ENVIRONMENTAL SCAN

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Advancing Measurement of Diagnostic Excellence for Better Healthcare

ENVIRONMENTAL SCAN REPORT
SEPTEMBER, 2024

ABOUT THE NATIONAL QUALITY FORUM

The National Quality Forum (NQF) works with members of the healthcare community to drive measurable health improvements together. NQF is a not-for-profit, membership-based organization that gives all healthcare stakeholders a voice in advancing quality measures and improvement strategies that lead to better outcomes and greater value. Learn more at www.qualityforum.org.

ABOUT THE GORDON AND BETTY MOORE FOUNDATION

This report is funded by the Gordon and Betty Moore Foundation.

RECOMMENDED CITATION

The National Quality Forum. Advancing Measurement of Diagnostic Excellence for Better Healthcare: Environmental Scan; 2024.

DISCLAIMER

Impact Statement

Purpose
NQF conducted this environmental scan to describe the landscape of diagnostic excellence quality measurement for accountability purposes, and to identify challenges and solutions to diagnostic excellence measurement.

Key Findings
Our scan identified significant gaps in existing measures of diagnostic excellence and significant obstacles hindering the development of new ones, including development of measures using patient-reported information, which is regarded as the gold standard for understanding the impact of the diagnostic process on patients. Notably, little progress has been made toward developing measures of diagnostic equity. Measure developers reported that difficulty detecting variations in care, a lack of agreement on optimal diagnostic processes, inherent uncertainty in diagnosis for specific conditions, and the lack of symptom data standards limits their ability to assess quality of diagnostic care, and thus develop new measures of diagnostic excellence.

Applications
These scan findings are being used to guide NQF's Advancing Measurement of Diagnostic Excellence for Better Healthcare initiative in helping to address some of the most difficult challenges and move the field of diagnostic excellence quality measurement forward. The findings also update the field on gaps in diagnostic excellence measurement, identifying steps in the diagnostic process that are not tracked or assessed for quality through measurement.

Table of Contents

Executive Summary

Accurate and timely diagnoses are critical to delivering high-quality, safe, and effective healthcare. Despite the importance of diagnostic excellence, it remains difficult to measure. Measurement challenges limit opportunities to assess and improve diagnostic performance and to avoid the significant harms associated with diagnostic errors and delays. The National Quality Forum's (NQF) Advancing Measurement of Diagnostic Excellence for Better Healthcare initiative aims to identify critical barriers to measurement methods central to diagnostic excellence and drive consensus on solutions to overcome current barriers. This work is funded by the Gordon and Betty Moore Foundation (Moore Foundation) in its effort to improve diagnostic performance to reduce harm, improve health outcomes, and save lives.

As part of our work, NQF conducted an environmental scan to describe the landscape of diagnostic excellence quality measurement for accountability purposes, and to identify challenges and solutions to diagnostic excellence measurement. The scan was conducted from March 2023 through April 2023, and we did not apply filters for date of publication or location. We scanned publicly available sources for diagnostic excellence measures; reviewed the literature to identify relevant, existing publications about diagnostic excellence measurement; and reviewed reports from measure developers funded by the Moore Foundation.

Measure developers reported the inability to detect variations in care to be a significant obstacle in creating measures. For many conditions, uncertainty due to lack of agreement on optimal diagnostic processes and timeframes hinders the creation of standards, which are needed to support measurement. Some authors stressed the need for clinical guidelines to identify the standardized steps in the diagnostic process that would define more consistent "standards of care" against which quality can be assessed.

We identified several significant gaps in existing measures of diagnostic excellence. The National Academy of Medicine (NAM) defines the beginning of the diagnostic process as the time period when the patient experiences a health problem and engages with the healthcare system. However, there is a lack of measures that address this time period, which limits the ability to assess the quality of care in this area. The scan also revealed limited progress on the measurement of diagnostic equity. Most of the measures in development faced barriers to specification and testing that precluded integrating equity concerns into the measure's design. We found that the most significant challenge was insufficient data, which limits our understanding of quality for populations of interest (e.g., through measure stratification).

There are also significant obstacles to developing patient-reported outcome performance measures (PRO-PMs). The field recognizes patients as the gold standard source for determining diagnostic excellence, underscoring the importance of measures using patient-reported information. However, we identified only two new PRO-PMs from the review of Moore Foundation-funded measure developer grants. More work is needed in this area to close this gap in quality measurement.

We found that the most prevalent challenges experienced by measure developers were related to data standards and interoperability, inconsistent coding, or uncertainty around diagnoses for specific conditions. We noted a lack of specificity in data standards and clinical documentation requirements, significant differences in coding practices across different health systems, and limited evidence

regarding the relationship of signs and symptoms to a given condition, which hinders the ability to define evidenced-based measure specifications.

There is growing research into and use of artificial intelligence (AI) for quality measurement to address some of these challenges. AI methods can make more data usable for the field of healthcare quality measurement specifically. Several Moore Foundation-funded grantees reported using AI methods in measure development and testing. Developers used machine learning (ML) and natural language processing (NLP) to extract and make inferences from unstructured information, such as free-text notes in the medical record. These data could then be used for measure calculation without adding to the burden of this process. Measure specifications with AI that require a human reviewer augmented the measure calculation process and helped ensure transparency. This was important because developers stressed that being transparent with both internal and external stakeholders about the challenges of AI, and having a collaborative approach to addressing them, are essential.

Overall, we identified 127 measures relevant to diagnostic excellence. Four Moore Foundation-funded measures have gone through the consensus-based entity (CBE) endorsement process but are not yet in federal use. Two of these measures were endorsed. One measure was submitted to the Measures Under Consideration (MUC) List, an annual list of measures being considered for use in federal programs by the Centers for Medicare & Medicaid Services (CMS). The measure has received conditional support from the multistakeholder group reviewing measures on the MUC List.

These scan findings are being used to guide NQF's Advancing Measurement of Diagnostic Excellence for Better Healthcare initiative in helping to address some of the most difficult challenges and move the field of diagnostic excellence quality measurement forward. Future work as part of this initiative will include a report to advance measurement methodologies, to be issued spring 2025. That report will address two major challenges of the measurement of diagnostic excellence by documenting and describing the solutions to those challenges. We also will develop a call to action identifying specific actors to help actualize these solutions and continue to improve capabilities around measurement of diagnostic excellence, to be issued winter 2025.

Introduction

Accurate and timely diagnoses are critical to delivering high-quality, safe, and effective healthcare. Despite the importance of diagnostic excellence, it remains difficult to measure. Measurement challenges leave the field without meaningful opportunities to assess and improve performance and to avoid the significant harms associated with diagnostic errors and delays.[1] The National Quality Forum (NQF) is addressing these challenges through the **Advancing Measurement of Diagnostic Excellence for Better Healthcare** initiative. This multi-year effort is funded by the Gordon and Betty Moore Foundation (the Moore Foundation). NQF conducted this environmental scan to describe the landscape of diagnostic excellence quality measurement for accountability purposes and to identify challenges and solutions to the measurement of diagnostic excellence.

For this environmental scan, we inventoried existing measures of diagnostic excellence, conducted a literature review focused on measuring diagnostic excellence for accountability purposes, and reviewed reports of Moore Foundation-funded grantees who have developed or are developing measures of diagnostic excellence. This report provides background on diagnostic excellence measurement definitions and scope, describes our methods for the environmental scan, presents our results, highlights key findings, and discusses their implications for tackling the barriers to measuring and improving diagnosis.

This environmental scan is one of several initiative products. NQF will also produce a report to advance diagnostic excellence measurement, presenting the challenges and solutions to diagnostic excellence measurement identified through this work, and a call-to-action report, which will specify the stakeholders needed to implement the solutions, the timeline for implementation, and the steps for implementation of solutions. NQF convened the Diagnostic Excellence Committee to inform and provide feedback on project work and products (for the roster, see Appendix F). More details on the initiative can be found on the project webpage.

Background

Quality measurement is a potentially powerful way to facilitate diagnostic excellence.[2] When used with quality improvement methodologies, measurement can be effective in improving quality of care and preventing diagnostic errors. Experts recognize a dearth of measurement of diagnostic performance and many barriers that hinder the widespread adoption of quality measurement for diagnostic excellence.[2] In its landmark 2015 report, *Improving Diagnosis in Health Care*, the National Academy of Medicine (NAM, then called the Institute of Medicine) defined diagnostic error as, "the failure to (a) establish an accurate and timely assessment of the patient's health problem(s) or (b) communicate that explanation to the patient."[1] The 2015 NAM report emphasized that "errors in diagnosis are a major threat to achieving high-quality care" and acknowledged that "current pushes for accountability neglect diagnostic performance." However, they also concluded that "it would be premature either to adopt an accountability framework or to assume that the traditional accountability frameworks for public reporting and payment will be effective in reducing diagnostic error,"[1]

To advance measurement as a tool for improving diagnoses, the Moore Foundation started its Diagnostic Excellence Initiative, with its primary strategy being to "strengthen accountability for diagnostic excellence by helping to develop and validate new measures for diagnostic performance."[3] In

contrast to the prior focus on diagnostic errors, the Moore Foundation put forth a working definition of diagnostic excellence in a *JAMA Viewpoint*: *"*An optimal process to attain an accurate and precise explanation about a patient's condition. An optimal process would be timely, cost-effective, convenient, and understandable to the patient."[6] This definition incorporates the NAM conceptual framework's process orientation and further focuses on patient-centeredness, including the process of obtaining and clearly communicating accurate information about a patient's condition. The Moore Foundation's Diagnostic Excellence Initiative asserts that accountability measures and programs are essential for promoting transparency in diagnostic care quality, holding providers accountable for their performance, and publicly comparing care quality across providers.

Between 2019 and 2022, the Moore Foundation funded four cohorts of measure developers to develop and test measures of diagnostic excellence. Developers encountered several challenges during measure development and testing. The Moore Foundation selected NQF for its multistakeholder, consensus-driven measurement expertise to propel the healthcare quality community beyond currently intractable real-world challenges to high-impact measures of diagnostic quality. This project builds on NQF's prior work in this area on improving diagnostic quality and safety and reducing diagnostic errors, funded by the Centers for Medicare & Medicaid Services (CMS). In 2017, NQF convened a committee to develop a conceptual framework for measuring diagnostic quality and safety and to identify priorities for future measure development. In 2019, NQF convened a new multistakeholder committee to build on the previous committee's work by developing four use cases to support the practical application of the framework and identify comprehensive solutions to specific types of diagnostic errors. Now, a new committee has been formed for NQF's current initiative, which seeks to build upon its prior work in the area, address challenges experienced by measure developers, and advance solutions through multistakeholder engagement.

Early challenges that developers encountered in the development and implementation of diagnostic performance measures and their potential solutions were summarized in a report, *Diagnostic Excellence Initiative: Measure Implementation Challenge-to-Action Brief*, prepared by Battelle, a support contractor to the Moore Foundation. Barriers identified included lack of data specificity, data interoperability, and mechanisms for linking data across care settings. It also noted the challenges of using natural language processing (NLP) and of attributing outcomes to providers in the setting of shared patient management.[7] The findings highlighted the need to advance both data accessibility and measurement methods to better support quality measures of diagnostic error and safety. NQF's initiative with the Moore Foundation seeks to address these needs, identify gaps in current measures, and build from prior diagnostic excellence measurement efforts with multistakeholder input.

