CBE ID

4290

Title

Measuring the Value-Functions of Primary Care: Comprehensiveness of Care

Project

Primary Prevention

Endorsement Status

Not Endorsed

Is Under Review

No

Previous Endorsement Cycle

Spring 2025

Steward

American Board of Family Medicine

1.0 New or Maintenance

New

1.1 Measure Structure

Single Measure

1.3 Electronic Clinical Quality Measure (eCQM)

No

1.6 Measure Description

This measure evaluates the extent primary care physicians (PCPs) provide care-based and procedural-based services core to primary care. For each PCP, the resulting value reflects an average of the weighted proportion of services within each category provided during the measurement period.

Primary care providers (PCPs) caring for at least 30 patients per measurement period (the performance year and the 12 months prior to the performance year) score between 0 and 100 in scaled scores of comprehensiveness. Scores are based on weighted averages of 19 care-based and 20 procedural-based core primary care services.

1.7 Composite Measure

No

1.7 Measure Type

Process

1.8 Level of Analysis

Clinician: Individual

1.9 Care Setting

Ambulatory Care: Clinic, Ambulatory Care: Clinician Office, Ambulatory Care: Office, Clinician Office/Clinic, Hospital: Outpatient

1.10 Measure Rationale

The aim of this physician-level measure is to assess the extent to which a primary care physician provides services that are considered core to comprehensive primary care. Comprehensiveness of care – the "provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs" – is one of the key defining features of primary care.[1] Existing studies show that more comprehensive care by primary care physicians lowers patients' health care costs, prevents hospitalization and emergency department visits.[2,3] There is a strong consensus that comprehensiveness of care is one of the key ingredients in providing high-quality primary care to individuals, families, and communities. [1,4–6] Comprehensiveness is a multi-faceted concept that includes both the scope of services offered and the depth and breadth of health conditions managed by the PCP pending the needs of the population.[7] Although this physician-level measure assesses one aspect of comprehensiveness – provision of a range of services – there is empirical evidence of its positive association with patient outcomes.[2,3] Nonetheless, this measure is flexible enough to implement in most data environments (e.g. administrative claims, EHR or registry), is straight-forward to interpret, and preliminary analysis suggests positive association with patient outcomes.

The methodology for arriving at the list of core primary care services involved a high level of primary care insight from clinicians, patients, educators, and policy makers. The multi-stage process is outlined below:

1. The comprehensive list of primary care services and procedures was obtained from the published literature on scope of practice among family physicians and general practitioners (since they provide care to all age groups) and electronic health records (EHR) of a nationally-representative sample of primary care practices. For the literature, we closely followed Schultz and Glazier (2017) paper and a series of scope of practice papers by the research team at the American Board of Family Medicine (ABFM). [8–11] Based on the data available in the EHR dataset, included services were narrowed down based on relevance to an office-based or outpatient setting. Non-physicians were excluded from the measure solely due to limited data availability (there was unreliable specialty and credential information for these providers).

2. The services and procedures included are part of primary care medical education, fall under what primary care clinicians are board certified to perform, and data that the American Board of

Family Medicine (ABFM) collects about family medicine scope of practice practiced across the United States in primary care practices. The core primary care services were divided into two categories: care-based and procedure-based. The care-based category included a broad range of areas of care activities associated with certain modality, population, health condition, or healthcare setting, while the procedure-based category focused on specific procedures provided in primary care. Initially, there were 29 care-based services and 28 procedure-based services.

3. A technical expert panel (TEP) consisting of primary care physicians, researchers, health system executives, Federally Qualified Health Center (FQHC) representatives, family medicine advocacy groups and educators, and patient/caregivers (see TEP participants below) was held to finalize the list of core primary care services. Weights were assigned by the TEP in degrees of importance (0 = not important, and 100 = very important), then these weights were averaged. Along with creating the weighting, the TEP also reduced the list of services.

4. The weights within each category were then re-scaled so that their sum equaled the total number of services in that category (19 for care type, 20 for procedure). Categories were separated so that they could be reflective of relative importance to other services within that category (i.e. the weights in the care services reflect relative importance to other care services, while the weights in the procedure services reflect relative importance to other procedure services).

Our TEP participants included:

Tyler Barreto MD, MPH - Sea Mar CHC, Seattle, WA

Sanjay Basu MD, PhD - Center for Primary Care, Harvard Medical School

Andrew Bazemore MD, MPH – The American Board of Family Medicine Center for Professionalism and Value in Health Care

Reid Blackwelder MD, FAAFP - Quillen College of Medicine, ETSU

Hoon Byun DrPH – American Academy of Family Physicians- The Robert Graham Center

Yoonie Chung PhD – American Academy of Family Physicians-The Robert Graham Center

Julea Garner MD - Baptists Health-UAMS Family Medicine Residency

Jackson Griggs MD - Heart of Texas Community Health Center Inc.

Yalda Jabbarpour MD – American Academy of Family Physicians - The Robert Graham Center

Vivian Jiang MD - University of Colorado Medicine

Susan Lowe - Patient Representative

Amy Mullins MD - American Academy of Family Physicians

Ann O'Malley MD MPH- Mathematica

References:

1. Institute of Medicine (US), Division of Health Care Services, Committee on the Future of Primary Care. Primary Care: America's Health in a New Era. (Donaldson MS, Yordy KD, Lohr KN, Vanselow NA, eds.). National Academies Press (US); 1996.
2. Bazemore A, Petterson S, Peterson LE, Phillips RL. More Comprehensive Care Among Family Physicians is Associated with Lower Costs and Fewer Hospitalizations. Ann Fam Med. 2015;13(3):206-213. doi:10.1370/afm.1787
3. O'Malley AS, Rich EC, Shang L, et al. New approaches to measuring the comprehensiveness of primary care physicians. Health Serv Res. 2019;54(2):356-366. doi:10.1111/1475-6773.13101
4. World Health Organization. Primary Health Care. Geneva: World Health Organization. 1978.
5. Haggerty JL, Beaulieu MD, Pineault R, et al. Comprehensiveness of care from the patient perspective: comparison of primary healthcare evaluation instruments. Healthc Policy Polit Sante. 2011;7(Spec Issue):154-166.
6. Bitton A. The Necessary Return of Comprehensive Primary Health Care. Health Serv Res. 2018;53(4):2020-2026. doi:10.1111/1475-6773.12817
7. O'Malley AS, Rich EC. Measuring Comprehensiveness of Primary Care: Challenges and Opportunities. J Gen Intern Med. 2015;30 Suppl 3: S568-575. doi:10.1007/s11606-015-3300-z
8. Schultz SE, Glazier RH. Identification of physicians providing comprehensive primary care in Ontario: a retrospective analysis using linked administrative data. CMAJ Open. 2017;5(4):E856E863. doi:10.9778/cmajo.20170083
9. O'Neill T, Peabody MR, Blackburn BE, Peterson LE. Creating the Individual Scope of Practice (ISOP) scale. J Appl Meas. 2014;15(3):227-239.
10. Coutinho AJ, Cochrane A, Stelter K, Phillips RL, Peterson LE. Comparison of Intended Scope of Practice for Family Medicine Residents with Reported Scope of Practice Among Practicing Family Physicians. JAMA. 2015; 314(22):2364-2372. doi:10.1001/jama.2015.13734
11. Peterson LE, Fang B, Pu er JC, Bazemore AW. Wide Gap between Preparation and Scope of Practice of Early Career Family Physicians. J Am Board Fam Med. 2018;31(2):181-182. doi:10.3122/jabfm.2018.02.170359

1.11 Measure Webpage

https://professionalismandvalue.org/measures/the-comprehensiveness-of-care-meas...

1.13 Data Dictionary

Attached

1.13a Attach Data Dictionary

Code Tables_v4_02.25.25.xlsx

1.14 Numerator

The numerator is a weighted sum of core primary care services that the PCP performed at least once during the performance period.

1.14a Numerator Details

The numerator is a weighted sum of core primary care services in both categories (19 care-based and 20 procedure- based services) that the PCP provides. The numerator ranges from 0 to 380. See "Calculation of Measure Score" section for more details.

The detailed inclusion criteria for care-based and procedure-based services are provided in the data dictionary in Tables 2 and 3.

1.15 Denominator

The denominator will always be 380 (19 care-based services X 20 procedure-based services.) The denominator is algebraically derived to combine the two weighted core primary care services categories of the measure. This is a result of the formula itself and the algebra employed to calculate performance: when you add a weighted proportion of the 19 care-bases services to a weighted proportion of the 20 procedure-based services, the common denominator will be the product of 19 and 20, or 380.

