NATIONAL QUALITY FORUM
Driving measurable health improvements together

Patient Safety, Fall 2021 Cycle: CDP Report

DRAFT REPORT FOR COMMENT
MARCH 31, 2022

This report is funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001.

https://www.qualityforum.org

Contents

Executive Summary

Patient safety has long been a central goal of the National Quality Forum (NQF), and patient safety measurement efforts over the last two decades have focused on quality improvement in healthcare organizations to improve care delivery and outcomes for patients. Examples include reductions in central line-associated blood stream infections (CLABSIs), catheter-associated urinary tract infections (CAUTIs), falls, pressure ulcers, inpatient mortality, and vital care processes for sepsis, medication reconciliation, and others. NQF's Patient Safety Standing Committee, a multistakeholder group consisting of patient safety clinical leaders, patient representatives, healthcare quality experts, and other thought leaders, carefully reviews new and existing patient safety measures and makes recommendations for endorsement.

During this cycle, the Patient Safety Standing Committee evaluated four newly submitted measures and one maintenance measure against NQF's measure evaluation criteria. Measures focused on unintended weight loss, coronavirus disease 2019 (COVID-19) vaccination coverage, and excessive radiation exposure from computed tomography (CT) scans. The Standing Committee recommended all five measures for endorsement.

The Standing Committee recommended the following measures:

- NQF #0689 Percent of Residents Who Lose Too Much Weight (Long Stay) (Acumen/Centers for Medicare & Medicaid Services [CMS])
- NQF #3636 Quarterly Reporting of COVID-19 Vaccination Coverage Among Healthcare Personnel (Centers for Disease Control and Prevention [CDC])
- NQF #3633e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level) (Alara Imaging/University of California, San Francisco [UCSF])
- NQF #3662e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level) (Alara Imaging/UCSF)
- NQF #3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Facility Level) (Alara Imaging/UCSF)

Brief summaries of the measures and their evaluations are included in the body of the report; detailed summaries of the Standing Committee's discussion and ratings of the criteria for each measure are in Appendix A.

Introduction

On a global level, the World Health Organization (WHO) estimates that one of the top 10 causes of patient mortality and morbidity worldwide is healthcare that does not adequately protect the safety of the patient. WHO also estimates that nearly 10 percent of patients in high-income countries suffer harm while receiving hospital care, of which almost half of the cases are preventable.[1] Recent studies place the number of preventable deaths of hospitalized patients in the United States (U.S.) at approximately 22,000 a year. These deaths are largely due to diagnostic errors, errors in surgery or other procedures, and poor management of medical conditions.[2]

NQF's Patient Safety Standing Committee makes recommendations for endorsing NQF's portfolio of structure, process, and outcome measures pertaining to patient safety. These measures have been used in various accountability and public reporting programs nationally and have led to lower rates of complications, medical errors, and mortality, among others. These measures also span various settings and are focused on care delivered in hospitals, rehabilitation centers, skilled nursing facilities, outpatient clinics, and delivered by health plans.

The measures reviewed this cycle focused on three clinical areas: unintended weight loss in nursing home residents, COVID-19 vaccination rates, and excessive radiation exposure from CT scans.

Unintentional Weight Loss

While avoiding weight gain and obesity are topics commonly discussed with aging adults to prevent the onset of chronic disease, older adults are also at risk of nutritional deficiencies or malnutrition and can experience unintentional weight loss.[3,4] Unintentional weight loss in the elderly can occur in any living situation but can be especially pronounced in hospitals or institutional settings[5] and can also lead to complications, including various types of functional decline, frailty, and mortality.

Vaccination of Healthcare Personnel

WHO recommends 10 vaccinations for healthcare personnel (HCP) and urges HCP to be fully vaccinated according to the vaccination schedule at use in their respective countries.[8] Within the U.S., the Centers for Disease Control and Prevention (CDC) urge HCP to reduce the chance of acquiring vaccine-preventable diseases by keeping their personal vaccination records up to date.[9] In the context of the current COVID-19 pandemic, WHO identifies HCP as critical members of the pandemic response effort who are at higher risk of contracting COVID-19 due to their role. It recommends including HCP in the list of priority vaccinations, along with older people and those with chronic health conditions.

Excessive Radiation Exposure

High and moderate levels of radiation exposure are shown to be linked to increased risk of leukemia, and more recent studies have connected cumulative exposure to low doses with increased risk of leukemia as well.[12] Studies show that doses of radiation as low as 10 millisieverts (mSV) from acute exposures and 50 mSv from prolonged exposures can increase the risk of cancer. While a variety of

environmental exposures can result in exposures that exceed these numbers, certain cancer treatments or diagnostic scans commonly result in acute exposures greater than this threshold.

NQF Portfolio of Performance Measures for Patient Safety Conditions

The Patient Safety Standing Committee (Appendix C) oversees NQF's portfolio of Patient Safety measures (Appendix B), which includes measures for medication safety, healthcare-associated infections, perioperative safety, falls, mortality, venous thromboembolism, pressure ulcers, workforce safety, and radiation safety. This portfolio contains 47 measures: 21 outcome and resource use measures, 19 process measures, three composite measures, three structure measures, and one intermediate outcome measure.

Additional measures relevant to patient safety have been assigned to other portfolios. These include care coordination measures (Geriatrics and Palliative Care), imaging efficiency measures (Cost and Efficiency), and a variety of condition- or procedure-specific outcome measures (Cardiovascular, Cancer, Renal, etc.).

Patient Safety Measure Evaluation

On February 16, 2022 the Patient Safety Standing Committee evaluated four new measures and one measure undergoing maintenance review against NQF's standard measure evaluation criteria.

Table 1. Patient Safety Measure Evaluation Summary

Measure	Maintenance	New	Total
Measures under review for endorsement	1	4	5
Measures recommended for endorsement	1	4	5

Scientific Methods Panel Measure Evaluation

Prior to the Standing Committee's review, the Scientific Methods Panel (SMP) reviewed four complex measures in this topic area. The SMP passed three measures but did not reach consensus on validity for the remaining measure during its measure evaluation. Measures that passed the SMP's review or for which the SMP did not reach consensus were reviewed by the Standing Committee.

A meeting summary detailing the SMP's measure evaluation for the fall 2021 cycle is available on the SMP webpage.

Comments Received Prior to Standing Committee Evaluation

NQF accepts comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). In addition, NQF solicits comments for a continuous 16-week period during each evaluation cycle via an online tool located on the project webpage. For this evaluation cycle, the commenting period opened on December 6, 2021, and pre-meeting commenting closed on January 16,

2022. As of January 16, 2022, 67 comments were submitted and shared with the Standing Committee prior to the measure evaluation meeting(s) (Appendix D).

NQF members had the opportunity to express their support ("support" or "do not support") for each measure submitted for endorsement consideration to inform the Standing Committee's recommendations during the commenting period. This expression of support (or not) during the commenting period replaces the member voting opportunity that was previously held subsequent to the Standing Committee's deliberations. NQF #3636 received one expression of "support." NQF #3633e, NQF #3662e, and NQF #3663e each received two expressions of "support" and one of "do not support."

Summary of Measure Evaluation

The following brief summaries of the measure evaluation highlight the major issues that the Standing Committee considered. Details of the Standing Committee's discussion and ratings of the criteria for each measure are included in Appendix A.

NQF #0689 Percent of Residents Who Lose Too Much Weight (Long Stay) (Acumen/CMS)

Description: This measure captures the percentage of long-stay nursing home residents with a target Minimum Data Set (MDS) 3.0 assessment (OBRA, PPS, or discharge) that indicates a weight loss of 5% or more of the baseline weight in the last 30 days, or 10% or more of the baseline weight in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen; **Measure Type**: Outcome: Intermediate Clinical Outcome; **Level of Analysis**: Facility; **Setting of Care**: Post-Acute Care; **Data Source**: Assessment Data from the MDS 3.0

This facility-level measure was originally endorsed in 2011 and maintained endorsement in 2015. Additionally, this measure is publicly reported nationally in Care Compare and the Provider Data Catalog. The Standing Committee noted that the developer provided updated evidence in support of the measure. The Standing Committee had no concerns and passed the measure on evidence. For future maintenance review, the Standing Committee recommended that the measure developer include evidence on whether the full list of risks associated with weight loss either are or are not modifiable by facilities. When reviewing the performance gap, the Standing Committee noted that patients over 85 years of age and those who are White had a slightly higher risk of losing too much weight. The Standing Committee agreed that a gap exists and passed the measure on performance gap. For future maintenance review, the Standing Committee suggested that the developer present a stratified analysis of the measure scores by facility characteristics/types, disease areas, and different subpopulations of interest to further examine disparities by subgroups that are known to have differing outcomes.

The SMP reviewed this measure and passed it on reliability but did not reach consensus on validity. The Standing Committee agreed that the specifications were reasonable after confirming that the measure excludes residents either under hospice care or with a life expectancy of less than six months; it also agreed that the reliability testing was sufficient. The Standing Committee accepted the SMP's rating of moderate for reliability. Based on the SMP's feedback, the Standing Committee discussed whether certain MDS items might warrant a risk adjustment strategy. The developer reported that they reviewed the suggested variables and observed low to moderate correlations between diagnostic options on the

MDS and weight loss but none that changed the facility's measure score or rank; therefore, the measure was intentionally not risk-adjusted. The Standing Committee members agreed that risk adjustment would not be appropriate for this measure for conceptual and empirical reasons and passed the measure on validity. However, the Standing Committee recognized that specialized facilities that have greater concentrations of high-risk patients may be disadvantaged on this measure. For future maintenance review, the Standing Committee suggested that the developer examine how their risk adjustment strategy might affect scores at highly specialized facilities (e.g., those that take mechanically ventilated patients). The Standing Committee did not raise any concerns about feasibility, use, or usability and passed the measure on these three criteria and overall suitability for endorsement; however, it recommended that the developer present a review of performance changes since first use in 2011 at the measure's next maintenance review.

NQF #3636 Quarterly Reporting of COVID-19 Vaccination Coverage Among Healthcare Personnel (CDC)

Description: This quarterly measure identifies the average percentage of healthcare personnel (HCP) who have ever received a primary COVID-19 vaccination course among the total number of HCP who regularly work in the facility; **Measure Type**: Process; **Level of Analysis**: Facility; **Setting of Care**: Post-Acute Care; **Data Source**: Varies (National Healthcare Safety Network)

This facility-level measure was newly submitted for endorsement. The measure is publicly reported nationally as part of the National Healthcare Safety Network (NHSN). The Standing Committee discussed whether the evidence provided in the submission supported the contention that measuring COVID-19 vaccination rates among HCP would lead to an increase in vaccination rates, and ultimately, a decrease in cases. The developer noted that while systematic reviews of evidence surrounding vaccination of HCP for COVID-19 were not yet available at the time of measure's submission, several studies have since been published showing a decrease in case rates in facilities that had high vaccination rates for HCP and the impact that the reporting of vaccination rates at a facility had on those rates. Several Standing Committee members noted additional evidence in support of the measure, such as the significant reductions in the spread of COVID-19 when HCP are vaccinated. A few Standing Committee members expressed concern that members of the public might inappropriately equate low COVID-19 vaccination rates at a facility with poor quality of care at that facility. Other Standing Committee members countered that while the quality of care provided might be otherwise good, the vaccination status of HCP at that facility also has the potential to impact the patients cared for at the facility and should be public knowledge for evaluating care facilities. The Standing Committee acknowledged that the evidence the developer provided was sound, especially considering it was gathered amid an emerging global pandemic. Therefore, the Standing Committee passed the measure on evidence. The Standing Committee noted large gaps in performance between the lowest- and highest-performing nursing homes and a large difference in vaccination rates according to the type of HCP and passed the measure on performance gap.

The Standing Committee had no concerns with the measure's reliability and voted to pass the measure on this criterion. The Standing Committee then reviewed the validity testing of the measure, as well as how the developer addressed any potential threats to validity. The Standing Committee expressed some concerns with the optional reporting category of contract personnel included in the denominator,

stating that it seems facilities would report this category when it improves their score and not report it when it does not. The developer clarified that the denominator was created to mirror the denominator of NQF #0431, the currently NQF-endorsed influenza vaccination of HCP measure, which also does not require the reporting of contract personnel. The Standing Committee stressed that contract personnel have become a much greater percentage of HCP since the pandemic began and urged the developer to consider making this reporting category a requirement for future maintenance reviews. Ultimately, the Standing Committee had no concerns and passed the measure on validity.

The Standing Committee discussed whether collecting data for this measure was more feasible amid the pandemic when it was critically relevant and whether it would pose a reporting burden at a later date when the threat may have waned. The developer explained that they chose quarterly reporting to mitigate extremes and make reporting less burdensome than weekly but more immediately useful than annually. Ultimately, the Standing Committee passed the measure on feasibility. The Standing Committee expressed no concerns about use and usability and passed the measure on use, usability, and overall suitability for endorsement.

The Standing Committee reviewed one related measure for NQF #3636: NQF #0431 *Influenza Vaccination Coverage Among Healthcare Personnel*. As stated above, the denominator for NQF #3636 was harmonized to mirror that of NQF #0431; however, the data collection time frame for each measure is different. The Standing Committee acknowledged that not enough information is known yet about the potential seasonality of COVID-19 infections to make any additional recommendations for harmonization at this time. The Standing Committee also noted that future vaccine mandates may affect the alignment of these measures upon maintenance review.

Excessive Radiation Exposure

NQF #3633e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level) (Alara Imaging/UCSF)

Description: This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible; **Measure Type**: Outcome: Intermediate Clinical Outcome; **Level of Analysis**: Clinician: Individual; **Setting of Care**: Ambulatory Care, Inpatient/Hospital, Outpatient Services; **Data Source**: Electronic health records

This individual clinician-level measure was newly submitted for endorsement. During the discussion on evidence, the Standing Committee expressed concern that much of the data in the studies came from a pediatric population. In response, the developer explained that while the systematic reviews of radiation dosing and CT scans all focus on children, there are many papers that focus on adults, which similarly show that patients with an increased exposure to CT scans have an increased risk of developing cancer. The Standing Committee passed the measure on evidence. The Standing Committee agreed that

there was room for improvement and that disparities existed, specifically for those living with a higher level of poverty, and passed the measure on performance gap.

The SMP reviewed this measure and passed it on reliability and validity. The Standing Committee asked for clarity on how the reliability thresholds were determined and their impact on how many radiologists might be excluded from the measure as a result. The developer clarified that the vast majority of radiologists perform at least the minimum number of scans; therefore, the threshold leads to the exclusion of very few radiologists from the measure. The Standing Committee voted to accept the SMP's rating of high for reliability. The Standing Committee also discussed several topics related to the validity of the measure, such as whether the developer had considered additional clinical care factors outside of body mass index (BMI) that might affect dosing. In response, the developer stated that no other factors had a strong impact on the measure, and none were significant enough to warrant risk adjustment. The developer also addressed the possibility of misclassification in creating their dosing strata within CT categories but stated that they worked extensively with clinicians and radiologists to understand the dosing needs for various types of patients and erred on the side of allowing for the possibility of a higher dose when they were asked to by clinical experts. The Standing Committee also asked whether there were any validity issues based on provider attribution and specifically asked the developer how attribution is assigned. The developer stated that for the current measure, at the individual-clinician level, it is the radiologist who bills for the exam and who is held accountable. The Standing Committee had no further questions and voted to accept the SMP's rating of high for validity.

Regarding feasibility, the Standing Committee questioned what the effect might be of having only one vendor who can pull these data. The developer replied that they created this vendor organization to respond to a request from the Centers for Medicare & Medicaid Services (CMS) to manage nationwide implementation and reporting because no other alternative had presented itself. The developer added that the measure specifications are publicly available, and all collected data are already in the electronic health record (EHR), billing claims, or other frequently used data systems. Therefore, the fact that there is currently only one vendor who can report this measure does not preclude other vendors from also doing so. In addition, clinicians and hospitals can report on the measure at no cost using a web interface. The Standing Committee had no other concerns with feasibility. Likewise, the Standing Committee had no concerns with use and voted to pass the measure on both feasibility and use.

The Standing Committee asked how frequent the need was for additional scans due to low quality. The developer replied that in a quality study using a sample of 700+ scans, which included an overrepresentation of low-dose scans (in which poor image quality would be most likely), only 11 percent were considered unacceptable. The developer acknowledged the need to pay attention to this issue as a possible unintended consequence of encouraging lower-dose scans; they intend to monitor it closely once the measure is implemented and adjust the thresholds if needed. The Standing Committee passed the measure on usability and overall suitability for endorsement.

NQF #3662e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level) (Alara Imaging/UCSF)

Description: This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk

factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible; **Measure Type**: Outcome: Intermediate Clinical Outcome; **Level of Analysis**: Clinician: Group/Practice; **Setting of Care**: Inpatient/Hospital, Outpatient Services; **Data Source**: Electronic health records

This clinician group-level measure was newly submitted for endorsement. The Standing Committee noted that the evidence and performance gap information provided for NQF #3662e was the same as that provided for NQF #3633e. They had no further concerns or discussion beyond what had been addressed during the previous measure discussion. The Standing Committee passed the measure on both evidence and performance gap.

The SMP reviewed this measure and passed it on both reliability and validity. The numerator, denominator, and exclusions for NQF #3662e were identical to that of NQF #3633e. The Standing Committee asked for confirmation of whether the threshold for the number of scans performed to achieve sufficient reliability for inclusion in the measure was the same for the clinician group as it was for the individual clinician level, which the developer confirmed. At the group level, this would exclude very few, if any, practices. The Standing Committee accepted this response with no further concerns and then accepted the SMP's rating of high for reliability. The data element validity testing was conducted at the individual clinician level and was identical to NQF #3633e, as were the face validity results. The Standing Committee questioned whether there might be some attribution concerns that persist at the group level or whether the group level mitigated most of the concerns it had with attribution at the individual clinician level. Ultimately, the Standing Committee decided the measure was valid and accepted the SMP's rating of high for validity.

The Standing Committee reiterated that the feasibility, use, and usability criteria were essentially the same for NQF #3662e as what was previously reviewed and discussed for NQF #3633e and passed NQF #3662e on all three criteria and on overall suitability for endorsement.

A summary of the discussion of measures related to NQF #3662e can be found below, following the summary of discussion and voting for NQF #3663e.

NQF #3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Facility Level) (Alara Imaging/UCSF)

Description: This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient and hospital outpatient care settings are eligible; **Measure Type**: Outcome: Intermediate Clinical Outcome; **Level of Analysis**: Facility Level; **Setting of Care**: Ambulatory Care, Inpatient/Hospital, Outpatient Services; **Data Source**: Electronic health records

This facility-level measure was newly submitted for endorsement. The Standing Committee noted that the evidence and opportunity for improvement for NQF #3663e were the same as for NQF #3633e and NQF #3662e and voted to pass the measure on evidence and performance gap.