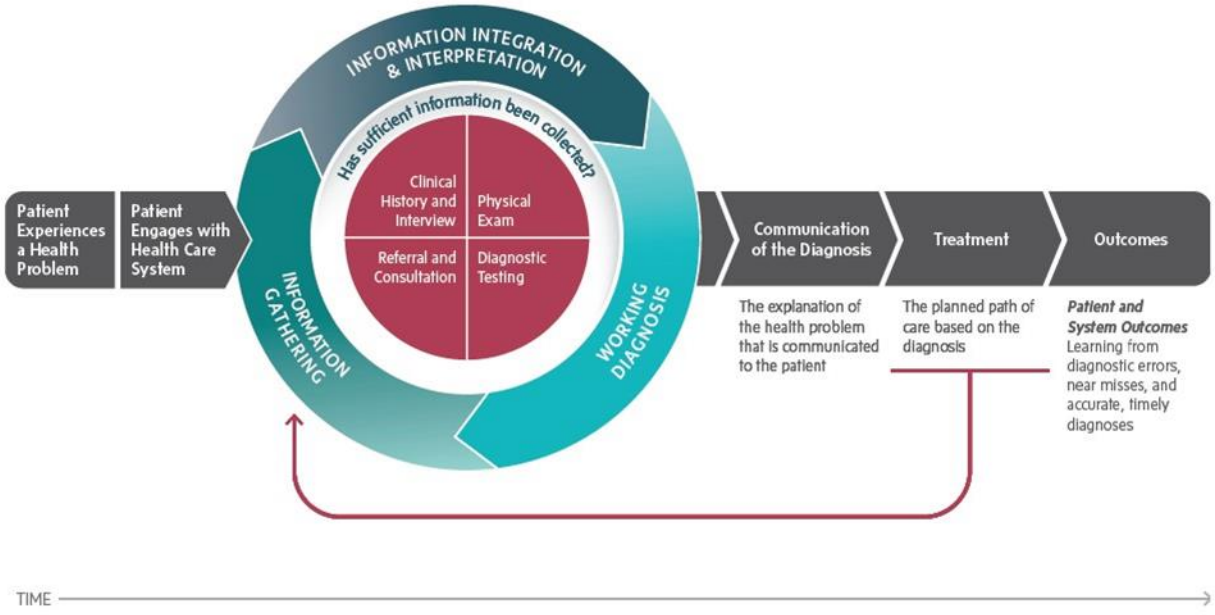
To inform the focus of this work and guide the activities of the Diagnostic Excellence Committee, we conducted this environmental scan.

The Scope of Diagnostic Excellence Accountability Measurement

To guide our approach to the review and inform our assessment of the challenges in measuring diagnostic excellence, including the identification of measure gaps, we considered the scope of the entire diagnostic process as defined in the 2015 NAM *Improving Diagnosis in Health Care* report (see Figure 1 below).[1]

NAM's conceptual model broadened the field's focus beyond diagnostic errors. By documenting the full diagnostic process, NAM created a framework applicable not only to errors but also to diagnostic excellence as a whole. As seen in the figure below, the diagnostic process begins with a patient experiencing a health problem and then engaging the healthcare system for care. The next portion of the process encompasses information gathering about the condition, developing and iterating on working diagnoses, and information gathering to develop a working diagnosis through physical exams, diagnostic testing, and/or specialty consults. The next step is communicating with the patient to explain the health problem in terms of a working diagnosis. Then the patient receives treatment, which may uncover other potential diagnostic possibilities (e.g., the diagnosis is incorrect or needs refinement). As a result of the treatment phase, the clinician may circle back to the information-gathering phase. Finally, the patient experiences outcomes, which are an essential part of the diagnostic journey and also provide data to inform and improve diagnosis in the context of a learning health system (i.e., a health system that systematically gathers internal data and integrates findings into its own practices to improve care quality and outcomes).[8]

Figure 1: The Diagnostic Process



As described in the Methods section below, we categorized measures identified during the environmental scan according to the steps in the NAM diagnostic process figure. NQF previously developed a conceptual framework in 2017 for measuring diagnostic quality and safety that was intended to "facilitate systematic identification and prioritization of measure gaps and to help guide efforts to fill those gaps through measure development and endorsement."[9] As the focus for the current initiative is to identify challenges in the measurement of diagnostic excellence across the entire diagnostic process, we chose to categorize measures using the steps in the diagnostic process identified by NAM. This approach allowed us to identify which parts of the diagnostic process are not currently assessed by measures for accountability purposes and to identify opportunities for such measurement in the future. As the results demonstrate below, it is much harder to develop measures for some steps in the process than others.

Methods

We conducted three analyses to understand the state of measurement of diagnostic excellence. First, we scanned for diagnostic excellence measures that were reviewed by CBEs (NQF and Battelle) or used in accountability programs. Second, we reviewed the literature to identify what has been published about the use of diagnostic excellence measurement for accountability purposes, the development of measures in this field, and the challenges and potential solutions related to measurement. Third, we reviewed grantee reports from Moore Foundation-funded measure developers to understand in greater detail the challenges they have encountered and the solutions they have implemented. We describe methods for each analysis in detail below.

MEASURES SCAN

To identify measures of diagnostic excellence that have already been developed and could be used for accountability purposes, NQF conducted a measures scan through April 2023 using five measure sources:

1. Centers for Medicare & Medicaid Services (CMS) Measures Inventory Tool (CMIT);
2. NQF Quality Positioning System (QPS), which contains measures submitted to NQF for endorsement;
3. The National Committee for Quality Assurance (NCQA) set of health plan measures, the Healthcare Effectiveness Data and Information Set (HEDIS);
4. The measures used by Qualified Clinical Data Registries (QCDRs) as part of the Merit-based Incentive Payment System (MIPS); and,
5. The Joint Commission (TJC) set of measures.

We chose these measure sources because two are inventories of measures that are often used or considered for national programs (CMIT and QPS), and the last three are national programs that apply to three separate, major sectors of the health care system (HEDIS – health plans, MIPS – clinicians, and TJC – hospitals).*

We developed a set of search terms for each data source, adapted slightly for each source's search protocol (see Appendix A, tables 1–3). Using the search terms, we pulled an initial list of measures potentially relevant to diagnostic excellence. Two reviewers then analyzed each measure's relevance to diagnostic excellence quality measurement. A third reviewer confirmed the findings, and reviewers discussed any areas of disagreement. We then removed duplicate measures to compose a final list.

To identify measure gaps (the types of measures still needed to fully assess diagnostic excellence for accountability purposes), we categorized each of the final measures into one of the NAM's diagnostic process steps. Then, because some of the NAM steps were associated with many measures (e.g., "diagnostic testing"), we grouped measures together based on the measure topic or approach. For

* In addition to the measure sources listed above, we considered including work from The Leapfrog Group, also funded by the Moore Foundation. Through this national initiative, the Leapfrog Group will "publicly report and recognize hospitals for preventing patient harm due to diagnostic errors." At the time we did our measures scan, however, The Leapfrog Group had released a pilot survey and recommended practices, but had not yet released measures or measure results, so we did not include this work.

example, two of the groupings for measures in the "diagnostic testing" process step were "further classification/severity of disease" and "overuse/appropriate use"). These steps and groupings can be found in Table 1. Because the diagram encompasses the entire diagnostic journey, measures categorized into it will help quantify those areas of the process with fewer or no measures.

LITERATURE REVIEW FOR BACKGROUND AND MEASURES

To further review progress in the field and gather innovative thinking on approaches to measurement, we conducted a literature review in PubMed for articles published through April 2023. We did not apply filters for date of publication or location. We focused on articles relevant to measures for accountability purposes (e.g., for accreditation, public reporting, or pay-for-performance programs) consistent with the goals of the Moore Foundation's Initiative. Measures used for accountability programs present specific challenges, as their scores must be valid and reliable across many varied healthcare system providers.

We used search terms related to diagnostic excellence, quality, error, and safety; misdiagnosis, missed diagnosis, and wrong diagnosis; and quality, accountability, and performance measures (see Appendix B for a full list of search terms). We did not limit the search by study or article type or the location of the research; we included opinion pieces as well as original research. We limited the search to articles written in English. Two reviewers assessed the resulting abstracts for their relevance to diagnostic excellence accountability measurement. Reviewers discussed any areas of disagreement. For articles deemed relevant, we pulled full-text articles. Two reviewers then assessed the full-text articles. A third reviewer confirmed the findings, and the reviewers discussed any discrepancies.

Two reviewers then abstracted data from the full-text articles deemed as relevant into a data chart and discussed any areas of disagreement (see Appendix C for the full list of relevant articles). We captured information on measure numerators and denominators; measure status at time of publication (i.e., in use, developed, or conceptual); challenges related to measure development, testing, and implementation; solutions to identified challenges; and whether the measure addresses care coordination or health equity, or focuses on capturing the patient voice. Additionally, we collected background on how diagnostic excellence is defined in the literature.

In a supplemental targeted review, we incorporated key articles related to diagnostic excellence quality measurement and the Moore Foundation's published work in this area, regardless of whether they emerged in our initial search. These included a series of *Viewpoint* articles from the *Journal of the American Medical Association (JAMA)*. This series was produced in partnership with the Moore Foundation and explored diagnostic excellence from multiple perspectives, including challenges and methods for improving diagnostic excellence. In addition, we identified other articles published by Moore Foundation-funded measure developers on work relevant to their diagnostic excellence measures. One reviewer assessed these additional articles for relevance based on their abstracts, and a second reviewer confirmed the relevancy of articles identified. Reviewers discussed any discrepancies.

MOORE FOUNDATION-FUNDED MEASURE DEVELOPER GRANTEE REPORT ABSTRACTION FOR CURRENT MEASURE DEVELOPMENT

We sought to capture the Moore Foundation-funded grantees' recent learnings by reviewing grantee reports prepared by April 2023. The Moore Foundation shared grantee reports with NQF for all four cohorts of measure developers (see Appendix E for a full list of grant titles). The grantee reports for the

first two cohorts of developers contained information on the measures developed, challenges encountered by the developers, and learnings from the measure development/testing process. The grantee reports for the third and fourth cohorts provided an initial description of the grantee's project and the measure(s) they intended to pursue and were more limited in scope. Two reviewers abstracted information from the grantee reports for the first two cohorts of developers, and reviewers discussed any discrepancies. One reviewer abstracted those from the third and fourth cohorts.

This work augments the Moore Foundation-funded work done by Battelle mentioned earlier. The *Challenge-to-Action Brief* was used as a starting point for this abstraction. Our scan adds to those findings with additional, independent sources of information and updates the findings with more current Moore Foundation-funded measure developer grantee information. Our abstraction focused on challenges to measure development, testing, and implementation; measure specifications; measure data sources; any discussion of addressing health equity through measurement; and whether the grantee engaged with the CBE endorsement process. We categorized the developers' challenges according to the themes Battelle identified: 1) alignment of measures to evidence; 2) data specificity; 3) sensitivity/specificity thresholds; 4) use of NLP and machine learning (ML) approaches; 5) defining target populations; and 6) system fragmentation). We then elaborated on the specific challenges in these areas based on the abstracted information. We also identified and described additional challenge themes.

