclusions in the table below (i.e., episode length less than one year, no defined episode sub-group, and no attributed clinician (TIN-NPI level)), these patient cohorts are excluded from the measure in order to assess episodes in the intended setting and by the measure's intended attribution approach.

Table 3: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost			O/E Cost Ratio		
			Mean	Percentile		Mean	Percentile	
	#	%		10th	90th		10th	90th
All Episodes Meeting Triggering Logic	6,918,490	100.00%	$9,217	$502	$24,036	1.25	0.12	2.73
Episode Length Less Than One Year	209,685	3.03%	$48,077	$2,213	$123,238	4.33	0.26	10.50
No Defined Measure Sub-Group	106,845	1.54%	$13,610	$405	$33,776	1.00	1.00	1.00
No Attributed Clinician (TIN-NPI Reporting Only)	1,272,121	18.39%	$12,650	$721	$34,191	1.38	0.15	3.02
Hospice Care	317,580	4.59%	$25,935	$1,226	$65,310	2.84	0.18	6.92
Outlier Cases	124,301	1.80%	$26,805	$1,385	$52,029	4.57	0.07	12.38
Final Episodes (TIN)	5,913,209	85.47%	$6,620	$481	$17,856	0.99	0.12	2.24
Final Episodes (TIN-NPI)	4,104,029	59.32%	$6,020	$463	$16,029	0.99	0.12	2.19

*This table does not include all measure exclusions.

3.4.3 Interpretation

The statistical results indicate that the majority of excluded episodes differ substantially in mean observed cost, mean O/E cost ratio, and/or cost (or O/E cost ratio) variation compared to the final set of episodes. These results support the exclusion of these episodes to ensure a comparable patient cohort that will yield meaningful information to attributed clinicians. Further discussion of the results for exclusions applied based on the clinical validity of the study population are provided below.

Episodes where the patient received hospice care: As expected, these episodes have higher costs and higher O/E cost ratios than the final set of episodes. The mean observed cost for these episodes is $25,935, compared to $6,620 at the TIN level and $6,020 at the TIN-NPI

[39] CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

level. These episodes also have a high mean O/E cost ratio (2.84), compared to final episodes at the TIN and TIN-NPI levels (0.99 each). These discrepancies in O/E cost ratios become more noticeable at the 90th percentile, where the O/E cost ratio for these episodes is 6.92, compared to 2.24 and 2.19 at the TIN and TIN-NPI levels, respectively.

Episodes classified as outlier cases: These episodes have a mean observed cost of $26,805, which is substantially higher than the mean observed costs for final episodes at both the TIN and TIN-NPI levels. The O/E cost ratio for outlier cases ranges from 0.07 at the 10th percentile to 12.38 at the 90th percentile, indicating that the risk adjustment model is currently unable to account for the patient characteristics associated with these high- and low-cost outlier episodes. Excluding outliers based on risk-adjusted cost eliminates the episodes that deviate most from expected spending levels based on patient characteristics.

3.5 Risk Adjustment or Stratification

3.5.1 Method of Controlling for Differences

Differences in case mix are controlled for using a statistical risk model with 128 risk factors and stratification by 4 risk categories. These 4 risk categories account for the 2 episode sub-groups, both of which are stratified by Part D enrollment status (either enrolled or not in Medicare Part D during the episode window).

The risk adjustment model for the Diabetes measure broadly follows the CMS-HCC risk adjustment methodology, which is derived from Medicare Parts A and B claims and is used in the Medicare Advantage (MA) program. Patient age is included as 1 of 12 age categorical variables derived from the MA risk adjustment model's age/sex variables. Severity of illness is measured using HCCs, indicators of enrollment and long-term care status, and disease interactions. The risk adjustment model also includes dual Medicare and Medicaid eligibility status and variables for factors identified by the expert clinician workgroup as affecting resource use.

The model includes 79 HCC indicators derived from the patient's Parts A and B claims during the period 120 days prior to the episode trigger claim and are specified in the CMS-HCC Version 22 (V22) 2016 model. Episodes for patients without a full 120-day lookback period are excluded from the measure. This 120-day period is used to measure patient health status and ensures that each patient's claims record contains sufficient fee-for-service data both for measuring spending levels and for risk adjustment purposes.

In addition, the risk adjustment model includes status indicator variables for whether the patient qualifies for Medicare through Disability or ESRD. The model also includes an indicator of whether the patient recently required long-term care, defined as 90 days in a long-term care facility without being discharged to community for 14 days. Patients who need to reside in long-term care facilities typically require more intensive care than patients who live in the community. These enrollment and long-term care status variables are non-diagnostic indicators of severity of illness.

The model also accounts for disease interactions between HCCs and/or enrollment status variables included in the MA model. These interactions are included because certain combinations of comorbidities increase costs more than is predicted by the HCC indicators alone.

Furthermore, the risk adjustment model includes measure-specific factors intended to further isolate costs that attributed clinicians can reasonably influence, informed by expert clinician input and empirical analyses. The following variables were added to avoid potential unintended consequences, including whether the patient:

- Had dementia
- Had a recent all-cause admission in prior 90 days
- Had an amputation
- Has an intravitreal Bevacizumab injection
- Had a prior intravitreal Bevacizumab injection
- Had a coronary artery bypass graft (CABG)
- Had a prior carotid endarterectomy/stent
- Has or had continuous glucose monitoring or an insulin pump
- Had gastric bypass/bariatric surgery
- Had prior peripheral vascular interventions
- Had a prior percutaneous coronary intervention
- Has an intravitreal Ranibizumab or Aflibercept injection
- Had a prior intravitreal Ranibizumab or Aflibercept injection

The risk adjustment approach for this measure uses an ordinary least squares linear regression model for each episode sub-group and Medicare Part D enrollment status combination to ensure fair comparison. The episode group's annualized observed costs are winsorized at the 1st and 99th percentiles prior to the regression for each model to handle extreme observations. Then, the predicted, or expected, cost is winsorized at 0.5th percentile to make sure episodes with unusually small predicted cost, which would lead to abnormally large O/E cost ratios, do not dominate certain clinicians' final score. The winsorized expected costs are renormalized to ensure the average expected episode cost is the same before and after winsorizing. Then, as presented in the exclusions analysis above, extremely low- or high-cost outlier episodes with residuals below the 1st percentile or above the 99th percentile are excluded to reduce the effect of episodes that deviate the most from their expected values in absolute terms. The expected cost after excluding these outliers is again renormalized to ensure that average expected costs are the same after outlier removal.