1.15a Denominator Details

List of Core Primary Care Services:

Care-based services: 19 total services

1. Adult outpatient care: 1.13 weight
2. Behavioral health care: 1.11
3. Chronic disease management: 1.15
4. Chronic Pain Management: 0.99
5. Complementary and Alternative medicine: 0.66
6. End of life care: 1.08
7. Geriatric outpatient care: 1.13
8. Home visits: 0.80
9. Newborn Care: 0.86
10. Office Surgery/Minor Surgery: 0.94
11. Orthopedic Care/Musculoskeletal Care: 1.01
12. Pediatric outpatient care: 1.04
13. Prenatal Care: 0.88
14. Preventive care: 1.14
15. Smoking cessation: 1.03
16. Substance use disorder care: 0.97
17. Sports medicine Care: 0.95
18. Urgent care/Acute care:1.01
19. Gynecological and Reproductive Health:1.12

Procedure-based services: 20 total services

1. Allergy shots: 0.801 weight
2. Cardiac stress tests: 0.682
3. Colposcopy: 0.748
4. Endometrial biopsy: 0.868
5. Vision acuity test: 1.092
6. Immunization – Flu: 1.224
7. Immunization – Others: 1.25
8. Implantable long-acting reversible contraception insertion or removal: 1.118
9. Intrauterine device insertion or removal: 1.118
10. Joint and tendon aspiration or injection: 1.171
11. Neonatal circumcision: 0.669
12. Office lab procedures: 1.21
13. Office skin procedures: 1.184
14. Pap Smear/Cervical Cancer Screening: 1.263
15. Simple fracture care: 1.079
16. Ultrasound – Musculoskeletal: 0.947
17. Ultrasound - Other point-of-care: 0.96
18. Ultrasound – Prenatal: 0.868
19. Vasectomy: 0.617
20. Warfarin therapy management: 1.131

1.15b Denominator Exclusions

None

1.15c Denominator Exclusions Details

None

1.15d Age Group

Children (0-17 years), Adults (18-64 years), Older Adults (65 years and older)

1.16 Type of Score

Continuous variable, e.g., average

1.17 Measure Score Interpretation

Better performance = Higher score

1.18 Calculation of Measure Score

The measure calculation reflects an average of 2 weighted proportions. We present here how to calculate the measure using “numerator” and “denominator” components, and the algebra that was performed to derive those components.

Step 1: Identify primary care physicians (PCPs) who provided care to at least 30 patients during the performance period and the 12 months prior to the performance period.

Step 2: For each PCP, use the inclusion criteria to identify which care-based and procedure-based services were performed (Table 2 and 3) during the performance period and the 12 months prior to the performance period.

Step 3: Use the Numerator equation shown in the measure score calculation diagram attachment to calculate the numerator.

1. For the care services, sum up the weights of core primary care services performed by the PCP using Table 1. Then multiply that result by 9.5 (0.5 x 19). This is the care services portion of the numerator.
2. For the procedure services, repeat that process (a), except multiply the sum of the weights by 10 (0.5 x 20).
3. Sum those two components to obtain the complete numerator.

Step 4: Divide the numerator calculated in Step 3 by 380 (the denominator) and multiply by 100. This represents the average of the weighted proportion of core primary care services provided by the PCP, scaled to a range of 0 to 100.

1.18a Attach measure score calculation diagram

Measure Score Calculation Diagram.pdf

1.19 Measure Stratification Details

The measure is not stratified.

1.20 Types of Data Sources

Claims Data, Electronic Health Records, Registries

1.25 Data Source Details

The measure can be used in any data source that captures individual-level health care encounters across care settings (such as medical claims) and/or clinical data (such as from electronic health records or a registry) regarding demographics, services and procedures from primary care visits. ABFM tested the measure using two separate data sources: claims data and EHR data. Information on the two tested data sources (MarketScan for Claims data and American Family Cohort for EHR data) can be found below:

MarketScan (Claims):

The MarketScan Research database captures individual-level healthcare utilization, expenditures, and enrollment across the patient care setting spectra. Claims for inpatient, outpatient, prescriptions drugs in the outpatient settings, and carve-out services are included in these data.

The data come from a large selection of employers, health plans, and government/public organizations. The annual medical database includes private-sector health data from approximately 350 payers representing 20 billion service records. The data also represent the health care experiences of insured employees and their dependents of active employees, early retirees, and those receiving COBRA benefits (Consolidated Omnibus Budget Reconciliation Act). Included in these data are Medicare-eligible retirees that have employer-sponsored Medicare Supplemental plans.

The database has limited geographical location information: Specific zip code level information and address are not available for either the patient or provider in this dataset. These data have been reviewed by their internal and external statisticians and found to be incompliance with HIPPA Privacy Rules. The data contained for all enrolled beneficiaries are considered de-identified. The ABFM has a signed data use agreement with MarketScan which grants ABFM access to their claims extracts via the Stanford Department of Population Health Sciences (PHS), whose Data Core team reviews and approves access for the individual user. For those included and enrolled during the measurement period, all encounters with primary care physicians should be captured and included in these data.

American Family Cohort (EHR):

Derived from the ABFM PRIME Registry, the American Family Cohort is a clinical data repository of electric health record data from more than 2,000 primary care clinicians across the United States, with the electronic medical record data elements pulled from visits in the primary care setting. These data were established by the American Board of Family Medicine with the mission to measure clinical quality as well as develop specific clinical measure for quality reports and dashboards for primary care practices. The data contains a convenience sample of 1,000 practices representing over 5 million patients with the focus of care received by family physicians, general internists, general pediatricians, and advanced clinical practitioners, such as nurse practitioners and physician assistants. In the American Family Cohort, which is the research database representing extracts from the original PRIME Registry, there exist detailed demographics, diagnosis codes, procedures, some laboratory results, medication data including prescriptions, and free-text clinical notes.

1.26 Minimum Sample Size

Minimum sample of at least 30 patients per measurement period.

2.1 Attach Logic Model

Comprehensiveness Logic Model.pdf

2.2 Evidence of Measure Importance

In 2023, Baughman and colleagues published a review titled “Defining Comprehensiveness in

Primary Care: A Scoping Review.[1]" The review pulls from 25 individual articles that explore comprehensiveness and its components, including whole-person care (also referred to as "person-centered care"), range of services (also referred to "depth-and-breadth of care), and referral to specialty care (also referred to as "coordination of care" or "integration of care"). Whole-person care is identified as "the provision of individualized care to patients with respect to their physical, emotional, and social aspects of their lives," and should consider socio-cultural, spiritual, and societal/environmental factors. Range of services highlights the need for PCPs to offer "several categories of health care, including preventative, curative, rehabilitative, and palliative" that address both chronic and acute conditions. Referral to specialty care is thought of as an integrated component of comprehensiveness and requires that the PCP correctly recognizes "the balance between depth and breadth of care within the limitations of primary care scope." So, referrals are done when it is appropriate, and then coordinated and integrated "in a mutually supportive referral network" through utilizing EHRs, care management teams, and practice affiliations.

The 25 articles included in the review reflect a number of types of studies. While none are RCTs, there were several observational studies and/or interview-based studies that were included. Some surveyed hundreds of patients[2] or interviewed a combination of patients and providers[3], while others performed in-depth qualitative interviews, distributed surveys to patients, providers, and organizations, and analyzed claims data[4]. Sample sizes vary, as do the methods used to gather the information (surveys, interviews, validated tools, etc.), and not all the reviewed studies were US-based. In addition, there were multiple commentaries describing the need and/or framework of components of comprehensive care, with publication dates ranging from 2009 to 2020. These commentaries are consistent in their recommendations for the need and general definition of comprehensiveness.

There were also several review articles included in the scoping review. These reviews compiled information from other sources related to the definition and impact of comprehensiveness. For example, O'Malley et al.[5], discuss "whole-person care," which involves the knowledge of patients' medical history, preferences, and family and cultural orientation. The authors claim that elements of this type of care "…have been associated with improved patient self-management for chronic conditions, adherence to physicians' advice, and self-reported health status improvements. Greater whole-patient centered communication has also been associated with better patient recovery from discomfort, few patient concerns, and fewer diagnostic tests and referrals." The authors go on to state the following:

"Comprehensive primary care, which involves meeting the large majority of each patient's physical and mental health care needs, has been associated with better health outcomes provided at lower cost, lower hospitalization rates for ambulatory care-sensitive conditions, improved health and better self-reported health outcomes, and greater equity (i.e., reduced disparities in disease severity as a result of earlier detection and prevention across different populations)."

A review by Jimenez et al.[6] promotes a definition of comprehensiveness that is consistent with what O'Malley and others use, namely as "…the scope of services offered and its capacity to manage the most common health conditions, at any stage of a person's life," and go on to describe dimensions of comprehensiveness such as the scope or range of services and the depth/breadth of those offerings. They cite literature that links comprehensiveness and better patient outcomes: "…research indicates that more comprehensive PC is associated with greater efficiency, better health, and lower costs."

In a 2021 review, Jonas and Rosenbaum[7] examined 21 articles and identified evidence that "whole-person integrated care" (i.e., comprehensiveness) is associated with better health outcomes (managing chronic pain, improvements in physical and mental health, work productivity, well-being, improved medication adherence, HbA1c levels, etc.), higher patient satisfaction, lower costs, and improved clinician experience.[7]

Comprehensiveness is being more widely recognized as an important measure of quality care. In an editorial, Asif Bitton echoes others when describing the evidence linking higher comprehensiveness to clinical outcomes, but goes on to note the importance of more study of comprehensiveness as a measure.[8] Notably, recent work has demonstrated the proposed comprehensive measure to be valid and reliable;[9] together, these underscore the importance of implementing a comprehensive measure both to assess quality and to better understand how the aspects that are encompassed by comprehensiveness influence patient outcomes.