Results

MEASURES SCAN

NQF identified a total of 520 measures across the five measure sources. As the same measure could appear in multiple sources (e.g., if the measure had been submitted to NQF for endorsement and used in a CMS accountability program, it would appear in QPS and CMIT), we deduplicated the measures list. Because the relevant measures we pulled from CMIT are actively used in CMS accountability programs, we deduplicated measures from the other sources against the list from CMIT. After reviewing measures and performing a deduplication, we identified 127 measures as relevant to diagnostic excellence for accountability purposes. We identified 86 unique measures of diagnostic excellence used in accountability programs (i.e., measures from CMIT, HEDIS, or TJC, or reported by QCDRs).

We then categorized the 127 measures into the NAM steps and grouped similar measure types together as described above. Table 1 describes how many measures we assigned to each NAM step in the diagnostic process and each grouping within the NAM step. A small number of measures corresponded to more than one NAM area. See Appendix D for a complete list of relevant measures categorized by the NAM step in the diagnostic process and groupings.

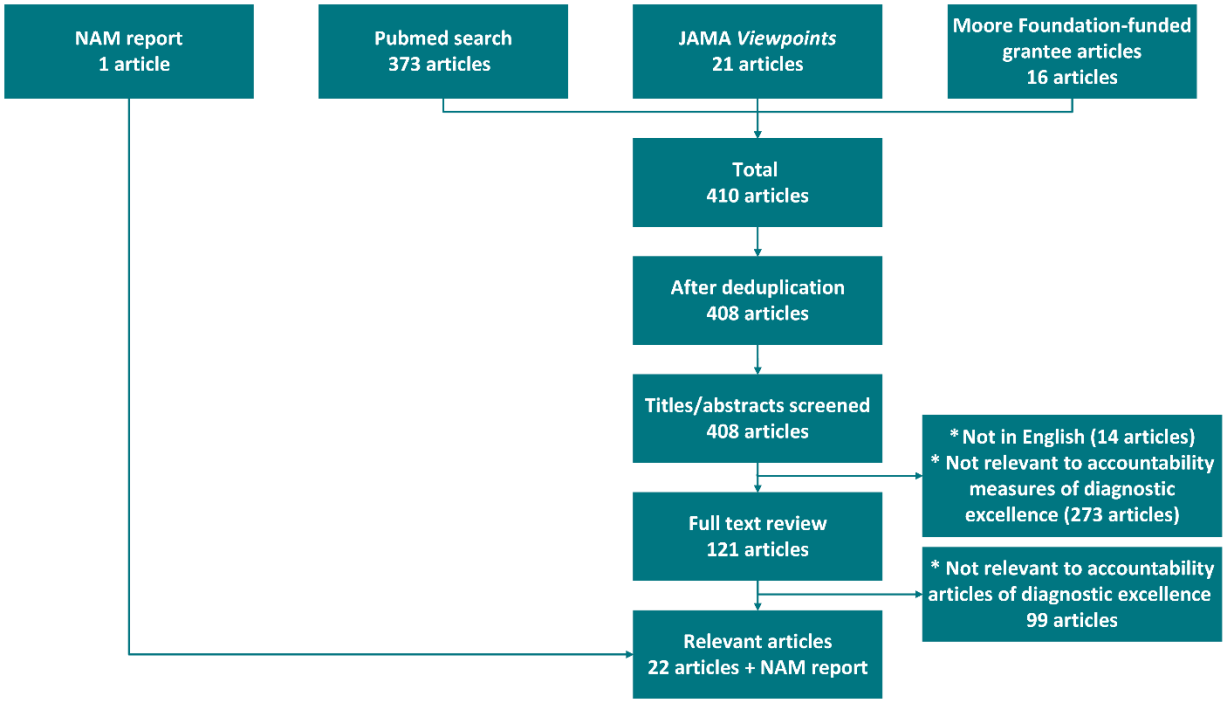
Table 1. Number of Measures by NAM Diagnostic Process Step and Grouping of Similar Measurement Approaches

NAM Step in the Diagnostic Process (Total Number)	Number of Measures by Grouping of Similar Measurement Approaches
Patient Experiences a Health Problem (0)	0 measures
Patient Engages with Health Care System (0)	0 measures
Clinical History, Interview, and Assessment (7)	Assessment in patients with a risk factor (5) Further classification/severity of disease (2)
Physical Exam (2)	In patients with a risk factor (2)
Referral, Consultation, and Follow-up (29)	Cross-provider communication (5) Closing the loop (2) Documentation in radiology/pathology reports (18) Follow-up with patient (1) Use of automated reminder systems (3)
Diagnostic Testing (72)	Complete or guideline-based testing (2) Confirmation of diagnosis (2) Further classification/severity of disease (7) Overuse/appropriate use (20) Testing in patients with a risk factor (15) Testing in patients with symptoms (2) Testing prior to or at baseline for treatment (11) Timeliness of testing (13)
Communication of the Diagnosis (6)	Documentation/records (2) Patient centeredness (3) Timely communication (1)
Treatment (9)	Positive screening with follow-up plan (9)
Outcomes (3)	Diagnosis was correct (2) Diagnosis was timely (1)

LITERATURE REVIEW

The PubMed search generated 373 articles. We also identified 21 additional articles from the *JAMA Viewpoint* Moore Foundation series on diagnostic excellence, as well as 16 articles published by measure developers related to their work funded by the Moore Foundation. Based on abstract/title reviews from these sources, we pulled 123 articles for full-text review. We reviewed the full-text articles, and our final list contained 22 articles deemed relevant to diagnostic excellence quality measurement for accountability purposes (see Appendix C for a complete list of relevant articles). We also included the NAM report as relevant literature as it first drew attention to the issue of diagnosis in medicine and guided our analysis as described above.

Figure 2 – Literature Review Flow Chart



Most articles were excluded because the measures described by the articles were not proposed for accountability purposes. For example, we excluded articles researching frequencies or incidence rates of diagnostic errors or misses as they were not researching these rates for accountability purposes, but rather to describe the extent of this issue. We also excluded many articles about measurement of diagnostic excellence if they were focused on measuring for research or quality improvement, and not for accountability. This is because measurement at a national level for accountability introduces more challenges related to reliability, validity, and comparability than measures used for quality improvement. Additionally, we excluded a number of articles because they contained ideas for measures in their conclusions but did not contain a full measure concept with sufficient detail to understand the planned target disease areas, units of measurement, or population. We excluded a small number of articles because they did not relate to medical disciplines, were focused on laboratory quality, or were focused on demonstrating the reliability or validity of a specific measure. Finally, many articles contained information not about diagnostic excellence, but rather treatment of a condition.

The relevant articles contained information on an additional 99 diagnostic excellence measures beyond the measures identified by the measures scan, with varying levels of detail on the measures.

MOORE FOUNDATION-FUNDED MEASURE DEVELOPER GRANTEE REPORT ABSTRACTION

Across the four cohorts of measure developers, NQF reviewed information on 73 different measures from 29 measure developers. Grantee reports had varying levels of detail for each measure. Sixteen reports had detailed information for abstraction.

Key Findings from the Measures Scan, Literature Review, and Grantee Abstraction

EVOLVING MEASUREMENT OF DIAGNOSTIC EXCELLENCE OVER TIME

In the measures scan, earlier measures are primarily process measures (e.g., Measure [M] 6, M10-14, M20-21, M41-46, M54-58). These earlier measures largely focused on the adequacy of screening or diagnostic testing, although there are some measures that capture whether results were communicated to a primary care provider.[M12-16] Over time, more intermediate clinical outcome[M61-63, M100-105, M107] and outcome measures (e.g., inappropriate diagnosis)[M125-127] were developed. Beyond measure type, in the more recent measures, there is a continued focus on screening and diagnostic testing, but more measures begin to evaluate the patient-centeredness of communications,[M114] overuse/appropriate use of testing,[M49-53, M59-63] assessment in patients with a risk factor,[M79-80] and documentation in radiology/pathology. More specifically, these newer measures assess not only whether radiology/pathology documentation was done but whether that documentation was done properly and with sufficient detail.[M22, M26, M31-34, M115]

The literature review and an examination of the measures scan show steady progress over time toward more complex quality measurement methods and improvement in measuring diagnostic outcomes. For example, earlier (i.e., older than 10 years) articles refer to a complete absence of measurement of diagnostic error. Further, early articles primarily focused on diagnostic error and did not define or aim to measure excellence, as opposed to error.[4,5] Additionally, measures in early articles are primarily related to screenings for disease.[11-13] More recent articles have integrated concepts of timeliness, accuracy, and adequacy of communication with patients into their concept of diagnostic excellence.[14-16] There are more measures meant to capture the adequacy of care coordination.[10,17] However, what has not changed is that the most recent articles still call for more substantial measurement and standards for diagnostic error and excellence to motivate further change.[18,19] Finally, two articles point out that there are still difficulties with measuring outcomes rather than processes due to an inability to measure differences attributed to quality rather than chance and because of resource constraints.[20,21]

GAPS IN DIAGNOSTIC EXCELLENCE MEASURES

Our measures scan identified at least one measure for seven of the nine NAM steps in the diagnostic process illustrated in Figure 1. Two areas had no identified measures: “patient experiences a health problem” and “patient engages with health care system.” Without measures for these two steps, there is a risk that diagnostic measurement and improvement efforts miss patients who have a complaint but do not or cannot continue to engage through the remainder of the diagnostic process. Some groups most at risk of diagnostic delays or misses have limited encounters with providers due to inadequate insurance coverage, transportation, availability of providers in their community, language barriers, or other risk factors. Therefore, they are often missing from diagnostic measures because their data are not part of those collected for measurement. Including people who have limited access to care in diagnostic excellence measures is challenging. For example, it requires obtaining and linking their data on signs and symptoms, but such data may not be accessible if they are not “in the system.” The lack of measures in these two areas of the NAM-defined process demonstrates a significant gap in measurement and represents an opportunity for measure development.

There were other steps in the diagnostic process for which we identified only a few measures: for "physical exam" we found two measures,M9, M10 for "outcomes" we found three measures,M125, M126, M127 and for "communication of the diagnosis" to the patient, we found six measures.M110, M111, M112, M113, M114, M115 As there is with other areas of measurement, we found a lack of outcome measures specific to diagnostic excellence. This will be an important area for future measure development. As noted in one of the *Viewpoint* articles, "diagnostic measurement for accountability would optimally focus on system-level performance for outcomes that matter to patients and clinicians."[2] We found no patient-reported outcome performance measures (PRO-PMs) during the measures scan and no measures specifically addressing diagnostic equity.