During the discussion on reliability, the Standing Committee noted that at the hospital level, the developer obtained CT scans during inpatient hospitalizations and conducted a split-sample analysis, for which the intraclass correlation coefficient was very high (greater than 0.99 within each hospital). The Standing Committee noted that this measure also applies to outpatient scans and asked the developer to comment on whether there are any technical differences between the two settings. The developer clarified that the indications would not be identical, and inpatient settings would likely have more trauma and stroke scans; nonetheless, the results were identical. The Standing Committee accepted this explanation, asked no further questions, and voted to accept the SMP's rating of high for reliability. Likewise, the Standing Committee had no questions or concerns about the measure's validity and voted to accept the SMP's rating of high for validity.

The Standing Committee noted that the feasibility, use, and usability information provided for NQF #3663e was identical to what had been provided for NQF #3633e and NQF #3662e and was previously discussed by the Standing Committee. It had no concerns and passed the measure on all three criteria and on overall suitability for endorsement.

Following this recommendation for endorsement, the Standing Committee held a discussion about measures related to the three adult radiology measures: NQF #3633e, NQF #3662e, and NQF #3663e. It first discussed how these three measures relate to each other. The Standing Committee noted that measurement at the facility represents a very accurate reflection of radiology practice and structures; however, additional information is needed from widespread use of the measure to make a final determination on this matter. The Standing Committee questioned whether the facility level may sufficiently capture the necessary quality data and might alone be sufficient and whether the individual clinician and group measures might be combined and then harmonized with the facility-level measure, thus creating two total measures. The developer stated that each measure captures an important component of responsibility and care quality, and one cannot be prioritized over the other two. The developer added that no additional work is needed to assemble the data between the various levels of analysis addressed by these three measures; therefore, they permit attribution at different levels using the same amount of effort. The Standing Committee acknowledged these comments and requested that when bringing the measures back for maintenance review, the developer should examine whether these measures could be further harmonized or combined as they review the real-world data. The Standing Committee reviewed two additional related measures to NQF #3633e, NQF #3662e, and NQF #3663e: NQF #2820 *Pediatric Computed Tomography (CT) Radiation Dose* and NQF #3621 *Composite Weighted Average for CT Exam Types*. The Standing Committee asked whether NQF #2820 could be incorporated into the three adult radiology measures. The developer shared their plans to update NQF #2820 after further research on quality thresholds for pediatric patients and to move towards a second-generation eCQM version of NQF #2820 rather than a claims-based measure. The Standing Committee had no additional comments about NQF #3621.

References

1 Patient Safety. https://www.who.int/news-room/fact-sheets/detail/patient-safety. Last accessed March 2022.

2 Rodwin BA, Bilan VP, Merchant NB, et al. Rate of Preventable Mortality in Hospitalized Patients: a Systematic Review and Meta-analysis. *J Gen Intern Med*. 2020;35(7):2099-2106.

3 Gill LE, Bartels SJ, Batsis JA, et al. Weight Management in Older Adults. *Curr Obes Rep*. 2015;4(3):379-388.

4 Wells JL, Dumbrell AC. Nutrition and aging: assessment and treatment of compromised nutritional status in frail elderly patients. *Clin Interv Aging*. 2006;1(1):67-79.

5 Allepaerts S, De Flines J, Paquot N, et al. Nutrition in the elderly. *Rev Med Liege*. 2014;69(5-6):244-250.

6 Kergoat MJ. Weight loss in elderly residents in nursing homes and extended care facilities. *Rev Med Suisse Romande*. 2000;120(11):853-861.

7 Gabrovec B, Veninšek G, Samaniego LL, et al. The role of nutrition in ageing: A narrative review from the perspective of the European joint action on frailty - ADVANTAGE JA. *Eur J Intern Med*. 2018;56:26-32.

8 World Health Organization (WHO). Table 4: Summary of WHO Position Papers – Immunization of Health Care Workers. September 2020. https://www.who.int/immunization/policy/Immunization_routine_table4.pdf. Last accessed March 2022.

9 Recommended Vaccines for Healthcare Workers | CDC. https://www.cdc.gov/vaccines/adults/rec-vac/hcw.html. Published August 4, 2021. Last accessed March 2022.

10 COVID-19 advice for the public: Getting vaccinated. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines/advice. Last accessed March 2022.

11 WHO's COVID-19 guidance and information for Health workers and administrators. https://www.who.int/teams/risk-communication/health-workers-and-administrators. Last accessed March 2022.

12 Little MP, Wakeford R, Borrego D, et al. Leukaemia and myeloid malignancy among prospectively analysed cohorts of persons exposed to low doses (<100 mSv) of ionising radiation at ages under 21 years. *Lancet Haematol*. 2018;5(8):e346-e358.

13 Brenner DJ, Doll R, Goodhead DT, et al. Cancer risks attributable to low doses of ionizing radiation: Assessing what we really know. *Proc Natl Acad Sci U S A*. 2003;100(24):13761-13766.

14 Patel SA, Qureshi MM, Sahni D, et al. Identifying an Optimal Adjuvant Radiotherapy Dose for Extremity and Trunk Merkel Cell Carcinoma Following Resection. *JAMA Dermatol*. 2017;153(10):1007-1014.

Appendix A: Details of Measure Evaluation

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable

NQF ensures that quorum is maintained for all live voting. Quorum is 66 percent of active Standing Committee members minus any recused Standing Committee members. Due to the exclusion of recused Standing Committee members from the quorum calculation, the required quorum for live voting may vary among measures. Quorum (16 out of 23 Standing Committee members for NQF #0689 and NQF #3636 and 15 out of 22 Standing Committee members for NQF #3633e, NQF #3662e, and NQF #3663e) was reached and maintained during the full measure evaluation meeting on February 16, 2022. Vote totals may differ between measure criteria and between measures as Standing Committee members may have joined the meeting late, stepped away for a portion of the meeting, or had to leave the meeting before voting was complete. The vote totals listed below reflect Standing Committee members present and eligible to vote at the time of the vote. Voting results are provided below.

A measure is recommended for endorsement by the Standing Committee when greater than 60 percent of voting members select a passing vote option (Pass, High and Moderate, Yes) on all must-pass criteria and overall suitability for endorsement. A measure is not recommended for endorsement when less than 40 percent of voting members select a passing vote option on any must-pass criteria or overall suitability for endorsement. If a measure does not pass a must-pass criterion, voting during the measure evaluation meeting stops. The Standing Committee does not re-vote on the measures during the post-comment meeting unless the Standing Committee decides to reconsider the measure(s) based on submitted comments or a formal reconsideration request from the developer. During the measure evaluation meeting, the Standing Committee has not reached consensus on a measure if between 40 and 60 percent of voting members select a passing vote option on any must-pass criteria or overall suitability for endorsement. The Standing Committee will re-vote on criteria that did not reach consensus and potentially overall suitability for endorsement during the post-comment web meeting.

Measures Recommended

NQF #0689 Percent of Residents Who Lose Too Much Weight (Long-Stay)

Measure Worksheet | Specifications

Description: This measure captures the percentage of long-stay nursing home residents with a target Minimum Data Set (MDS) 3.0 assessment (OBRA, PPS, or discharge) that indicates a weight loss of 5% or more of the baseline weight in the last 30 days, or 10% or more of the baseline weight in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen. The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment. Long-stay nursing facility residents are identified as those who have had 101 or more cumulative days of nursing facility care.

Numerator Statement: The numerator is the number of long-stay nursing home residents with a selected target assessment indicating a weight loss of 5% or more of the baseline weight in the last 30 days or 10% or more of the baseline weight in the last 6 months who were not on a physician-prescribed weight-loss regimen (K0300 = [2]). The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment.

Denominator Statement: The denominator includes all long-stay residents in the nursing home who have a target assessment (OBRA, PPS or discharge) during the selected quarter and who do not meet the exclusion criteria.

Exclusions: There are four exclusions applied to the denominator: (1) the target assessment is an OBRA admission assessment (A0310A = [01]) or a PPS 5-day assessment (A0310B = [01]), (2) having a prognosis of life expectancy of less than six months (J1400 = [1]) or the six-month prognosis item is missing (J1400 = [-]) on the target assessment,

(3) receiving hospice care (O0100K2 = [1]) or the hospice care item is missing (O0100K2 = [-]) on the target assessment, or/and (4) the weight loss item is missing (K0300 = [-]) on the target assessment. Only 1,551 episodes in the 2019 (Q1-Q4) long stay resident sample were excluded from the denominator for this measure due to missing responses on the prognosis of life expectancy being less than 6 months, which accounts for 0.04% of the total episodes. Additionally, only 7,948 (0.241%) episodes in the 2019 (Q1-Q4) long stay residents sample were excluded due to missing responses for the Hospice care item, and only 30,854 (0.935%) episodes were excluded due to missing responses for the weight loss item. f the facility sample includes fewer than 20 residents after all other resident-level exclusions are applied, then the facility is suppressed from public reporting because of small sample size.

Adjustment/Stratification: No risk adjustment or stratification

Level of Analysis: Facility

Setting of Care: Post-Acute Care

Type of Measure: Outcome

Data Source: Assessment Data, Electronic Health Records: Electronic Health Records

Measure Steward: Centers for Medicare & Medicaid Services

STANDING COMMITTEE MEETING [February 16, 2022]

1. Importance to Measure and Report:

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Total Votes-20; Pass-18; No Pass-2**; 1b. Performance Gap: **Total Votes-19; H-4; M-13; L-2; I-0**

Rationale:

- The Standing Committee noted that the developer provided updated evidence in support of the measure, including new evidence on several actions that nursing home staff and facilities can take to prevent unintended weight loss.
- The Standing Committee recommended that the measure developer include additional evidence in their next submission regarding whether the full list of risks associated with weight loss either are or are not modifiable by facilities. Ultimately, the Standing Committee had no immediate concerns and passed the measure on evidence.
- The Standing Committee noted that the mean performance was 5.2 percent with a standard deviation of 3.1 percent and range of 1.6 percent to 9.2 percent. The developer also noted that the interquartile range (IQR) of 3.9 percent and the small number of facilities with "perfect" scores (2.6 percent) indicate room for improvement.
- The Standing Committee noted that patients over the age of 85 and those who are White had a slightly higher risk of losing too much weight.
- The Standing Committee passed the measure on performance gap and also suggested that the developer present a stratified analysis of the measure scores for consideration in future reviews, such as by facility characteristics/types, disease areas, and different subpopulations of interest.

2. Scientific Acceptability of Measure Properties:

(2a. Reliability – precise specifications, testing; 2b. Validity – testing, threats to validity)

2a. Reliability: **Total Votes-19; Yes-19; No-0**; 2b. Validity: **Total Votes-20; H-1; M-15; L-3; I-1**

Rationale:

- The SMP passed this measure on reliability but did not reach consensus on validity, expressing concern with the decision not to risk-adjust and the correlations not being strong enough to demonstrate validity. The SMP asked the Standing Committee to further discuss this matter.
- The Standing Committee noted that in the 2015 submission, the developer reported that the kappa for gold-standard nurse assessment to facility nurse assessment of weight loss item was 0.918. The kappa for gold-standard nurse assessment to facility nurse assessment of the six-month prognosis item was 0.964.

- The Standing Committee noted that for accountable-entity level validity testing, the developer hypothesized a number of relationships to test the measure. Tests of convergent validity and variation by state, seasonality, stability analysis, and confidence interval analysis were run to demonstrate the validity of the measure. The Standing Committee noted that the developer reported statistically significant negative correlations between NQF #0689 *Percent of Residents Who Lose Too Much Weight* and the following: Overall Facility Five-Star Ratings (0.108), Quality Ratings (0.143), Staffing Ratings (0.029), and Registered Nurse Staffing Ratings (0.011), as expected.
- The proportion of variation explained by the state in which the facilities are located was small but statistically significant ($p < 0.001$).
- The Standing Committee discussed whether certain MDS items might warrant a risk adjustment strategy after noting the SMP's concerns. The developer reported that they reviewed the suggested variables and observed low to moderate correlations between diagnostic options on the MDS and weight loss but none that changed the facility's measure score or rank; therefore, the measure was intentionally not risk-adjusted.

3. Feasibility: Total Votes-20; H-15; M-5; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- The Standing Committee noted that the measure relies on data from the MDS 3.0, which is mandatory for all Medicare- and Medicaid-certified nursing facilities.
- The Standing Committee noted that all data are generated during the provision of care, and all data elements are in defined fields in electronic clinical data.
- The Standing Committee noted that 1.216 percent of data were missing from episodes in 2019, and these were excluded from the denominator. The missingness did not warrant concern with regard to the feasibility or bias of the measure.

4. Usability and Use:

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Total Votes-19; Pass-19**; 4b. Usability: **Total Votes-19; H-10; M-9; L-0; I-0**

Rationale:

- The Standing Committee noted that the measure is publicly reported to both measured facilities and the public via the following: Care Compare, Provider Data Catalog, Certification and Survey Provider Enhanced Reports (CASPER). The developer did not report plans to use the measure in other accountability programs.
- The Standing Committee noted that the measure developer analyzed inquiries submitted to their support inbox. The developer reported that they have not received any feedback or concerns from those being measured, measure users, or implementers since October 2019.
- The Standing Committee observed that decreasing scores over time demonstrated an improvement in the quality of care.
- The Standing Committee noted that based on the literature, it was unexpected that White residents were at greater risk of unintended weight loss than non-White residents. The developer conducted testing to assess whether this unexpected result was due to differences in quality of care, chance, or another explanatory factor and concluded that age explained the difference.

5. Related and Competing Measures

- No related or competing measures were noted.

6. Standing Committee Recommendation for Endorsement: Total Votes: 19; Yes-18; No-1

7. Public and Member Comment

- No public comments were received for this measure.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision

9. Appeals

NQF #3636 Quarterly Reporting of COVID-19 Vaccination Coverage Among Healthcare Personnel

Measure Worksheet | Specifications

Description: This quarterly measure identifies the average percentage of healthcare personnel (HCP) who have ever received a primary COVID-19 vaccination course among the total number of HCP who regularly work in the facility.The measure is reported for a quarter (3-month period). The quarterly COVID-19 vaccination coverage is determined by selecting one week per month and calculating the percentage of HCP who have ever received a primary COVID-19 vaccination course, then averaging 3 weekly percentages (one week from each of the 3 months in the quarter).

Numerator Statement: The numerator for this measure consists of the cumulative number of HCP in the denominator population, who:

1. have received a complete vaccination course against COVID-19 administered at the healthcare facility; or

2. reported in writing (paper or electronic) or provided documentation that a complete vaccination course against COVID-19 was received elsewhere

Denominator Statement: The target population is the number of healthcare personnel (HCP) eligible to work in the healthcare facility for at least one day during the one-week data collection reporting period, excluding persons with contraindications/exclusions to COVID-19 vaccination. The quarterly reported measure includes at least one week of data collection a month for each of the 3 months in a quarter.

The denominators are reported by aggregating categories below:

1. Employees: all persons who receive a direct paycheck from the reporting facility (i.e., on the facility's payroll).

2. Licensed independent practitioners: include physicians (MD, DO), advanced practice nurses, and physician assistants only who are affiliated with the reporting facility who do not receive a direct paycheck from the reporting facility.

3. Adult students/trainees and volunteers include all students/trainees and volunteers aged 18 or over who do not receive a direct paycheck from the reporting facility.

4. Other contract personnel: Facilities may also report on individuals who are contract personnel. However, reporting for this category is optional. Contract personnel are defined as persons providing care, treatment, or services at the facility through contract who do not fall into any of the above-mentioned denominator categories.

Exclusions: Exclusions include individuals with contraindications to COVID-19 vaccination and individuals for whom the COVID-19 vaccine is not authorized or recommended.

Adjustment/Stratification: No risk adjustment or stratification

Level of Analysis: Facility

Setting of Care: Post-Acute Care

Type of Measure: Process

Data Source: Other (specify)

Measure Steward: Surveillance Branch, Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention

STANDING COMMITTEE MEETING [February 16, 2022]

1. Importance to Measure and Report:

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Votes-18; H-N/A; M-12; L-0; I-6**; 1b. Performance Gap: **Total Votes-18; H-11; M-6; L-1; I-0**

Rationale:

- The Standing Committee noted that that evidence for this measure derives from the Advisory Committee on Immunization Practices' (ACIP) recommendations for allocation of COVID-19 vaccines as presented to the Director of the Centers for Disease Control and Prevention (CDC).
- The ACIP COVID-19 Vaccines Workgroup considered evidence related to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemiology, vaccination program implementation, and ethical principles in developing the interim recommendation on the allocation of the initial supply of COVID-19 vaccines (Phase 1a of vaccine distribution).
- The Standing Committee discussed whether the evidence provided in the submission supported the contention that measuring COVID-19 vaccination among HCP would lead to an increase in vaccination rates, and ultimately, a decrease in cases.
- The developer replied that while systematic reviews of evidence of the surrounding vaccinations of HCP for COVID-19 were not yet available at the time of measure submission, several studies have since been published showing a decrease in case rates in facilities that had high vaccination rates of HCP and that the reporting of vaccination rates at a facility had an impact on those rates.
- The Standing Committee raised a concern that members of the public might inappropriately equate low COVID-19 vaccination rates at a facility with poor quality of care at that facility; ultimately, it acknowledged that while the quality of care provided might be otherwise good, the vaccination status of HCP at that facility also has the potential to impact the patients cared for at the facility and should be public knowledge for evaluating care facilities.
- The Standing Committee stated that the evidence the developer provided was sound, especially considering it was gathered amid an emerging global pandemic, and passed the measure on evidence.
- The Standing Committee noted lower COVID-19 vaccination coverage rates among certain HCP categories (i.e., nurses and aides) and among facilities located in zip codes with indicators of social vulnerability.
- Other research has identified lower vaccination coverage among nurses and support staff and among Black and Hispanic HCP as well as higher vaccination acceptance among doctoral-degree personnel. Various studies have found decreased likelihood of vaccine acceptance among HCP identified as Black, Latinx, female, or having lower educational attainment.
- The Standing Committee had no concerns and passed the measure on performance gap.

2. Scientific Acceptability of Measure Properties:

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Total Votes-18; H-N/A; M-15; L-2; I-1**; 2b. Validity: **Total Votes-18; H-8; M-10; L-0; I-0**

Rationale:

- The Standing Committee noted that the developer conducted reliability testing at the patient/encounter level, and the overall Pearson correlation coefficient for the number of HCP who received COVID-19 vaccinations as reported to the NHSN (measure numerator) compared to the number of COVID-19 vaccinations administered by the Pharmacy Partnership for Long-Term Care Program (PPP) (independent comparator) was 0.846 ($p<0.0001$ [869 Facilities]). The developer stated that this correlation is both linear and high, showing that the numerator is strongly associated with the data from the independent comparator.
- The Standing Committee passed the measure on reliability.
- The Standing Committee expressed some concerns with the optional reporting category of contract personnel included in the denominator, stating that it seems facilities would report this category when it improves their score and not report it when it does not. The developer clarified that the denominator was created to mirror the denominator of NQF #0431, the currently NQF-endorsed influenza vaccination of

HCP measure, which also does not require the reporting of contract personnel. The Standing Committee stressed that contract personnel have become a much greater percentage of HCP since the pandemic began and urged the developer to consider making this reporting category a requirement in the future.