References:

1. Baughman D, Nasir R, Ngo L, Bazemore A. Defining comprehensiveness in primary care: a scoping review. J Prim Health Care. 2023 Sep;15(3):253-261. doi: 10.1071/HC23067. PMID: 37756243.
2. Haggerty JL, Beaulieu MD, Pineault R, Burge F, Lévesque JF, Santor DA, Bouharaoui F, Beaulieu C. Comprehensiveness of care from the patient perspective: comparison of primary healthcare evaluation instruments. Health Policy. 2011 Dec;7(Spec Issue):154-66. PMID: 23205042; PMCID: PMC3399439.
3. Tarrant C, Windridge K, Boulton M, Baker R, Freeman G. How important is personal care in general practice? BMJ. 2003 Jun 14;326(7402):1310. doi: 10.1136/bmj.326.7402.1310. PMID: 12805168; PMCID: PMC161634.
4. Rissi JJ, Gelmon S, Saulino E, Merrithew N, Baker R, Hatcher P. Building the foundation for health system transformation: Oregon's Patient-Centered Primary Care Home program. J Public Health Manag Pract. 2015 Jan-Feb;21(1):34-41. doi: 10.1097/PHH.0000000000000083. PMID: 25414954.
5. O'Malley AS, Rich EC, Maccarone A, DesRoches CM, Reid RJ. Disentangling the Linkage of Primary Care Features to Patient Outcomes: A Review of Current Literature, Data Sources,

and Measurement Needs. J Gen Intern Med. 2015 Aug;30 Suppl 3(Suppl 3):S576-85. doi: 10.1007/s11606-015-3311-9. PMID: 26105671; PMCID: PMC4512966.
6. Jimenez G, Matchar D, Koh GCH, Tyagi S, van der Kleij RMJJ, Chavannes NH, Car J. Revisiting the four core functions (4Cs) of primary care: operational definitions and complexities. Prim Health Care Res Dev. 2021 Nov 10;22:e68. doi: 10.1017/S1463423621000669. PMID: 34753531; PMCID: PMC8581591.
7. Jonas WB, Rosenbaum E. The Case for Whole-Person Integrative Care. Medicina (Kaunas). 2021 Jun 30;57(7):677. doi: 10.3390/medicina57070677. PMID: 34209250; PMCID: PMC8307064.
8. Bitton A. The Necessary Return of Comprehensive Primary Health Care. Health Serv Res. 2018 Aug;53(4):2020-2026. doi: 10.1111/1475-6773.12817. Epub 2017 Dec 29. PMID: 29285762; PMCID: PMC6051987.
9. Kamdar N, Garvert D, Yasui O, Winget M, Phillips R, Shuemaker J. Reliability and Validity of a Comprehensiveness of Care Measure in Primary Care, A Case Study of the PRIME Registry
10. The Annals of Family Medicine Nov 2024, 22 (Supplement 1) 6653; DOI: 10.1370/afm.22.s1.6653

2.3 Anticipated Impact

The implementation of this measure could result in benefits for patients, clinicians, physician practices, and payers. These benefits reflect clinical outcomes as well as financial and value-based outcomes (the business case).

In the review by O'Malley et al.[1], the authors state "Comprehensive care…has been associated with better health outcomes provided at lower cost, lower hospitalization rates for ambulatory care-sensitive conditions, improved health and better self-reported health outcomes and greater equity (i.e., reduced disparities in disease severity as a result of earlier detection and prevention across different populations)." This reflects improvements to patient health and well-being as well as lower costs to payers by reduced utilization. The improvement in health equity is also notable.

Jonas and Rosenbaum[2] cite several sources that show the impact of programs around increased comprehensiveness can reduce hospital admissions, hospital days, outpatient surgeries/procedures, and drug costs, as well as total medical costs. The authors also found a study where appropriate referrals to community-based public assistance programs (e.g., housing services, utility assistance) and social needs were met resulted in overall medical cost savings. In addition to patient benefits, this review also identified evidence that practices with more comprehensive care have lower levels of physician burnout and reduced employee turnover.

An analysis of Medicare claims found that greater comprehensiveness in primary care physicians was associated with fewer hospitalizations and lower costs among patients cared for by these physicians.[3]

The business case for this measure can be quantified through reduced care utilization and costs for patient care, as well as reduced clinician burnout and turnover. The cost of implementation is likely to be minimal, given that it does not typically require additional training, more technology or infrastructure, or additional staff process or resources.

References:

1. O'Malley AS, Rich EC, Maccarone A, DesRoches CM, Reid RJ. Disentangling the Linkage of Primary Care Features to Patient Outcomes: A Review of Current Literature, Data Sources, and Measurement Needs. J Gen Intern Med. 2015 Aug;30 Suppl 3(Suppl 3):S576-85. doi: 10.1007/s11606-015-3311-9. PMID: 26105671; PMCID: PMC4512966.
2. Jonas WB, Rosenbaum E. The Case for Whole-Person Integrative Care. Medicina (Kaunas). 2021 Jun 30;57(7):677. doi: 10.3390/medicina57070677. PMID: 34209250; PMCID: PMC8307064.
3. Henry, T.L., Petterson, S., Phillips, R.S. et al. Comparing Comprehensiveness in Primary Care Specialties and Their Effects on Healthcare Costs and Hospitalizations in Medicare Beneficiaries. J GEN INTERN MED 34, 2708–2710 (2019). https://doi.org/10.1007/s11606-019-05338-3

2.4 Performance Gap

MarketScan (Claims):

Year 2019 (2018-2019):

N = 120,594 Providers

Mean (SD): 0.36 (0.16)

Percentiles:

Min: 0.00

10th: 0.14

20th: 0.21

30th: 0.28

40th: 0.33

50th: 0.37

60th: 0.40

70th: 0.44

80th: 0.48

90th: 0.55

Max: 1.00

Year 2021 (2020-2021):

N = 115,844 Providers

Mean (SD): .33 (.16)

Percentiles:

Min: 0.00

10th: 0.11

20th: 0.17

30th: 0.25

40th: 0.31

50th: 0.34

60th: 0.38

70th: 0.41

80th: 0.45

90th: 0.51

Max: 1.00

The measured entity is the individual physician as indicated by their encrypted National Provider Identification (NPI) number. We considered the calendar years 2019 and 2021 as individual performance measurement years. We restricted to primary care physicians. We identified 120,594 individual PCP NPIs in 2019 and 115,844 in 2021 who had provided care to a minimum of 30 unique patients during the year. These samples of individual PCP NPIs for each calendar year represent a cross-sectional sample of all those providers that meet the 30 unique patients seen in that specific calendar year. For this measure, we also utilized the antecedent years (for 2019, we utilized 2018 claims; for 2021, we utilized 2020 claims).

We identified primary care physicians based on the following inclusion and exclusion criteria. All professional claims in MarketScan for years 2019 (or the performance year) with specialty types identified as Internal Medicine, Pediatric Medicine, Family Medicine, and Geriatric Medicine based on the documentation in the claims. There did not exist a broad provider and taxonomy specialty code master lookup, and these providers could not be linked to external data sources since provider National Provider numbers (NPIs) were encrypted and disallowed for external linkages. NPIs that had submitted a claim with those specialty types were considered primary care physicians (PCPs). To ensure sufficient sample size of patients for the performance year, all NPIs (providers) must have seen at least 30 or more unique patients in the outpatient setting. Any claims that took place in the inpatient, emergency department, and laboratory settings (e.g. in MarketScan, these are based on the STDPLAC code: 21, 23, 26, 31, 51, 52, 61, and 81) were excluded in determining the 30 patients criteria. While a PCP could see more patients across other patient care settings, the focus was on primary care visits. Among NPIs, we classified hospitalists who were primary care physicians that practiced in the inpatient setting by calculating the fraction of claims that took place in the inpatient setting based on place of service codes (STDPLAC: 21, 31, 51, 52, 61). Using a data-driven approach, those who met the threshold >= 0.9 were considered hospitalists. The same approach was applied in determining hospital-based primary care physicians based on STDPLAC code of 21, 22, 23, 31, 51, 52, 61 with the same threshold of claims >= 0.9.

American Family Cohort (EHR):

Year 2019 (2018-2019):

N = 1075 Providers

Mean (SD): 0.564 (0.189)

Percentiles:

Min: 0

10^{th}: 0.312

20^{th}: 0.446

30^{th}: 0.507

40^{th}: 0.560

50th: 0.604

60th: 0.639

70th: 0.671

80th: 0.708

90th: 0.764

Max: 0.936

Year 2021 (2020-2021):

N = 1075 Providers

Mean (SD): 0.554 (0.174)

Percentiles:

Min: 0.021

10th: 0.308

20th: 0.438

30th: 0.494

40th: 0.546

50th: 0.589

60th: 0.623

70th: 0.651

80th: 0.693

90th: 0.749

Max: 0.93

Recognizing potential temporal effects related to differences in primary care service utilization before and after COVID-19, we selected both pre-COVID-19 (2019) and post-COVID-19 (2022) years for data analysis. In order to compare performance, we ensured that the practices and the associated providers (physicians and advanced practice practitioners) were included in the analysis in both performance measurement years.