Finally, the risk adjustment model outlined above is stratified for each of the 2 Diabetes measure sub-groups, which are based on the patient's diabetes type, below:

- Type 1 Diabetes
- Type 2 Diabetes

Once patients have been sub-grouped, episode sub-groups are stratified by a patient's Medicare Part D enrollment status (either enrolled or not enrolled in Part D). This means that for each measure-specific sub-group, a separate risk adjustment model is run for patients with and without Part D enrollment. This is done to account for differences in patient populations and their associated cost with and without Part D enrollment, and stratifying by Part D enrollment improves model fit compared to not stratifying by enrollment status.

Full details of the risk adjustment model are in the draft Measure Codes List File.[40]

3.5.2 Conceptual, Clinical, and Statistical Methods

We selected the CMS-HCC model based on previous studies evaluating its appropriateness for use in risk adjusting Medicare claims data. This model was developed specifically for use in the Medicare population, meaning that it accounts for conditions found in the Medicare population and is calibrated on Medicare fee-for-service beneficiaries. In addition, the CMS-HCC model is

[40] CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

routinely updated for changes in coding practices (e.g., the transition from the 9th revision of the International Statistical Classification of Diseases and Related Health Problems, or ICD-9, to ICD-10 codes) and is exhaustive on these code sets. Because the CMS-HCC model has already been extensively tested, we focus our testing on how the CMS-HCC model was adapted to the Diabetes measure methodology.

The Clinician Expert Workgroup provided input on measure-specific risk adjustors after reviewing empirical analyses on subpopulations of interest to assess whether and if so, how, particular factors should be accounted for in the model. These could include patient characteristics, factors outside of the reasonable influence of the clinician, or any other factors that would help prevent unintended consequences. These additional risk adjustors are listed in the section above.

As previously noted, the risk adjustment model is run on episodes stratified into episode sub-groups, which may qualify as "ordering" of risk factors. Episode sub-groups were also determined based the workgroup's input, with the goal of ensuring clinical comparability among episodes so that the cost measure fairly compares clinicians with similar patient case-mix. The episode sub-groups are listed in the above section. Patients are categorized into these 2 episode sub-groups, because patients with either type 1 or type 2 diabetes comprise 2 clinically distinct patient populations.

3.5.3 Conceptual Model of Impact of Social Risks

Our conceptual model of the impact of social risk factors is informed by both published external research and our own data analysis.[41,42,43]

3.5.4 Statistical Results

The literature has extensively tested the use of the HCC model as applied to Medicare claims data. Although the variables in the HCC model were chosen to predict annual cost, CMS has also used this risk adjustment model in a number of other settings (e.g., accountable care organizations (ACOs), previous physician Quality and Resource Use Reports (QRUR) programs, and other measures such as the National Quality Forum (NQF) #3512: Knee Arthroplasty, NQF #3509: Routine Cataract Removal with Intraocular Lens (IOL) Implantation, NQF #3510: Screening/Surveillance Colonoscopy, and NQF #2158: MSPB-Hospital cost measure). Recalling that the risk model relies on the existing CMS-HCC model, testing results for factors included in the CMS-HCC V22 2016 model can be found in the Evaluation of the CMS-HCC Risk-Adjustment Model report[44] and the Report to Congress: Risk Adjustment in Medicare Advantage[45]. For measure-specific factors not included in the CMS-HCC model, we sought expert clinician input through the workgroup, which provided recommendations on additional risk adjustors and episode sub-groups.

---

[41] Assistant Secretary of Health and Human Services for Planning and Evaluation. Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs. Washington, D.C. December 2016.
[42] Chen LM, Epstein AM, Orav EJ, Filice CE, Samson LW, Joynt Maddox KE. Association of Practice-Level Social and Medical Risk With Performance in the Medicare Physician Value-Based Payment Modifier Program. JAMA. 2017; 318(5):453-461.
[43] Medicare Payment Advisory Commission. Beneficiaries Dually Eligible for Medicare and Medicaid. 2018; https://www.macpac.gov/publication/data-book-beneficiaries-dually-eligible-for-medicare-and-medicaid-3/.
[44] Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.
[45] CMS, "Report to Congress: Risk Adjustment in Medicare Advantage," https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf.

3.5.5 Analyses and Interpretation in Selection of Social Risk Factors

Acumen analyzed sex, dual status, income, education, and unemployment as social risk factors (more information on these variables can be found in Section 3.1.7). Patient sex and dual status were obtained from the EDB and CME. Information on income, education, and unemployment was obtained from ACS data and linked to episodes by census block group where possible to provide a more granular level of analysis than ZIP code. Patients without geographic information necessary to obtain ACS data were excluded, representing less than 2% of episodes.

The percentage of female patients range from 47.99% to 53.93% across the 2 episode sub-groups, stratified by Part D enrollment status, in this measure. The majority of the patients (71.45% - 99.41%) have non-dual status. Income level is categorized into high, medium, and low from the continuous average income variable in ACS; therefore, each category has 33% of observations. While 1.26% to 3.65% of patients are classified as having below a high school education level, the overwhelming majority of episodes are classified at a high school level or greater. Finally, 16.61% to 19.94% of patients have high unemployment designation (>10%).