- The Standing Committee noted that the developer conducted validity testing at the accountable-entity level. The overall Pearson correlation coefficient between the quarterly COVID-19 coverage measure for Q3 2021 and annual influenza vaccination coverage (NQF #0431) was 0.4169 ($p<0.0001$ [1,654 facilities]), indicating a "medium" correlation using the generally accepted range for medium correlation: 0.30–0.49.
- The Standing Committee also noted that the data presented represent a medium correlation when stratified by facility size (0.457 for the third quartile [94-131 HCP] and 0.450 for the fourth quartile [>132 HCP]).
- The Standing Committee had no concerns regarding the validity testing of the measure or how the developer addressed any potential threats to validity and passed the measure on this criterion.

3. Feasibility: Total Votes-17; H-10; M-7; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- The Standing Committee noted that the data source is not specified because it may vary by facility. Data may be collected from electronic sources or paper-based sources, or it may be obtained from existing records or a system specifically designed for COVID-19 vaccination tracking. The data are then reported to the NHSN.
- The Standing Committee discussed whether collecting data for this measure was more feasible amid the pandemic when it was critically relevant and whether it would pose a reporting burden at a later date when the threat may have waned. The developer explained that they chose quarterly reporting to mitigate extremes and make reporting less burdensome than weekly, which is the current practice among many institutions, but more immediately useful than annually.
- The Standing Committee passed the measure on feasibility.

4. Usability and Use:

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Total Votes-16; Pass-16; No Pass-0**; 4b. Usability: **Total Votes-17; H-8; M-8; L-0; I-1**

Rationale:

- The Standing Committee noted that this measure is currently in use in public reporting, public health/disease surveillance, and regulatory and accreditation programs.
- The Standing Committee noted that this measure was submitted to the Measure Applications Partnership (MAP) for 2020-2021 consideration for implementing measures in federal programs. MAP offered conditional support for rulemaking for this measure and encouraged the developer to fully specify the measure as soon as possible.
- The Standing Committee expressed no concerns and passed the measure on use and usability.

5. Related and Competing Measures

- This measure is related to the following measures:
    - NQF #0431 Influenza Vaccination Coverage Among Healthcare Personnel
        - The Standing Committee noted that the denominator for NQF #3636 was harmonized to mirror that of NQF #0431; however, the data collection time frame for each measure is different. The Standing Committee acknowledged that not enough information is known yet about the potential seasonality of COVID-19 infections to make any additional recommendations at this time.

6. Standing Committee Recommendation for Endorsement: **Total Votes: 17; Yes-16; No-1**

7. Public and Member Comment

- The measure developer submitted a public comment summarizing new peer-reviewed evidence and systematic literature reviews that have been published since the measure was submitted that support this measure.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision

9. Appeals

NQF #3633e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level)

Measure Worksheet | Specifications

Description: This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible.

Numerator Statement: Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

Denominator Statement: All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

Exclusions: Denominator exclusions are CT exams that simultaneously include multiple body regions outside of four commonly encountered multiple region groupings (specified as LOINC code 96914-7, CT Dose and Image Quality Category, Full Body). Denominator exclusions are also CT exams with missing patient age, missing size-adjusted radiation dose, or missing global noise. These are technical exclusions (“missing data”) from the initial population. Technical exclusions will be flagged, corrected whenever possible, and tracked at the level of the accountable entity.

Adjustment/Stratification: Stratification by risk category (specify number of categories), Statistical risk model with risk factors (specify number of risk factors)

The means by which a CT examination is determined to be “out-of-range" with respect to radiation dose is measured by observing whether its patient size-adjusted radiation dose exceeds a pre-determined evidence-based threshold. The value of this size-adjusted radiation dose is calculated with the following equation for any given exam:

$D[A] = D[R] * \exp(-(d-d[k]) * \beta[k])$

Where...

D[A]is the size-adjusted radiation dose of the exam

D[R]is the radiation dose of the exam, without adjustment

d is the diameter of the anatomic area being examined

d[k] is the “expected diameter” of the CT category associated with the exam. This “expected diameter” is equal to the median diameter of all exams associated with the CT category in the UCSF International CT Dose Registry containing 6.5 million exams from 161 institutions.

β[k] is the “size-adjustment coefficient” of the CT category associated with the exam. This “size-adjustment coefficient” is the slope parameter of a collection of log-transformed linear regression models fit using the UCSF Registry. A total of 18 models were fit, each using data from one of the CT Dose and Image Quality Categories. The models are parametrized such that, in the kth model and associated dataset, for the jth observation, from the ith hospital, we define: log({D[R]}[ij]) = {β[0]}[k] + β[k] * d[ij] + {z[i]}[k] + ε[ij]

Where D[R] and d are respectively the radiation dose without adjustment and diameter of the anatomic area being examined, β[0]is an intercept term, z is a random effect indicating variation due to the hospital at which the exam was performed, and ε is the residual variation. We restrict the value of β[k]to be greater than 0; when it is less than 0, it is set to 0 and no adjustment is performed. For the estimated values of β[k] across CT categories (strata), please see 2b.30 below.

The intended interpretation of D[A]is the “expected radiation dose of the exam if the diameter of the anatomic area being examined were equal to the population-level median.”

Level of Analysis: Clinician: Individual

Setting of Care: Outpatient Services, Inpatient/Hospital, Ambulatory Care

Type of Measure: Outcome: Intermediate Clinical Outcome

Data Source: Electronic Health Records, Electronic Health Data

Measure Steward: Alara Imaging

STANDING COMMITTEE MEETING [February 16, 2022]

1. Importance to Measure and Report:

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Total Votes-17; H-1; M-11; L-3; I-2**; 1b. Performance Gap: **Total Votes-17; H-7; M-9; L-1; I-0**

Rationale:

- The Standing Committee noted that the developer cited two systematic reviews and two studies in support of the measure but expressed concern that much of the data in the studies came from a pediatric population. The developer replied that while the systematic reviews of radiation dosing and CT scans all focus on children, there are many papers that focus on adults, which similarly shows that patients with an increased exposure to CT scans have an increased risk of developing cancer.
- The Standing Committee noted that this measure was tested in seven health systems and one vertically integrated organization, including 42,493 CT exams interpreted by 606 physicians between 2020 and 2021. The mean performance score was 30 percent, with a standard deviation of 21 percent and a range of 0–100 percent.
- The Standing Committee noted that the developer also examined differences based on age and sex and found minimal variation.
- The Standing Committee noted that studies have found that most social risk factors are not predictive of radiation dose for CT exams; however, patients living in poverty are at higher risk for comorbid conditions associated with exposure to multiple scans over time and increased cumulative exposure to ionizing radiation from diagnostic imaging.
- The Standing Committee passed the measure on evidence and performance gap.

2. Scientific Acceptability of Measure Properties:

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Total Votes-18; Yes-18; No-0**; 2b. Validity: **Total Votes-17; Yes-14; No-3**

Rationale:

- This measure was deemed as complex and was evaluated by the SMP.

- The Standing Committee noted that the developer conducted a signal-to-noise analysis using EHRs from 606 clinicians within seven health systems and one vertically integrated organization from February 2020 to April 2021.
- The number of exams per clinician in the one month of data used for testing ranged from 1 to 604 (mean=77); the predicted reliability for 12 months exceeded 0.90 for 89 percent of participating clinicians.
- The estimated mean split-half intraclass correlation coefficient (ICC) using 47,635 CT exams collected from 606 individual clinicians was 0.99 (following the exclusion of clinicians who read only one scan during the test month and a Spearman-Brown adjustment to a 12-month data collection period).
- The Standing Committee asked for clarity on the potential impact of the reliability thresholds of the number of scans needed to reach reliability on the measure and how many clinicians might be excluded as a result. The developer stated that the vast majority of radiologists perform at least the minimum number of scans (i.e., 28); therefore, the threshold results in the exclusion of very few radiologists from the measure.
- The Standing Committee accepted the SMP's rating of high for reliability (**Total SMP Votes-11; H-9; M-2; L-0; I-0**).
- The Standing Committee noted that validity testing was conducted at the patient/encounter level and that the developer examined CT category, patient size, radiation dose, size-adjusted radiated dose, global noise, and thresholds for "out-of-range" values to define the numerator. The results, weighted by the distribution of CT categories in the University of California, San Francisco (UCSF) International CT Dose Registry, showed a sensitivity of 0.86 and a specificity of 0.96 (n=978 CT exams). When tested across the 606 individual clinicians, the correct classification rate of the assignment of CT exams to CT category in field-testing was 95 percent on average.
- The Standing Committee also noted that validity testing was conducted at the accountable-entity level. The eCQM was compared against the medical record review using field-testing data collected from eight health systems/vertically integrated organizations. The Standing Committee concluded that the results indicate a correct and robust implementation of the measure logic.
- The Standing Committee noted that the developer conducted face validity testing, and 94–100 percent agreed that implementation of the measure in federal programs would lead to a reduction in average CT radiation dose while maintaining adequate CT image quality.
- The Standing Committee noted that the SMP expressed concerns about missing data only focusing on the "radiation dose" aspect of the measure.
- The Standing Committee discussed whether the developer had considered additional clinical care factors outside of BMI that might affect dosing; the developer replied that no other factors had a strong impact on the measure, and none were significant enough to warrant risk adjustment. The developer also addressed the possibility of misclassification in creating their dosing strata within CT categories but stated that they worked extensively with clinicians and radiologists to understand the dosing needs for various types of patients and erred on the side of allowing for the possibility of a higher dose when they were asked to by clinical experts. The Standing Committee also asked whether any validity issues emerged based on provider attribution and specifically asked the developer how attribution is assigned. The developer stated that for the current measure at the individual-clinician level, it is the radiologist who bills for the exam and who is responsible for quality.
- The Standing Committee accepted the SMP's rating of moderate for validity (**Total SMP Votes-11; H-5; M-6; L-0; I-0**).

3. Feasibility: Total Votes-17; H-13; M-3; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- The Standing Committee noted that data elements for this measure are in defined fields in a combination of electronic sources generated or collected by and used by HCP during the provision of care.

- The Standing Committee noted that the Feasibility Scorecard indicated that no data elements have issues with accuracy, and 100 percent coverage was achieved in simulated data unit tests.
- The Standing Committee questioned what the effect might be of having only one vendor who can pull these data (i.e., Alara Imaging). The developer replied that they created this vendor organization to respond to a request from CMS to manage nationwide implementation and reporting because no other alternative had presented itself. The developer added that the measure specifications are publicly available, and all collected data are already in the EHR, billing claims, or other frequently used data systems. Therefore, the fact that there is currently only one vendor who can report this measure does not preclude other vendors from also doing so. In addition, clinicians and hospitals can report on the measure at no cost using a web interface. The Standing Committee passed the measure on feasibility.

4. Usability and Use:

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Total Votes-18; Pass-17; No Pass-1**; 4b. Usability: **Total Votes-18; H-2; M-15; L-1; I-0**

Rationale:

- The Standing Committee noted that this measure is not currently in use in any quality improvement or accountability programs.
- The developer stated that this measure will be submitted to the CMS Merit-based Incentive Payment System (MIPS). MIPS measures are publicly reported on Care Compare.
- The Standing Committee noted that although the measure is scored and reported at the aggregate level, users requested that feedback be more nuanced to make that feedback actionable.
- One unexpected finding was the lack of consistency among facilities saving Radiation Dose Structured Reports (RDSRs). The developer worked with sites to modify their systems to save the RDSR to capture 94 percent of dose reports.
- Because the goal of this measure is to reduce patient exposure to radiation, the developer noted a concern that radiation dose reduction might result in deteriorated image quality but did not find any evidence of poor image quality in the results. The developer stated that this potential issue will be monitored annually.
- The Standing Committee asked how frequent the need was for additional scans due to low quality. The developer replied that in a quality study using a sample of 700+ scans, which included an overrepresentation of low-dose scans (in which poor image quality would be most likely), only 3 percent were considered low quality, and another 8 percent were considered moderate quality but still unacceptable.
- The developer noted the need to pay attention to this issue as a possible unintended consequence of encouraging lower-dose scans; they intend to monitor it closely once the measure is implemented and adjust the thresholds if needed.
- The Standing Committee passed the measure on use and usability.

5. Related and Competing Measures

- This measure is related to the following measures:
    - NQF #2820 Pediatric Computed Tomography Radiation Dose
        - The Standing Committee asked whether this measure could be incorporated into the three adult radiology measures. The developer shared their plans to update NQF #2820 to further harmonize it with the current measures for future maintenance reviews.
    - NQF #3621 Composite Weighted Average for 3 CT Exam Types: Overall Percent of CT Exams for Which Dose Length Product Is at or Below the Size-Specific Diagnostic Level
        - The Standing Committee had no additional comments about NQF #3621.
    - NQF #3662e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level) (Alara Imaging/UCSF)

- NQF #3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Facility Level) (Alara Imaging/UCSF)
    - The Standing Committee held a discussion about how NQF #3633e, NQF #3662e, and NQF #3663e relate to each other.
    - The Standing Committee questioned whether the facility level sufficiently captured the necessary quality data and might alone be sufficient or whether the individual-clinician and group measures might be combined and then harmonized with the facility-level measure, thus creating two total measures. The developer stated that each measure captures an important component of responsibility and care quality, and one cannot be prioritized over the other two. The developer added that no additional work is needed to assemble the data between the various levels of analysis addressed by these three measures; therefore, they permit attribution at different levels using the same amount of effort.
    - The Standing Committee acknowledged these comments and requested that the developer continue to examine whether these measures could be further harmonized or combined as they review the real-world data collected before the measures undergo maintenance review.

6. Standing Committee Recommendation for Endorsement: **Total Votes: 19; Yes-15; No-4**

7. Public and Member Comment

- Twenty-five pre-evaluation comments were received for this measure.
    - Seventeen comments were in support of this measure.
        - Comments stated that this measure will meaningfully improve physicians' abilities to monitor the equipment used in these scans, increase their quality, and reduce dose variability, which should lead to a decline in cumulative radiation dose.
        - Comments stated the strength of this measure: It is based on the clinical indication for imaging rather than the type of examination a radiologist chooses to perform.
        - Comments stated that the measure was highly feasible: There were few barriers to the successful implementation of the measure and very little missing data.
        - Comments stated the importance to patients that providers use the lowest-appropriate dose for specific diagnostic or follow-up exams.
        - Comments stated that this measure can reduce not only excessive, high doses, but also suboptimal low doses by identifying outliers and increasing awareness of protocol optimization.
        - Comments stated that the measure feedback is actionable, and users have been very satisfied with the feedback they have received on their measure performance.
    - Four comments were not in support of this measure.
        - Comments expressed concerns that this measure conflates the choice of protocol for the clinical indication with radiation dose optimization, thus making improvement on the measure more challenging.
        - Comments expressed concerns with the assessment of image quality, radiation risk, subjectivity, patient size, and image rendition; underaddressing exam components and exam diversity; and not providing sufficient guidance for compliance regarding outlier exams.
        - Comments expressed concern that the specifications for the measure have not been validated, specifically the method of determining the classification of dosing studies.
        - Comments expressed concern that the measure deviates from international standards for diagnostic reference levels and lacks consensus on defining global noise.

- Comments expressed concern for the unintended consequences of using too low a dose and possibly missing a disease diagnosis, also resulting in a “wasted dose with no medical benefit.”

○ The measure developer submitted four comments specifically addressing the concerns submitted in the pre-evaluation public comments.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision

9. Appeals

NQF #3662e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level)

Measure Worksheet | Specifications

Description: This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible.

Numerator Statement: Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.;

Denominator Statement: All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value

Exclusions: Denominator exclusions are CT exams that simultaneously include multiple body regions outside of four commonly encountered multiple region groupings (specified as LOINC code 96914-7, CT Dose and Image Quality Category, Full Body). Denominator exclusions are also CT exams with missing patient age, missing size-adjusted radiation dose, or missing global noise. These are technical exclusions (“missing data”) from the initial population. Technical exclusions will be flagged, corrected whenever possible, and tracked at the level of the accountable entity.

Adjustment/Stratification: None

Statistical risk model with risk factors (specify number of risk factors), Stratification by risk category (specify number of categories)

The means by which a CT examination is determined to be “out-of-range" with respect to radiation dose is measured by observing whether its patient size-adjusted radiation dose exceeds a pre-determined evidence-based threshold. The value of this size-adjusted radiation dose is calculated with the following equation for any given exam: $D[A] = D[R] * \exp(-(d-d[k]) * \beta[k])$

Where...

D[A]is the size-adjusted radiation dose of the exam

D[R]is the radiation dose of the exam, without adjustment

d is the diameter of the anatomic area being examined

d[k] is the “expected diameter” of the CT category associated with the exam. This “expected diameter” is equal to the median diameter of all exams associated with the CT category in the UCSF International CT Dose Registry containing 6.5 million exams from 161 institutions.

β[k] is the "size-adjustment coefficient" of the CT category associated with the exam. This "size-adjustment coefficient" is the slope parameter of a collection of log-transformed linear regression models fit using the UCSF Registry. A total of 18 models were fit, each using data from one of the CT Dose and Image Quality Categories. The models are parametrized such that, in the kth model and associated dataset, for the jth observation, from the ith hospital, we define: log({D[R]}[ij]) = {β[0]}[k] + β[k] * d[ij] + {z[i]}[k] + ε[ij]

Where D[R] and d are respectively the radiation dose without adjustment and diameter of the anatomic area being examined, β[0]is an intercept term, z is a random effect indicating variation due to the hospital at which the exam was performed, and ε is the residual variation. We restrict the value of β[k]to be greater than 0; when it is less than 0, it is set to 0 and no adjustment is performed. For the estimated values of β[k] across CT categories (strata), please see 2b.30 below. The intended interpretation of D[A]is the "expected radiation dose of the exam if the diameter of the anatomic area being examined were equal to the population-level median."

Level of Analysis: Clinician: Group/Practice

Setting of Care: Ambulatory Care, Inpatient/Hospital, Outpatient Services

Type of Measure: Outcome: Intermediate Clinical Outcome

Data Source: Electronic Health Data, Electronic Health Records

Measure Steward: Alara Imaging

STANDING COMMITTEE MEETING [February 16, 2022]

1. Importance to Measure and Report:

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Total Votes-18; H-0; M-16; L-0; I-2**; 1b. Performance Gap: **Total Votes-18; H-8; M-10; L-0; I-0**

Rationale:

- The Standing Committee noted that the evidence and performance gap for this measure are identical to NQF #3633e and passed the measure on both criteria.