We restricted to providers that existed in only one practice in both performance year. We also restricted to this initial distribution of scores based on providers that have at least 30 patients seen in both performance years. Comprehensiveness of care in the 2-year period considers services that are performed in the year antecedent to it, but the patient denominator restrictions are based strictly on the calendar year of the measure. Therefore, if we consider the year 2019, for the 1-year measure, all care-based and procedure-based services for the measure are only those being applied to that provider in 2019. In contrast, for the 2-year measure, all procedure-based and care-based services are based on the combination of 2018 and 2019; however, the patient volume threshold is based on 2019 performance year. Since we wanted the comprehensiveness of care estimates to be sufficiently stable, we also restricted to those providers that had the following criteria:

1. Provider must exist in one practice in both 2019 and 2022
2. Provider must contribute at least 30 patients in both 2019 and 2022, respectively
3. Provider can have the same or different patients contributing to the score in both 2019 and 2022, respectively
4. Procedure-based and care-based services under the measure specification were calculated for each provider based on 2-year measures required 2018-2019 services for the 2019 performance year, and 2021-2022 for the 2022 performance year

2.5 Health Care Quality Landscape

Comprehensiveness of care is one of the key defining features of primary care.[1] And yet, there are no existing quality measures that attempt to measure comprehensiveness. Existing studies show that more comprehensive care by primary care physicians lowers patients' health care costs, prevents hospitalization and emergency department visits.[2,3] There is a strong consensus that comprehensiveness of care is one of the key ingredients in providing high-quality primary care to individuals, families, and communities. [1,4-6]

While other measures exist to evaluate the performance of specialists in the treatment of specific conditions, primary care often involves the management of multiple conditions. As such, comprehensiveness is a multi-faceted concept that includes both the scope of services offered and the depth and breadth of health conditions managed by the PCP pending the needs of the population.[7]

The current measure is the first to attempt to assess this complex concept in physicians, and while the measure assesses one aspect of comprehensiveness – provision of a range of services – there is empirical evidence of its positive association with patient outcomes.[2,3] Additionally, this measure is flexible enough to implement in most data environments (e.g. administrative claims, survey, or

EHR), is straight-forward to interpret, and preliminary analysis suggests positive association with patient outcomes.

Additionally, the methodology for arriving at the list of core primary care services involved a high level of primary care insight from clinicians. The multi-stage process is described in the rationale for the measure.

In conclusion, this clinician-level comprehensiveness of care quality measure will contribute to directly measuring the quality of primary care provided to the patient panel. Moreover, this measure will facilitate comparative physician performance across along a spectrum of comprehensiveness.

References:

1. Institute of Medicine (US), Division of Health Care Services, Committee on the Future of Primary Care. *Primary Care: America's Health in a New Era*. (Donaldson MS, Yordy KD, Lohr KN, Vanselow NA, eds.). National Academies Press (US); 1996.
2. Bazemore A, Petterson S, Peterson LE, Phillips RL. More Comprehensive Care Among Family Physicians is Associated with Lower Costs and Fewer Hospitalizations. *Ann Fam Med*. 2015;13(3):206-213. doi:10.1370/afm.1787
3. O'Malley AS, Rich EC, Shang L, et al. New approaches to measuring the comprehensiveness of primary care physicians. *Health Serv Res*. 2019;54(2):356-366. doi:10.1111/1475-6773.13101
4. World Health Organization. *Primary Health Care*. Geneva: World Health Organization. 1978.
5. Haggerty JL, Beaulieu MD, Pineault R, et al. Comprehensiveness of care from the patient perspective: comparison of primary healthcare evaluation instruments. *Healthc Policy Polit Sante*. 2011;7(Spec Issue):154-166.
6. Bitton A. The Necessary Return of Comprehensive Primary Health Care. *Health Serv Res*. 2018;53(4):2020-2026. doi:10.1111/1475-6773.12817
7. O'Malley AS, Rich EC. Measuring Comprehensiveness of Primary Care: Challenges and Opportunities. *J Gen Intern Med*. 2015;30 Suppl 3: S568-575. doi:10.1007/s11606-015-3300-z
8. Schultz SE, Glazier RH. Identification of physicians providing comprehensive primary care in Ontario: a retrospective analysis using linked administrative data. *CMAJ Open*. 2017;5(4):E856-E863. doi:10.9778/cmajo.20170083
9. O'Neill T, Peabody MR, Blackburn BE, Peterson LE. Creating the Individual Scope of Practice (I-SOP) scale. *J Appl Meas*. 2014;15(3):227-239.
10. Coutinho AJ, Cochrane A, Stelter K, Phillips RL, Peterson LE. Comparison of Intended Scope of Practice for Family Medicine Residents with Reported Scope of Practice Among Practicing Family Physicians. *JAMA*. 2015; 314(22):2364-2372. doi:10.1001/jama.2015.13734
11. Peterson LE, Fang B, Puffer JC, Bazemore AW. Wide Gap between Preparation and Scope of Practice of Early Career Family Physicians. *J Am Board Fam Med*. 2018;31(2):181-182.

doi:10.3122/jabfm.2018.02.170359

2.6 Meaningfulness to Target Population

We surveyed 289 patients from across the US on how valuable it would be for them to know if their primary care doctor provided Comprehensiveness of Care in their practice. 63% (n=182) stated that it was "**very valuable**" and 33% (n=95) stated that it was "**somewhat valuable**" to know their primary care doctor provided Comprehensiveness of Care in their practice (using a Likert scale: very valuable, somewhat valuable, not so valuable, not at all valuable).

We also asked how much it would influence their choice if they knew that one doctor provided more primary care services compared to another doctor. 57% (n=164) stated that it would influence their choice "**quite a lot"** and 36% (n=103) stated that it would influence their choice "**somewhat"** (using a Likert scale: quite a lot, somewhat, not very much, not at all).

Face validity of 50-60% is considered acceptable with 80% optimal (ref: Council for Medical Specialty Societies).

3.1 Contributions Towards Closing Care Gaps

There are a variety of ways that this measure may contribute toward advancing health equity. In general, there are racial disparities in preventive services. For example, Black, Hispanic, and Asian populations experience lower rates of cancer screenings than white patients.[1] As another example, among people with diabetes, those of racial and ethnic minority groups are more likely to experience foot ulcers and amputations.[2,3] These are just two examples, but are illustrative of disparities that may be closely linked to access to comprehensive primary care.

More comprehensive care can increase the use of preventive services and early detection of cancers and diabetic complications through improved patient-provider relationships and trust.

Our methodology involves exploring comprehensiveness by various characteristics, such as demographic distribution of patient population and geographic region. Such analyses would allow one to identify patient characteristics that may be associated with higher or lower physician comprehensiveness. Incentivizing improvement could reduce health disparities within conditions where screening and preventive care is typically provided by primary care physicians.

References

1. Tong M, Hill LO, Artiga S. Racial disparities in cancer outcomes, screening, and treatment. Kaiser Family Foundation. 03 Feb 2022. Available at: https://www.kff.org/racial-equity-and-health-policy/issue-brief/racial-…
2. Clayton EO, Njoku-Austin C, Scott DM, Cain JD, Hogan MV. Racial and Ethnic Disparities in the Management of Diabetic Feet. Curr Rev Musculoskelet Med. 2023 Nov;16(11):550-556. doi: 10.1007/s12178-023-09867-7. Epub 2023 Sep 21. Erratum in: Curr Rev Musculoskelet Med. 2024 Oct;17(10):436. doi: 10.1007/s12178-024-09882-2. PMID: 37733148; PMCID: PMC10587034.
3. Marvellous A. Akinlotan, Kristin Primm, Jane N. Bolin, Abdelle L. Ferdinand Cheres, JuSung Lee, Timothy Callaghan, Alva O. Ferdinand; Racial, Rural, and Regional Disparities in Diabetes-Related Lower-Extremity Amputation Rates, 2009–2017. Diabetes Care 1 September 2021; 44 (9): 2053–2060. https://doi.org/10.2337/dc20-3135

4.1a Data Structure and Availability

All required data elements are routinely generated and used during care delivery and are available in electronic sources. All data elements are in structured fields.

Claims data is subject to coding errors or inaccuracies, and it is also possible that a service can be provided but failed to billed for. However, because the services included in this measure are tied to reimbursement, we are confident that missing data and inaccuracies are minimal in these data.

Similarly, EHR data reflects real-time capture of services provided and is directly tied to coding for reimbursement. Therefore, it is unlikely that there are many occurrences of missing data or inaccuracies in these data.

Further, because the measure looks for any occurrence of a service provided during the evaluation period, a single instance of an omission or inaccurate capture of a service provided would only negatively impact the accuracy of the measure if it was the only occurrence of that service being provided during the entire evaluation period. That is, assuming that in most cases providers who offer a service do so multiple times during the evaluation period, it's only required that one of those instances be accurately captured for the measure to correctly reflect that provider's comprehensiveness regarding that service. Therefore, the measure is very robust to missing or inaccurate data.

4.1b Implementation Costs and Burden

Neither clinicians nor patients would incur any costs or burden related to the implementation of this measure.

4.1c Confidentiality

There is no threat to patient confidentiality.