Acumen examined the impact of including social risk factors into our risk adjustment model by running goodness of fit tests when different risk factors are added and compared to the base risk adjustment model, where the base risk adjustment model refers to the full standard set of risk adjustment variables from the CMS-HCC V22 2016 model, disability status, ESRD status, interaction variables, recent long-term care use, and measure-specific clinical risk adjustors. Acumen ran a step-wise regression to include the following additional social risk factors on top of the adapted CMS-HCC model:

- Sex
- Dual status
- Sex + dual status
- Sex + dual status + race
- Sex + dual status + income + education + unemployment
- Sex + dual status + AHRQ SES index score
- Sex + dual status + race + income + education + unemployment
- Sex + dual status + race + AHRQ SES index score

The step-wise regressions help evaluate individual as well as joint significance of the social risk factors. We examined the impact of including social risk factors into our risk adjustment model with T-test of individual significance and F-test of joint significance.

Our analysis of the correlation between Diabetes measure scores calculated with and without the social risk factors found that measure scores calculated with and without these social factors were highly correlated at the TIN level, with a Spearman correlation coefficient of 0.95, and the TIN-NPI level, with a correlation coefficient of 0.96.

3.5.6 Methods for Statistical Model or Stratification Development

To analyze the validity of the current risk adjustment model, we examined 2 analyses: (1) R-squared and adjusted R-squared for the regression models, and (2) predictive ratios and O/E cost ratios to examine the fit of the models at different levels of patient complexity.

1) R-squared and adjusted R-squared were calculated for the measure. These results should be evaluated in the context of the measure's service assignment rules which are intended to ensure only clinically associated costs are grouped to episodes. This is an important distinction from all-cost measures as service assignment leaves less variation for the risk adjustment model to explain. In this context, a low R-squared may indicate the effectiveness of the service assignment rules. These results are provided in Section 3.5.7.

2) Predictive ratios and O/E cost ratios were calculated for each "risk decile" for the episode group. A "risk decile" is based on the risk scores, which indicate how costly episodes are expected to be, as predicted through risk adjustment. After arranging episodes into deciles based on their risk score, we calculated the predictive ratios and average O/E cost ratios for each decile. The predictive ratio aims to examine the fit of the model at different levels of patient complexity to examine the model's ability to predict both very low and high cost episodes, and is calculated using the formula of average (expected cost)/average (observed cost) for all episodes in each decile. Similarly, the O/E cost ratio demonstrates the model's prediction accuracy, and is calculated using the formula of average (observed cost/expected cost) for all episodes in each decile. These are discussed in Sections 3.5.8 and 3.5.9.

3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Diabetes cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.27 The adjusted R-squared is 0.27 More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.[46]

3.5.8 Statistical Risk Model Calibration Statistics

We interpret calibration as how accurately the risk model's predictions match the actual episode cost. We calculate the average O/E cost ratio for each risk decile to demonstrate the model's prediction accuracy. Across all episodes, the average O/E cost ratio is 1.02, with average ratios ranging from 0.95 (5thrisk decile) to 1.12 (2nd risk decile). The 1st through 4th risk deciles have average O/E cost ratios ranging from 1.01 to 1.12, while the 5th through 9th risk deciles have average O/E cost ratios ranging from 0.95 to 0.99, and the 10th risk decile has an average O/E cost ratio of 1.02. This indicates that the model moderately under-predicts observed episode cost for the lowest risk episodes. Full results are available in the National Summary Data Report (NSDR) addendum on the MACRA Feedback Page.[47]

3.5.9 Statistical Risk Model Calibration – Risk Decile

Analysis of predictive ratios by risk decile for the measure shows that the model has moderate variation in predictive ratios across risk score deciles, as predictive ratios range from 0.90 (1st risk decile) to 1.06 (6th risk decile). This variation is largely being driven by the first 3 and 10th risk deciles (ranging from 0.90 to 0.92); removing these deciles would reduce the range to 0.08 (0.98 to 1.06). These results indicate that the model moderately under-predicts low cost episodes in the lowest risk deciles.

3.5.10 Interpretation

The R-squared values for the model, which measure the percentage of variation in results predicted by the model, are higher than the values presented in similar analyses of risk adjustment models.[48] As noted in Section 3.5.6, these results should be interpreted alongside service assignment rules, which remove clinically unrelated services. In addition to selecting a relatively homogenous population, the resulting variation is more likely to reflect the variation driven by factors within a clinician's reasonable influence.

---

[46] Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.
[47] CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.
[48] Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

As demonstrated in Section 3.5.8, the average O/E cost ratios are highest in the lowest risk deciles. Furthermore, as demonstrated in Section 3.5.9, the predictive ratios are lowest in the lowest risk deciles. These results indicate that the model under-predicts observed episode costs for the least risky episodes (risk deciles 1-4), while it better predicts observed episode costs for more risky episodes.

3.6 Identification of Meaningful Differences in Performance

3.6.1 Method

Our method of determining clinically meaningful differences in episode-based cost measure performance consists of stratifying clinician measure O/E cost ratios by meaningful characteristics and investigating the clinician O/E cost ratio distribution by percentile. The cost measure score numerator is the sum of the O/E cost ratio for all episodes attributed to a clinician. This sum is then multiplied by the national average observed episode cost to generate a dollar figure. The denominator is the total number of episodes from the attributed to a clinician. Using O/E cost ratios allows for direct comparisons of performance at the episode sub-group level since a dollar figure cannot be calculated for those episodes using the national average observed episode cost. Stratification is performed for each of the following characteristics: urban/rural, census division, census region, risk score, and the number of episodes attributed to the clinician or clinician group. We analyze the distribution of measure O/E cost ratios for clinicians defined by these characteristics.