2. Scientific Acceptability of Measure Properties:

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Total Votes-18; Yes-18; No-0**; 2b. Validity: **Total Votes-19; Yes-16; No-3**

Rationale:

- This measure was deemed complex and was evaluated by the SMP.
- The Standing Committee noted that a signal-to-noise analysis was conducted using EHRs from 16 groups within seven health systems and one vertically integrated organization from February 2020 to April 2021.
- The clinician groups ranged in size from 31 to 109 physicians (mean=27). The number of exams per clinician group in the one month of data used for testing ranged from 56 to 14,312 (mean=3,031).
- The estimated mean split-half ICC using 48,500 CT exams was 0.99 (after a Spearman-Brown adjustment to a 12-month data collection period).
- The developer stated that a minimum of 28 CT exams are required to achieve 90 percent reliability based on this method.
- The clinician groups ranged in size from 31 to 109 physicians (mean=27). The number of exams per clinician group in the one month of data used for testing ranged from 56 to 14,312 (mean=3,031). The estimated mean split-half ICC using 48,500 CT exams collected from 606 individual clinicians was 0.99 (following the exclusion of clinicians who read only one scan in the test month and a Spearman-Brown adjustment to a 12-month data collection period).
- The Standing Committee asked for confirmation of whether the threshold for the number of scans performed to achieve sufficient reliability for inclusion in the measure was the same for the clinician group as it was for the individual-clinician level, which the developer confirmed. At the group level, this would exclude very few, if any, practices.
- The Standing Committee accepted the SMP's rating of high for reliability (**Total SMP Votes: 11; H-8; M-3; L-0; I-0**).

- The Standing Committee noted that validity testing was conducted at the patient/encounter level and that the developer examined CT category, patient size, radiation dose, size-adjusted radiated dose, global noise, and thresholds for "out-of-range" values to define the numerator. An International Classification of Diseases, Tenth Revision (ICD-10)-based algorithm to assign the CT category was compared to chart review as the gold standard. The results, weighted by the distribution of CT categories in the UCSF International CT Dose Registry, showed a sensitivity of 0.86 and a specificity of 0.96 (n=978 CT exams). When tested across the 16 clinician groups, the correct classification rate of the assignment of CT exams to CT category in field-testing was 92 percent on average and varied from 88–97 percent across the 16 clinician groups.
- The Standing Committee also noted that validity testing was conducted at the accountable-entity level. The eCQM was compared against the medical record review at the accountable-entity level using field-testing data collected from eight health systems/vertically integrated organizations. The Standing Committee concluded that the results indicate a correct and robust implementation of the measure logic.
- The Standing Committee noted that the developer conducted face validity testing, and 94–100 percent agreed that implementation of the measure in federal programs would lead to a reduction in average CT radiation dose while maintaining adequate CT image quality.
- The Standing Committee questioned whether there might be some attribution concerns that persist at the group level or whether the group level mitigated most of the concerns it had with attribution at the clinician level.
- The Standing Committee accepted the SMP's rating of moderate for validity (**Total SMP Votes-11; H-7; M-4; L-0; I-0**).

3. Feasibility: Total Votes-18; H-11; M-7; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- The Standing Committee noted that the feasibility for this measure is identical to NQF #3633e and passed the measure on this criterion.

4. Usability and Use:

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Total Votes-19; Pass-18; No Pass-1**; 4b. Usability: **Total Votes: 18; H-2; M-15; L-1; I-0**

Rationale:

- The Standing Committee noted that the usability and use for this measure are identical to NQF #3633e and passed the measure on both criteria.

5. Related and Competing Measures

- This measure is related to the following measures:
    - NQF #2820 Pediatric Computed Tomography Radiation Dose
        - The Standing Committee asked whether this measure could be incorporated into the three adult radiology measures. The developer shared their plans to update NQF #2820 to further harmonize it with the current measures for future maintenance reviews.
    - NQF #3621 Composite Weighted Average for 3 CT Exam Types: Overall Percent of CT Exams for Which Dose Length Product Is at or Below the Size-Specific Diagnostic Level
        - The Standing Committee had no additional comments about NQF #3621.
    - NQF #3633e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level) (Alara Imaging/ UCSF)

- NQF #3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Facility Level) (Alara Imaging/UCSF)
    - The Standing Committee held a discussion about how NQF #3633e, NQF #3662e, and NQF #3663e relate to each other.
    - The Standing Committee questioned whether the facility level sufficiently captured the necessary quality data and might alone be sufficient or whether the individual-clinician and group measures might be combined and then harmonized with the facility-level measure, thus creating two total measures. The developer stated that each measure captures an important component of responsibility and care quality, and one cannot be prioritized over the other two. The developer added that no additional work is needed to assemble the data between the various levels of analysis addressed by these three measures; therefore, they permit attribution at different levels using the same amount of effort.
    - The Standing Committee acknowledged these comments and requested that the developer continue to examine whether these measures could be further harmonized or combined as they review the real-world data collected before the measures undergo maintenance review.

6. Standing Committee Recommendation for Endorsement: **Total Votes: 18; Yes-15; No-3**

7. Public and Member Comment

- Twenty-five pre-evaluation comments were received for this measure.
    - Seventeen comments were in support of this measure.
        - Comments stated that this measure will meaningfully improve physicians' abilities to monitor the equipment used in these scans, increase their quality, and reduce dose variability, which should lead to a decline in cumulative radiation dose.
        - Comments stated the strength of this measure: It is based on the clinical indication for imaging rather than the type of examination a radiologist chooses to perform.
        - Comments stated that the measure was highly feasible: There were few barriers to the successful implementation of the measure and very little missing data.
        - Comments stated the importance to patients that providers use the lowest-appropriate dose for specific diagnostic or follow-up exams.
        - Comments stated that this measure can reduce not only excessive, high doses, but also suboptimal low doses by identifying outliers and increasing awareness of protocol optimization.
        - Comments stated that the measure feedback is actionable, and users have been very satisfied with the feedback they have received on their measure performance.
    - Four comments were not in support of this measure.
        - Comments expressed concerns that this measure conflates the choice of protocol for the clinical indication with radiation dose optimization, thus making improvement on the measure more challenging.
        - Comments expressed concerns with assessing image quality, radiation risk, subjectivity, patient size, and image rendition; underaddressing exam components and exam diversity; and not providing sufficient guidance for compliance regarding outlier exams.
        - Comments expressed concern that the specifications for the measure have not been validated, specifically the method of determining the classification of dosing studies.
        - Comments expressed concern that the measure deviates from international standards for diagnostic reference levels and lacks consensus on defining global noise.

- Comments expressed concern for the unintended consequences of using too low a dose and possibly missing a disease diagnosis, also resulting in a “wasted dose with no medical benefit.”

○ The measure developer submitted four comments specifically addressing the concerns submitted in the pre-evaluation public comments.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision

9. Appeals

NQF #3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Facility Level)

Measure Worksheet | Specifications

Description: This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient and hospital outpatient care settings are eligible.

Numerator Statement: Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

Denominator Statement: All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

Exclusions: Denominator exclusions are CT exams that simultaneously include multiple body regions outside of four commonly encountered multiple region groupings (specified as LOINC code 96914-7, CT Dose and Image Quality Category, Full Body). Denominator exclusions are also CT exams with missing patient age, missing size-adjusted radiation dose, or missing global noise. These are technical exclusions (“missing data”) from the initial population. Technical exclusions will be flagged, corrected whenever possible, and tracked at the level of the accountable entity.

Adjustment/Stratification: None

Statistical risk model with risk factors (specify number of risk factors), Stratification by risk category (specify number of categories)

The means by which a CT examination is determined to be “out-of-range" with respect to radiation dose is measured by observing whether its patient size-adjusted radiation dose exceeds a pre-determined evidence-based threshold. The value of this size-adjusted radiation dose is calculated with the following equation for any given exam: $D[A] = D[R] * \exp(-(d-d[k]) * \beta[k])$

Where...

D[A]is the size-adjusted radiation dose of the exam

D[R]is the radiation dose of the exam, without adjustment

d is the diameter of the anatomic area being examined

d[k] is the “expected diameter” of the CT category associated with the exam. This “expected diameter” is equal to the median diameter of all exams associated with the CT category in the UCSF International CT Dose Registry containing 6.5 million exams from 161 institutions.

β[k] is the “size-adjustment coefficient” of the CT category associated with the exam. This “size-adjustment coefficient” is the slope parameter of a collection of log-transformed linear regression models fit using the UCSF Registry. A total of 18 models were fit, each using data from one of the CT Dose and Image Quality Categories. The models are parametrized such that, in the kth model and associated dataset, for the jth observation, from the ith hospital, we define: log({D[R]}[ij]) = {β[0]}[k] + β[k] * d[ij] + {z[i]}[k] + ε[ij]

Where D[R] and d are respectively the radiation dose without adjustment and diameter of the anatomic area being examined, β[0]is an intercept term, z is a random effect indicating variation due to the hospital at which the exam was performed, and ε is the residual variation. We restrict the value of β[k]to be greater than 0; when it is less than 0, it is set to 0 and no adjustment is performed. For the estimated values of β[k] across CT categories (strata), please see 2b.30 below.

The intended interpretation of D[A]is the “expected radiation dose of the exam if the diameter of the anatomic area being examined were equal to the population-level median.”

Level of Analysis: Facility

Setting of Care: Outpatient Services, Inpatient/Hospital

Type of Measure: Outcome: Intermediate Clinical Outcome

Data Source: Electronic Health Records, Electronic Health Data

Measure Steward: Alara Imaging

STANDING COMMITTEE MEETING [February 16, 2022]

1. Importance to Measure and Report:

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Total Votes-17; H-1; M-14; L-1; I-1**; 1b. Performance Gap: **Total Votes-17; H-7; M-10; L-0; I-0**

Rationale:

- The Standing Committee noted that the evidence and performance gap for this measure are identical to NQF #3633e and passed the measure on both criteria.

2. Scientific Acceptability of Measure Properties:

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Total Votes-16; Yes-16; No-0**; 2b. Validity: **Total Votes-16; Yes-15; No-1**

Rationale:

- This measure was deemed complex and was evaluated by the SMP.
- The Standing Committee noted that the developer conducted a signal-to-noise analysis using EHRs from 16 hospitals within seven health systems and one vertically integrated organization from February 2020 to April 2021.
- The number of CT exams obtained during inpatient hospitalizations (n=15) in the one month of testing data ranged from 134-1,568 (mean 715); thus, the number of CT exams from inpatient settings per hospital is estimated to vary from 1,608–18,816 for a 12-month period.
- The estimated mean split-half ICC using 37,172 CT exams was 0.99. The number of exams per hospital in the one month of data used for testing ranged from 625 to 6,157 (mean=2,323); the predicted reliability for 12 months exceeded 0.99 for every hospital.
- For the individual hospitals, the predicted reliability for 12 months of inpatient CT exams exceeded 0.99 for every hospital during the testing phase.
- The Standing Committee noted that at the hospital level, the developer obtained CT scans during inpatient hospitalizations and conducted a split-sample analysis, for which the ICC was very high (greater than 0.99 within each hospital). The Standing Committee noted that this measure also applies to outpatient scans and asked the developer to comment on whether any technical differences exist between the two settings. The developer clarified that the indications would not be identical, and inpatient settings would likely have more trauma and stroke scans; nonetheless, the results were identical.

- The Standing Committee accepted the SMP's rating of high for reliability (**Total SMP Votes-11; H-9; M-2; L-0; I-0**).
- The Standing Committee noted that validity testing was conducted at the patient/encounter level, and the developer examined CT category, patient size, radiation dose, size-adjusted radiated dose, global noise, and thresholds for "out-of-range" values to define the numerator. An ICD-10-based algorithm to assign the CT category was compared to a chart review as the gold standard. The results, weighted by the distribution of CT categories in the UCSF International CT Dose Registry, showed a sensitivity of 0.86 and a specificity of 0.96 (n=978 CT exams). When tested across the 16 hospitals, the correct classification rate of the assignment of CT exams to CT category in field-testing was 92 percent on average and varied from 88–97 percent across the 16 hospitals.
- The Standing Committee also noted that validity testing was conducted at the accountable-entity level. The measure score was compared against the medical record review at the accountable-entity level using field-testing data collected from eight health systems/vertically integrated organizations. The Standing Committee concluded that the results indicate a correct and robust implementation of the measure logic.
- The Standing Committee noted that the developer conducted face validity testing, and 94–100 percent agreed that implementation of the measure in federal programs would lead to a reduction in average CT radiation dose while maintaining adequate CT image quality.
- The Standing Committee noted that the SMP expressed concerns about missing data only focusing on the "radiation dose" aspect of the measure.
- The Standing Committee discussed whether the developer had considered additional clinical care factors outside of BMI that might affect dosing; the developer replied that no other factors had a strong impact on the measure, and none were significant enough to warrant risk adjustment. The developer also addressed the possibility of misclassification in creating their dosing strata within CT categories but stated that they worked extensively with clinicians and radiologists to understand the dosing needs for various types of patients and erred on the side of allowing for the possibility of a higher dose when they were asked to by clinical experts.
- The Standing Committee accepted the SMP's rating of high for validity (**Total SMP Votes-11; H-6; M-5; L-0; I-0**).

3. Feasibility: Total Votes-17; H-12; M-4; L-1; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- The Standing Committee noted that the feasibility for this measure is identical to NQF #3633e and passed the measure on this criterion.

4. Usability and Use:

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Total Votes-17; Pass-16; No Pass-1**; 4b. Usability: **Total Votes-17; H-2; M-14; L-1; I-0**

Rationale:

- The Standing Committee noted that the usability and use for this measure are identical to NQF #3633e and passed the measure on both criteria.

5. Related and Competing Measures

- This measure is related to the following measures:
    - NQF #2820 Pediatric Computed Tomography Radiation Dose
        - The Standing Committee asked whether this measure could be incorporated into the three adult radiology measures. The developer shared their plans to update NQF #2820 to further harmonize it with the current measures for future maintenance reviews.

- NQF #3621 Composite Weighted Average for 3 CT Exam Types: Overall Percent of CT Exams for Which Dose Length Product Is at or Below the Size-Specific Diagnostic Level
    - The Standing Committee had no additional comments about NQF #3621.
- NQF #3662e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level) (Alara Imaging/UCSF)
- NQF #3633e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level) (Alara Imaging/UCSF)
    - The Standing Committee held a discussion about how NQF #3633e, NQF #3662e, and NQF #3663e relate to each other.
    - The Standing Committee noted that measurement at the facility level seems to represent the most accurate reflection of radiology practice and structures and questioned whether the facility level might alone be sufficient or whether the individual-clinician and group measures might be combined and then harmonized with the facility-level measure, thus creating two total measures. The developer stated that each measure captures an important component of responsibility and care quality, and one cannot be prioritized over the other two. The developer added that no additional work is needed to assemble the data between the various levels of analysis addressed by these three measures; therefore, they permit attribution at different levels using the same amount of effort.
    - The Standing Committee acknowledged these comments and requested that the developer continue to examine whether these measures could be further harmonized or combined as they review the real-world data collected before the measures undergo maintenance review.

6. Standing Committee Recommendation for Endorsement: **Total Votes: 17; Yes-15; No-2**

7. Public and Member Comment

- Twenty-five pre-evaluation comments were received for this measure.
    - Seventeen comments were in support of this measure.
        - Comments stated that this measure will meaningfully improve physicians' abilities to monitor the equipment used in these scans, increase their quality, and reduce dose variability, which should lead to a decline in cumulative radiation dose.
        - Comments stated the strength of this measure: It is based on the clinical indication for imaging rather than the type of examination a radiologist chooses to perform.
        - Comments stated that the measure was highly feasible: There were few barriers to the successful implementation of the measure and very little missing data.
        - Comments stated the importance to patients that providers use the lowest-appropriate dose for specific diagnostic or follow-up exams.
        - Comments stated that this measure can reduce not only excessive, high doses, but also suboptimal low doses by identifying outliers and increasing awareness of protocol optimization.
        - Comments stated that the measure feedback is actionable, and users have been very satisfied with the feedback they have received on their measure performance.
    - Four comments were not in support of this measure.
        - Comments expressed concern that this measure conflates the choice of protocol for the clinical indication with radiation dose optimization, thus making improvement on the measure more challenging.
        - Comments expressed concerns with the assessment of image quality, radiation risk, subjectivity, patient size, and image rendition; underaddressing exam components and

exam diversity; and not providing sufficient guidance for compliance around outlier exams.
    - Comments expressed concern that the specifications for the measure have not been validated, specifically the method of determining the classification of dosing studies.
    - Comments expressed concern that the measure deviates from international standards for diagnostic reference levels and lacks consensus on defining global noise.
    - Comments expressed concern for the unintended consequences of using too low a dose and possibly missing a disease diagnosis, also resulting in a "wasted dose with no medical benefit."
- The measure developer submitted four comments specifically addressing the concerns submitted in the pre-evaluation public comments.