4.3 Feasibility Informed Final Measure

As part of extensive efforts to identify core services for comprehensive care in primary care, a preliminary list of services was created and then reduced to only include services that could be found in structured fields in EHR data to ensure feasibility of data collection. The measure was shown to be feasible by mapping data elements to physician systems and confirming that the data elements were available and extractable, and results could be calculated.

4.4 Proprietary Information

Not a proprietary measure and no proprietary components

5.1.1 Data Used for Testing

ABFM tested the measure utilizing two data sources:

- 2018-2019 and 2020-2021 MarketScan was used to test using claims data
- 2018-2019 and 2021-2022 American Family Cohort was used to test using EHR data

More detail about the data source can be found in 1.25 Data Source Details.

We considered the following for reliability testing:

1. We tested at varying patient volume requirements in both 2019 and 2022 years, respectively. This included patient volumes of: 30, 40, 50, 75, 100, 150, 300, 500, 1000, and 1250. We concluded that 30 patients were the minimum requirement for reliability and that there is an inflection point where the reliability metric attenuates at 300+ which would be optimal for testing validity.
2. We restricted to providers that had only one practice represented in both performance years.
3. We estimated reliability using the measure performance year and the year antecedent to it.

For validity testing, we conducted the following analysis:

1. We restricted to providers that satisfied sufficient patient volume in both of the performance years with the specific chronic condition of interest(Type 2 Diabetics (N >= 20 in each performance year)).
2. We restricted to providers that had a sufficient level of documentation for the outcome measure that would be used for validity testing. All providers must have sufficient level of documentation of 80% or 90% with a hemoglobin A1C value in the appropriate units of the

diabetic patient panel for every provider for both performance years. The remainder of diabetics in the patient panel who did not have a documented A1C were imputed to the provider-year mean of the known diabetics with documented A1C. If more than one A1C was collected for a patient, they were assigned the most recent measure in the performance year per the initial measure specification.

5.1.1a Dates of Testing Data

Not a required field for Spring 2025.

5.1.2 Differences in Data

The data reflects both claims and EHR data; these two types of data are different in the detail and type of information they provide (claims data includes items necessary for billing, such as details about the encounter, while EHR data may include more granular information regarding diagnoses, lab values, etc.). However, both provide information on the number and date(s) of primary care visits and allow one to determine the purpose of the visit (e.g., for primary care). Therefore, they are both reasonable for this measure.

Claims data are pulled from 2018-2019 and 2020-2021 while EHR data are pulled from 2018-2019 and 2021-2022.

5.1.3 Characteristics of Measured Entities

MarketScan (Claims):

When using the MarketScan data, the measured entity is the individual physician as indicated by their encrypted National Provider Identification (NPI) number. We considered the calendar years 2019 and 2021 as individual performance measurement years. We restricted to primary care physicians. We identified 120,594 individual PCP NPIs in 2019 and 115,844, in 2021 who had provided care to a minimum of 30 unique patients during the year. These samples of individual PCP NPIs for each calendar year represent a cross-sectional sample of all those providers that meet the 30 unique patients seen in that specific calendar year. For the measure, we also utilized the antecedent years (for 2019, we utilized 2018 claims; for 2021, we utilized 2020 claims). We identified a subsect of 63,759 individual NPIs in 2019 and 2021 with a minimum of 30 patients seen in both 2019 and 2021. We also restricted to 14,840 physicians with at least 100 patients in both 2019 and 2021, and 5,344 with 200 or more patients in 2019 and 2021 as well to assess sample size for measurement. We conducted additional sensitivity analyses that varied the number of patients seen to identify potential inflection points on reliability.

American Family Cohort (EHR):

Within the American Family Cohort, we restricted to providers that existed in only one practice in both performance year. We also restricted to this initial distribution of scores based on providers that have at least 30 patients seen in both performance years. Comprehensiveness of care in the 2-year period considers services that are performed in the year antecedent to it, but the patient denominator restrictions are based strictly on the calendar year of the measure. Therefore, if we consider the year 2019, for the 1-year measure, all care-based and procedure-based services for the measure are only those being applied to that provider in 2019. In contrast, for the 2-year measure, all procedure-based and care-based services are based on the combination of 2018 and 2019; however, the patient volume threshold is based on 2019 performance year. Since we wanted the comprehensiveness of care estimates to be sufficiently stable, we also restricted to those providers that had the following criteria:

1. Provider must exist in one practice in both 2019 and 2022
2. Provider must contribute at least 30 patients in both 2019 and 2022, respectively
3. Provider can have the same or different patients contributing to the score in both 2019 and 2022, respectively
4. Procedure-based and care-based services under the measure specification were calculated for each provider based on services provided in 2018-2019 for the 2019 performance year, and in 2021-2022 for the 2022 performance year

5.1.4 Characteristics of Units of the Eligible Population

MarketScan (Claims):

In the MarketScan data, over 7 million individual patients were included in each performance measure year as being treated by at least one primary care clinician in the data analyzed. Age and sex distributions of these patients are provided below (race information is not available in the MarketScan data):

Year: 2019

Total Patients: 7,698,414

Sex, n (%)

Female: 4,240,840 (55.1)

Male: 3,457,574 (44.9)

Age, n (%)

0 – 17: 1,835,236 (23.8)

18 – 34: 1,763,370 (22.9)

35 – 54: 2,433,437 (31.6)

55 – 64: 1,362,016 (17.7)

65+: 304,355 (4.0)

Year: 2021

Total Patients: 7,390,825

Sex, n (%)

Female: 3,980,210 (53.9)

Male: 3,410,615 (46.1)

Age, n (%)

0 – 17: 1,576,408 (21.3)

18 – 34: 1,732,909 (23.4)

35 – 54: 2,261,798 (30.6)

55 – 64: 1,214,026 (16.4)

65+: 605,684 (8.2)

American Family Cohort (EHR):

Year: 2019

Total Patients: 3,155,272

Sex, n (%)

Female: 1,778,597 (56.4)

Male: 1,376,675 (43.6)

Age, n (%)

0-17: 445,433 (14.1)

8-34: 506,156 (16.0)

35-54: 819,326 (26.0)

55-64: 551,214 (17.5)

65+: 810,410 (25.7)

NA*: 22,733 (0.7)

Race, n (%)

American Indian Or Alaska Native: 27,241 (0.9)

Asian: 59,884 (1.9)

Black Or African American: 212,973 (6.7)

Hispanic Or Latino: 290,676 (9.2)

Multiple Races: 1,141 (0.0)

Native Hawaiian Or Other Pacific Islander: 6,095 (0.2)

Unknown, 526427 (16.7), 291216 (15.6)

White, 2008155 (63.6), 1205743 (64.6)

NA*: 22680 (0.7), 25446 (1.4)

Social Deprivation Index Quintile, n (%)

SDI 1: 686,038 (21.7)

SDI 2: 653,226 (20.7)

SDI 3: 626,795 (19.9)

SDI 4: 607,985 (19.3)

SDI 5: 558,548 (17.7)

NA*: 22680 (0.7), 25446 (1.4)

Year: 2022

Total Patients: 1,867,270

Sex, n (%)

Female: 1,044,086 (55.9)

Male: 823,184 (44.1)

Age, n (%)

0-17: 239,477 (12.8)

18-34: 263,647 (14.1)

35-54: 463,976 (24.8)

55-64: 324,516 (17.4)

65+: 550,199 (29.5)

NA*: 25,455 (1.4)

Race, n (%)

American Indian or Alaska Native: 10,606 (0.6)

Asian: 36,646 (2.0)

Black or African American: 128,614 (6.9)

Hispanic or Latino: 164,502 (8.8)

Multiple Races: 1,681 (0.1)

Native Hawaiian or Other Pacific Islander: 2,816 (0.2)

Unknown: 291,216 (15.6)

White: 1,205,743 (64.6)

NA*: 25,446 (1.4)

Social Deprivation Index Quintile, n (%)

SDI 1: 312,265 (16.7)

SDI 2: 344,577 (18.5)

SDI 3: 325,572 (17.4)

SDI 4: 439,752 (23.6)

SDI 5: 419,658 (22.5)

NA*: 25,446 (1.4)

*NA indicates the information was not available

5.2.1 Level(s) of Reliability Testing Conducted

Person or encounter level (i.e., data element) (e.g., inter–abstractor reliability), Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)

5.2.2 Method(s) of Reliability Testing

PERSON OR ENCOUNTER LEVEL:

Previous evidence has demonstrated the reliability and validity of claims data using ICD-10 and CPT codes of the critical data elements, including the numerator and denominator, using acceptable methodologies. [1, 2, 3]

Previous evidence has also demonstrated the reliability and validity of EHR data.[4, 5, 6] PRIME also conducts the following validation process for all the critical data elements for accuracy:

- Physicians agreed that the data provided includes the necessary administrative and patient-level data documentation to comply with quality measure data elements.
- During a series of meetings, account managers and technicians reviewed the data elements with clinicians and worked with them to map each individual data element.
- Clinicians reviewed Completed data element mappings in the context of the measure and refined the data mapping as needed to ensure accuracy.
- Physicians attest that the data is accurate and valid.

151 individual physicians and 95 groups attested that the critical data elements used for calculating this measure were accurate and valid.