The purpose of this analysis is to ensure that there is a sufficiently large difference in measure O/E cost ratios among clinicians to determine a meaningful difference in performance. In addition, this analysis looks to confirm that the measure behaves as expected with respect to meaningful clinician characteristics.

3.6.2 Statistical Results

Key findings show that, generally, there is a large performance difference among clinicians in the Diabetes measure:

(i) The 99th percentile of the measure O/E cost ratio is approximately 4 times the measure O/E cost ratio at the 1st percentile for both the TIN level and TIN-NPI levels; and
(ii) The O/E cost ratio at the 90th percentile is approximately 98.53% and 119.62% greater than the O/E cost ratio at the 10th percentile at both the TIN and TIN-NPI levels, respectively.

These results in conjunction with the measure reliability results presented in Section 3.2.3 that show most of the variation is among providers, indicate there is a large potential for Medicare costs savings.

In terms of regional difference in clinician O/E cost ratio, clinicians in urban areas seem to perform comparable to those in rural areas. Similarly, the mean O/E cost ratios for clinicians across the 4 census regions (excluding ‘Unknown’) are within a 0.04 or less range (0.98-1.02 at the TIN level and 0.94-0.97 at the TIN-NPI level), indicating minimal to no variation. Additionally, the mean O/E cost ratios for clinicians across 9 census divisions (excluding ‘Unknown’) are within a 0.12 range at the TIN level (0.96-1.08) and a 0.14 range at the TIN-NPI level (0.90-1.04), indicating moderate variation.

In terms of other clinician characteristics, analysis of clinicians by number of episodes indicates that clinicians with more episodes perform similarly to those who have fewer episodes. The exception is at the TIN-NPI level, where clinicians with either 200-299 episodes (mean O/E ratio: 1.04) or 300+ episodes (mean O/E cost ratio: 1.08) have a larger mean O/E cost ratio than the rest of the categories that have a range of 0.95-0.98. However, these large mean O/E cost

ratios are likely driven by relatively low clinician counts in each of those 2 categories (672 TIN-NPIs had 200-299 episodes; 251 TIN-NPIs had 300+ episodes). We also analyzed clinicians by risk score decile, as variation by risk score decile could indicate that the risk adjustment model is over- or under-correcting for clinicians with systematically riskier patients. Measure O/E cost ratios show moderate variation by risk score decile, with a range in mean TIN O/E cost ratio of 0.92 to 1.19 and a range in mean TIN-NPI O/E cost ratio of 0.90 to 1.11.

Tables 4-A and 4-B below present the distribution of cost measure O/E cost ratios by a range of clinician/clinician group characteristics, allowing a comparison of O/E cost ratio distributions for these breakdowns. The cost measure O/E cost ratios are presented at the TIN level and the TIN-NPI.

Table 4-A: Diabetes TIN Level Cost Measure O/E Ratios

Characteristic	# of TINs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
All TINs	38,996	1.00	0.46	0.68	0.98	1.35	1.85
Episode Sub-group							
Type 1 Diabetes	19,701	1.01	0.08	0.38	0.88	1.73	3.57
Type 2 Diabetes	38,996	1.00	0.45	0.67	0.97	1.36	1.87
Urban/Rural							
Urban	32,249	1.00	0.45	0.68	0.98	1.35	1.84
Rural	6,738	1.00	0.47	0.69	0.97	1.35	1.93
Unknown	9	0.79	0.39	0.39	0.84	1.18	1.18
Census Region							
Northeast	7,545	0.98	0.45	0.67	0.96	1.30	1.73
Midwest	6,448	1.02	0.48	0.71	1.00	1.36	1.85
South	16,906	1.02	0.48	0.69	0.99	1.36	1.89
West	7,872	0.99	0.43	0.65	0.95	1.36	1.86
Unknown	225	0.69	0.28	0.41	0.63	1.04	1.37
Census Division							
New England	1,438	0.96	0.40	0.65	0.92	1.30	1.79
Middle Atlantic	6,107	0.98	0.46	0.67	0.96	1.30	1.73
East North Central	5,069	1.02	0.48	0.70	1.00	1.35	1.82
West North Central	1,379	1.03	0.48	0.72	1.00	1.38	2.00
South Atlantic	8,930	0.99	0.47	0.68	0.96	1.33	1.83
East South Central	2,929	1.00	0.46	0.69	0.97	1.33	1.88
West South Central	5,047	1.08	0.51	0.73	1.05	1.45	1.96
Mountain	2,308	1.02	0.49	0.69	0.98	1.39	1.97
Pacific	5,564	0.97	0.41	0.63	0.94	1.35	1.82
Unknown	225	0.69	0.28	0.41	0.63	1.04	1.37
TIN risk score decile							
1st	3,899	0.92	0.38	0.58	0.88	1.29	1.82
2nd	3,900	0.92	0.41	0.61	0.89	1.25	1.76
3rd	3,900	0.94	0.42	0.64	0.91	1.27	1.80
4th	3,899	0.95	0.44	0.66	0.92	1.26	1.75
5th	3,900	0.97	0.48	0.67	0.94	1.28	1.74
6th	3,900	1.00	0.49	0.69	0.98	1.32	1.79
7th	3,899	1.02	0.49	0.72	1.01	1.33	1.83

Characteristic	# of TINs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
8th	3,900	1.04	0.52	0.73	1.02	1.36	1.85
9th	3,900	1.09	0.55	0.77	1.06	1.41	1.90
10th	3,899	1.19	0.64	0.87	1.16	1.56	2.03
Number of episodes							
10-19 Episodes	0	-	-	-	-	-	-
20-39 Episodes	14,087	1.01	0.40	0.61	0.96	1.45	2.03
40-59 Episodes	7,502	1.00	0.47	0.67	0.97	1.36	1.80
60-79 Episodes	4,308	1.00	0.51	0.70	0.97	1.33	1.73
80-99 Episodes	2,942	1.00	0.55	0.72	0.98	1.30	1.61
100-199 Episodes	5,558	1.00	0.56	0.74	0.99	1.27	1.57
200-299 Episodes	1,617	1.02	0.61	0.78	1.01	1.26	1.55
300+ Episodes	2,982	1.01	0.68	0.82	0.99	1.20	1.46