Major Gap: Measures of Diagnostic Equity

All three parts of our scan demonstrated limited progress on measures of equity and disparities in diagnostic excellence. As noted in one of the *Viewpoint* articles, measures would ideally allow us to quantify gaps in excellence, especially for historically disadvantaged populations. In addition, measures that rely on patient-reported data would be particularly valuable for this effort, as long as they are developed in concert with the measured populations.[22] Despite the importance of measuring diagnostic equity, in the measures scan we did not identify accountability measures that could be used for these purposes. There was one article in which the authors noted that they were able to find information on age and sex to build a sufficiently large sample to compare performance in a valid manner across these subpopulations.[23] However, other literature review articles cited limited opportunities to stratify measures. In one, the authors noted their stratification strategy would allow for the detection of potential selection biases in care decision making; however, difficulties with data availability made it hard to find sufficient numbers of encounters, limiting interpretation of stratified results.[24]

The Moore Foundation-funded measure developer grantee reports demonstrated limited progress on the measurement of diagnostic equity. However, most grantees were still testing and finalizing measure specifications at the time of this review and had not yet attempted stratification. Several grantees found that they had insufficient data to compare and assess any differences in measure calculation across groups (i.e., measure stratification). For example, one grantee was focused on developing measures of timely, closed-loop communication and follow-up testing on actionable incidental radiological findings (i.e., referral and testing for radiological findings not related to the initial reason for imaging). A lack of available patient risk-factor data, as well as a lack of detail on type of imaging study, prevented stratifying measure results by these categories for further analysis.

Major Gap: Patient-Reported Outcome Performance Measures

Our environmental scan found a lack of PRO-PMs in diagnostic excellence quality measurement. We identified no PRO-PMs in the measures scan or literature review. The literature highlighted this gap often by noting the importance of experience measures to diagnostic excellence in particular.[2,4,10,25,26] Several authors identified patients as the gold standard source for whether excellence happened.[2,4,25,26] [2,25] Articles from the literature review also suggested that measures capturing the patient voice and experience could be part of measuring equity. Only two of the Moore Foundation grants funded PRO-PM development. One was a patient-reported measure of diagnostic excellence for use by clinicians and administrators in the emergency department, and the second was a patient-reported outcome measure of the timeliness and quality of cancer diagnosis.

In all, while the measures scan, literature review, and abstraction of Moore Foundation-funded grantee reports found that some gaps remain, we also found that some of the Moore-funded grantee measures focused on testing in patients with signs or symptoms, missed diagnoses, and adequate communication, which would help to close some of the gaps identified by the measures scan.

USE OF ARTIFICIAL INTELLIGENCE METHODS IN QUALITY MEASUREMENT

The measures scan and literature review did not identify any measures utilizing artificial intelligence (AI) methods in the measure. However, several Moore Foundation-funded grantees were able to use AI methods in the development and testing of their measures. Developers used ML and NLP to extract and make inferences from unstructured data for measure calculation. The challenges that measure developers experienced regarding AI largely overlapped with challenges related to system fragmentation and data specificity. AI was used to assess symptoms and other unstructured information in measures of inaccurate, missed, and delayed diagnoses. The grantees found that the use of AI in measures required several rounds of both training the model and testing the model. It also required using data from diverse health systems in order to develop a model that was both specific enough to accurately capture unstructured data and could anticipate and correctly process the variety of language content it would encounter across diverse provider systems to minimize the need for customization in multiple health systems. Potential measure users interviewed by a Moore Foundation-funded grantee expressed some doubt about the reliability of measures using AI methods. However, some of these concerns were alleviated when the developer revised the measure specifications to require that a human reviewer augmented the process. Additionally, two grantees identified that a key solution was to be transparent about the challenges with AI, both internally and externally, and to be collaborative in addressing them.

IMPLEMENTATION OF DIAGNOSTIC EXCELLENCE MEASURES FOR ACCOUNTABILITY PURPOSES

From the measures scan, 86 measures are in use in accountability programs in the United States (U.S.), 54 of which are in use in federal programs. An additional 50 measures from the literature review are in use in U.S. and international programs; only one of these 50 is used in a U.S. accountability program.[14] Four Moore Foundation-funded measures have gone through the CBE endorsement process but are not yet in use in an accountability program. Two of these were endorsed: CBE #3690 Inappropriate diagnosis of urinary tract infection [M126] and CBE #3671 Inappropriate diagnosis of community-acquired pneumonia.[M125] Two were not endorsed: CBE #3716 CVD Risk Assessment Measure - Proportion of Pregnant/Postpartum Patients That Receive CVD Risk Assessment with a Standardized Tool and CBE #3735 CVD Risk Follow-up Measure - Proportion of patients with a positive CVD risk assessment who receive follow-up care. One of these measures was also submitted to the Measures Under Consideration (MUC) List, a list that CMS releases each year for measures it is considering adding to quality programs: Cardiovascular Disease (CVD) Risk Assessment Measure - Proportion of Pregnant/Postpartum Patients that Receive CVD Risk Assessment with a Standardized Instrument (MUC2022-048). It received conditional support from the multistakeholder group reviewing measures on the MUC List.

MEASURE DEVELOPMENT CHALLENGES EXPERIENCED

As described above, to further elaborate on the challenges documented in Battelle's *Measure Implementation Challenge-to-Action Brief*, we categorized the challenges experienced by the Moore Foundation grantees by the six categories described in the brief. Below, we assess the challenges

experienced by measure developers (as identified through the grantee report abstraction) grouped by the six Battelle categories and highlight how our research adds detail to the challenges described by Battelle. The most prevalent challenges cited by measure developers in their reports were related to data specificity, system fragmentation, or sensitivity/specificity thresholds (i.e., a measure's ability to correctly classify patients as having had a timely and correct diagnosis versus a delayed or incorrect diagnosis). We provide specific examples abstracted from the grantee reports below.

Sensitivity/Specificity Thresholds

Of the six challenges identified by the Battelle brief, the most common challenge experienced by the Moore Foundation-funded measure developers was the sensitivity/specificity thresholds challenge. Nine developers experienced this challenge across 25 measures. Although it was common, it was also the challenge that developers were best able to solve; for 18 out of 25 measures (72 percent), developers found a solution to this challenge. These solutions often involved adjusting definitions to capture more accurate populations or adding different diagnostic tests to the specifications to sharpen the measure's ability to accurately identify whether a patient truly had or did not have the diagnosis of interest.

Data Specificity

Data specificity challenges were the second most frequent, with 15 grantees experiencing them across 34 measures. Developers found solutions for twenty-one measures (61 percent). This challenge manifested primarily as a lack of specificity in coding languages, such as a lack of specific codes for fetal congenital heart defects, and was solved by introducing more detailed codes, for example, Logical Observation Identifiers Names and Codes (LOINC) for germline versus biomarker testing.

More precisely, we found a lack of specificity in data standards and clinical documentation requirements, as well as insufficient guidelines for documentation to ensure the data were captured with adequate specificity. Measure developers cited difficulties with capturing whether a test was "ordered" or "performed," obtaining accurate testing dates without manual chart review, and needing to reconcile date of diagnosis discrepancies between cancer registries and electronic health records (EHRs). Additionally, there were limited standards for where some medical information, such as a new cancer diagnosis, should be updated in the EHR during the diagnostic journey. For example, some providers put this information in the "problem list," while others put it in the "history."

System Fragmentation

Measure developers were least able to solve system fragmentation challenges such as limited interoperability or access to data across the care continuum. Eight grantees experienced this across 25 measures, and solutions were identified for only six measures (24 percent). An example of this challenge was demonstrated by a measure of interval colorectal cancers (i.e., cancers identified after a screening/surveillance exam in which no cancer was identified and before the date of the next recommended exam). The measure required data from sources outside the main health system setting when colonoscopies were done elsewhere. Reports scanned into the EHR were not consistently dated. Some reflected the date on which the report was scanned into the EHR, not the date of the colonoscopy. Additionally, across different health systems, developers found significant differences in coding practices. For a measure of diagnostic delay of lung cancer, there were variations in coding across sites for identifying diagnostic procedures and complications. The lack of standardized, interoperable data elements also limited developers' opportunities to build measures for accountability that are reliable across the healthcare system.

Alignment of Measures to Evidence

Nine grantees developing 25 measures experienced this challenge, and developers identified solutions for six measures (24 percent). Developers reported limited evidence regarding the relationship of signs and symptoms to a condition, limiting their abilities to define evidenced-based specifications. One grantee's measure sought to produce a single composite measure encompassing pneumonia, urinary tract infection, skin and soft tissue infections, and sepsis to assess diagnostic divergence in the emergency department between diagnosis recorded and diagnosis predicted. However, the developer's chart reviewers and technical expert panel found that the measure results were difficult to confirm and interpret because for some diseases, a diagnostic error involving the syndrome or pathogen makes no therapeutic difference, as long as the right antimicrobial treatment is given. Therefore, the technical expert panel agreed there was insufficient evidence for this type of composite measure, which would limit its acceptance and use. Instead, the panel suggested that measures focus on single diseases.

For another measure examining the timeliness of lung cancer diagnosis, there was limited evidence in the U.S. of the clinical features associated with early stages of lung cancer prior to the developer's work. The developer demonstrated that symptoms (e.g., unexplained weight loss or fatigue) and minor abnormalities in common blood tests (e.g., slightly raised platelets) could be used to prompt the patient and clinician to consider additional, more definitive diagnostic tests for lung cancer. However, while research indicates that individuals later diagnosed with cancer exhibit evidence of differences in clinical features/healthcare system contacts compared to those individuals who are not diagnosed with cancer, the developer was concerned about balancing the methods for identifying these cancer cases earlier without overwhelming systems of care, given that these symptoms and abnormal lab values are fairly common. This uncertainty also impaired developers' efforts to correctly differentiate poor from good quality diagnoses. The impact of limited evidence also affected another measure, which identified instances of diagnostic discordance between initial and discharge diagnosis for pneumonia after release from the emergency department. After chart review, the developer found a low prevalence of medical error among the discordances, but ample opportunities for learning. This showed that often the discordances did not necessarily indicate a lack of diagnostic excellence. Rather, they showed the large uncertainty surrounding pneumonia diagnosis in this setting.

Use of Natural Language Processing and Machine Learning Approaches

Six measure developers used AI in their measures, and all experienced challenges with this new methodology in all 12 of their measures. Developers found solutions for eight measures (67 percent). As mentioned in the section above, these challenges with AI largely overlapped with challenges related to system fragmentation and data specificity. Solutions primarily relied upon increasing both internal and external transparency about algorithms.

Defining Target Populations

Three Moore-funded grantee measure developers had challenges with defining target populations for six separate measures. Developers found solutions for four of these six measures (67 percent). They involved respecifying measures with assistance from clinical experts or conducting additional research to revise measure specifications.