References:

1. Stausberg, J., Lehmann, N., Kaczmarek, D., & Stein, M. (2008). Reliability of diagnoses coding with ICD-10. International Journal of Medical Informatics, 77(1), 50-57. doi:10.1016/j.ijmedinf.2006.11.005
2. *American Medical Association. The CPT Code Process. 2025. Accessed May 5, 2025 at: https://www.ama-assn.org/about/cpt-editorial-panel/cpt-code-process*
3. Ammann EM, Kalsekar I, Yoo A, Scamuffa R, Hsiao CW, Stokes AC, Morton JM, Johnston SS. Assessment of obesity prevalence and validity of obesity diagnoses coded in claims data for selected surgical populations: A retrospective, observational study. Medicine (Baltimore). 2019 Jul;98(29):e16438. doi: 10.1097/MD.0000000000016438. PMID: 31335698; PMCID: PMC6709187.
4. Ehrenstein V, Kharrazi H, Lehmann H, et al. Obtaining Data From Electronic Health Records. In: Gliklich RE, Leavy MB, Dreyer NA, editors. Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2 [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2019 Oct. Chapter 4. Available from: https://www.ncbi.nlm.nih.gov/books/NBK551878/
5. Hao S, Tao G, Pearson WS, Rochlin I, Phillips RL, Rehkopf DH, Kamdar N. Treatment of Chlamydia and Gonorrhea in Primary Care and Its Patient-Level Variation: An American Family Cohort Study. Ann Fam Med. 2025 Mar 24;23(2):136-144. doi: 10.1370/afm.240164. PMID: 40127987; PMCID: PMC11936364.
6. Velásquez EE, Kamdar NS, Rehkopf DH, Saydah S, Bull-Otterson L, Hao S, Vala A, Chu I, Bazemore AW, Phillips RL, Boehmer T. Post-COVID Conditions in US Primary Care: A PRIME Registry Comparison of Patients With COVID-19, Influenza-Like Illness, and Wellness Visits. Ann Fam Med. 2024 Jul-Aug;22(4):279-287. doi: 10.1370/afm.3131. PMID: 39038980; PMCID: PMC11268691.

ACCOUNTABLE ENTITY LEVEL:

MarketScan (Claims) and American Family Cohort (EHR):

ABFM assessed reliability using the variance partition coefficient (VPC), which uses a hierarchical regression model with comprehensiveness of care index as the dependent variable in accordance with Merlo, et al., and also performed for reliability adjustment approaches as done by Bamdad, Brown, Kamdar, et al.

Using the VPC, we fit a hierarchical linear regression model with random intercepts (variance components) for physician-level intercepts. The VPC ranges from 0 to 1 and is a proportion that measures the level-specific (e.g. physician-level) contribution towards the total outcome variance. For a null model that does not contain any fixed effects, we estimated the VPC, which is equivalent to the intraclass correlation coefficient (ICC).

Random Intercept Model: Using the random intercept model approach, we limited to providers that met the minimum patient eligibility criteria of >= 30 patients in both 2019 and 2022. Each provider had a calculated comprehensiveness of care score for 2019 and 2022, yielding 2 observations per provider (physician-year level). This structure allowed for both physician-years to be included in the reliability analysis to obtain an estimate of within-provider variance.

MarketScan (Claims):

For MarketScan, we also applied a beta-binomial model, in accordance with approaches by Adams, et al., in “The Reliability of Provider Profiling” (3). The beta-binomial model assumes a numerator representing the weighted sum of the services for each provider in the performance year. The denominator represents the total number of services available per the aforementioned comprehensiveness of care definition. Due to limitations in the domain of values that the beta-binomial model takes, the numerator assumes an integer value, but we surmise this does not contribute to substantial error in estimation of the comprehensiveness of care index for the purposes of reliability estimation.

As a variant of the reliability testing for the far more restrictive patient volume requirements with the random intercept model (e.g. requiring both performance years to satisfy the patient volume criteria), we used this approach for each performance year to estimate reliability. Therefore, we varied patient volume for each of 2019 and 2021 performance years, respectively.

References:

1. Bamdad MC, Brown CS, Kamdar N, Weng W, Englesbe MJ, Lussiez A. Patient, Surgeon, or Hospital: Explaining Variation in Outcomes after Colectomy. J Am Coll Surg. 2022 Mar 1;234(3):300-309. doi: 10.1097/XCS.0000000000000063. PMID: 35213493; PMCID: PMC10369366.

2. Austin PC, Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis. Stat Med. 2017 Sep 10;36(20):3257-3277. doi: 10.1002/sim.7336. Epub 2017 May 23. PMID: 28543517; PMCID: PMC5575471.

3. Adams, John L., The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND Corporation, 2009. https://www.rand.org/pubs/technical_reports/TR653.html.

5.2.3 Reliability Testing Results

For both data sources, reliability for this measure was tested at various minimum patient panel sizes. Sample size varies based on the minimum requirement. A summary of the reliability results is listed below. A full breakdown of the results at all tested sample sizes can be found in the attached Additional Reliability Testing Results.

MarketScan (Claims):

Random Intercept Model:

Intraclass coefficients (ICCs) ranged from 0.759 to 0.870 (across patient panel sizes). At the minimum sample size of 30+, the ICC was 0.759.

Beta-Binomial Model:

Total Patients in 2019: 30+; Number of providers: 120,594; Min: 0.830; 10^{th}: 0.832; 25^{th}: 0.837; 50^{th}: 0.850; 75^{th}: 0.877; 90^{th}: 0.920; Max: 1.000; Mean: 0.863; SD: 0.037

Total Patients in 2021: 30+; Number of providers: 115,844; Min: 0.836; 10^{th}: 0.838; 25^{th}: 0.845; 50^{th}: 0.860; 75^{th}: 0.890; 90^{th}: 0.935; Max: 1.000; Mean: 0.873; SD: 0.040

American Family Cohort (EHR):

For the random intercept model, intraclass coefficients (ICCs) ranged from 0.762 to 0.851 (across patient panel sizes). At the minimum sample size of 30+, the ICC was 0.762.

5.2.3a Attach Additional Reliability Testing Results

Additional Reliability Testing Results.pdf

5.2.4 Interpretation of Reliability Results

MarketScan (Claims):

At the minimum sample size of 30+, the ICC was 0.759 and the minimum reliability from the beta-binomial model was 0.863 for 2019 and .873 for 2021. These are above the generally accepted threshold of 0.70 for acceptable reliability.

American Family Cohort (EHR):

At the minimum sample size of 30+, the ICC was 0.762. This is also above 0.70 and suggests good reliability.

5.3.1 Level(s) of Validity Testing Conducted

Person or encounter level (i.e., data element) (e.g., sensitivity and specificity), Accountable entity level (i.e., measure score) (e.g., criterion validity)

5.3.2 Type of Accountable Entity Level Validity Testing Conducted

Empirical validity testing at the accountable entity-level (e.g., criterion validity, construct validity, known groups analysis), Systematic assessment of face validity of the measure's performance score as an indicator of quality or resource use

5.3.3 Method(s) of Validity Testing

PERSON OR ENCOUNTER LEVEL:

Previous evidence has demonstrated the reliability and validity of claims data using ICD-10 and CPT codes of the critical data elements, including the numerator and denominator, using acceptable methodologies. [1, 2, 3]

Previous evidence has also demonstrated the reliability and validity of EHR data.[4, 5, 6] PRIME also conducts the following validation process for all the critical data elements for accuracy:

- Physicians agreed that the data provided includes the necessary administrative and patient-level data documentation to comply with quality measure data elements.
- During a series of meetings, account managers and technicians reviewed the data elements with clinicians and worked with them to map each individual data element.
- Clinicians reviewed Completed data element mappings in the context of the measure and refined the data mapping as needed to ensure accuracy.
- Physicians attest that the data is accurate and valid.

151 individual physicians and 95 groups attested that the critical data elements used for calculating this measure were accurate and valid.

References:

1. Stausberg, J., Lehmann, N., Kaczmarek, D., & Stein, M. (2008). Reliability of diagnoses coding with ICD-10. International Journal of Medical Informatics, 77(1), 50-57. doi:10.1016/j.ijmedinf.2006.11.005
2. *American Medical Association. The CPT Code Process. 2025. Accessed May 5, 2025 at: https://www.ama-assn.org/about/cpt-editorial-panel/cpt-code-process*
3. Ammann EM, Kalsekar I, Yoo A, Scamuffa R, Hsiao CW, Stokes AC, Morton JM, Johnston SS. Assessment of obesity prevalence and validity of obesity diagnoses coded in claims data for selected surgical populations: A retrospective, observational study. Medicine (Baltimore). 2019 Jul;98(29):e16438. doi: 10.1097/MD.0000000000016438. PMID: 31335698; PMCID: PMC6709187.
4. Ehrenstein V, Kharrazi H, Lehmann H, et al. Obtaining Data From Electronic Health Records. In: Gliklich RE, Leavy MB, Dreyer NA, editors. Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2 [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2019 Oct. Chapter 4. Available from: https://www.ncbi.nlm.nih.gov/books/NBK551878/
5. Hao S, Tao G, Pearson WS, Rochlin I, Phillips RL, Rehkopf DH, Kamdar N. Treatment of Chlamydia and Gonorrhea in Primary Care and Its Patient-Level Variation: An American

Family Cohort Study. Ann Fam Med. 2025 Mar 24;23(2):136-144. doi: 10.1370/afm.240164. PMID: 40127987; PMCID: PMC11936364.
6. Velásquez EE, Kamdar NS, Rehkopf DH, Saydah S, Bull-Otterson L, Hao S, Vala A, Chu I, Bazemore AW, Phillips RL, Boehmer T. Post-COVID Conditions in US Primary Care: A PRIME Registry Comparison of Patients With COVID-19, Influenza-Like Illness, and Wellness Visits. Ann Fam Med. 2024 Jul-Aug;22(4):279-287. doi: 10.1370/afm.3131. PMID: 39038980; PMCID: PMC11268691.