8. Consensus Standards Approval Committee (CSAC) Endorsement Decision

9. Appeals

Appendix B: Patient Safety Portfolio—Use in Federal Programs*

Measure #	Measure Title	Federal Programs (Finalized or Implemented)
0022	Use of High-Risk Medications in Older Adults (DAE)	Merit-Based Incentive Payment System (MIPS) Program
0097	Medication Reconciliation Post-Discharge	Medicare Part C Star Rating Physician Compare
0101	Falls: Screening, Risk-Assessment, and Plan of Care to Prevent Future Falls	Merit-Based Incentive Payment System (MIPS) Program Medicare Shared Savings Program
0138	National Healthcare Safety Network (NHSN) Catheter-Associated Urinary Tract Infection (CAUTI) Outcome Measure	Hospital-Acquired Condition Reduction Program Hospital Compare Hospital Value-Based Purchasing Inpatient Rehabilitation Facility Quality Reporting Long-Term Care Hospital Quality Reporting Inpatient Rehabilitation Facility Compare Prospective Payment System-Exempt Cancer Hospital Quality Reporting
0139	National Healthcare Safety Network (NHSN) Central Line-Associated Bloodstream Infection (CLABSI) Outcome Measure	Hospital-Acquired Condition Reduction Program Hospital Compare Hospital Value-Based Purchasing Long-Term Care Hospital Quality Reporting Prospective Payment System-Exempt Cancer Hospital Quality Reporting Long-Term Care Hospital Compare
0204	Skill Mix (Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN], Unlicensed Assistive Personnel [UAP], and Contract)	None
0205	Nursing Hours per Patient Day	None
0468	Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Pneumonia Hospitalization	Hospital Compare Hospital Value-Based Purchasing
0500	Severe Sepsis and Septic Shock: Management Bundle	None
0531	Patient Safety Indicator (PSI) 90: Patient Safety and Adverse Events Composite	Hospital Compare Hospital-Acquired Condition Reduction Program Hospital Compare
0537	Multifactor Fall Risk Assessment Conducted for All Patients Who Can Ambulate	Home Health Compare

Measure #	Measure Title	Federal Programs (Finalized or Implemented)
0541	Proportion of Days Covered (PDC): 3 Rates by Therapeutic Category	Marketplace Quality Rating System (QRS)
0553	Care for Older Adults (COA) – Medication Review	Medicare Part C Star Rating
0555	INR Monitoring for Individuals on Warfarin	Marketplace Quality Rating System (QRS)
0674	Percent of Residents Experiencing One or More Falls With Major Injury (Long Stay)	Home Health Compare Nursing Home Compare Nursing Home Quality Initiative Long-Term Care Hospital Quality Reporting Skilled Nursing Facility Quality Reporting Long-Term Care Hospital Compare
0679	Percent of High-Risk Residents With Pressure Ulcers (Long Stay)	Nursing Home Compare Nursing Home Quality Initiative
0684	Percent of Residents With a Urinary Tract Infection (Long Stay)	Nursing Home Compare Nursing Home Quality Initiative
0686	Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)	Nursing Home Compare Nursing Home Quality Initiative
0687	Percent of Residents Who Were Physically Restrained (Long Stay)	Nursing Home Compare Nursing Home Quality Initiative
0689	Percent of Residents Who Lose Too Much Weight (Long Stay)	Nursing Home Compare Nursing Home Quality Initiative
0753	American College of Surgeons – Centers for Disease Control and Prevention (ACS-CDC) Harmonized Procedure-Specific Surgical Site Infection (SSI) Outcome Measure	Hospital Value-Based Purchasing Hospital Acquired Condition Reduction Program Hospital Compare Prospective Payment System-Exempt Cancer Hospital Quality Reporting
1716	National Healthcare Safety Network (NHSN) Facility-Wide Inpatient Hospital-Onset Methicillin-Resistant Staphylococcus Aureus (MRSA) Bacteremia Outcome Measure	Hospital Value-Based Purchasing Hospital-Acquired Condition Reduction Program Hospital Compare Prospective Payment System-Exempt Cancer Hospital Quality Reporting

Measure #	Measure Title	Federal Programs (Finalized or Implemented)
1717	National Healthcare Safety Network (NHSN) Facility-Wide Inpatient Hospital-Onset Clostridium Difficile Infection (CDI) Outcome Measure	Hospital-Acquired Condition Reduction Program Hospital Compare Hospital Value-Based Purchasing Inpatient Rehabilitation Facility Quality Reporting Long-Term Care Hospital Quality Reporting Long-Term Care Hospital Compare Inpatient Rehabilitation Facility Compare Prospective Payment System-Exempt Cancer Hospital Quality Reporting
1893	Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization	Hospital Value-Based Purchasing Hospital Compare
2456	Medication Reconciliation: Number of Unintentional Medication Discrepancies per Medication per Patient	None
2720	National Healthcare Safety Network (NHSN) Antimicrobial Use Measure	None
2723	Wrong-Patient Retract-and-Reorder (Wrong Patient-RAR) Measure	None
2726	Prevention of Central Venous Catheter (CVC)-Related Bloodstream Infections	Merit-Based Incentive Payment System (MIPS) Program
2820	Pediatric Computed Tomography (CT) Radiation Dose	Marketplace QRS HEDIS Quality Measure Rating System
2940	Use of Opioids at High Dosage in Persons Without Cancer	None
2950	Use of Opioids From Multiple Providers in Persons Without Cancer	HEDIS Quality Measure Rating System
2951	Use of Opioids From Multiple Providers and at High Dosage in Persons Without Cancer	HEDIS Quality Measure Rating System
2988	Medication Reconciliation for Patients Receiving Care at Dialysis Facilities	None
2993	Potentially Harmful Drug-Disease Interactions in Older Adults (DDE)	None

Measure #	Measure Title	Federal Programs (Finalized or Implemented)
3025	Ambulatory Breast Procedure Surgical Site Infection (SSI) Outcome Measure	None
3136	GAPPS: Rate of Preventable Adverse Events per 1,000 Patient-Days Among Pediatric Inpatients	None
3215	Adult Inpatient Risk-Adjusted Sepsis Mortality	None
3316e	Safe Use of Opioids – Concurrent Prescribing	Hospital Inpatient Quality Reporting (IQR) Medicare and Medicaid Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals
3389	Concurrent Use of Opioids and Benzodiazepines (COB)	Medicaid
3450	Practice Environment Scale - Nursing Work Index (PES-NWI) (Composite and Five Subscales) (previously NQF #0206 - Undergoing Maintenance)	None
3501e	Hospital Harm – Opioid-Related Adverse Events	None
3502	Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure	None
3503e	Hospital Harm – Severe Hypoglycemia	None
3504	Claims-Only Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure	None
3533e	Hospital Harm – Severe Hyperglycemia	None
3558	Initial Opioid Prescribing for Long Duration (IOP-LD)	None
3621	Composite Weighted Average for Three CT Exam Types: Overall Percent of CT Exams for Which Dose Length Product Is at or Below the Size-Specific Diagnostic Reference Level (for CT Abdomen-Pelvis With Contrast/Single Phase Scan, CT Chest Without Contrast/Single	None

* CMS Measures Inventory Tool Last Accessed on March 2, 2022.

Appendix C: Patient Safety Standing Committee and NQF Staff

STANDING COMMITTEE

John James, PhD (Co-Chair)
Founder, Patient Safety America
Houston, TX

Donald Yealy, MD, FACEP (Co-Chair)
Professor and Chair, University of Pittsburgh-Department of Emergency Medicine
Pittsburgh, PA

Emily Aaronson, PhD
Assistant Chief Quality Officer, Massachusetts General Hospital
Boston, MA

Joel Bundy, MD, FACP, FASN, CPE
Vice President, Chief Quality & Safety Officer, Sentara Healthcare
Norfolk, VA

Elissa Charbonneau, DO, MS
Chief Medical Officer, Encompass Health Corporation
Birmingham, AL

Curtis Collins, PharmD, MS
Specialty Pharmacist, Infectious Diseases, St. Joseph Mercy Health System
Ann Arbor, MI

Theresa Edelstein, MPH, LNHA
Vice President, New Jersey Hospital Association
Princeton, NJ

Terry Fairbanks, MD, MS, FACEP
Vice President, Quality & Safety, MedStar Health
Washington, DC

Jason Falvey, DPT, PhD
Assistant Professor, University of Maryland School of Medicine, Department of Epidemiology and Public Health
Baltimore, MD

Robert Green, MD, MPH, MA
Vice President of Quality & Patient Safety, New York Presbyterian Healthcare System
New York, NY

Sara Hawkins, PhD, RN, CPPS
Director of Patient Safety & Risk, Eastern Idaho Regional Medical Center (EIRMC)

Idaho Falls, ID

Bret Jackson
President, The Economic Alliance for Michigan
Novi, MI

Laura Kinney MA, BSN, RN
Director of Clinical Quality, Teladoc Health
Louisville, KY

Arpana Mathur, MD, MBA
Medical Director, Physician Services, CVS Health
Naperville, IL

Raquel Mayne, MS, MPH, RN
Senior Quality Management Specialist, Hospital for Special Surgery
New York City, NY

Anne Myrka, RPh, MAT
Director, Drug Safety, Island Peer Review Organization (IPRO)
Lake Success, NY

Edward Pollak, MD
Chief Quality Officer, Henry Ford Health System
Detroit, MI

Jamie Roney, DNP, NPD-BC, CCRN-K
Covenant Health Texas Regional Research Coordinator, Covenant Health System
Lubbock, TX

Nancy Schoenborn, MD
Geriatric Medicine Specialist, American Geriatrics Society
Baltimore, MD

David Seidenwurm, MD, FACR
Quality and Safety Director, Sutter Health
Sacramento, CA

Geeta Sood, MD, ScM
Assistant Professor of Medicine,
Johns Hopkins University School of Medicine
Baltimore, MD

Iona Thraen, PhD, ACSW
Patient Safety Director, Utah Hospital and Health Clinics Adjunct Assistant Professor, University of Utah, School of Medicine, Department of Biomedical Informatics
Salt Lake City, UT

Yanling Yu, PhD
Physical Oceanographer and Patient Safety Advocate, Washington Advocate for Patient Safety
Seattle, WA

NQF STAFF

Elizabeth Drye, MD, MS
Chief Scientific Officer, Measurement Science and Application

Tricia Elliott, DHA, MBA, CPHQ, FNAHQ
Senior Managing Director, Measurement Science and Application

Poonam Bal, MHSA
Senior Director, Measurement Science and Application

Matthew K. Pickering, PharmD
Senior Director, Measurement Science and Application

Tamara H. Funk, MPH
Director, Measurement Science and Application

Erin Buchanan, MPH
Senior Manager, Measurement Science and Application

Hannah Ingber, MPH
Manager, Measurement Science and Application

Sean Sullivan, MA
Associate, Measurement Science and Application

Yemsrach Kidane, PMP
Senior Project Manager, Program Operations

Jesse Pines, MD, MBA, MSCE
Consultant

Appendix D: Measure Specifications

NQF #0689 Percent of Residents Who Lose Too Much Weight (Long-Stay)

STEWARD

Centers for Medicare & Medicaid Services

DESCRIPTION

This measure captures the percentage of long-stay nursing home residents with a target Minimum Data Set (MDS) 3.0 assessment (OBRA, PPS, or discharge) that indicates a weight loss of 5% or more of the baseline weight in the last 30 days, or 10% or more of the baseline weight in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen. The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment. Long-stay nursing facility residents are identified as those who have had 101 or more cumulative days of nursing facility care.

TYPE

Outcome

DATA SOURCE

Assessment Data, Electronic Health Records: Electronic Health Records The data source is the Minimum Data Set (MDS) 3.0, and the collection instrument is the Resident Assessment Instrument (RAI). For MDS 3.0 item sets used to calculate the quality measure, please see "MDS3.0_Final_Item_Sets_v1.17.2 for October 1 2020 zip (ZIP)" under the "Downloads" section of the following webpage: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIMDS30TechnicalInformation

LEVEL

Facility

SETTING

Post-Acute Care

NUMERATOR STATEMENT

The numerator is the number of long-stay nursing home residents with a selected target assessment indicating a weight loss of 5% or more of the baseline weight in the last 30 days or 10% or more of the baseline weight in the last 6 months who were not on a physician-prescribed weight-loss regimen (K0300 = [2]). The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment.

NUMERATOR DETAILS

Residents are counted in the numerator if they are long-stay residents, defined as residents who have had 101 or more cumulative days of nursing facility care. Note that the count of cumulative days of nursing facility care continues upon an anticipated reentry within 30 days to the same facility. For example, residents who return to the nursing home following a hospital discharge would not have their length of stay within the episode of care reset to zero if the residents return to the nursing home within 30 days of the prior discharge when return was anticipated. The cumulative days count would resume from the last day of their prior stay. The target population includes all long-stay residents with a target assessment (assessments may be an OBRA quarterly, annual or significant change/correction assessment (A0310A = [02, 03, 04, 05, 06]); or discharge assessment with or without anticipated return (A0310F = [10, 11])), except those with exclusions (specified in sp.16 and sp.17). Note that

the PPS assessment schedule changed with the implementation of the Patient Driven Payment Model (PDPM), and PPS 14-, 30-, 60-, and 90-day assessments (A0310B = [02, 03, 04, 05]) are no longer used for target assessments after October 1, 2019. This change may impact the type of target assessment selected for a very small share of long-stay residents who are under SNF care. These residents are still included in the measure denominator, but their target assessment would likely be an OBRA quarterly assessment instead.

The numerator is the number of long-stay residents in the denominator sample with a selected target assessment that indicates a weight loss of 5% or more of the baseline weight in the last month or 10% or more of the baseline weight in the last six months and the resident was not on a physician-prescribed weight loss regimen (K0300=[2]). The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment.

For every calendar quarter (3-month period), the Centers for Medicare & Medicaid Services (CMS) select target assessments conducted during that quarter from each nursing facility to calculate the measure. For any resident with multiple episodes of care during the quarter, only the latest episode will be counted.

A target assessment is defined as the latest assessment that meets the following criteria: (a) it is contained within the resident's selected episode, (b) it has a qualifying reason for assessment, and (c) its target date is no more than 120 days before the end of the episode.

An episode is defined as a period of time spanning one or more stays. An episode begins with an admission and ends with either (a) a discharge, or (b) the end of the target period, whichever comes first. Data are publicly reported on the Nursing Home Compare webpage and are weighted on an average of four target periods.

DENOMINATOR STATEMENT

The denominator includes all long-stay residents in the nursing home who have a target assessment (OBRA, PPS or discharge) during the selected quarter and who do not meet the exclusion criteria.

DENOMINATOR DETAILS

Residents are counted in the denominator if they are long-stay residents, defined as residents whose length of stay is 101 days or more. Residents who return to the nursing home following a hospital discharge may not have their length of stay within the episode of care reset to zero. The denominator is the number of long-stay residents with a selected target assessment (assessment types include: OBRA quarterly, annual or significant change/correction assessment (A0310A = [02, 03, 04, 05, 06]); or discharge assessment with or without anticipated return (A0310F = [10, 11])) during the selected quarter, except those with exclusions (specified in sp.16 and sp.17).

EXCLUSIONS

There are four exclusions applied to the denominator: (1) the target assessment is an OBRA admission assessment (A0310A = [01]) or a PPS 5-day assessment (A0310B = [01]), (2) having a prognosis of life expectancy of less than six months (J1400 = [1]) or the six-month prognosis item is missing (J1400 = [-]) on the target assessment, (3) receiving hospice care (O0100K2 = [1]) or the hospice care item is missing (O0100K2 = [-]) on the target assessment, or/and (4) the weight loss item is missing (K0300 = [-]) on the target assessment. Only 1,551 episodes in the 2019 (Q1-Q4) long stay resident sample were excluded from the denominator for this measure due to missing responses on the prognosis of life expectancy being less than 6 months, which accounts for 0.04% of the total episodes. Additionally, only 7,948 (0.241%) episodes in the 2019 (Q1-Q4) long stay residents sample were excluded due to missing responses for the Hospice care item, and only 30,854 (0.935%) episodes were excluded due to

missing responses for the weight loss item. If the facility sample includes fewer than 20 residents after all other resident-level exclusions are applied, then the facility is suppressed from public reporting because of small sample size.

EXCLUSION DETAILS

A long-stay resident is excluded from the denominator if:

1. Target assessment is an OBRA Admission assessment (A0310A= [01]) or a PPS 5-Day assessment (A0310B= [01])

2. Prognosis of life expectancy is less than 6 months (J1400 = [1]) or the Prognosis item is missing (J1400 = [-]) on the target assessment.

3. Receiving Hospice care (O0100K2 = [1]) or the Hospice care item is missing (O0100K2 = [-]) on the target assessment.

4. Weight loss item is missing (K0300= [-]) on the target assessment.

If the facility sample includes fewer than 20 residents after all other resident-level exclusions are applied, then the facility is suppressed from public reporting because of small sample size.

RISK ADJUSTMENT

Yes - Additional risk adjustment analysis is included No risk adjustment or stratification Not applicable. This measure is not risk-adjusted.

STRATIFICATION

This measure is not stratified.

TYPE SCORE

Rate/proportion

Better quality = Lower score

ALGORITHM

The Percent of Residents Who Lose Too Much Weight (NQF 0689) is primarily publicly reported as a four-quarter measure, which is based on a rolling four-quarter weighted average that is updated quarterly on Care Compare (https://www.medicare.gov/care-compare/). The four-quarter measure score is computed as follows:

Where QM[Q1], QM[Q2], QM[Q3], and QM[Q4] correspond to the QM values for the four quarters, and D[Q1], D[Q2], D[Q3] and D[Q4] are the denominators (number of eligible residents) for the four quarters.Outlined below are the steps for calculating the quarterly score for this measure.

Step 1: Identify the total number of long-stay residents who have a target assessment (OBRA, PPS, or discharge) during quarter and don't meet the exclusion criteria.

Step 2: Starting with the set of residents identified in Step 1, determine the number of long-stay residents who have experienced weight loss of 5% or more in the last month or 10% or more in the last six months and the weight loss was not planned or prescribed by a physician (K0300=[02]).

Step 3: Divide the result of Step 2 by the result of Step 1.

Step 4: Multiply the result of step 3 by 100 to obtain a percent value.

A description of the time period for the data included in this measure is provided in sp.13 above.

COPYRIGHT / DISCLAIMER

N/A

NQF #3636 Quarterly Reporting of COVID-19 Vaccination Coverage Among Healthcare Personnel

STEWARD

Surveillance Branch, Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention

DESCRIPTION

This quarterly measure identifies the average percentage of healthcare personnel (HCP) who have ever received a primary COVID-19 vaccination course among the total number of HCP who regularly work in the facility. The measure is reported for a quarter (3-month period). The quarterly COVID-19 vaccination coverage is determined by selecting one week per month and calculating the percentage of HCP who have ever received a primary COVID-19 vaccination course, then averaging 3 weekly percentages (one week from each of the 3 months in the quarter).

TYPE

Process

DATA SOURCE

Other (specify) Data are collected using the National Healthcare Safety Network of the U.S. Centers for Disease Control and Prevention - https://www.cdc.gov/nhsn/index.html

LEVEL

Facility

SETTING

Post-Acute Care

NUMERATOR STATEMENT

The numerator for this measure consists of the cumulative number of HCP in the denominator population, who:

1. have received a complete vaccination course against COVID-19 administered at the healthcare facility; or

2. reported in writing (paper or electronic) or provided documentation that a complete vaccination course against COVID-19 was received elsewhereNUMERATOR DETAILS

This quarterly measure identifies the average percentage of healthcare personnel (HCP) who have ever received a primary COVID-19 vaccination course among the total number of HCP who regularly work in the facility.
The measure is reported for a quarter (3-month period). The quarterly COVID-19 vaccination coverage is determined by selecting one week per month and calculating the percentage of HCP who have ever received a primary COVID-19 vaccination course, then averaging 3 weekly percentages (one week from each of the 3 months in the quarter).

NUMERATOR DETAILS

The time period for data collection is one week. A week always begins at 12:01 AM on a Monday and ends on the following Sunday at midnight.

Collect the cumulative number of healthcare personnel (HCP) who have received a primary vaccination course against COVID-19 vaccines at this facility or elsewhere since December 2020. Data sources may include HCP health records and paper and/or electronic documentation of vaccinations given at the healthcare facility or elsewhere. vaccinations elsewhere should provide documentation of the vaccination, which includes the vaccine type.

A completed primary COVID-19 vaccine series is defined by the FDA authorization for use COVID-19 Vaccines | FDA (https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines) and recommendations made by the Advisory Committee on Immunization Practices ACIP COVID-19 Vaccine Recommendations | CDC (https://www.cdc.gov/vaccines/hcp/acip-recs/vacc-specific/covid-19.html) which are reviewed and, if adopted by CDC and the Department of Health and Human Services, published in the Morbidity and Mortality Weekly Report (MMWR).

These recommendations are further described Interim Clinical Considerations for Use of COVID-19 Vaccines Currently Approved or Authorized in the United States (https://www.cdc.gov/vaccines/covid-19/clinical-considerations/covid-19-vaccines-us.html) . As of November 1, 2021, completion of a primary vaccination series is receipt of two doses of mRNA vaccines (manufactured by Pfizer-BioNTech or Moderna) or one dose of viral vector vaccine (manufactured by Janssen).