ADDITIONAL MEASURE DEVELOPMENT CHALLENGES IDENTIFIED IN THE LITERATURE

The literature review also demonstrated additional difficulties with measuring diagnostic excellence for accountability purposes. First, there are difficulties in measuring cases in which the diagnosis is uncertain.[2,26] The concept of uncertainty is important given the length of the diagnostic process; it is expected that as testing is performed and symptoms are documented, more certainty is gained, but this takes time and depends upon many factors such as the condition, the patient's access to healthcare, and the patient's insurance status. Second, as with non-diagnostic excellence measures, quality measures can only identify differences in quality with sufficient statistical certainty when there is sufficient variation in care or outcomes to distinguish differences across providers. Measures of diagnostic excellence have also suffered from their reliance on burdensome medical record reviews and poor interrater reliability.[27] Two articles noted developers' difficulties in creating measures due to lack of substantial variation in care or an inability to detect variations in care with the measures as specified.[12,23] Finally, some authors noted that measures are most useful when provider performance can be judged against clinical guidelines, but this requires that guidelines identify the standardized steps in the diagnostic process that define a consistent "standard of care" against which quality can be assessed. There is still disagreement for many conditions about what is an optimal diagnostic process or timeframe. No standards can be set without agreement in these areas.[2,10,27] Hence, guidelines are an important enabler of measurement. This is especially important because measures should be "disease- and clinical context-specific" to be actionable.[27]

Summary and Conclusions

The diagnostic excellence quality measurement space has progressed since 2015, when the NAM 2015 report acknowledged a lack of accountability for diagnostic performance. Measures have expanded from earlier process measures focused specifically on diagnostic error to the broader measurement of diagnostic excellence, including measures that evaluate the patient-centeredness of communications, overuse/appropriate use of testing, intermediate outcomes, and outcomes along the diagnostic process. A small number of recently developed diagnostic excellence measures have even been endorsed and proposed for use in accountability programs. However, we found major gaps in measures that the field needs to address. There are still a limited number of outcome measures, including PRO-PMs, and measures that focus on equity. Stratifying measures to assess disparities remains a challenge. In addition, we found no measures that evaluate the beginning of the diagnostic process, including when a patient experiences a health problem and engages with the healthcare system for diagnostic care.

Measure developers funded by the Moore Foundation are starting to close these gaps; however, they still face significant barriers to building diagnostic excellence measures for accountability. Our environmental scan findings are consistent with those of Battelle's 2022 brief but provide more detail about the challenges it identified. We found substantial barriers to measure development with a lack of specificity in data standards and clinical documentation, significant differences in coding practices, and a lack of interoperable data elements upon which to build measures for accountability across the healthcare system. Our findings further highlight that uncertainty in the diagnostic process is a particular challenge to quality measurement.

To solve some challenges associated with developing measures, developers are implementing AI methods. While there are also challenges with using AI, this technology holds promise for filling gaps in

diagnostic excellence measurement, as it gives developers the ability to extract and make inferences from unstructured data for measure calculation. Developers funded by the Moore Foundation have used AI to start building outcome measures and other important measures of diagnostic excellence.

As part of NQF's Advancing Measurement of Diagnostic Excellence for Better Healthcare initiative, NQF and the Diagnostic Excellence Committee will prioritize several of the challenges highlighted by these findings and develop solutions for addressing the challenges. We will describe these solutions in a report to advance measurement methodologies. We will also develop a call to action that will advance our abilities to measure diagnostic excellence for accountability purposes. Future work will be available on our project webpage.

References

1. National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care.* The National Academies Press. https://doi.org/10.17226/21794.
2. Burstin H, Cosby K. Measuring Performance of the Diagnostic Process. *JAMA*. 2022;328(2):143-144. doi:10.1001/jama.2022.10166
3. The Gordon and Betty Moore Foundation. Diagnostic Excellence Initiative. www.moore.org/initiative-strategy-detail?initiativeId=diagnostic-excellence
4. Graber ML, Trowbridge R, Myers JS, Umscheid CA, Strull W, Kanter MH. The next organizational challenge: finding and addressing diagnostic error. *Jt Comm J Qual Patient Saf*. 2014;40(3):102-110. doi:10.1016/s1553-7250(14)40013-8
5. Graber ML, Wachter RM, Cassel CK. Bringing diagnosis into the quality and safety equations. *JAMA*. 2012;308(12):1211-1212. doi:10.1001/2012.jama.11913
6. Yang D, Fineberg HV, Cosby K. Diagnostic Excellence. *JAMA*. 2021;326(19):1905-1906.
7. Battelle. *Diagnostic Excellence Initiative: Measure Implementation Challenge-to-Action Brief*; 2022. https://www.battelle.org/insights/white-papers/diagnostic-excellence-initiative-measure-implementation-challenge-to-action-brief.
8. Agency for Healthcare Research and Quality. About Learning Health Systems. Published Content last reviewed May 2019. https://www.ahrq.gov/learning-health-systems/about.html
9. The National Quality Forum. *Improving Diagnostic Quality and Safety*; 2017. https://www.qualityforum.org/Publications/2017/09/Improving_Diagnostic_Quality_and_Safety_Final_Report.aspx.
10. Meyer AN, Scott TMT, Singh H. Adherence to National Guidelines for Timeliness of Test Results Communication to Patients in the Veterans Affairs Health Care System. *JAMA Netw Open*. 2022;5(4):e228568. doi:10.1001/jamanetworkopen.2022.8568
11. Chiarelli AM, Halapy E, Nadalin V, Shumak R, O'Malley F, Mai V. Performance measures from 10 years of breast screening in the Ontario Breast Screening Program, 1990/91 to 2000. *Eur J Cancer Prev*. 2006;15(1):34-42. doi:10.1097/01.cej.0000195713.02567.36
12. Saini SD, Schoenfeld P, Vijan S. Can the adenoma detection rate reliably identify low-performing endoscopists? Results of a modeling study. *Dig Dis Sci*. 2013;58(7):1856-1862. doi:10.1007/s10620-013-2592-2
13. Keck PE, Kessler RC, Ross R. Clinical and economic effects of unrecognized or inadequately treated bipolar disorder. *J Psychiatr Pract*. 2008;14 Suppl 2:31-38. doi:10.1097/01.pra.0000320124.91799.2a
14. Gurgel RK, Briggs SE, Dhepyasuwan N, Rosenfeld RM. Quality Improvement in Otolaryngology-Head and Neck Surgery: Age-Related Hearing Loss Measures. *Otolaryngol Head Neck Surg*. 2021;165(6):765-774. doi:10.1177/01945998211000442
15. Hameed AB, Tarsa M, Graves CR, et al. Cardiovascular Risk Assessment as a Quality Measure in the Pregnancy and Postpartum Period. *JACC: Advances*. 2023;2(1):100176. doi:10.1016/j.jacadv.2022.100176
16. Kocher B, Emanuel EJ. Aligning Incentives for Improving Diagnostic Excellence. *JAMA*. 2022;327(16):1543-1544. doi:10.1001/jama.2022.4594
17. Kadom N, Venkatesh AK, Shugarman SA, Burleson JH, Moore CL, Seidenwurm D. Novel Quality Measure Set: Closing the Completion Loop on Radiology Follow-up Recommendations for Noncritical Actionable Incidental Findings. *J Am Coll Radiol*. 2022;19(7):881-890. doi:10.1016/j.jacr.2022.03.017
18. Agha L, Skinner J, Chan D. Improving Efficiency in Medical Diagnosis. *JAMA*. 2022;327(22):2189-2190. doi:10.1001/jama.2022.8587
19. Combs CA, Hameed AB, Friedman AM, Hoskins IA. Special statement: Proposed quality metrics to assess accuracy of prenatal detection of congenital heart defects. *Am J Obstet Gynecol*. 2020;222(6):B2-B9. doi:10.1016/j.ajog.2020.02.040

20. Abel G, Saunders CL, Mendonca SC, Gildea C, McPhail S, Lyratzopoulos G. Variation and statistical reliability of publicly reported primary care diagnostic activity indicators for cancer: a cross-sectional ecological study of routine data. *BMJ Qual Saf*. 2018;27(1):21-30. doi:10.1136/bmjqs-2017-006607
21. Porter JB, Rosenthal EL, Winget M, et al. Improving Care With a Portfolio of Physician-Led Cancer Quality Measures at an Academic Center. *J Oncol Pract*. 2017;13(8):e673-e682. doi:10.1200/JOP.2017.021139
22. McDonald KM. Achieving Equity in Diagnostic Excellence. *JAMA*. 2022;327(20):1955-1956. doi:10.1001/jama.2022.7252
23. le Clercq CM, Mooi RJ, Winkens B, et al. Temporal trends and variability of colonoscopy performance in a gastroenterology practice. *Endoscopy*. 2016;48(3):248-255. doi:10.1055/s-0041-111117
24. Hoffmeister L, Lavados PM, Comas M, Vidal C, Cabello R, Castells X. Performance measures for in-hospital care of acute ischemic stroke in public hospitals in Chile. *BMC Neurol*. 2013;13:23. doi:10.1186/1471-2377-13-23
25. Kumar PR, Nash DB. An Update to "Understanding Ambulatory Care Practices in the Context of Patient Safety and Quality Improvement". *Am J Med Qual*. 2021;36(3):185-196. doi:10.1177/1062860620938762
26. Singh H, Graber ML, Hofer TP. Measures to Improve Diagnostic Safety in Clinical Practice. *J Patient Saf*. 2019;15(4):311-316. doi:10.1097/PTS.0000000000000338
27. Newman-Toker DE. Where Is the "Low-Hanging Fruit" in Diagnostic Quality and Safety? *Qual Manag Health Care*. 2018;27(4):234-236. doi:10.1097/QMH.0000000000000184

Appendix A: Search Terms – Measures Scan

Table 1. Search Terms Used for the Centers for Medicare & Medicaid Services (CMS) Measures Inventory Tool (CMIT) and National Quality Forum (NQF) Quality Positioning System (QPS)

Search Term
Delayed diagnosis
Diagnosis accuracy
Diagnosis delay
Diagnosis error
Diagnosis interval
Diagnosis safety
Diagnosis timeliness
Diagnostic accuracy
Diagnostic error
Diagnostic excellence
Diagnostic interval
Diagnostic quality
Diagnostic safety
Disease detection
Failure to diagnose
Inaccurate diagnosis
Missed diagnosis
Timely diagnosis
Wrong diagnosis

Table 2. Search Terms Used for Qualified Clinical Data Registries (QCDRs) and The Joint Commission

Search Term
Accuracy
Error
Excellence
Safety
Timeliness
Timely Diagnosis

Table 3. Search Terms Used for the National Committee for Quality Assurance (NCQA) Healthcare Effectiveness Data and Information Set (HEDIS)

Search Term
Accuracy
Cost
Detection
Diagnosis
Diagnostic
Error
Excellence

Search Term
Interval
Safety
Screening
Testing
Timely

Appendix B: Search Terms – Literature Review

((("diagnostic excellence"[tiab]) OR ("diagnosis error"[tiab] OR "diagnosis errors"[tiab] OR "diagnostic error"[tiab] OR "diagnostic errors"[tiab] OR "misdiagnosis"[tiab] OR "misdiagnoses"[tiab] OR "missed diagnosis"[tiab] OR "missed diagnoses"[tiab] OR "wrong diagnosis"[tiab] OR "wrong diagnoses"[tiab] OR "inaccurate diagnosis"[tiab] OR "inaccurate diagnoses"[tiab] OR "delayed diagnosis"[tiab] OR "delayed diagnoses"[tiab] OR "diagnosis delay"[tiab] OR "diagnosis delays"[tiab] OR "diagnostic delay"[tiab] OR "diagnostic delays"[tiab] OR "failure to diagnose"[tiab] OR "diagnostic interval"[tiab] OR "diagnostic intervals"[tiab] OR (Delayed diagnosis[mh]) OR (diagnos*[tiab] AND delay*[tiab]))) AND (("performance measure*"[tiab] OR "accountability measure*"[tiab]) OR "quality measure*"[tiab])) OR (("Quality Indicators, Health Care"[Mesh]) AND (("Delayed Diagnosis"[Mesh]) OR ("Diagnostic Errors"[Mesh])))