ACCOUNTABLE ENTITY LEVEL:

FACE VALIDITY:

26 Primary Care Physicians in the PRIME registry which represents physicians across the United States were asked: “Based on the measure description and the specifications, would you agree with this statement: The scores obtained from the Comprehensiveness of Care measure can be used to distinguish good quality of care from poor quality. In other words, a score of 95% for clinician A versus a score of 80% for clinician B would suggest that clinician A provides higher Comprehensiveness of care."

EMPIRICAL VALIDITY TESTING AT THE ACCOUNTABLE ENTITY LEVEL:

MarketScan (Claims):

For the MarketScan data, we performed validity testing by associating the physician-specific comprehensiveness of care index with emergency department visits as a patient-level outcome. This was done using both descriptive analysis (first described below) and multivariable hierarchical logistic regression (subsequently described in this sub-section further below). Therefore, for each patient, we identified the provider-specific comprehensiveness of care index.

Some initial descriptive analyses included:

1. Examining the relationship between age group and at least one ED visit throughout the entire performance period (2018-2019). Performed the same relationship for 2020-2021.
2. Examining the relationship between age group and at least one ED visit throughout the entire performance period with a revised definition of 2019 only. Performed the same relationship for 2021 only.
3. Quantifying the relationship between deciles of comprehensiveness of care and patients having at least one ED visit during the performance period 2019 (and 2021).

To further test validity, we used hierarchical logistic regression with physician (encrypted NPI-level) random effects to estimate the unadjusted and adjusted odds of at least one ED visit during the performance period, using comprehensive of care as a main exposure variable. ED visits were chosen on the basis that higher levels of comprehensiveness of care as an attribute of the physicians when they see the patients would result in lower levels of ED utilization. We did three different models.

1. Model 1: Unadjusted bivariate logistic regression with dependent variable of ED visit and independent variable of comprehensiveness of care.
2. Model 2: Adjusted multivariable logistic regression with dependent variable of ED visit and independent variable of comprehensiveness of care, adjusted for age and sex.
3. Model 3: Adjusted multivariable logistic regression with dependent variable of ED visit and independent variable of comprehensiveness of care, adjusted for age, sex, and provider heterogeneity index.

Block inclusion of variables would help ascertain if comprehensiveness of care is still associated with ED visits after controlling for salient, observable patient-level or contextual-level (e.g. provider heterogeneity index) characteristics. Our hypothesis is that, even with the inclusion of adjustment variables, there will still be a significant association between comprehensiveness of care and ED visit.

American Family Cohort (EHR):

For validity testing, we primarily applied the principles of convergent validity. By formal definition, convergent validity refers to the degree in which multiple measures/indicators of a single underlying concept are interrelated. For instance, we aimed to calculate the magnitude of the measure score of interest (e.g. comprehensiveness of care) with another indicator of processes related to a target outcome or multiple target outcomes with similar processes (e.g. chronic disease management with diabetes hemoglobin A1C in the primary care setting). This would establish, for our purposes, a process-process measure correlation or association.

We conducted convergent validity using the following approach:

1. Descriptive analysis relating comprehensiveness of care quartile for the provider versus the proportion of diabetics with uncontrolled hemoglobin A1C that meet >=20 diabetics using either:
    1. Visit diagnosis code
    2. Visit diagnosis code and/or problem list diagnosis
2. Proposed negative binomial or Poisson regression to estimate unadjusted associations as well as adjusted for patient panel characteristics. Multivariable models will indicate if unadjusted associations show significance between comprehensiveness of care and poorly controlled diabetes A1C at the provider-level, or well controlled blood pressure for hypertensive patients. Since the provider must appear across multiple performance years, a measure is taken in each year; therefore, a repeated measures analysis has been considered with an associated covariance structure to determine the best fit model.

5.3.4 Validity Testing Results

FACE VALIDITY:

80.77% of providers surveyed agreed that scores obtained from the Comprehensiveness of Care measure can be used to distinguish good quality of care from poor quality

EMPIRICAL VALIDITY TESTING AT THE ACCOUNTABLE ENTITY LEVEL:

MarketScan (Claims):

2018-2019 Analysis:

We separated the comprehensiveness of care index by binning into 0.1 increments, and related these deciles to the ED visit rate for 2018-2019 (as the use case).

Comprehensiveness of Care	No ED Visits	At Least One ED Visit	Total
0.0 – <0.1	457,949 (54.1%)	388,862 (45.9%)	846,811 (5.2%)
0.1– <0.2	989,185 (59.1%)	684,982 (40.9%)	1,674,167 (10.2%)
0.2 – <0.3	990,064 (63.5%)	568,340 (36.5%)	1,558,404 (9.5%)
0.3 – <0.4	1,874,354 (69.9%)	808,443 (30.1%)	2,682,797 (16.4%)
0.4 – <0.5	2,399,277 (71.8%)	944,154 (28.2%)	3,343,431 (20.4%)
0.5 – <0.6	1,537,599 (70.4%)	645,275 (29.6%)	2,182,874 (13.3%)
0.6 – <0.7	938,920 (70.8%)	386,930 (29.2%)	1,325,850 (8.1%)
0.7 – <0.8	633,706 (70.5%)	265,322 (29.5%)	899,028 (5.5%)
0.8 – <0.9	694,818 (71.5%)	276,927 (28.5%)	

			971,745 (5.9%)
0.9 – 1.0	641,651 (72.0%)	248,950 (28.0%)	890,601 (5.4%)
Total	11,157,523 (68.1%)	5,218,185 (31.9%)	16,375,708 (100.0%)

Our findings from the above table show a definitive decrease in the ED visit rates as comprehensiveness of care increases, as expected.

To further substantiate that these findings are valid by varying the outcome, we also restricted the ED rate to only 2019 (rather than 2018-2019), which would represent only the performance year (2019 only). Similar results were obtained:

Comprehensiveness of Care	No ED Visits	At Least One ED Visit	Total
0.0 – <0.1	524,891 (62.0%)	321,920 (38.0%)	846,811 (5.2%)
0.1– <0.2	1,143,856 (68.3%)	530,311 (31.7%)	1,674,167 (10.2%)
0.2 – <0.3	1,137,094 (73.0%)	421,310 (27.0%)	1,558,404 (9.5%)
0.3 – <0.4	2,107,059 (78.5%)	575,738 (21.5%)	2,682,797 (16.4%)
0.4 – <0.5	2,684,372 (80.3%)	659,059 (19.7%)	3,343,431 (20.4%)
0.5 – <0.6	1,729,205 (79.2%)	453,669 (20.8%)	2,182,874 (13.3%)
0.6 – <0.7	1,053,643 (79.5%)	272,207 (20.5%)	1,325,850 (8.1%)
0.7 – <0.8	713,170 (79.3%)	185,858 (20.7%)	899,028 (5.5%)
0.8 – <0.9	778,951 (80.2%)	192,794 (19.8%)	971,745 (5.9%)

0.9 – 1.0	718,534 (80.7%)	172,067 (19.3%)	890,601 (5.4%)
Total	12,590,775 (76.9%)	3,784,933 (23.1%)	16,375,708 (100.0%)

2020-2021 Analysis:

Rate of Emergency Department Visits by Decile of Comprehensiveness of Care Score

Comprehensiveness of Care	No ED Visits	At Least One ED Visit	Total
0.0 – <0.1	650,343 (52.9%)	579,387 (47.1%)	1,229,730 (7.5%)
0.1– <0.2	1,059,304 (60.4%)	693,917 (39.6%)	1,753,221 (10.7%)
0.2 – <0.3	1,103,459 (67.3%)	536,115 (32.7%)	1,639,574 (10.0%)
0.3 – <0.4	2,191,523 (73.0%)	811,721 (27.0%)	3,003,244 (18.3%)
0.4 – <0.5	2,323,660 (73.4%)	841,523 (26.6%)	3,165,183 (19.2%)
0.5 – <0.6	1,415,832 (72.0%)	550,878 (28.0%)	1,966,710 (12.0%)
0.6 – <0.7	785,990 (73.4%)	284,396 (26.6%)	1,070,386 (6.5%)
0.7 – <0.8	668,817 (74.6%)	227,989 (25.4%)	896,806 (5.5%)
0.8 – <0.9	775,942 (74.3%)	267,915 (25.7%)	1,043,857 (6.3%)
0.9 – 1.0	502,589 (74.5%)	172,086 (25.5%)	674,675 (4.1%)
Total	11,477,459 (69.8%)	4,965,927 (30.2%)	16,443,386 (100.0%)