DENOMINATOR STATEMENT

The target population is the number of healthcare personnel (HCP) eligible to work in the healthcare facility for at least one day during the one-week data collection reporting period, excluding persons with contraindications/exclusions to COVID-19 vaccination. The quarterly reported measure includes at least one week of data collection a month for each of the 3 months in a quarter.

The denominators are reported by aggregating categories below:

1. Employees: all persons who receive a direct paycheck from the reporting facility (i.e., on the facility's payroll).

2. Licensed independent practitioners: include physicians (MD, DO), advanced practice nurses, and physician assistants only who are affiliated with the reporting facility who do not receive a direct paycheck from the reporting facility.

3. Adult students/trainees and volunteers include all students/trainees and volunteers aged 18 or over who do not receive a direct paycheck from the reporting facility.

4. Other contract personnel: Facilities may also report on individuals who are contract personnel. However, reporting for this category is optional. Contract personnel are defined as persons providing care, treatment, or services at the facility through contract who do not fall into any of the above-mentioned denominator categories.DENOMINATOR DETAILS

This quarterly measure identifies the average percentage of healthcare personnel (HCP) who have ever received a primary COVID-19 vaccination course among the total number of HCP who regularly work in the facility.
The measure is reported for a quarter (3-month period). The quarterly COVID-19 vaccination coverage is determined by selecting one week per month and calculating the percentage of HCP who have ever received a primary COVID-19 vaccination course, then averaging 3 weekly percentages (one week from each of the 3 months in the quarter).

DENOMINATOR DETAILS

To identify all healthcare personnel (HCP) eligible to work during the reporting week.

1. Include all HCP who were eligible to have worked at this healthcare facility for at least 1 day during the reporting week, regardless of clinical responsibility or patient contact.

2. HCP who are eligible to have worked include those who are scheduled to work in the facility at least 1 day of the week. Working any part of 1 day is considered as working 1 day.

3. Include HCP even if they are on temporary leave during the reporting week. Temporary leave is defined as less than or equal to 2 weeks in duration. Examples of temporary leave may include sick leave or vacation. In instances where temporary leave extends past two weeks, the healthcare worker should not be included for the current week of data collection.

4. Include persons who worked full-time and part-time.

5. Each person should be counted only once in the denominator.

6. HCP categories should be mutually exclusive. Do not count a person in more than one category.

7. If HCP were eligible to have worked in two or more facilities, each facility should include such personnel in their denominator.

8. Count HCP as individuals rather than full-time equivalents.

9. Data sources for determining eligibility may include payroll, attendance, or other records.

EXCLUSIONS

Exclusions include individuals with contraindications to COVID-19 vaccination and individuals for whom the COVID-19 vaccine is not authorized or recommended.

EXCLUSION DETAILS

Medical contraindications are listed in a vaccine's FDA authorization or labeling and include severe allergic reaction. The most up-to-date list of contraindications as well as exclusions may be found at https://www.cdc.gov/vaccines/covid-19/clinical-considerations/covid-19-vaccines-us.html and includes:

1. Contraindications include severe allergic reaction (e.g., anaphylaxis) after a previous dose or to a component of the COVID-19 vaccine or immediate allergic reaction of any severity to a previous dose or known (diagnosed) allergy to a component of the vaccine.

Individuals for whom the COVID-19 vaccine is not authorized or recommended include the following:

1. COVID-19 vaccines are not currently authorized for individuals 11 years of age or younger.

2. COVID-19 vaccination should be deferred for at least 90 days for individuals who received monoclonal antibodies or convalescent plasma as part of COVID-19 treatment. https://www.cdc.gov/vaccines/covid-19/clinical-considerations/covid-19-vaccines-us.html

COVID-19 vaccines may be administered without regard to the timing of other vaccines. This includes simultaneous administration of the COVID-19 vaccine and other vaccines on the same day. It is not known if the reactogenicity of COVID-19 vaccines is increased with coadministration, including with other vaccines known to be more reactogenic, such as adjuvanted vaccines. When deciding whether to administer an(other) vaccine(s) with a COVID-19 vaccine, vaccination providers should consider whether the patient is behind or at risk of becoming behind on recommended vaccines, their risk of vaccine-preventable disease (e.g., during an outbreak or occupational exposures), and the reactogenicity profile of the vaccines. https://www.cdc.gov/vaccines/covid-19/clinical-considerations/covid-19-vaccines-us.html

RISK ADJUSTMENT

No risk adjustment or stratification

N/A

STRATIFICATION

N/A

TYPE SCORE

Rate/proportion

Better quality = Higher score

ALGORITHM

Data Collection:

1. Identify all healthcare personnel (HCP) eligible to work during the selected week. The week always begins on a Monday at 12:00 midnight and ends on Sunday at 11:59 pm.

2. Categorize all eligible HCP into one of four HCP categories (a – d)

3. Among eligible HCP, identify those who have received a primary COVID-19 vaccination course administered at the healthcare facility or elsewhere.

4. Among eligible HCP who have not received a primary COVID-19 vaccination course, identify those who have a contraindication or exclusion to vaccination.

5. Among eligible HCP who have not received any COVID-19 vaccines and who do not have a contraindication or exclusion to vaccination, identify those who have refused or declined vaccination.

6. Among eligible HCP who have not received any COVID-19 vaccines, identify those whose COVID-19 vaccination status can not be determined.

Measure Calculation:

1. For each one week period, tabulate the denominator by summing the number of HCP in each of the categories of HCP minus the number of HCP with contraindications or exclusions to COVID-19 vaccination.

2. Calculate the weekly COVID-19 vaccination coverage percentage by dividing the number of HCP in the denominator who have received a complete COVID-19 vaccination course by the number of HCP in the denominator and multiplying by 100.

Report quarterly COVID-19 vaccination coverage by averaging 3 weekly coverage percentages (one week from each of the 3 months in the quarter).

If facilities calculate COVID-19 vaccination coverage more than one week per month, the last full week in the reporting month should be used.

COPYRIGHT / DISCLAIMER

N/A

NQF #3633e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level)

STEWARD

Alara Imaging

DESCRIPTION

This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible.

TYPE

Outcome: Intermediate Clinical Outcome

DATA SOURCE

Electronic Health Records, Electronic Health Data The measure derives standardized data elements from structured fields within the EHR and the radiology electronic clinical data systems including the Radiology Information System (RIS) and the Picture Archiving and Communication System (PACS). Primary imaging data stored in structured fields in the radiology electronic clinical data systems have been historically inaccessible using the existing eCQM framework. Thus, the eCQM cannot consume CT images and Radiation Dose Structured Reports (RDSR, which contain the radiation dose) in their original DICOM formats. These primary data, listed below, must be processed to create "calculated" data elements that can then be ingested by the eCQM. The measure developers have created software (available to all users to install locally by agreement, or made accessible through a web interface) to access and process primary data elements from these electronic systems to calculate variables that the eCQM uses to calculate the measure score.

The following primary data elements, their sources, and how they are used in the measure, are illustrated in Table sp-2 below. The steps for how these data elements are accessed, ingested, and processed by the eCQM are described in sp.22.

1. Diagnostic Study, Performed: Categorized CT Exams. All diagnostic CT exams performed during the measurement period, including the type of exam performed (derived from procedure (CPT®) codes associated with the exam bill) and the reason for study (derived from diagnosis (ICD-10-CM) codes associated with the exam order and with the exam bill). A validated algorithm uses combinations of diagnosis and procedure codes to generate the CT Dose and Image Quality Category ("CT category") that specifies the radiation dose and image quality thresholds for each CT exam. (CPT Copyright 2017 American Medical Association. All rights reserved. CPT® is a registered trademark of the American Medical Association.)

2. Diagnostic Study, performed: CT Studies with Radiation Dose Result. Radiation dose is derived from the Radiation Dose Structured Report (RDSR), a DICOM structured element generated by the CT machine for every exam, giving the total radiation dose delivered by the exam (measured as dose length product, mGy-cm). This is used to generate Calculated CT Size-Adjusted Dose ("size-adjusted radiation dose").

3. Diagnostic Study, performed: CT Studies with Image Quality Result. CT image pixel data are generated by the CT machine for every CT exam and stored as DICOM structured data. They are used to measure patient size (measured as diameter on mid-scan axial or coronal images, in mm), which is used in generating the final data element Calculated CT Size-Adjusted Dose. They are also used to generate the final data element Calculated CT Global Noise ("global noise," measured in Hounsfield units).

4. Birth date, to confirm the patient is 18 years of age or older.

5. Supplemental data elements: payer, race, ethnicity, and sex. Table sp-2. Primary data elements are accessed and combined to generate final data elements. “Radiology Electronic Clinical Data Systems” are the core information systems for data storage and practice management that are nearly universal in radiology practices, including the Picture Archiving and Communication System (PACS) and Radiology Information System (RIS).

Data source

Primary Accessed Data Element

Primary Accessed Data Element

Code System

Calculated

Data Element

Calculated Data Element

Code System

Calculated Data Element

Description

Electronic Health Record (EHR), or

Radiology Electronic Clinical Data Systems (non-EHR)

Diagnostic Study, performed: CT Studies ICD-10-CM CT Dose and Image Quality Category LOINC Reflects the type of exam performed based on body region and clinical indication. Each CT category has a specific set of dose and image quality thresholds.

CPT®

Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies

Result attribute: Radiation Dose Structured Report (RDSR) DICOM Calculated CT Size-Adjusted Dose LOINC Reflects the total radiation dose received during CT, risk-adjusted by patient size. The size-adjusted radiation dose thresholds vary by the CT category.

Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies

Result attribute: Image Pixel Data DICOM

Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies

Result attribute: Image Pixel Data DICOM Calculated CT Global Noise LOINC Reflects the image quality (represented by global noise) of the CT. The global noise thresholds vary by the CT category. The measure adjusts global noise measurement by slice thickness. Electronic Health Record (EHR) Birth Date LOINC Birth Date LOINC MM-DD-YYYY, to confirm the patient is eligible.

LEVEL

Clinician: Individual

SETTING

Outpatient Services, Inpatient/Hospital, Ambulatory Care

NUMERATOR STATEMENT

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NUMERATOR DETAILS

The numerator represents the total number of out-of-range (i.e. failed) exams.

Through this application, these LOINC variable names will be shortened for brevity, as follows:

Calculated CT Size-Adjusted Dose = size-adjusted radiation dose

Calculated CT Global Noise = global noise

CT Dose and Image Quality Category = CT category

Definitions

Size-adjusted radiation dose reflects the total radiation dose delivered during a CT, risk-adjusted for patient size. The total radiation dose is recorded for each CT exam using the standardized metric of dose length product (ACR–AAPM–SPR: Practice parameter, European Commission, Radiation Protection No. 185, ICRP Publication 135, Kanal 2017, Smith-Bindman 2019). The patient size is defined as the effective diameter of the anatomic area scanned in millimeters, computed on the mid-slice of the scan. Where axial images are available showing the entire anatomic area, the patient size is computed as the average effective patient diameter on the axial image (Cheng 2013). If axial images showing the entire anatomic area are unavailable, the effective diameter is computed on the coronal localizer image (Christianson 2012). The dose length product is adjusted for patient size using log-transformed linear regression models. The size-adjusted radiation dose value is compared with thresholds that vary by the CT category.

Global noise reflects the image quality of the CT exam. Noise is the most widely used measure of CT image quality. (Catalano 2007, Christianson 2012, Malkus 2017, Schindera 2009, Smith 2008, Szczykutowicz 2017, Szczykutowicz 2021, Willemink 2014) Noise represents differences in the appearance of homogenous areas of tissue that is not a result of inherent tissue composition, but rather of the quality due to imaging technique. In general, image noise in CT reflects the number of x-ray photons hitting the detector, and this will be influenced by the x-ray tube voltage and tube current, as well as patient factors such as the patient's body habitus, the body

region being evaluated, and other scanning parameters such as the slice thickness. Different clinical questions require different values of noise, yet in general, the greater the noise, the worse the image quality and the poorer the diagnostic accuracy, although this is not a simple linear relationship. Diagnostic accuracy may be acceptable

for a large range of noise values, but unacceptable only at a high value. Noise can be quantified in CT images by positioning standard elliptical regions of interest in a known density structure (e.g. water, air, soft tissue) and measuring the standard deviation of the measured values in Hounsfield units. (Catalano 2007). Noise as defined in this measure is calculated on every CT image within a scan (a single irradiating event), and the global noise value for each scan is the mean value across all images. For CT exams that have multiple scans (for example a scan without contrast, followed by a scan with contrast, followed by a delayed scan), the exam is assigned the "best" global noise value across all scans, i.e. the highest quality scan. The global noise value for each scan is also standardized to a 3 mm slice thickness. (Alshipli 2017) The global noise value is compared with thresholds that vary by the CT category.

Details needed to calculate the numerator

To calculate the numerator, the size-adjusted radiation dose and global noise for each CT exam are compared against the following evidence-based thresholds specific to the CT Category (Table sp-1). If a CT exam has a size-adjusted radiation dose and/or global noise value exceeding these thresholds, the exam is considered out-of-range (i.e. "failed") and is counted in the numerator.

Table sp-1. Size-adjusted radiation dose and global noise thresholds by CT category.

CT Category

Size-Adjusted Radiation Dose THRESHOLD (Dose length product, mGy-cm)

Global Noise THRESHOLD (Hounsfield units)

Abdomen and Pelvis Low Dose

598

64

Abdomen and Pelvis Routine Dose

644

29

Abdomen and Pelvis High Dose

1260

29

Cardiac Low Dose

93

55

Cardiac Routine Dose

576

32

Chest Low Dose

377

55

Chest Routine Dose

377

49
Cardiac High Dose or Chest High Dose
1282
49
Head Low Dose
582
115
Head Routine Dose
1025
115
Head High Dose
1832
115
Extremity
320
73
Neck or Cervical Spine
1260
25
Thoracic or Lumbar Spine
1260
25
Simultaneous Chest and Abdomen and Pelvis
1637
29
Simultaneous Thoracic and Lumbar Spine
2520
25
Simultaneous Head and Neck Routine Dose
2285
25
Simultaneous Head and Neck High Dose
3092
25

References

ACR–AAPM–SPR: Practice parameter for diagnostic reference levels and achievable doses in medical x-ray imaging. Revised 2018.Alshipli M and Kabir NA, Effect of slice thickness on image noise and diagnostic content of single-source-dual energy computed tomography 2017 J. Phys.: Conf. Ser. 851 012005

Catalano C, Francone M, Ascarelli A, Mangia M, Iacucci I, Passariello R. Optimizing radiation dose and image quality. Eur Radiol. 2007;17 Suppl 6:F26-32.

Cheng PM. Automated estimation of abdominal effective diameter for body size normalization of CT dose. J Digit Imaging. 2013 Jun;26(3):406-11.

Christianson O, Li X, Frush D, Samei E. Automated size-specific CT dose monitoring program: assessing variability in CT dose. Med Phys. 2012 Nov;39(11):7131-9.

Christianson O, Winslow J, Frush DP, Samei E. Automated Technique to Measure Noise in Clinical CT Examinations. AJR Am J Roentgenol. 2015 Jul;205(1):W93-9.

European Commission, Radiation Protection No. 185, European guidelines on diagnostic reference levels for paediatric imaging. 2018.

ICRP, 2017. Diagnostic reference levels in medical imaging. ICRP Publication 135. Ann. ICRP 46(1).

Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-133.

Malkus A, Szczykutowicz TP. A method to extract image noise level from patient images in CT. Med Phys. 2017 Jun;44(6):2173-2184. doi: 10.1002/mp.12240. Epub 2017 Apr 25.

Schindera ST, Nelson RC, Yoshizumi T, et al. Effect of automatic tube current modulation on radiation dose and image quality for low tube voltage multidetector row CT angiography: phantom study. Acad Radiol. 2009;16(8):997-1002.

Smith AB, Dillon WP, Lau BC, et al. Radiation dose reduction strategy for CT protocols: successful implementation in neuroradiology section. Radiology. 2008;247(2):499-506.

Smith-Bindman R, Wang Y, Chu P, et al. International variation in radiation dose for computed tomography examinations: prospective cohort study. BMJ. 2019;364:k4931.

Szczykutowicz TP, DuPlissis A, Pickhardt PJ. Variation in CT Number and Image Noise Uniformity According to Patient Positioning in MDCT. AJR Am J Roentgenol. 2017 May;208(5):1064-1072. doi: 10.2214/AJR.16.17215. Epub 2017 Mar 7.

Szczykutowicz TP, Nett B, Cherkezyan L, et al. Protocol Optimization Considerations for Implementing Deep Learning CT Reconstruction. AJR American journal of roentgenology. 2021;216(6):1668-1677.

Willemink MJ, Takx RA, de Jong PA, et al. Computed tomography radiation dose reduction: effect of different iterative reconstruction algorithms on image quality. J Comput Assist Tomogr. 2014;38(6):815-823..

DENOMINATOR STATEMENT

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

DENOMINATOR DETAILS

Target population

The target population includes all diagnostic CT exams of specified anatomic sites performed on adults during the measurement period.

On a practical level, to be included, the exam must have an assigned CT category and must have a size-adjusted radiation dose value and a global noise value (meaning the relevant CT data must be available to allow calculation of patient size and image quality.)

CT exams performed in conjunction with nuclear medicine (such as SPECT and PET-CT), biopsies, procedures related to an intervention, assessments of bone mineral density, where the body region is not specified, or where no primary images were obtained, are not included as they are not diagnostic CT.

Definitions

CT Dose and Image Quality Category (short term: "CT category"): reflects the type of exam performed based on the body region and the clinical indication for the exam. Each CT category has a specific set of radiation dose and global noise thresholds. The categories are:

1. Abdomen and Pelvis Low Dose
2. Abdomen and Pelvis Routine Dose
3. Abdomen and Pelvis High Dose
4. Cardiac Low Dose
5. Cardiac Routine Dose
6. Chest Low Dose
7. Chest Routine Dose
8. Cardiac High Dose or Chest High Dose
9. Head Low Dose
10. Head Routine Dose
11. Head High Dose
12. Extremity
13. Neck or Cervical Spine
14. Thoracic or Lumbar Spine
15. Simultaneous Chest and Abdomen and Pelvis
16. Simultaneous Thoracic and Lumbar Spine
17. Simultaneous Head and Neck Routine Dose
18. Simultaneous Head and Neck High Dose

Time period for data collection

One calendar year, although shorter periods can be used for high-volume entities

Codes

LOINC codes representing the data elements required for this measure are published in the Value Set Authority Center (VSAC). They are attached in section sp.11. The data elements themselves and data sources are described in section sp.29.

EXCLUSIONS

Denominator exclusions are CT exams that simultaneously include multiple body regions outside of four commonly encountered multiple region groupings (specified as LOINC code 96914-7, CT Dose and Image Quality Category, Full Body). Denominator exclusions are also CT exams with missing patient age, missing size-adjusted radiation dose, or missing global noise. These are technical exclusions ("missing data") from the initial population. Technical exclusions will be flagged, corrected whenever possible, and tracked at the level of the accountable entity.