Appendix C: List of Relevant Literature Review Articles

Table 1. List of Relevant Literature Review Articles (listed from most recent to least)

Title	Author(s)	Date	Country	Link
Cardiovascular Risk Assessment as a Quality Measure in the Pregnancy and Postpartum Period	Hameed et al	2023	U.S.	https://www.jacc.org/doi/10.1016/j.jacadv.2022.100176
Aligning Incentives for Improving Diagnostic Excellence	Kocher, Emanuel	2022	U.S.	https://jamanetwork.com/journals/jama/article-abstract/2791103
Improving Efficiency in Medical Diagnosis	Agha, Skinner, Chan	2022	U.S.	https://jamanetwork.com/journals/jama/article-abstract/2792808
Measuring Performance of the Diagnostic Process	Burstin, Cosby	2022	U.S.	https://pubmed.ncbi.nlm.nih.gov/35737397/
Achieving Equity in Diagnostic Excellence	McDonald	2022	U.S.	https://pubmed.ncbi.nlm.nih.gov/35522307/
Adherence to National Guidelines for Timeliness of Test Results Communication to Patients in the Veterans Affairs Health Care System	Meyer et al	2022	U.S.	https://pubmed.ncbi.nlm.nih.gov/35452111/
Novel Quality Measure Set: Closing the Completion Loop on Radiology Follow-up Recommendations for Noncritical Actionable Incidental Findings	Kadom et al	2022	U.S.	https://pubmed.ncbi.nlm.nih.gov/35606263/

Title	Author(s)	Date	Country	Link
An Update to "Understanding Ambulatory Care Practices in the Context of Patient Safety and Quality Improvement"	Kumar & Nash	2021	U.S.	https://pubmed.ncbi.nlm.nih.gov/32691608/
Quality Improvement in Otolaryngology-Head and Neck Surgery: Age-Related Hearing Loss Measures	Gurgel et al	2021	U.S.	https://pubmed.ncbi.nlm.nih.gov/33752512/
Diagnostic Excellence	Yang et al	2021	U.S.	https://pubmed.ncbi.nlm.nih.gov/34709367/
Special statement: Proposed quality metrics to assess accuracy of prenatal detection of congenital heart defects	Combs et al	2020	U.S.	https://pubmed.ncbi.nlm.nih.gov/32114082/
Measures to Improve Diagnostic Safety in Clinical Practice	Singh et al	2019	U.S.	https://pubmed.ncbi.nlm.nih.gov/27768655/
Variation and statistical reliability of publicly reported primary care diagnostic activity indicators for cancer: a cross-sectional ecological study of routine data	Abel et al	2018	UK	https://pubmed.ncbi.nlm.nih.gov/28847789/
Where Is the "Low-Hanging Fruit" in Diagnostic Quality and Safety?	Newman-Toker	2018	U.S.	https://pubmed.ncbi.nlm.nih.gov/30260932/
Improving Care With a Portfolio of Physician-Led Cancer Quality Measures at an Academic Center	Porter et al	2017	U.S.	https://pubmed.ncbi.nlm.nih.gov/28727487/
Temporal trends and variability of colonoscopy performance in a gastroenterology practice	le Clercq et al	2016	Netherlands	https://pubmed.ncbi.nlm.nih.gov/26808394/
Improving Diagnosis in Health Care	National Academies of Sciences, Engineering, and Medicine	2015	U.S.	https://doi.org/10.17226/21794

Title	Author(s)	Date	Country	Link
The next organizational challenge: finding and addressing diagnostic error	Graber et al	2014	U.S.	https://pubmed.ncbi.nlm.nih.gov/24730205/
Performance measures for in-hospital care of acute ischemic stroke in public hospitals in Chile	Hoffmeister et al	2013	Chile	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3599613/pdf/1471-2377-13-23.pdf
Can the adenoma detection rate reliably identify low-performing endoscopists? Results of a modeling study	Saini et al	2013	U.S.	https://pubmed.ncbi.nlm.nih.gov/23456494/
Bringing diagnosis into the quality and safety equations	Graber et al	2012	U.S.	https://pubmed.ncbi.nlm.nih.gov/23011708/
Clinical and economic effects of unrecognized or inadequately treated bipolar disorder	Keck et al	2008	U.S.	https://pubmed.ncbi.nlm.nih.gov/18677197/
Performance measures from 10 years of breast screening in the Ontario Breast Screening Program, 1990/91 to 2000	Chiarelli et al	2006	Canada	https://pubmed.ncbi.nlm.nih.gov/16374227/

Appendix D: List of Relevant Measures from Measures Scan

Table 1. List of Relevant Measures from the Measures Scan Organized by NAM Step in the Diagnostic Process and Grouping of Similar Measurement Approaches

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M1	Adult Major Depressive Disorder (MDD): Suicide Risk Assessment	Clinical History, Interview, and Assessment	Assessment in patients with a risk factor	CMIT
M2	Child and Adolescent Major Depressive Disorder (MDD): Suicide Risk Assessment	Clinical History, Interview, and Assessment	Assessment in patients with a risk factor	CMIT
M3	Cognitive Impairment or Dysfunction Assessment for Patients with Parkinson's Disease	Clinical History, Interview, and Assessment	Assessment in patients with a risk factor	CMIT
M4	Maternity Care: Postpartum Follow-up and Care Coordination	a. Clinical History, Interview, and Assessment b. Diagnostic Testing	a. Assessment in patients with a risk factor b. Testing in patients with a risk factor	CMIT
M5	Parkinson's Disease: Psychiatric Symptoms Assessment for Patients with Parkinson's Disease	Clinical History, Interview, and Assessment	Assessment in patients with a risk factor	CMIT
M6	Bipolar Disorder and Major Depression: Assessment for Manic or Hypomanic Behaviors	Clinical History, Interview, and Assessment	Assessment prior to treatment	QPS
M7	Child and Adolescent Major Depressive Disorder: Diagnostic Evaluation	Clinical History, Interview, and Assessment	Further classification/severity of disease	QPS
M8	Comprehensive Cognitive Assessment Assists with Differential Diagnosis	Clinical History, Interview, and Assessment	Further classification/severity of disease	QCDR
M9	Diabetes Mellitus: Diabetic Foot and Ankle Care, Peripheral Neuropathy Neurological Evaluation	Physical Exam	In patients with a risk factor	CMIT
M10	Diabetic Foot & Ankle Care, Peripheral Neuropathy – Neurological Evaluation	Physical Exam	In patients with a risk factor	QPS
M11	Biopsy Follow-Up	Referral, Consultation, and Follow-up	Closing the loop	CMIT

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M12	Communication with the Physician or Other Clinician Managing On-going Care Post Fracture for Men and Women Aged 50 Years and Older	Referral, Consultation, and Follow-up	Cross-provider communication	QPS
M13	Diabetic Retinopathy: Communication with the Physician Managing Ongoing Diabetes Care	Referral, Consultation, and Follow-up	Cross-provider communication	CMIT
M14	Emergency Transfer Communication Measure	Referral, Consultation, and Follow-up	Cross-provider communication	QPS
M15	Skin Cancer: Biopsy Reporting Time Pathologist to Clinician	Referral, Consultation, and Follow-up	Cross-provider communication	CMIT
M16	Transitions of Care (TRC) Between the Inpatient and Outpatient settings	a. Referral, Consultation, and Follow-up b. Referral, Consultation, and Follow-up	a. Cross-provider communication b. Follow-up with patient	CMIT
M17	All Final Reports for Male Patients Aged 18 Years and Older Undergoing Prostate MRI for Prostate Cancer Screening or Surveillance	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QCDR
M18	Appropriate Cervical Spine Radiography and CT Imaging in Trauma	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QPS
M19	Appropriate Follow-up Imaging for Incidental Abdominal Lesions	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT
M20	Barrett's Esophagus	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT
M21	Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QPS

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M22	Low Dose Cancer Screening Recommendation for Computed Tomography (CT) and Computed Tomography Angiography (CTA) of Chest with Diagnosis of Emphysema	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QCDR
M23	Lung Cancer Reporting (Biopsy/Cytology Specimens)	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT
M24	Lung Cancer Reporting (Resection Specimens)	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT
M25	Melanoma Reporting	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT
M26	Mismatch Repair (MMR) or Microsatellite Instability (MSI) Biomarker Testing Status in Colorectal Carcinoma, Endometrial, Gastroesophageal, or Small Bowel Carcinoma	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT, QPS
M27	Nuclear Medicine: Correlation with Existing Imaging Studies for All Patients Undergoing Bone Scintigraphy	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT
M28	Optimizing Patient Exposure to Ionizing Radiation: Appropriateness: Follow-up CT Imaging for Incidentally Detected Pulmonary Nodules According to Recommended Guidelines	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT
M29	Prostate Cancer Reporting Best Practices	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QCDR
M30	Radical Prostatectomy Pathology Reporting	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	CMIT
M31	Surveillance Imaging for Liver Nodules Less Than 10mm in Patients at Risk for Hepatocellular Carcinoma (HCC)	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QCDR

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M32	Use of Quantitative Criteria for Oncologic FDG PET Imaging	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QCDR
M33	Use of Structured Reporting in Prostate MRI	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QCDR
M34	Use of Thyroid Imaging Reporting & Data System (TI-RADS) in Final Report to Stratify Thyroid Nodule Risk	Referral, Consultation, and Follow-up	Documentation in radiology/pathology reports	QCDR
M35	Diagnostic Imaging: Reminder System for Screening Mammograms	Referral, Consultation, and Follow-up	Use of automated reminder systems	QPS
M36	Melanoma: Continuity of Care Recall System	Referral, Consultation, and Follow-up	Use of automated reminder systems	CMIT
M37	Tracking of Clinical Results Between Visits	Referral, Consultation, and Follow-up	Use of automated reminder systems	QPS
M38	Photo Documentation of Cecal Intubation	Diagnostic Testing	Complete or guideline-based testing	CMIT
M39	Quantitative HER2 Evaluation by Immunohistochemistry (IHC) Uses the System Recommended by the ASCO/CAP Guidelines	Diagnostic Testing	Complete or guideline-based testing	QPS
M40	Penicillin Allergy: Appropriate Removal or Confirmation	Diagnostic Testing	Confirmation of diagnosis	QCDR
M41	Use of Spirometry Testing in the Assessment and Diagnosis of COPD	Diagnostic Testing	Confirmation of diagnosis	QPS, HEDIS
M42	Age-Related Macular Degeneration (AMD): Dilated Macular Examination	Diagnostic Testing	Further classification/severity of disease	CMIT
M43	Hematology: Myelodysplastic Syndrome (MDS) and Acute Leukemias: Baseline Cytogenetic Testing Performed on Bone Marrow	Diagnostic Testing	Further classification/severity of disease	QPS
M44	Hepatitis C: Confirmation of Hepatitis C Viremia	Diagnostic Testing	Further classification/severity of disease	QPS