Table 4-B: Diabetes TIN-NPI Cost Measure O/E Ratios

Characteristic	# of TIN-NPIs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
All TIN-NPIs	81,786	0.96	0.41	0.61	0.92	1.34	1.84
Episode Sub-group							
Type 1 Diabetes	31,532	0.99	0.07	0.35	0.83	1.74	3.89
Type 2 Diabetes	81,786	0.96	0.41	0.61	0.92	1.35	1.86
Urban/Rural							
Urban	68,646	0.96	0.41	0.61	0.92	1.34	1.82
Rural	13,134	0.96	0.43	0.62	0.92	1.35	1.91
Unknown	6	0.91	0.39	0.39	0.95	1.37	1.37
Census Region							
Northeast	15,284	0.94	0.41	0.60	0.91	1.30	1.75
Midwest	17,869	0.96	0.41	0.61	0.92	1.35	1.83
South	35,216	0.97	0.44	0.63	0.94	1.35	1.88
West	13,186	0.94	0.38	0.58	0.91	1.34	1.82
Unknown	231	0.68	0.31	0.41	0.63	1.03	1.36
Census Division							
New England	4,070	0.90	0.38	0.56	0.86	1.28	1.73
Middle Atlantic	11,214	0.96	0.43	0.62	0.93	1.32	1.77
East North Central	13,336	0.96	0.41	0.61	0.93	1.34	1.80
West North Central	4,533	0.96	0.41	0.59	0.92	1.37	1.93
South Atlantic	19,236	0.94	0.43	0.62	0.91	1.30	1.79
East South Central	6,114	0.96	0.44	0.62	0.92	1.32	1.81
West South Central	9,866	1.04	0.47	0.67	1.01	1.45	2.02
Mountain	4,578	0.96	0.41	0.60	0.93	1.36	1.89
Pacific	8,608	0.93	0.37	0.57	0.89	1.32	1.77
Unknown	231	0.68	0.31	0.41	0.63	1.03	1.36

Characteristic	# of TIN-NPIs	Mean O/E Ratio	O/E Percentile				
			1st	10th	50th	90th	99th
TIN risk score decile							
1st	8,178	0.90	0.37	0.55	0.85	1.30	1.86
2nd	8,179	0.90	0.39	0.57	0.86	1.27	1.79
3rd	8,179	0.90	0.41	0.58	0.86	1.26	1.78
4th	8,178	0.91	0.40	0.58	0.87	1.27	1.78
5th	8,179	0.91	0.40	0.59	0.87	1.28	1.80
6th	8,179	0.93	0.41	0.60	0.89	1.31	1.83
7th	8,178	0.97	0.45	0.63	0.93	1.34	1.83
8th	8,179	1.01	0.46	0.66	0.99	1.37	1.84
9th	8,179	1.06	0.49	0.72	1.04	1.41	1.88
10th	8,178	1.11	0.55	0.78	1.08	1.46	1.93
Number of episodes							
10-19 Episodes	0	-	-	-	-	-	-
20-39 Episodes	41,816	0.95	0.38	0.57	0.91	1.39	1.96
40-59 Episodes	19,081	0.96	0.45	0.63	0.93	1.31	1.75
60-79 Episodes	9,364	0.96	0.49	0.66	0.93	1.28	1.66
80-99 Episodes	4,900	0.96	0.51	0.67	0.94	1.27	1.61
100-199 Episodes	5,702	0.98	0.54	0.70	0.97	1.27	1.55
200-299 Episodes	672	1.04	0.55	0.75	1.05	1.27	1.54
300+ Episodes	251	1.08	0.57	0.81	1.10	1.27	1.48

3.6.3 Interpretation

The results in Tables 4-A and 4-B above indicate that there is no notable variation in the mean cost measure O/E cost ratio across episode sub-groups, the urban/rural divide, census region, or episode volume at both the TIN and TIN-NPI levels. For each of these characteristics, the largest difference in the mean O/E cost ratio across categories was 0.03 or less. The only exception was episode volume with moderate variation in the mean O/E cost ratio of 0.13 among TIN-NPIs, which is driven by clinicians with 200-299 episodes and 300+ episodes. . Generally, this indicates that the risk adjustment model is functioning as intended for these characteristics; it is adjusting cost performance such that there are no substantive differences across the categories for these characteristics. For episode sub-groups, the model is run separately for each sub-group by Part D enrollment status to account for a more fair comparison across episodes in the Type 1 Diabetes and Type 2 Diabetes episode sub-groups. These results also support that there is meaningful variation in cost performance, even after risk adjustment, across these characteristics. Overall, these results indicate that there is large potential for saving Medicare spending and that there are no notable systemic differences across census region, episode sub-groups, and episode volume.

For TIN or TIN-NPI risk score decile, the difference in mean O/E cost ratio across categories was 0.27 at both the TIN level (range: 0.92 to 1.19) and the TIN-NPI level (range: 0.90 to 1.11). The lower values within the ranges of measure O/E cost ratios by risk score decile generally appear in the lower risk deciles at the TIN and TIN-NPI levels, and the higher values appear in the higher risk deciles at the TIN and TIN-NPI levels, specifically the 10th risk decile. This means that at both reporting levels, as the risk score decile increases, the mean O/E cost ratio also increases which may indicate a meaningful difference in provider performance for more difficult episodes. This variation indicates that the current risk adjustment model may not adequately

capture the impact of certain risk factors on clinician or clinician group performance, particularly among clinicians or clinician groups with especially low- and high-risk patient populations.