EXCLUSION DETAILS

Exclusions

CT exams that cannot be placed into a CT category because they are simultaneous include exams of multiple body regions outside of four commonly encountered multiple region groupings are excluded. The four commonly encountered multiple region groupings are: (1) Simultaneous Chest and Abdomen and Pelvis; (2) Simultaneous Thoracic and Lumbar Spine; (3) Simultaneous Head and Neck Routine Dose; and (4) Simultaneous Head and Neck High Dose. Simultaneous exams of the abdomen and lower extremity are already included as a subset of exams included as part of the "Abdomen and Pelvis High Dose" category. Chest and cardiac are not considered separate body regions for purposes of determining whether the exam contains multiple body regions.

Technical exclusions

CT exams missing any of the four data elements required to calculate measure score are considered technical exclusions: CT category; size-adjusted radiation dose; global noise; birth date.

EXCLUSION DETAILS

Stratification by risk category (specify number of categories), Statistical risk model with risk factors (specify number of risk factors) The means by which a CT examination is determined to be “out-of-range" with respect to radiation dose is measured by observing whether its patient size-adjusted radiation dose exceeds a pre-determined evidence-based threshold. The value of this size-adjusted radiation dose is calculated with the following equation for any given exam:

$D[A] = D[R] * \exp(-(d-d[k]) * \beta[k])$

Where...

D[A]is the size-adjusted radiation dose of the exam

D[R]is the radiation dose of the exam, without adjustment

d is the diameter of the anatomic area being examined

d[k] is the “expected diameter” of the CT category associated with the exam. This “expected diameter” is equal to the median diameter of all exams associated with the CT category in the UCSF International CT Dose Registry containing 6.5 million exams from 161 institutions.

β[k] is the “size-adjustment coefficient” of the CT category associated with the exam. This “size-adjustment coefficient” is the slope parameter of a collection of log-transformed linear regression models fit using the UCSF Registry. A total of 18 models were fit, each using data from one of the CT Dose and Image Quality Categories. The models are parametrized such that, in the kth model and associated dataset, for the jth observation, from the ith hospital, we define:

$\log(\{D[R]\}[ij]) = \{\beta[0]\}[k] + \beta[k] * d[ij] + \{z[i]\}[k] + \varepsilon[ij]$

Where D[R] and d are respectively the radiation dose without adjustment and diameter of the anatomic area being examined, β[0]is an intercept term, z is a random effect indicating variation due to the hospital at which the exam was performed, and ε is the residual variation. We restrict the value of β[k]to be greater than 0; when it is less than 0, it is set to 0 and no adjustment is performed. For the estimated values of β[k] across CT categories (strata), please see 2b.30 below.

The intended interpretation of D[A]is the “expected radiation dose of the exam if the diameter of the anatomic area being examined were equal to the population-level median.”

Technical exclusions

CT exams missing any of the four data elements required to calculate measure score are considered technical exclusions: CT category; size-adjusted radiation dose; global noise; birth date.

RISK ADJUSTMENT

Stratification by risk category (specify number of categories), Statistical risk model with risk factors (specify number of risk factors) The means by which a CT examination is determined to be “out-of-range" with respect to radiation dose is measured by observing whether its patient size-adjusted radiation dose exceeds a pre-determined evidence-based threshold. The value of this size-adjusted radiation dose is calculated with the following equation for any given exam:

D[A] = D[R] * exp(-(d-d[k]) * β[k])

Where...

D[A]is the size-adjusted radiation dose of the exam

D[R]is the radiation dose of the exam, without adjustment

d is the diameter of the anatomic area being examined

d[k] is the “expected diameter” of the CT category associated with the exam. This “expected diameter” is equal to the median diameter of all exams associated with the CT category in the UCSF International CT Dose Registry containing 6.5 million exams from 161 institutions.

β[k] is the “size-adjustment coefficient” of the CT category associated with the exam. This “size-adjustment coefficient” is the slope parameter of a collection of log-transformed linear regression models fit using the UCSF Registry. A total of 18 models were fit, each using data from one of the CT Dose and Image Quality Categories. The models are parametrized such that, in the kth model and associated dataset, for the jth observation, from the ith hospital, we define:

log({D[R]}[ij]) = {β[0]}[k] + β[k] * d[ij] + {z[i]}[k] + ε[ij]

Where D[R] and d are respectively the radiation dose without adjustment and diameter of the anatomic area being examined, β[0]is an intercept term, z is a random effect indicating variation due to the hospital at which the exam was performed, and ε is the residual variation. We restrict the value of β[k]to be greater than 0; when it is less than 0, it is set to 0 and no adjustment is performed. For the estimated values of β[k] across CT categories (strata), please see 2b.30 below.

The intended interpretation of D[A]is the “expected radiation dose of the exam if the diameter of the anatomic area being examined were equal to the population-level median.”

STRATIFICATION

The only stratification variable is the CT category, which is constructed using International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) diagnosis codes and CPT® (Current Procedural Terminology) procedure codes from the billing entity’s claim (or other mapped fields in the electronic health record).

CT categories were constructed to reflect various body regions and different clinical indications for imaging, since different amounts of radiation and image quality are needed to create images sufficient for diagnosis depending on these factors. The framework for creating these categories

took an image-quality informed approach, which first relied on categorizing CT exams into 10 body regions. In five of these regions (extremities, neck [including cervical spine], thoraco-lumbar spine [reflecting either thoracic spine or lumbar spine], combined chest-abdomen, and combined thoraco-lumbar spine [reflecting both thoracic and lumber spine]), clinical indications for scanning do not play a substantial role in altering the amount of radiation needed to produce required images; thus, there is a single CPT®-determined category for each of these body regions. In five other body regions (head, chest, cardiac, abdomen, and combined head and neck), clinical indications do affect the optimal radiation dose, thus these regions were sub-divided based on ICD-10-CM/CPT® defined clinical indications into low, routine, or high radiation dose categories. The "combined head and neck" category was divided into routine and high dose. The approach to determining low, routine, or high radiation doses within these categories was informed by: 1) a review of the published literature; 2) consultation with radiologists with specialty expertise; 3) input from a Technical Expert Panel; and 4) empirical evaluation of about 4.5 million consecutive CT exams from 161 imaging facilities that contribute to the UCSF International CT Dose Registry (January 1, 2016 to December 31, 2019). The categories had face validity as assessed by the Technical Expert Panel, and a manuscript describing this work is under resubmission review in Radiology. The strategy in creating the logic to assign exams to CT categories was to identify indications that were exceptions to the routine radiation dose category, rather than to identify every indication for scanning within the routine category. For example, lung cancer screening is the only defined indication for low-dose chest CT, and evaluation for suspected aortic rupture or dissection (or, more generally, a patient in acute shock) is the only defined indication for high-dose chest CT, leaving all other chest CTs in the routine-dose category. As in this example, all strata were constructed to mimic clinical decision-making

TYPE SCORE

Rate/proportion

Better quality = Lower score

ALGORITHM

At a high level, the following steps occur for each CT exam assessed during the reporting period for the reporting entity:

1. The CT exam is assigned to a CT category using diagnosis (ICD-10-CM) and procedure (CPT®) codes.

2. The patient's size is calculated from DICOM (pixel) data included with the CT exam.

3. The size-adjusted radiation dose is calculated from DICOM data, including the Radiation Dose Structured Report (RDSR) and image pixel data, stored with the CT exam.

4. The global noise is calculated from DICOM (pixel) data stored with the CT exam.

5. The size-adjusted radiation dose and global noise are compared with allowable thresholds, and if either (or both) exceed the allowable thresholds, the CT exam is considered out-of-range (failed).

6. The measure score for the reporting entity is calculated as the proportion of out-of-range CT exams for the reporting entity.

As described in section sp.29, the measure derives standardized data elements from structured fields within the EHR and the radiology electronic clinical data systems including the Radiology Information System (RIS) and the Picture Archiving and Communication System (PACS).

In its existing framework, the eCQM cannot consume primary imaging data in its original format and thus cannot access the requisite data for measure calculation. UCSF and Alara Imaging, Inc. have developed software to access and process primary data elements from the electronic systems to calculate the three variables required by the measure – CT category, size-adjusted radiation dose, and global noise – which can then be ingested by the eCQM for calculating the measure score. The calculation of these variables is broadly described as "pre-processing."

This approach was tested across diverse EHR and PACS platforms. The software is installed at imaging facilities or hospitals within the firewall and functions as an edge device, drawing in data from the specified sources and calculating the variables that can be ingested by the eCQM in a manner that minimizes burden. The software can be fully integrated locally into existing data flows using QDM or FHIR or can be available as a web interface for organizations that do not desire a fully integrated solution.

Consecutive, diagnostic CT exams over one calendar year will be evaluated by the eCQM. These exams may be submitted prospectively in real-time or batch-submitted retrospectively (daily, weekly, monthly). The following steps take place to ingest and calculate the measure score on consecutive CT exams:

Ingestion – Edge Device

1. Radiology electronic clinical data systems record and store information related to medical imaging studies. EHRs record and store information related to the patient and medical imaging encounters.

2. Radiology electronic clinical data systems are configured to automatically forward relevant CT studies with included RDSR reports via DICOM protocols to the edge device. Once the CT study is forwarded to the edge device, the edge device queries the EHR via FHIR or direct API calls for additional information that is then linked to the related exam.

Ingestion – Web Interface

3. For sites not using the integrated edge device, information can be exported from the EHR and radiology electronic clinical data systems via custom reports such as FHIR resources, CCDA documents, and DICOM studies. Relevant information can then be uploaded by sites through a web application for measure calculation. This service will be provided at cost, or free, to minimize burden on providers.

Calculation

1. CT category: The software categorizes the CT exam based on anatomic area (determined by the procedure (CPT®) codes on the exam claims data) and clinical indication (based on the diagnosis (ICD-10-CM) codes associated with the exam order).

2. Size-adjusted radiation dose: The software calculates patient size from image pixel data and receives radiation dose from the Radiation Dose Structured Report (RDSR). The software uses these variables to perform risk adjustment of radiation dose based on patient size. The output of this process is size-adjusted radiation dose.

3. Global noise: The software measures noise in pixel data on CT images. Noise varies by slice thickness, with thinner image slices having higher noise; thus, global noise is adjusted by slice thickness.

4. Software assesses the information for each CT exam for eligibility based on initial population assessment criteria and missing data. Missing data are flagged for the reporting entity and recovered when possible.

5. Remaining CT exams undergo pre-processing on the edge device software or web application, in which the three data elements needed for measure calculation are generated from primary data elements.

6. The eCQM receives all data elements.

7. The eCQM removes denominator exclusions (simultaneous CT exams of multiple body regions outside of four commonly encountered multiple region groupings).

8. For each individual CT exam, the eCQM compares size-adjusted radiation dose and global noise against allowable thresholds specific to the CT category. Exams exceeding dose or noise thresholds are considered failures (out-of-range).

9. The eCQM scores each CT exam in range (pass) or out-of-range (fail). The sum of all out-of-range exams constitutes the numerator for the measure at the patient or population level.

10. An overall measure score (i.e. proportion of CT exams that are out-of-range relative to all evaluated exams) is calculated and can be queried/aggregated at the level of the individual clinician.

For sites that wish to use existing EHR vendors for eCQM computation and submission, primary data elements are sent via the edge device or downloaded via the web interface for ingestion and storage by site EHRs either as a FHIR observation resource, or if FHIR is unavailable, through an integration with an EHR via API.

The measure score can be reported to CMS by the existing EHR vendor, or if preferred, the measure steward is also able to compute and submit measure results to CMS on behalf of sites. Either way, reporting will follow established CMS implementation guidelines.

Feedback will be provided to the individual clinician on the proportion of scans that are out-of-range and the reason these scans are out-of-range to encourage performance improvement.

COPYRIGHT / DISCLAIMER

Alara Imaging, Inc. is not responsible for any use of the Measure. Alara Imaging, Inc. makes no representations, warranties, or endorsement about the quality of any organization or physician that uses or reports performance measures and Alara Imaging, Inc. has no liability to anyone who relies on such measures or specifications.

The Measure can be reproduced and distributed, without modification, for noncommercial purposes (e.g., use by healthcare providers in connection with their practices). Commercial use is defined as the sale, licensing, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain. All commercial uses or requests for modification must be approved by Alara Imaging, Inc. and are subject to a license at the discretion of Alara Imaging, Inc. The Measure is not a clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications.

Alara Imaging, Inc., the University of California San Francisco, and its members and users shall not be responsible for any use or accuracy of the Measure or any code contained within the Measure. THE MEASURE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. Alara Imaging, Inc. disclaims all liability for use or accuracy of any third-party code contained in the specifications. CPT® contained in the Measure specifications is copyright 2004-2021 American Medical Association. LOINC® is copyright 2004-2021 Regenstrief Institute, Inc.

NQF #3662e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level)

STEWARD

Alara Imaging

DESCRIPTION

This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible.

TYPE

Outcome: Intermediate Clinical Outcome

DATA SOURCE

Electronic Health Data, Electronic Health Records The measure derives standardized data elements from structured fields within the EHR and the radiology electronic clinical data systems including the Radiology Information System (RIS) and the Picture Archiving and Communication System (PACS). Primary imaging data stored in structured fields in the radiology electronic clinical data systems have been historically inaccessible using the existing eCQM framework. Thus, the eCQM cannot consume CT images and Radiation Dose Structured Reports (RDSR, which contain the radiation dose) in their original DICOM formats. These primary data, listed below, must be processed to create "calculated" data elements that can then be ingested by the eCQM. The measure developers have created software (available to all users to install locally by agreement or made accessible through a web interface) to access and process primary data elements from these electronic systems to calculate variables that the eCQM uses to calculate the measure score.

The following primary data elements, their sources, and how they are used in the measure, are illustrated in Table sp-2 below. The steps for how these data elements are accessed, ingested, and processed by the eCQM are described in sp.22.

1. Diagnostic Study, Performed: Categorized CT Exams. All diagnostic CT exams performed during the measurement period, including the type of exam performed (derived from procedure (CPT®) codes associated with the exam bill) and the reason for study (derived from diagnosis (ICD-10-CM) codes associated with the exam order and with the exam bill). A validated algorithm uses combinations of diagnosis and procedure codes to generate the CT Dose and Image Quality Category ("CT category") that specifies the radiation dose and image quality thresholds for each CT exam. (CPT Copyright 2017 American Medical Association. All rights reserved. CPT® is a registered trademark of the American Medical Association.)

2. Diagnostic Study, performed: CT Studies with Radiation Dose Result. Radiation dose is derived from the Radiation Dose Structured Report (RDSR), a DICOM structured element generated by the CT machine for every exam, giving the total radiation dose delivered by the exam (measured as dose length product, mGy-cm). This is used to generate Calculated CT Size-Adjusted Dose ("size-adjusted radiation dose").

3. Diagnostic Study, performed: CT Studies with Image Quality Result. CT image pixel data are generated by the CT machine for every CT exam and stored as DICOM structured data. They are used to measure patient size (measured as diameter on mid-scan axial or coronal images, in mm), which is used in generating the final data element Calculated CT Size-Adjusted Dose. They are also used to generate the final data element Calculated CT Global Noise ("global noise," measured in Hounsfield units).

4. Birth date, to confirm the patient is 18 years of age or older.

5. Supplemental data elements: payer, race, ethnicity, and sex.

Table sp-2. Primary data elements are accessed and combined to generate final data elements. "Radiology Electronic Clinical Data Systems" are the core information systems for data storage and practice management that are nearly universal in radiology practices, including the Picture Archiving and Communication System (PACS) and Radiology Information System (RIS).

Data source

Primary Accessed Data Element

Primary Accessed Data Element

Code System

Calculated

Data Element

Calculated Data Element

Code System

Calculated Data Element

Description

Electronic Health Record (EHR),

or

Radiology Electronic Clinical Data Systems (non-EHR)

Diagnostic Study, performed: CT Studies ICD-10-CM CT Dose and Image Quality Category LOINC Reflects the type of exam performed based on body region and clinical indication. Each CT category has a specific set of dose and image quality thresholds.

CPT®

Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies

Result attribute: Radiation Dose Structured Report (RDSR) DICOM Calculated CT Size-Adjusted Dose LOINC Reflects the total radiation dose received during CT, risk-adjusted by patient size. The size-adjusted radiation dose thresholds vary by the CT category.

Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies

Result attribute: Image Pixel Data DICOM

Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies

Result attribute: Image Pixel Data DICOM Calculated CT Global Noise LOINC Reflects the image quality (represented by global noise) of the CT. The global noise thresholds vary by the CT category. The measure adjusts global noise measurement by slice thickness.

Electronic Health Record (EHR) Birth Date LOINC Birth Date LOINC MM-DD-YYYY, to confirm the patient is eligible

LEVEL

Clinician: Group/Practice

SETTING

Outpatient Services, Ambulatory Care, Inpatient/Hospital

NUMERATOR STATEMENT

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NUMERATOR DETAILS

The numerator represents the total number of out-of-range (i.e. failed) exams.

Through this application, these LOINC variable names will be shortened for brevity, as follows:

Calculated CT Size-Adjusted Dose = size-adjusted radiation dose

Calculated CT Global Noise = global noise

CT Dose and Image Quality Category = CT category

Definitions

Size-adjusted radiation dose reflects the total radiation dose delivered during a CT, risk-adjusted for patient size. The total radiation dose is recorded for each CT exam using the standardized metric of dose length product (ACR–AAPM–SPR: Practice parameter, European Commission, Radiation Protection No. 185, ICRP Publication135, Kanal 2017, Smith-Bindman 2019). The patient size is defined as the effective diameter of the anatomic area scanned in millimeters, computed on the mid-slice of the scan. Where axial images are available showing the entire anatomic area, the patient size is computed as the average effective patient diameter on the axial image (Cheng 2013). If axial images showing the entire anatomic area are unavailable, the effective diameter is computed on the coronal localizer image (Christianson 2012). The dose length product is adjusted for patient size using log-transformed linear regression models. The size-adjusted radiation dose value is compared with thresholds that vary

by the CT category.

Global noise reflects the image quality of the CT exam. Noise is the most widely used measure of CT image quality. (Catalano 2007, Christianson 2012, Malkus 2017, Schindera 2009, Smith 2008, Szczykutowicz 2017, Szczykutowicz 2021, Willemink 2014) Noise represents differences in the appearance of homogenous areas of tissue that is not a result of inherent tissue composition, but rather of the quality due to imaging technique. In general, image noise in CT reflects the number of x-ray photons hitting the detector, and this will be influenced by the x-ray tube voltage and tube current, as well as patient factors such as the patient's body habitus, the body region being evaluated, and other scanning parameters such as the slice thickness. Different clinical questions require different values of noise, yet in general, the greater the noise, the worse the image quality and the poorer the diagnostic accuracy, although this is not a simple linear relationship. Diagnostic accuracy may be acceptable for a large range of noise values, but unacceptable only at a high value. Noise can be quantified in CT images by positioning standard elliptical regions of interest in a known density structure (e.g. water, air, soft tissue) and measuring the standard deviation of the measured values in Hounsfield units. (Catalano 2007). Noise as defined in this measure is calculated on every CT image within a scan (a single irradiating event), and the global noise value for each scan is the mean value across all images. For CT exams that have multiple scans (for example a scan without contrast, followed by a scan with contrast, followed by a delayed scan), the exam is assigned the "best" global noise value across all scans, i.e., the highest quality scan. The global noise value for each scan is also standardized to a 3 mm slice thickness. (Alshipli 2017) The global noise value is compared with thresholds that vary by the CT category.