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M45	HER2 Testing for Overexpression or Gene Amplification in Patients with Breast Cancer	Diagnostic Testing	Further classification/severity of disease	QPS
M46	Laboratory Investigation for Secondary Causes of Fracture	Diagnostic Testing	Further classification/severity of disease	QPS
M47	Primary Open-Angle Glaucoma (POAG): Optic Nerve Evaluation	Diagnostic Testing	Further classification/severity of disease	CMIT
M48	Sleep Apnea: Severity Assessment at Initial Diagnosis	Diagnostic Testing	Further classification/severity of disease	CMIT
M49	Abdomen Computed Tomography (CT) Use of Contrast Material	Diagnostic Testing	Overuse/appropriate use	CMIT
M50	Appropriate Testing for Children with Pharyngitis	Diagnostic Testing	Overuse/appropriate use	QPS
M51	Appropriate Testing for Pharyngitis	Diagnostic Testing	Overuse/appropriate use	CMIT, HEDIS
M52	Appropriate Use of DXA Scans in Women Under 65 Who Do Not Meet the Risk Factor Profile	Diagnostic Testing	Overuse/appropriate use	CMIT
M53	Avoidance of Chest X-ray in Pediatric Patients with Asthma, Bronchiolitis, or Croup	Diagnostic Testing	Overuse/appropriate use	QCDR
M54	Cardiac Imaging for Preoperative Risk Assessment for Non-Cardiac, Low-Risk Surgery	Diagnostic Testing	Overuse/appropriate use	CMIT
M55	Cardiac Stress Imaging Not Meeting Appropriate Use Criteria: Preoperative Evaluation in Low-Risk Surgery Patients	Diagnostic Testing	Overuse/appropriate use	CMIT
M56	Cardiac Stress Imaging Not Meeting Appropriate Use Criteria: Routine Testing After Percutaneous Coronary Intervention (PCI)	Diagnostic Testing	Overuse/appropriate use	CMIT, QPS
M57	Cardiac Stress Imaging Not Meeting Appropriate Use Criteria: Testing in Asymptomatic, Low-Risk Patients	Diagnostic Testing	Overuse/appropriate use	CMIT, QPS

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M58	Colonoscopy Interval for Patients with a History of Adenomatous Polyps-Avoidance of Inappropriate Use	Diagnostic Testing	Overuse/appropriate use	QPS
M59	Emergency Department Utilization of Computed Tomography (CT) for Minor Blunt Head Trauma for Patients Aged 2 Through 17 Years	Diagnostic Testing	Overuse/appropriate use	QCDR
M60	Emergency Department Utilization of Computed Tomography (CT) for Minor Blunt Head Trauma for Patients Aged 18 Years and Older	Diagnostic Testing	Overuse/appropriate use	QCDR
M61	Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level)	Diagnostic Testing	Overuse/appropriate use	QPS
M62	Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level)	Diagnostic Testing	Overuse/appropriate use	QPS
M63	Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Facility Level)	Diagnostic Testing	Overuse/appropriate use	QPS
M64	MRI Lumbar Spine for Low Back Pain	Diagnostic Testing	Overuse/appropriate use	CMIT
M65	Overuse of Imaging for the Evaluation of Primary Headache	Diagnostic Testing	Overuse/appropriate use	CMIT
M66	Overutilization of Imaging Studies in Melanoma	Diagnostic Testing	Overuse/appropriate use	QPS
M67	Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients	Diagnostic Testing	Overuse/appropriate use	CMIT
M68	Sentinel Lymph Node Biopsy for Invasive Breast Cancer	Diagnostic Testing	Overuse/appropriate use	CMIT

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M69	Use of Imaging Studies for Low Back Pain	Diagnostic Testing	Overuse/appropriate use	CMIT, QPS, HEDIS
M70	Adult Kidney Disease: Laboratory Testing (Lipid Profile)	Diagnostic Testing	Testing in patients with a risk factor	QPS
M71	Annual Hepatitis C Virus (HCV) Screening for Patients who are Active Injection Drug Users	Diagnostic Testing	Testing in patients with a risk factor	CMIT
M72	Comprehensive Diabetes Care	Diagnostic Testing	Testing in patients with a risk factor	QPS
M73	Diabetes: Eye Exam	Diagnostic Testing	Testing in patients with a risk factor	CMIT
M74	Diabetes: Medical Attention for Nephropathy	Diagnostic Testing	Testing in patients with a risk factor	CMIT
M75	Diabetes Screening for People with Schizophrenia or Bipolar Disorder Who Are Using Antipsychotic Medications (SSD)	Diagnostic Testing	Testing in patients with a risk factor	CMIT
M76	Diabetes Screening for People with Schizophrenia or Bipolar Disorder Who Are Using Antipsychotic Medications (SSD)	Diagnostic Testing	Testing in patients with a risk factor	QPS
M77	Hepatitis C: Screening for Hepatocellular Carcinoma (HCC) in Patients with Cirrhosis	Diagnostic Testing	Testing in patients with a risk factor	CMIT
M78	HIV Screening for Patients with a Sexually Transmitted Infection (STI)	Diagnostic Testing	Testing in patients with a risk factor	CMIT
M79	Kidney Health Evaluation	Diagnostic Testing	Testing in patients with a risk factor	CMIT
M80	Metabolic Monitoring for Children and Adolescents on Antipsychotics	Diagnostic Testing	Testing in patients with a risk factor	QPS
M81	Pregnant Women That Had HBsAg Testing	Diagnostic Testing	Testing in patients with a risk factor	QPS
M82	Pregnant Women That Had HIV testing	Diagnostic Testing	Testing in patients with a risk factor	QPS
M83	Screening for Metabolic Disorders	Diagnostic Testing	Testing in patients with a risk factor	CMIT

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M84	GERD - Upper Gastrointestinal Study in Patients with Alarm Symptoms	Diagnostic Testing	Testing in patients with symptoms	QPS
M85	Ultrasound Determination of Pregnancy Location for Pregnant Patients with Abdominal Pain	Diagnostic Testing	Testing in patients with symptoms	CMIT
M86	Bone Density Evaluation for Patients with Prostate Cancer and Receiving Androgen Deprivation Therapy	Diagnostic Testing	Testing prior to or at baseline for treatment	CMIT
M87	Hepatitis B Safety Screening	Diagnostic Testing	Testing prior to or at baseline for treatment	QCDR
M88	Paired Measure: Hepatitis C Virus (HCV) Genotype Testing Prior to Treatment (paired with 0395)	Diagnostic Testing	Testing prior to or at baseline for treatment	QPS
M89	Inflammatory Bowel Disease (IBD): Assessment of Hepatitis B Virus (HBV) Status Before Initiating Anti-TNF (Tumor Necrosis Factor) Therapy	Diagnostic Testing	Testing prior to or at baseline for treatment	CMIT
M90	RAS Gene Mutation Testing Performed for Patients with Metastatic Colorectal Cancer Who Receive Anti-Epidermal Growth Factor Receptor Monoclonal Antibody Therapy	Diagnostic Testing	Testing prior to or at baseline for treatment	QPS
M91	Rheumatoid Arthritis New DMARD Baseline CBC	Diagnostic Testing	Testing prior to or at baseline for treatment	QPS
M92	Rheumatoid Arthritis New DMARD Baseline Liver Function Test	Diagnostic Testing	Testing prior to or at baseline for treatment	QPS
M93	Rheumatoid Arthritis New DMARD Baseline Serum Creatinine	Diagnostic Testing	Testing prior to or at baseline for treatment	QPS
M94	Rheumatoid Arthritis: Tuberculosis Screening (Recommended for eMeasure Trial Approval)	Diagnostic Testing	Testing prior to or at baseline for treatment	QPS
M95	Tuberculosis Screening Prior to First Course Biologic Therapy	Diagnostic Testing	Testing prior to or at baseline for treatment	CMIT

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M96	Tympanostomy Tube Hearing Test	Diagnostic Testing	Testing prior to or at baseline for treatment	QPS
M97	Breast Cancer Screening Recall Rates	Diagnostic Testing	Timeliness of testing	CMIT
M98	Head Computed Tomography (CT) or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke who Received Head CT or MRI Scan Interpretation Within 45 Minutes of ED Arrival	Diagnostic Testing	Timeliness of testing	CMIT
M99	Helicobacter Pylori Status and Turnaround Time	Diagnostic Testing	Timeliness of testing	QCDR
M100	Report Turnaround Time: Computed Tomography (CT)	Diagnostic Testing	Timeliness of testing	QCDR
M101	Report Turnaround Time: Mammography	Diagnostic Testing	Timeliness of testing	QCDR
M102	Report Turnaround Time: MRI	Diagnostic Testing	Timeliness of testing	QCDR
M103	Report Turnaround Time: PET	Diagnostic Testing	Timeliness of testing	QCDR
M104	Report Turnaround Time: Radiography	Diagnostic Testing	Timeliness of testing	QCDR
M105	Report Turnaround Time: Ultrasound (Excluding Breast Ultrasound)	Diagnostic Testing	Timeliness of testing	QCDR
M106	Severe Sepsis/Septic Shock: Management Bundle	Diagnostic Testing	Timeliness of testing	CMIT
M107	Total Number of Ultrasound Exams Completed (Excluding Breast Ultrasound)	Diagnostic Testing	Timeliness of testing	QCDR
M108	Turnaround Time (TAT) - Biopsies	Diagnostic Testing	Timeliness of testing	QCDR
M109	Urinary Bladder Biopsy Diagnostic Requirements for Appropriate Patient Management	Diagnostic Testing	Timeliness of testing	QCDR
M110	Transfer of Health Information to the Patient Post-Acute Care (PAC)	Communication of the Diagnosis	Documentation/records	CMIT
M111	Venous Thromboembolism Warfarin Therapy Discharge Instructions	Communication of the Diagnosis	Documentation/records	QPS

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M112	Consideration of Cultural-Linguistic and Demographic Factors in Cognitive Assessment	Communication of the Diagnosis	Patient centeredness	QCDR
M113	L2: Patients Receiving Language Services Supported by Qualified Language Services Providers	Communication of the Diagnosis	Patient centeredness	QPS
M114	Provision of Feedback Following a Cognitive or Mental Status Assessment with Documentation of Understanding of Test Results and Subsequent Healthcare Plan with Timely Transmission of Results	a. Communication of the Diagnosis b. Referral, Consultation, and Follow-up	a. Patient centeredness, documentation/records b. Closing the loop	QCDR
M115	Skin Cancer: Biopsy Reporting Time - Clinician to Patient	Communication of the Diagnosis	Timely communication	QCDR
M116	Clinical Depression Screening and Follow-Up	Treatment	Positive screening with follow-up plan	CMIT
M117	Cognitive Assessment with Counseling on Safety and Potential Risk	Treatment	Positive screening with follow-up plan	QCDR
M118	Depression Screening and Follow-Up for Adolescents and Adults	Treatment	Positive screening with follow-up plan	HEDIS
M119	Global Malnutrition Composite Score	Treatment	Positive screening with follow-up plan	QPS
M120	Postpartum Depression Screening and Follow-Up	Treatment	Positive screening with follow-up plan	HEDIS
M121	Prenatal Depression Screening and Follow-Up	Treatment	Positive screening with follow-up plan	HEDIS
M122	Preventive Care and Screening: Screening for Depression and Follow-Up Plan	Treatment	Positive screening with follow-up plan	QPS
M123	Preventive Care and Screening: Screening for High Blood Pressure and Follow-Up Documented	Treatment	Positive screening with follow-up plan	CMIT
M124	Unhealthy Alcohol Use Screening and Follow-Up	Treatment	Positive screening with follow-up plan	HEDIS