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

Since CMS uses Medicare claims data to calculate the Diabetes measure, Acumen expects a high degree of data completeness. To further ensure that we have complete and accurate data for each patient who opens an episode, Acumen excludes episodes where the patient does not appear in the EDB, the patient resided outside of the United States or its territories during the measurement period.

The Diabetes measure also excludes episodes where the patient is enrolled in Medicare Part C or has a primary payer other than Medicare in the 120-day lookback period and episode window. In such situations, Medicare Parts A and B claims data may not capture the complete clinical profile for the patient needed to capture the clinical risk of the patient in risk adjustment. Furthermore, Parts A and B claims data may not capture all Medicare resource use if some portion of the patient's care is covered under Medicare Part C.

3.7.2 Missing Data Analysis

The table below presents the frequency of missing data across the 4 categories of missing data which caused episodes to be excluded from the Diabetes measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as well as the number of TINs and TIN-NPIs who had at least one episode excluded due to missing data. The missing data categories are:

- Patient was not found in Medicare EDB
- Patient has a primary payer other than Medicare during the episode window or in the 120-day lookback period
- Patient was not enrolled in Medicare Parts A and B, or was enrolled in Part C, during the 120-day lookback period and episode window
- Patient resided outside of the United States or its territories during the episode window

Table 5: Missing Data Categories for the Diabetes Measure

Exclusion	# Episodes	# TINs	# TIN-NPIs
Not Found in Medicare EDB	*	*	*
Other Primary Payer	481,265	37,047	121,770
Not Continuously Enrolled	487,619	37,251	122,137
Resided Outside of U.S. or its Territories	4,530	2,692	3,582

* indicates that there were fewer than 11 episodes

3.7.3 Interpretation

As the Diabetes measure is calculated with Medicare claims data, Acumen expects a high degree of data completeness, which is supported by the limited frequency (relative to the overall scale of the measure) of missing data, as noted above. Acumen takes measures to ensure that missing or inaccurate information in claims data is not included in the cost measure.

4.0 Feasibility

4.1 Data Elements Generated as Byproduct of Care Processes

The data elements used in this measure are generated, collected and/or used by healthcare personnel during the provision of care (e.g., blood pressure, laboratory values, diagnosis, depression score). The data collected during care provision are then translated into the appropriate coding system (e.g. ICD-10 diagnoses, MS-DRGs) for use in Medicare claims.

4.2 Electronic Sources

All data elements are in defined fields in electronic claims.

4.3 Data Collection Strategy

4.3.1 Data Collection Strategy Difficulties

Lessons and associated modifications may be categorized into three types: data collection procedures, handling of missing data, and sampling data associated with beneficiaries who died during an episode of care.

4.3.1.1 Data Collection

Acumen receives claims data directly from the Common Working File (CWF) maintained at the CMS Baltimore Data Center. Medicare claims are submitted by healthcare providers to a Medicare Administrative Contractor (MAC), and are subsequently added to the CWF. However, these claims may be denied or disputed by the MAC, leading to changes to historical CWF data. In rare circumstances, finalizing claims may take many months, or even years. As a result, it is not practical to wait until all claims for a given month are finalized before calculating this measure. As such, there is a trade-off between efficiency (accessing the data in a timely manner) and accuracy (waiting until most claims are finalized) when determining the length of the time (i.e., the “claims run-out” period) after which to pull claims data. To determine the appropriate claims run-out period, Acumen has performed testing on the delay between claim service dates and claims data finalization. Based on this analysis, Acumen uses a run-out period of three months after the end of the calendar year to collect data for development and testing purposes. If this measure is used in a CMS program, calculation and reporting would be done in line with that program’s reporting practices.

4.3.1.2 Missing Data

This measure requires complete beneficiary information, and a small number of episodes with missing data are excluded to ensure completeness of data and accurate comparability across episodes. For example, episodes where the beneficiary was not enrolled in Medicare Parts A and B for the 120 days prior to the episode start date are not included in this measure. This enables the risk adjustment model to accurately adjust for the beneficiary’s comorbidities using data from the previous 120 days of Medicare claims. Additionally, the risk adjustment model includes a categorical variable for beneficiary age bracket, so episodes for which the beneficiary’s date of birth cannot be located are not included in this measure.

4.3.1.3 Sampling

During measure testing, Acumen noted that episodes in which the beneficiary died prior to the episode end date exhibited different cost distributions compared to other episodes. To avoid this effect’s potential impact on clinician scores, this measure does not include episodes for which the beneficiary’s date of death occurs prior to the end of the episode window.

5.0 Usability and Use

5.1 Use

5.1.1 Current and Planned Use

The measure was developed for potential use in MIPS, under a contract with CMS.

5.1.2 Feedback on the Measure and Development Process

5.1.2.1 Technical Assistance Provided During Development or Implementation

Development: Field Testing
Acumen and CMS conducted a national field test of 5 episode-based cost measures developed in 2019 and 2020, including the Diabetes measure, for a 5-week comment period (August 17 to September 18, 2020). We provided a Field Test Report to a sample of clinician groups and clinicians.[49] Field Test Reports were provided for each measure that a clinician or clinician group was attributed 10 or more acute inpatient medical condition and procedural episodes or 20 chronic condition episodes. This testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measures with as many stakeholders as possible. This sampling technique was used for field testing only and is not indicative of the case minimums used for any potential program implementation.

All stakeholders, including those who did not qualify to receive a Field Test Report, could review a series of mock reports that were representative of each measure and reporting type. Other public documentation posted during field testing included: measure specifications for each measure (comprising a Draft Cost Measure Methodology document and a Draft Measure Codes List file), a Measure Development Process document, a Frequently Asked Questions document, and a Fact Sheet.[50] During field testing, Acumen conducted education and outreach activities for stakeholders including multiple office hours sessions with specialty societies, a publicly posted field testing webinar recording, and Quality Payment Program Help Desk support.