If we restrict to 2021 only, rather than the entire 2020-2021 performance period, the following descriptives hold true (in terms of the ED outcome rate being restricted to 2021 only):

Rate of Emergency Department Visits by Decile of Comprehensiveness of Care Score

Comprehensiveness of Care	No ED Visits	At Least One ED Visit	Total
0.0 – <0.1	742,634 (60.4%)	487,096 (39.6%)	1,229,730 (7.5%)
0.1– <0.2	1,208,748 (68.9%)	544,473 (31.1%)	1,753,221 (10.7%)
0.2 – <0.3	1,237,506 (75.5%)	402,068 (24.5%)	1,639,574 (10%)
0.3 – <0.4	2,409,132 (80.2%)	594,112 (19.8%)	3,003,244 (18.3%)
0.4 – <0.5	2,559,436 (80.9%)	605,747 (19.1%)	3,165,183 (19.2%)
0.5 – <0.6	1,570,344 (79.8%)	396,366 (20.2%)	1,966,710 (12%)
0.6 – <0.7	866,153 (80.9%)	204,233 (19.1%)	1,070,386 (6.5%)
0.7 – <0.8	734,259 (81.9%)	162,547 (18.1%)	896,806 (5.5%)
0.8 – <0.9	851,774 (81.6%)	192,083 (18.4%)	1,043,857 (6.3%)
0.9 – 1.0	555,039 (82.3%)	119,636 (17.7%)	674,675 (4.1%)
Total	12,735,025 (77.4%)	3,708,361 (22.6%)	16,443,386 (100.0%)

To further test validity, we used hierarchical logistic regression with physician (encrypted NPI-level) random effects to estimate the unadjusted and adjusted odds of at least one ED visit during the performance period, using comprehensive of care as a main exposure variable. ED visits were chosen on the basis that higher levels of comprehensiveness of care as an attribute of the physicians when they see the patients would result in lower levels of ED utilization. We did three different models:

The tables below assume comprehensiveness of care is a continuous variable.

Model 1: Predicting ED claims (yes/no) using comprehensiveness of care score as fixed effect and NPI as random effect.

Fixed Effect	*Parameter Estimate*	*Standard Error*	*Z-value*	*p-value*	*Odds Ratio*	*Odds Ratio Confidence Interval*
Intercept	-0.645	0.006	-117.0	<.0001	0.525	(0.519, 0.530)
CoC	1.629	0.014	-117.5	<.0001	0.196	(0.191, 0.202)

Model 2: Predicting ED claims (yes/no) using comprehensiveness of care score, age, and sex as fixed effects and NPI as random effect.

Fixed Effect	*Parameter Estimate*	*Standard Error*	*Z-value*	*p-value*	*Odds Ratio*	*Odds Ratio Confidence Interval*
Intercept	-0.618	0.0065	-112.10	<.0001	0.539	(0.533, 0.545)
CoC	-1.612	0.014	-117.02	<.0001	0.200	(0.194, 0.205)
Sex =Male	-0.084	0.001	-65.19	<.0001	0.919	(0.917, 0.922)
Age	0.001	<0.001	35.61	<.0001	1.001	(1.001, 1.001)

Model 3: Predicting ED claims (yes/no) using comprehensiveness of care score, age, sex, and provider heterogeneity index as fixed effects and NPI as random effect.

Fixed Effect	*Parameter*	*Standard*	*Z-value*	*p-value*	*Odds Ratio*

Odds Ratio

Estimate *Error*

Confidence Interval

Intercept	-1.735	0.006	-299.99	<.0001	0.176	(0.174, 0.178)
CoC	-1.079	0.014	-79.07	<.0001	0.340	(0.331, 0.349)
Sex = Male	0.033	0.001	25.16	<.0001	1.034	(1.031, 1.037)
Age	-0.003	<0.001	-83.55	<.0001	0.997	(0.997, 0.997)
PHI: 2*	0.777	0.002	380.68	<.0001	2.176	(2.167, 2.184)
PHI: 3*	1.845	0.002	800.17	<.0001	6.329	(6.300, 6.358)

*PHI: 2 is 2-4 providers. PHI: 3 is 5+ providers.

American Family Cohort (EHR):

We first must describe the outcome measures appropriately. Therefore, we identified, in our provider cohort, those providers that have a sufficient diabetic patient volume. We limited the analysis to clinicians with at least 20 diabetic patients in each performance year and only those where at least one of their patients had information regarding diabetic measures. This resulted in us including 778 total clinicians.

Table. Distribution of average percentage of diabetic patients with A1C > 9 or undocumented A1C by Comprehensiveness of Care decile grouping and year for providers included in the validation modeling.

Comp of Care	2019		2022	
percentile group	n	Mean (SD)	n	Mean (SD)
0 – 10%	34	0.49 (0.35)	27	0.60 (0.32)
11 – 20%	48	0.36 (0.29)	58	0.53 (0.34)

21 – 30%	49	0.30 (0.22)	50	0.33 (0.25)
31 – 40%	81	0.34 (0.23)	75	0.31 (0.24)
41 – 50%	81	0.33 (0.23)	86	0.32 (0.24)
51 – 60%	88	0.27 (0.22)	93	0.29 (0.22)
61 – 70%	89	0.28 (0.18)	91	0.31 (0.26)
71 – 80%	94	0.25 (0.17)	102	0.31 (0.22)
81 – 90%	109	0.30 (0.22)	97	0.31 (0.22)
91 – 100%	105	0.23 (0.11)	99	0.27 (0.19)

5.3.5 Interpretation of Validity Results

FACE VALIDITY:

The results indicate strong face validity. Face validity of 50-60% is considered acceptable with 80% optimal (ref: Council for Medical Specialty Societies).

EMPIRICAL VALIDITY TESTING AT THE ACCOUNTABLE ENTITY LEVEL:

MarketScan (Claims):

These findings above indicate that there is a strong average protective effect, with odds ratios ranging from unadjusted OR 0.196 (95% CI: 0.191, 0.202) to OR of 0.340 (95% CI: 0.331, 0.349). As predicted, the comprehensiveness of care index moved towards the null (reference of 1) as a result of adjustment for age, sex, and especially, provider heterogeneity index. For every unit increase in comprehensiveness of care, the patient-specific odds for at least one ED visit is 0.340, which is significantly protective of adverse health events as we would expect.

American Family Cohort (EHR):

Among clinicians performing in the lowest decile of Comprehensiveness of Care, an average of 49% of their diabetic patients had uncontrolled diabetes (A1C > 9) in 2019; in 2020 the average was 60% of diabetic patients. As Comprehensiveness of Care improved, the average percentage of diabetic patients with uncontrolled diabetes reduced.

These results imply that physicians that score lowest on Comprehensiveness have the highest percentage of diabetic patients with uncontrolled diabetes, and that higher performance on

Comprehensiveness is associated with fewer uncontrolled diabetic patients (or greater levels of disease management).

5.4.1 Methods Used to Address Risk Factors

No risk adjustment or stratification

6.1.1 Current Status

In use

6.1.2 Current or Planned Use(s)

Payment Program

6.1.3 Program Details

Name of the program and sponsor
QCDR Measure through ABFM PRIME- CMS QPP MIPS
URL of the program
https://qpp.cms.gov/mips/quality-requirements
Purpose of the program
A QCDR measure is a quality metric developed by a QCDR to fulfill specific quality reporting requirements within the MIPS program. QCDR measures are approved annually during the self-nomination process and are not part of the Rulemaking process.
Geographic area and percentage of accountable entities and patients included
United States; 390 active practices with 1627 active clinicians and over 8 million patients
Applicable level of analysis and care setting

Clinician and Clinician-Group; all applicable care settings

6.2.1 Actions of Measured Entities to Improve Performance

Activities measured entities could take to improve performance include:

1. Gather data on performance
2. Provide previously referred services
3. Refresh knowledge of previously referred services to feel comfortable providing the service themselves
4. Update procedures and policies, and allocate resources to provide more services and reduce referrals
5. Implement EHR-integrated tools for automated follow-ups to ensure appropriate screening and preventive health cadence for patients
6. Have interdisciplinary meetings where clinicians can share knowledge on services they are providing or interested in providing

The level of difficulty in achieving these actions will depend on which services the provider is trying to add and what resources are already available to the provider.

6.2.2 Feedback on Measure Performance

Measure was recently approved as an QCDR measure. No feedback has been provided.

6.2.3 Consideration of Measure Feedback

No feedback has been provided.

6.2.4 Progress on Improvement

Measure was recently approved as an QCDR measure. More time is needed to observe a change in performance.

6.2.5 Unexpected Findings

No unexpected findings.

6.2.5a Potential Unintended Consequences

No unintended consequences identified.

Developer POC email

HIngber@theabfm.org

Measure Developer POC

Hannah Ingber
American Board of Family Medicine
United States

Measured/accountable entity (reliability and/or validity) methodology and results (if available)

Measured entity (reliability and validity) methodology and results (if available)

The measure developer is different from the measure steward

No

Steward Address

Poonam Bal
Washington, DC
United States

Steward Organization

American Board of Family Medicine

Steward Organization URL

https://professionalismandvalue.org/measures-that-matter/

Steward POC email

pbal@theabfm.org