Measure Reference Number*	Measure Title	NAM Step in the Diagnostic Process	Grouping of Similar Measurement Approaches	Source
M125	Inappropriate Diagnosis of Community-Acquired Pneumonia (CAP) in Hospitalized Medical Patients (abbreviated form: Inappropriate Diagnosis of CAP)	Outcomes	Diagnosis was correct	QPS
M126	Inappropriate Diagnosis of Urinary Tract Infection (UTI) in Hospitalized Medical Patients (abbreviated form: Inappropriate Diagnosis of UTI)	Outcomes	Diagnosis was correct	QPS
M127	Door to Diagnostic Evaluation by a Qualified Medical Personnel	Outcomes	Diagnosis was timely	QPS

**Measure numbers are for internal reference within this document only and do not refer to any formal measure numbering systems such as CMIT or CBE measure numbers.*

Appendix E: Gordon and Betty Moore Foundation Measure Developer Grantees and Measures

Table 1: Gordon and Betty Moore Foundation Measure Developer Grantees and Measures

Grantee	Measure Title
Acumen	Improving the Diagnostic Performance of Screening Tests for Breast Cancer
American Board of Family Medicine	Measuring the Value-Functions of Primary Care: Physician level Continuity of Care Measure
American College of Emergency Physicians (ACEP)	Improving the Diagnosis of Ruptured Abdominal Aortic Aneurysms
ACEP	Implementing Quality Measures to Improve the Diagnosis of Pulmonary Embolism
American College of Radiology	Closing the Loop—Improving Care Coordination Toward Early Disease Detection
American Institutes for Research	Improving the Diagnosis of Urinary Tract Infection in Women
American Society of Clinical Oncology	Biomarker and Genomic Testing to Inform Personalized Cancer Therapy
American Society of Hematology/Health Services Advisory Group	Implementing Clinical Pretest Probability into Practice to Improve the Diagnosis of Pulmonary Embolism
Baylor College of Medicine	Clinical Quality Measure to Improve Diagnosis of Cancer: The Safer Dx Cancer e-measure
Baylor College of Medicine/University of North Carolina	Clinical Quality Measures to Improve the Diagnosis for Gastrointestinal (GI) Cancers
Brigham and Women's Hospital (BWH)	Diagnostic Delay of Venous Thromboembolism in Primary Care: A Data Science and Machine Learning Approach
BWH	Testing, Endorsement, and Implementation of Two Quality Measures to Improve the Follow-Up of Abnormal Cancer Screening Tests
Hospital for Special Surgery	Development of Diagnostic Quality Metrics for Periprosthetic Joint Infection
Johns Hopkins University	A Patient-Reported Measure of Diagnostic Excellence
National Committee for Quality Assurance	Quality Measures for Screening and Diagnosis of Lynch Syndrome, an Inheritable Risk for Colon Cancer
RAND	A Patient-Reported Outcome Performance Measure to Improve the Timeliness and Quality of Cancer Diagnosis
Stanford University	Missed Outpatient Diagnosis of Incident Heart Failure
Stanford University	Diagnosis of Vascular Risk to Reduce Heart Attack and Strokes
University of California Irvine	Developing Cardiovascular Measures to Improve Diagnosis of Cardiovascular Disease in Pregnant and Postpartum Women
University of California San Diego (UCSD)	SEP1+: A Composite Measure to Accurately Assess Early Sepsis Management
UCSD	Structured Review of Cases to Identify Diagnostic Improvement Opportunities
University of Michigan/University of Utah	Quality Measures to Improve Diagnosis of Infections

Grantee	Measure Title
University of Pennsylvania	Advancing Diagnostic Quality of Lung Cancer Screening through Measurement
University of Utah	Measuring and Improving Diagnostic Excellence in Pneumonia
University of Utah	Measuring Misdiagnosis of Infections Using Machine Learning in a Flexible Framework
University of Wisconsin	Measuring Interval Colorectal Cancer Rates Across Health Care Systems Overall and By Screening Modality
Yale University	Diagnostic Excellence Index for Pulmonary Embolism
Yale University	Improving Equity of Lung Cancer Diagnosis Following Chest Imaging in the Emergency Department
Washington University	Lung Cancer Diagnosis Measurement Instrument

Appendix F: Related NQF Work

Improving Diagnostic Quality & Safety Final Report, 2017

In an effort to develop a measurement framework to assist in reducing diagnostic harm, the National Quality Forum (NQF) convened a multistakeholder expert Committee to develop a conceptual framework for measuring diagnostic quality and safety and identify priorities for future measure development. Utilizing the evidence, concepts, and models contained in the National Academies of Sciences, Engineering, and Medicine's Improving Diagnosis in Health Care report, the conceptual framework is intended to facilitate systematic identification and prioritization of measure gaps and to help guide efforts to fill those gaps through measure development and endorsement.

Reducing Diagnostic Error: Measurement Considerations Final Report, 2020

The National Quality Forum (NQF) convened a multistakeholder committee to identify recommendations for the practical application of the Diagnostic Process and Outcomes domain of the 2017 Diagnostic Quality and Safety Measurement Framework, and for measuring and reducing diagnostic error, and measuring and improving patient safety. The project culminated in a final report that outlines the recommendations through a series of four use cases. They depict resolutions to specific types of diagnostic errors and broad, comprehensive recommendations with applications to multiple populations and settings.

Appendix G: Diagnostic Excellence Committee Members, Federal Liaisons, Advisory Group Members, and NQF Staff

Diagnostic Excellence Committee Members

Matt Austin, PhD
Associate Professor, Johns Hopkins Armstrong Institute for Patient Safety and Quality
Baltimore, Maryland

Rosemary "Rosie" Bartel, MA
Patient Advisor, Institute for Patient- and Family-Centered Care
Chilton, Wisconsin

Sigall Bell, MD
Associate Professor of Medicine, Harvard Medical School
Boston, Massachusetts

Joseph Cerimele, MD, MPH
Professor, University of Cincinnati School of Medicine
Cincinnati, Ohio

Melissa "Missy" Danforth, BA
Vice President of Health Care Ratings, Leapfrog Group
Washington, District of Columbia

Christina Davidson, MD
Vice Chair of Quality, Patient Safety & Health Equity, Baylor College of Medicine
Houston, Texas

William Brendle Glomb, MD, FCCP, FAAP
Subject Matter Expert
String Prairie, Texas

Joseph A. Grubenhoff, MD, MSCS
Medical Director, Diagnostic Safety Program at Children's Hospital Colorado
Denver, Colorado

Carole Hemmelgarn, MS
Senior Director of Education, MedStar Institute for Quality and Safety
Denver, Colorado

Barbara Jones, MD, MSCI
Associate Professor, University of Utah
Salt Lake City, Utah

Michael Kanter, MD
Professor and Chair, Kaiser Permanente Bernard J. Tyson School of Medicine
Pasadena, California

Joseph Kunisch, PhD, RN-BC, CPHQ
Vice President of Quality Programs, Harris Health System
Houston, Texas

Prashant Mahajan, MD, MPH, MBA
Professor of Emergency Medicine and Pediatrics, Chief of Pediatric EM at University of Michigan
Ann Arbor, Michigan

Raquel Mayne, MPH, MSN, RN, CPHQ
Vice President of Quality Management, Phelps Hospital Northwell Health
Sleepy Hollow, New York

Nicholas Meo, MD
Associate Medical Director of Quality and Patient Safety, UW Medicine – Harborview Medical Center
Seattle, Washington

Matthew Nielsen, MD, MS, FACS
Chair, Science and Quality Council at American Urological Society
Durham, North Carolina

Irit R. Rasooly, MD, MSCE
Assistant Professor, Children's Hospital of Philadelphia, University of Pennsylvania
Philadelphia, Pennsylvania

Kimberly R. Rodgers
Certified Patient Advocate
Chicago, Illinois

John Sather, MD
Director of Patient Safety and Quality, Department of Emergency Medicine, Yale School of Medicine
New Haven, Connecticut

David Seidenwurm, MD, FACR
Physicians Alliance Chief Medical Officer, Sutter Health
Sacramento, California

Hardeep Singh, MD, MPH
Co-Chief, Health Policy, Quality & Informatics, VA Center of Innovation and Baylor College of Medicine
Houston, Texas

Julia Skapik, MD, MPH, FAMIA
Chief Medical Information Officer, National Association of Community Health Centers (NACHC)
Bethesda, Maryland

Alexis Snyder, BA
Patient and Stakeholder Engagement Specialist
Brookline, Massachusetts

Leslie Tucker, BA
Senior Policy Advisory, Society to Improve Diagnosis in Medicine
Alexandria, VA

Divvy Uphadyay, MD, MPH
Program Lead on Committee to Improve Clinical Diagnosis, Geisinger Health System
Danville, Pennsylvania

Valerie Vaughn, MD, MSc
Director of Hospital Medicine Research, University of Utah
Salt Lake City, Utah

Emily Volk, MD, MBA, FCAP
President, College of American Pathologists
Louisville, Kentucky

Kevin P. Wake, MS, SCDAA, CMR
President, Uriel E. Owens Sickle Cell Disease Association of the Midwest
Kansas City, Kansas

Scott Woller, MD, FACP, FCCP
Chair of Medicine, Intermountain Medical Center
Salt Lake City, Utah

Jennifer Woodward, MD, MPH
Medical Director, Quality & Performance Measurement, American Academy of Family Physicians
Kansas City, Kansas

Ronald Wyatt, MD, MHA
Chief Science Officer and Chief Medical Officer, Society to Improve Diagnosis in Medicine
Orange Beach, Alabama

Federal Liaisons

Andrea Benin, MD
Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention
Atlanta, Georgia

Stephanie Clark, MD, MPH, MSHP
Medical Officer, Centers for Medicare & Medicaid Services
Philadelphia, Pennsylvania

Romsai Tony Boonyasai, MD, MPH
Division of Quality Measurement and Improvement, Agency for Healthcare Research and Quality
Baltimore, Maryland

National Quality Forum Staff

Elizabeth Drye, MD, SM
Chief Scientific Officer

Kathryn McDonald, PhD
Bloomberg Distinguished Professor of Health Systems, Quality and Safety and project co-principal investigator

Jenna Williams-Bader, MPH
Managing Director

Chuck Amos, MBA
Managing Director

Laura Blum-Meisnere, MA
Senior Director

Deidra Smith-Fisher, MBA, PMP
Director

Hannah Ingber, MPH
Senior Manager

Carolee Lantigua, MPA
Senior Manager

Joelencia LeFlore, MPH
Analyst

Isabella Rivero
Analyst

Magdelana Stinnett, MSPH
Analyst

OUR MISSION

To be the trusted voice driving measurable health improvements

OUR VISION

Every person experiences high value care and optimal health outcomes

OUR VALUES

Collaboration • Excellence
Integrity • Leadership
Passion