5.1.2.2 Technical Assistance with Results

Field Testing
During the feedback period, 1,558 a Field Test Reports for episode-based cost measures were downloaded by 1,013 clinician groups (TINs) and 545 clinicians (TIN-NPIs). Stakeholder comments from field testing were summarized for the Clinician Expert Workgroup to consider in recommending refinements to the measures based on the testing data and feedback.

The following sections offer more details on the contents of each report and describe the education and outreach efforts associated with the field testing feedback period.

Data Provided During Field Testing
Each Field Test Report contained:

- Detailed performance results for the attributed measure, including metrics of cost measure score and a breakdown of episode cost compared to the national average and TIN/TIN-NPIs with a similar patient case mix (or risk profile)

[49] The field test reports were available for download from the Quality Payment Program website: https://qpp.cms.gov/login.
[50] The Measure Development Process, Frequently Asked Questions, and Fact Sheet documents are posted on the MACRA Feedback Page: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html.

- Drill-down detail for each measure, including more detailed information on potential cost drivers in the TIN/TIN-NPI's episodes. For example:
    - Analysis of utilization and cost for the measure by specific service categories (e.g., outpatient evaluation and management services, procedures, and therapy, hospital inpatient services, emergency room services, post-acute services)
    - Breakdown of costs for Physician/Supplier Part B and inpatient claims (e.g., top 5 most billed services and by risk bracket)
    - Accompanying episode-level Comma Separated Value (CSV) file with detailed information for all episodes attributed to the TIN/TIN-NPI. This file provides detailed information on every episode used to calculate your measure score, which includes winsorized observed cost, risk-adjusted cost, facilities and clinicians rendering care, the share of cost by service setting, and the patient relationship code (PRC) on the trigger/reaffirming claim line.

Mock Field Test Reports for each measure type that was field tested in 2020 were available for download by eligible clinicians and clinician groups from the CMS MACRA Feedback webpage.[51]

Education and Outreach

Acumen directly conducted outreach via email to tens of thousands of stakeholders using the stakeholder contact list developed through previous education and outreach and clinician engagement efforts, as well as CMS, Quality Payment Program listservs.

Acumen and CMS hosted two office hours sessions between July and August 2020, to provide an overview of field testing to specialty societies, discuss what information their members would be particularly interested in, and answer any questions. Across both office hours sessions, there were over 35 attendees from targeted specialty societies.

Acumen worked closely with Quality Payment Program Service Center to respond to stakeholder inquiries during field testing and continued to answer questions after the feedback period ended.

Acumen and CMS posted the MACRA Wave 3 Cost Measures Field Testing Webinar to the Quality Payment Program Webinar Library at the start of the field testing period.[52] The webinar recording, slides, and transcript were available for stakeholders to review throughout field testing. The webinar presentation outlined: (i) the cost measure field testing project (ii) the measure development and re-evaluation processes, and (iii) field testing activities. The webinar recording was viewed approximately 450 times during the field testing period.

5.1.2.3 Feedback on Measure Performance and Implementation

Field Testing

For the duration of field testing, stakeholders were invited to provide feedback by completing an online survey or submitting a comment letter. In total, Acumen received 24 survey responses and 13 comment letters, including from specialty societies representing large numbers of potentially attributed clinicians. An additional 22 comments from person and family representatives were received through the Cost Measures Questionnaire for Person and Family Input distributed by Acumen's project partner, PFCCpartners, to their Patient Family Advisory (PFA) network.

[51] CMS, "Mock Field Test Reports," MACRA Feedback Page, https://www.cms.gov/files/zip/macra-2020-cmft-mock-reports.zip.
[52] MACRA Wave 3 Cost Measures Field Testing Webinar materials are available on the Quality Payment Program Webinar Library: https://qpp.cms.gov/about/webinars.

Survey responses and comment letters were collected via an online survey, which contained general and detailed questions on the reports themselves, questions on the supplemental documentation, and questions on the measure specifications.

Pre-Rulemaking

CMS received 29 comments on the 5 episode-based cost measures included in the Measures Under Consideration List released in December 2020. This included 7 comments for the Diabetes measure. After the Measure Applications Partnership (MAP) Clinician Expert Workgroup meeting in January 2021, there was another public comment period on their preliminary recommendations, which received 25 comments across the 5 measures, with 6 comments specific to the Diabetes measure.[53] These public comment periods were facilitated by NQF. Stakeholders were able to submit their comments via the NQF website.

5.1.2.4 Feedback from Providers Being Measured

Field Testing

The Field Testing Feedback Summary Report presents stakeholder feedback gathered during the field testing period.[54] The following list synthesizes some of the key points that were raised across all field-tested measures through the field testing feedback period:

- Measure development approach
    - Stakeholders expressed appreciation for the opportunity to provide feedback during field testing and for the incorporation of previous suggestions in an effort to continually improve the measure development and field testing processes.
    - Stakeholders reported that the COVID-19 and wildfire public health emergencies presented challenges to participating in field testing. CMS's inclusion of telehealth services in the cost measures, partly in response to the COVID-19 pandemic, was seen as a positive step that should be continued going forward in an effort to expand access to vulnerable patient populations so long as CMS monitors for unintended consequences.
- Field Test Report access, format, and content
    - Stakeholders didn't report any issues accessing Field Test Reports during the field testing period. Feedback generally was positive regarding the Field Test Report that was updated for 2020 and the supplemental episode-level data file, though some stakeholders preferred the previous Excel format.
- Components of episode-based cost measures
    - Field testing feedback was generally not supportive of the inclusion of Part D drug costs in cost measures, with stakeholders expressing concern that clinicians could be held accountable for transactions that are out of their control or if patients require high-cost medications. Relatedly, stakeholders expressed concern about the lack of transparency for Part D costs.
    - Stakeholder input related to the development and testing of chronic condition measures was mixed. Some stakeholders reported that chronic condition cost measures represent an opportunity to reduce healthcare costs without impeding patient access, choice, or quality of care while others reported it was difficult to evaluate the new measures without measure reliability testing results.

[53] Measure Applications Partnership, *National Quality Forum*, https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx.
[54] CMS, "2020 Field Testing Feedback Summary Report," MACRA Feedback Page, https://www.cms.gov/files/document/macra-2020-ft-feedback-summary-report.pdf.

- Stakeholders maintain that resource use and patient health outcomes are influenced by the social determinants of health and that the cost measures aren't adequately adjusted for these differences when calculating cost measures performance scores.
- Stakeholders recognize the importance of linking cost and quality, including opportunities to do in the forthcoming MIPS Value Pathways (MVPs), to better evaluate clinician performance and improve patient health outcomes.

The summary report additionally contains measure-specific feedback, which was used as the basis for the post-field testing refinements though no refinements were made to the of the Diabetes measure specifications.

5.1.2.5 Feedback from Other Users

Pre-Rulemaking

In the 2020-2021 MAP review cycle, the MAP recommended "do not support with potential for mitigation" for the Diabetes measure. The MAP noted the following mitigation points: (i) explore the correlations between the cost measure and quality measures; (ii) NQF endorsement; (iii) explore the concern that good care may result in higher episode costs but with global cost savings; and (iv) evaluate the connection between upstream interventions and downstream cost savings. The MAP's final recommendations are available for review on their website.[55]

The CY 2021 PFS proposed rule includes a detailed discussion of each of the mitigation points raised by the MAP and the steps taken to address them. Additionally, empirical validity testing results for the Asthma/COPD measure are available on the MACRA Feedback Page.[56]

Person and Family Engagement

Acumen and CMS incorporated actionable input from patients and caregivers throughout the Diabetes measure development process. Throughout Wave 3 of measure development, we solicited and considered PFE input on (i) selection of episode groups for development, and (ii) a broad set of questions around constructing measures that will provide meaningful feedback on clinicians' resource use via service assignment, provider attribution, episode length, and more. We also sought comments through a questionnaire during field testing for person and family input. This input was shared with the Diabetes Clinician Expert Workgroup for their consideration as they developed the measure. A discussion of the PFE approach and specific feedback is available on the MACRA Feedback Page.[57]

5.1.2.6 Consideration of Feedback

Field Testing

Careful consideration was given to all feedback gathered during field testing, however no updates were required to the measure's specifications based on the recommendations of field testing commenters and the Clinician Expert Workgroup, comprised of subject matter and measure-development experts.

After completing field testing, Acumen compiled the feedback provided through the survey and comment letters into a measure-specific report, which was then provided to the Clinician Expert

[55] Measure Applications Partnership, *National Quality Forum,* http://www.qualityforum.org/Publications/2021/03/MAP_2020-2021_Considerations_for_Implementing_Measures_Final_Report_-_Clinicians,_Hospitals,_and_PAC-LTC.aspx.
[56] CMS, "Testing Updates for Wave 3 of Measure Development," MACRA Feedback Page, https://www.cms.gov/files/document/testing-updates-wave-3.pdf.
[57] CMS, Summary of Person and Family Engagement (PFE) and Input for Wave 3 Episode-based Cost Measure Development (March 2021). https://www.cms.gov/files/document/summary-person-and-family-engagement.pdf.

Workgroup, along with empirical analyses to inform their discussion and evaluation of any refinements needed to ensure that the measure is capturing what it was intended to capture.

5.2 Usability

5.2.1 Improvement

N/A. The measure has not yet been implemented, and as such has not had influence over performance.

5.2.2 Unexpected Findings

N/A. There were no unexpected findings during the development and testing of this measure.

5.2.3 Unexpected Benefits

N/A. There were no unexpected benefits during the development and testing of this measure.

Other Additional Information

Diabetes Clinician Expert Workgroup Members:
Akankasha Goyal, MD, Endocrine Society
Amandeep Sahota, MD, MSc, American College of Physicians, American Liver Foundation
Amisha Wallia, MD, MS, Endocrine Society, American Diabetes Association
Angela Thompson, DNP, FNP-C, BC-ADM, CDE, FAANP, American Association of Nurse Practitioners, American Diabetes Association
Caitlin Hicks, MD, MS, Society for Vascular Surgery
Devika Nair, MD, MSCI, American Society of Nephrology, American College of Physicians
Elisabeth Volpert, DNP, APRN, FNP-C, American Nurses Association, American Academy of Nurse Practitioners
Harlivleen Gill, MBA, RD, Academy of Nutrition and Dietetics
Ilona Lorincz, MD, MSHP, Endocrine Society
John Cook, MD, American College of Physicians
John Thompson, MD, American Society of Retina Specialists, American Academy of Ophthalmology
Kathleen Blake, MD, MPH, American Medical Association, American College of Cardiology
Lisa Ranes, RD, CDE, CSOWM, American Association of Diabetes Educators
Namirah Jamshed, MD, American Geriatrics Society
Phillip Ward, DPM, American Podiatric Medical Association, American Board of Foot and Ankle Surgery
Sharyl Magnuson Boyle, MD, American Academy of Family Physicians, AC Forum
Terry Lee Mills, MD, MMM, CPE, FAAFP, American Academy of Family Physicians
Varsha Vimalananda, MD, MPH, Endocrine Society

The Diabetes Clinician Expert Workgroup is composed from the larger Chronic Condition and Disease Management Clinical Subcommittee. The composition list of the Clinical Subcommittee is included in the Episode-Based Cost Measures Development Process document.[58]

[58] CMS, “2020 Episode-Based Cost Measure Field Testing Wave 3 Measure Development Process,” MACRA Feedback Page, https://www.cms.gov/files/document/macra-cmft-ebcm-process-2020.pdf.