Details needed to calculate the numerator

To calculate the numerator, the size-adjusted radiation dose and global noise for each CT exam are compared against the following evidence-based thresholds specific to the CT Category (Table sp-1). If a CT exam has a size-adjusted radiation dose and/or global noise value exceeding these thresholds, the exam is considered out-of-range (i.e., "failed") and is counted in the numerator.

Table sp-1. Size-adjusted radiation dose and global noise thresholds by CT category.

CT Category

Size-Adjusted Radiation Dose

THRESHOLD

(Dose length product, mGy-cm)

Global Noise

THRESHOLD

(Hounsfield units)

Abdomen and Pelvis Low Dose

598

64

Abdomen and Pelvis Routine Dose

644
29
Abdomen and Pelvis High Dose
1260
29
Cardiac Low Dose
93
55
Cardiac Routine Dose
576
32
Chest Low Dose
377
55
Chest Routine Dose
377
49
Cardiac High Dose or Chest High Dose
1282
49
Head Low Dose
582
115
Head Routine Dose
1025
115
Head High Dose
1832
115
Extremity
320
73
Neck or Cervical Spine
1260
25
Thoracic or Lumbar Spine
1260
25
Simultaneous Chest and Abdomen and Pelvis
1637
29
Simultaneous Thoracic and Lumbar Spine
2520
25
Simultaneous Head and Neck Routine Dose
2285

25
Simultaneous Head and Neck High Dose
3092
25

References

ACR–AAPM–SPR: Practice parameter for diagnostic reference levels and achievable doses in medical x-ray imaging. Revised 2018.

Alshipli M and Kabir NA, Effect of slice thickness on image noise and diagnostic content of single-source-dual energy computed tomography 2017 J. Phys.: Conf. Ser. 851 012005

Catalano C, Francone M, Ascarelli A, Mangia M, Iacucci I, Passariello R. Optimizing radiation dose and image quality. Eur Radiol. 2007;17 Suppl 6:F26-32.

Cheng PM. Automated estimation of abdominal effective diameter for body size normalization of CT dose. J Digit Imaging. 2013 Jun;26(3):406-11.

Christianson O, Li X, Frush D, Samei E. Automated size-specific CT dose monitoring program: assessing variability in CT dose. Med Phys. 2012 Nov;39(11):7131-9.

Christianson O, Winslow J, Frush DP, Samei E. Automated Technique to Measure Noise in Clinical CT Examinations. AJR Am J Roentgenol. 2015 Jul;205(1):W93-9.

European Commission, Radiation Protection No. 185, European guidelines on diagnostic reference levels for paediatric imaging. 2018.

ICRP, 2017. Diagnostic reference levels in medical imaging. ICRP Publication 135. Ann. ICRP 46(1).

Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-133.

Malkus A, Szczykutowicz TP. A method to extract image noise level from patient images in CT. Med Phys. 2017 Jun;44(6):2173-2184. doi: 10.1002/mp.12240. Epub 2017 Apr 25.

Schindera ST, Nelson RC, Yoshizumi T, et al. Effect of automatic tube current modulation on radiation dose and image quality for low tube voltage multidetector row CT angiography: phantom study. Acad Radiol. 2009;16(8):997-1002.

Smith AB, Dillon WP, Lau BC, et al. Radiation dose reduction strategy for CT protocols: successful implementation in neuroradiology section. Radiology. 2008;247(2):499-506.

Smith-Bindman R, Wang Y, Chu P, et al. International variation in radiation dose for computed tomography examinations: prospective cohort study. BMJ. 2019;364:k4931.

Szczykutowicz TP, DuPlissis A, Pickhardt PJ. Variation in CT Number and Image Noise Uniformity According to Patient Positioning in MDCT. AJR Am J Roentgenol. 2017 May;208(5):1064-1072. doi: 10.2214/AJR.16.17215. Epub 2017 Mar 7.

Szczykutowicz TP, Nett B, Cherkezyan L, et al. Protocol Optimization Considerations for Implementing Deep Learning CT Reconstruction. AJR American journal of roentgenology. 2021;216(6):1668-1677.

Willemink MJ, Takx RA, de Jong PA, et al. Computed tomography radiation dose reduction: effect of different iterative reconstruction algorithms on image quality. J Comput Assist Tomogr. 2014;38(6):815-823.

DENOMINATOR STATEMENT

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

DENOMINATOR DETAILS

Target population

The target population includes all diagnostic CT exams of specified anatomic sites performed on adults during the measurement period.

On a practical level, to be included, the exam must have an assigned CT category and must have a size-adjusted radiation dose value and a global noise value (meaning the relevant CT data must be available to allow calculation of patient size and image quality.)

CT exams performed in conjunction with nuclear medicine (such as SPECT and PET-CT), biopsies, procedures related to an intervention, assessments of bone mineral density, where the body region is not specified, or where no primary images were obtained, are not included as they are not diagnostic CT.

Definitions

CT Dose and Image Quality Category (short term: “CT category”): reflects the type of exam performed based on the body region and the clinical indication for the exam. Each CT category has a specific set of radiation dose and global noise thresholds. The categories are:

1. Abdomen and Pelvis Low Dose
2. Abdomen and Pelvis Routine Dose
3. Abdomen and Pelvis High Dose
4. Cardiac Low Dose
5. Cardiac Routine Dose
6. Chest Low Dose
7. Chest Routine Dose
8. Cardiac High Dose or Chest High Dose
9. Head Low Dose
10. Head Routine Dose
11. Head High Dose
12. Extremity
13. Neck or Cervical Spine
14. Thoracic or Lumbar Spine
15. Simultaneous Chest and Abdomen and Pelvis
16. Simultaneous Thoracic and Lumbar Spine
17. Simultaneous Head and Neck Routine Dose

18. Simultaneous Head and Neck High Dose

Time period for data collection

One calendar year, although shorter periods can be used for high-volume entities

Codes

LOINC codes representing the data elements required for this measure are published in the Value Set Authority Center (VSAC). They are attached in section sp.11. The data elements themselves and data sources are described in section sp.29.

EXCLUSIONS

Denominator exclusions are CT exams that simultaneously include multiple body regions outside of four commonly encountered multiple region groupings (specified as LOINC code 96914-7, CT Dose and Image Quality Category, Full Body). Denominator exclusions are also CT exams with missing patient age, missing size-adjusted radiation dose, or missing global noise. These are technical exclusions ("missing data") from the initial population. Technical exclusions will be flagged, corrected whenever possible, and tracked at the level of the accountable entity.

EXCLUSION DETAILS

Exclusions

CT exams that cannot be placed into a CT category because they are simultaneous include exams of multiple body regions outside of four commonly encountered multiple region groupings are excluded. The four commonly encountered multiple region groupings are: (1) Simultaneous Chest and Abdomen and Pelvis; (2) Simultaneous Thoracic and Lumbar Spine; (3) Simultaneous Head and Neck Routine Dose; and (4) Simultaneous Head and Neck High Dose. Simultaneous exams of the abdomen and lower extremity are already included as a subset of exams included as part of the "Abdomen and Pelvis High Dose" category. Chest and cardiac are not considered separate body regions for purposes of determining whether the exam contains multiple body regions.

Technical exclusions

CT exams missing any of the four data elements required to calculate measure score are considered technical exclusions: CT category; size-adjusted radiation dose; global noise; birth date.

EXCLUSION DETAILS

CT exams that cannot be placed into a CT category because they are simultaneous include exams of multiple body regions outside of four commonly encountered multiple region groupings are excluded. The four commonly encountered multiple region groupings are: (1) Simultaneous Chest and Abdomen and Pelvis; (2) Simultaneous Thoracic and Lumbar Spine; (3) Simultaneous Head and Neck Routine Dose; and (4) Simultaneous Head and Neck High Dose. Simultaneous exams of the abdomen and lower extremity are already included as a subset of exams included as part of the "Abdomen and Pelvis High Dose" category. Chest and cardiac are not considered separate body regions for purposes of determining whether the exam contains multiple body regions.

Technical exclusions

CT exams missing any of the four data elements required to calculate measure score are considered technical exclusions: CT category; size-adjusted radiation dose; global noise; birth date.

RISK ADJUSTMENT

Stratification by risk category (specify number of categories), Statistical risk model with risk factors (specify number of risk factors) The means by which a CT examination is determined to be “out-of-range" with respect to radiation dose is measured by observing whether its patient size-adjusted radiation dose exceeds a pre-determined evidence-based threshold. The value of this size-adjusted radiation dose is calculated with the following equation for any given exam:

D[A] = D[R] * exp(-(d-d[k]) * β[k])

Where...

D[A]is the size-adjusted radiation dose of the exam

D[R]is the radiation dose of the exam, without adjustment

d is the diameter of the anatomic area being examined

d[k] is the “expected diameter” of the CT category associated with the exam. This “expected diameter” is equal to the median diameter of all exams associated with the CT category in the UCSF International CT Dose Registry containing 6.5 million exams from 161 institutions.

β[k] is the “size-adjustment coefficient” of the CT category associated with the exam. This “size-adjustment coefficient” is the slope parameter of a collection of log-transformed linear regression models fit using the UCSF Registry. A total of 18 models were fit, each using data from one of the CT Dose and Image Quality Categories. The models are parametrized such that, in the kth model and associated dataset, for the jth observation, from the ith hospital, we define:

log({D[R]}[ij]) = {β[0]}[k] + β[k] * d[ij] + {z[i]}[k] + ε[ij]

Where D[R] and d are respectively the radiation dose without adjustment and diameter of the anatomic area being examined, β[0]is an intercept term, z is a random effect indicating variation due to the hospital at which the exam was performed, and ε is the residual variation. We restrict the value of β[k]to be greater than 0; when it is less than 0, it is set to 0 and no adjustment is performed. For the estimated values of β[k] across CT categories (strata), please see 2b.30 below.

The intended interpretation of D[A]is the “expected radiation dose of the exam if the diameter of the anatomic area being examined were equal to the population-level median.”

STRATIFICATION

The only stratification variable is the CT category, which is constructed using International Classification of Diseases, Revision, Clinical Modification (ICD-10-CM) diagnosis codes and CPT® (Current Procedural Terminology) procedure codes from the billing entity’s claim (or other mapped fields in the electronic health record). CT categories were constructed to reflect various body regions and different clinical indications for imaging, since different amounts of radiation and image quality are needed to create images sufficient for diagnosis depending on these factors. The framework for creating these categories took an image-quality informed approach, which first relied on categorizing CT exams into 10 body regions. In five of these regions (extremities, neck [including cervical spine], thoraco-lumbar spine [reflecting either thoracic spine or lumbar spine], combined chest-abdomen, and combined thoraco-lumbar spine [reflecting both thoracic and lumber spine]), clinical indications for scanning do not play a substantial role in altering the amount of radiation needed to produce required images; thus, there is a single CPT®-determined category for each of these body regions. In five other body regions (head, chest, cardiac, abdomen, and combined head and neck), clinical indications do affect the optimal radiation dose, thus these regions were sub-divided based on ICD-10-CM/CPT® defined clinical indications into low, routine, or high radiation dose categories. The “combined head and neck” category was divided into routine and high dose. The approach to determining low, routine, or high radiation doses within these categories was informed by: 1) a review of the published literature; 2) consultation with radiologists with specialty expertise; 3) input from a Technical Expert Panel; and 4) empirical evaluation of about 4.5 million consecutive

CT exams from 161 imaging facilities that contribute to the UCSF International CT Dose Registry (January 1, 2016 to December 31, 2019). The categories had face validity as assessed by the Technical Expert Panel, and a manuscript describing this work is under resubmission review in Radiology. The strategy in creating the logic to assign exams to CT categories was to identify indications that were exceptions to the routine radiation dose category, rather than to identify every indication for scanning within the routine category. For example, lung cancer screening is the only defined indication for low-dose chest CT, and evaluation for suspected aortic rupture or dissection (or, more generally, a patient in acute shock) is the only defined indication for high-dose chest CT, leaving all other chest CTs in the routine-dose category. As in this example, all strata were constructed to mimic clinical decision-making regarding the most appropriate imaging protocol and its associated radiation dose range. The logic and code table for assigning body regions and indications to CT categories is provided in sp.11. Size-adjusted radiation dose and global noise are assessed against thresholds specific to the CT category, as described further below. However, the measure score is binary (in-range or out-of-range), and the total number/proportion of out-of-range exams is summed for a reportable entity without need for separate stratified calculation or reporting. The measure is not weighted by the stratum, but rather, every CT exam contributes equally to overall score. An entity that performs CT exams within only a few strata has its exams judged against the thresholds for the exams that it performs.

TYPE SCORE

Rate/proportion

Better quality = Lower score

ALGORITHM

At a high level, the following steps occur for each CT exam assessed during the reporting period for the reporting entity:

1. The CT exam is assigned to a CT category using diagnosis (ICD-10-CM) and procedure (CPT®) codes.
2. The patient's size is calculated from DICOM (pixel) data included with the CT exam.
3. The size-adjusted radiation dose is calculated from DICOM data, including the Radiation Dose Structured Report (RDSR) and image pixel data, stored with the CT exam.
4. The global noise is calculated from DICOM (pixel) data stored with the CT exam.
5. The size-adjusted radiation dose and global noise are compared with allowable thresholds, and if either (or both) exceed the allowable thresholds, the CT exam is considered out-of-range (failed).
6. The measure score for the reporting entity is calculated as the proportion of out-of-range CT exams for the reporting entity.

As described in section sp.29, the measure derives standardized data elements from structured fields within the EHR and the radiology electronic clinical data systems including the Radiology Information System (RIS) and the Picture Archiving and Communication System (PACS).

In its existing framework, the eCQM cannot consume primary imaging data in its original format and thus cannot access the requisite data for measure calculation. UCSF and Alara Imaging, Inc. have developed software to access and process primary data elements from the electronic systems to calculate the three variables required by the measure – CT category, size-adjusted radiation dose, and global noise – which can then be ingested by the eCQM for calculating the measure score. The calculation of these variables is broadly described as "pre-processing."

This approach was tested across diverse EHR and PACS platforms. The software is installed at imaging facilities or hospitals within the firewall and functions as an edge device, drawing in data from the specified sources and calculating the variables that can be ingested by the eCQM

in a manner that minimizes burden. The software can be fully integrated locally into existing data flows using QDM or FHIR or can be available as a web interface for organizations that do not desire a fully integrated solution.

Consecutive, diagnostic CT exams over one calendar year will be evaluated by the eCQM. These exams may be submitted prospectively in real-time or batch-submitted retrospectively (daily, weekly, monthly). The following steps take place to ingest and calculate the measure score on consecutive CT exams:

Ingestion – Edge Device

1. Radiology electronic clinical data systems record and store information related to medical imaging studies. EHRs record and store information related to the patient and medical imaging encounters.

2. Radiology electronic clinical data systems are configured to automatically forward relevant CT studies with included RDSR reports via DICOM protocols to the edge device. Once the CT study is forwarded to the edge device, the edge device queries the EHR via FHIR or direct API calls for additional information that is then linked to the related exam.

Ingestion – Web Interface

3. For sites not using the integrated edge device, information can be exported from the EHR and radiology electronic clinical data systems via custom reports such as FHIR resources, CCDA documents, and DICOM studies. Relevant information can then be uploaded by sites through a web application for measure calculation. This service will be provided at cost, or free, to minimize burden on providers.

Calculation

4. Software assesses the information for each CT exam for eligibility based on initial population assessment criteria and missing data. Missing data are flagged for the reporting entity and recovered when possible.

5. Remaining CT exams undergo pre-processing on the edge device software or web application, in which the three data elements needed for measure calculation are generated from primary data elements.

A. CT category: The software categorizes the CT exam based on anatomic area (determined by the procedure (CPT®) codes on the exam claims data) and clinical indication (based on the diagnosis (ICD-10-CM) codes associated with the exam order).

B. Size-adjusted radiation dose: The software calculates patient size from image pixel data and receives radiation dose from the Radiation Dose Structured Report (RDSR). The software uses these variables to perform risk adjustment of radiation dose based on patient size. The output of this process is size-adjusted radiation dose.

C. Global noise: The software measures noise in pixel data on CT images. Noise varies by slice thickness, with thinner image slices having higher noise; thus, global noise is adjusted by slice thickness.

6. The eCQM receives all data elements.

7. The eCQM removes denominator exclusions (simultaneous CT exams of multiple body regions outside of four commonly encountered multiple region groupings).

8. For each individual CT exam, the eCQM compares size-adjusted radiation dose and global noise against allowable thresholds specific to the CT category. Exams exceeding dose or noise thresholds are considered failures (out-of-range).

9. The eCQM scores each CT exam in range (pass) or out-of-range (fail). The sum of all out-of-range exams constitutes the numerator for the measure at the patient or population level.

10. An overall measure score (i.e. proportion of CT exams that are out-of-range relative to all evaluated exams) is calculated and can be queried/aggregated at the level of the clinician group.

For sites that wish to use existing EHR vendors for eCQM computation and submission, primary data elements are sent via the edge device or downloaded via the web interface for ingestion

and storage by site EHRs either as a FHIR observation resource, or if FHIR is unavailable, through an integration with an EHR via API.

The measure score can be reported to CMS by the existing EHR vendor, or if preferred, the measure steward is also able to compute and submit measure results to CMS on behalf of sites. Either way, reporting will follow established CMS implementation guidelines.

Feedback will be provided to the clinician group on the proportion of scans that are out-of-range and the reason these scans are out-of-range to encourage performance improvement. The only stratification variable is the CT category, which is constructed using International Classification of Diseases, 10^th Revision, Clinical Modification (ICD-10-CM) diagnosis codes and CPT® (Current Procedural Terminology) procedure codes from the billing entity's claim (or other mapped fields in the electronic health record).

CT categories were constructed to reflect various body regions and different clinical indications for imaging, since different amounts of radiation and image quality are needed to create images sufficient for diagnosis depending on these factors. The framework for creating these categories took an image-quality informed approach, which first relied on categorizing CT exams into 10 body regions. In five of these regions (extremities, neck [including cervical spine], thoraco-lumbar spine [reflecting either thoracic spine or lumbar spine], combined chest-abdomen, and combined thoraco-lumbar spine [reflecting both thoracic and lumber spine]), clinical indications for scanning do not play a substantial role in altering the amount of radiation needed to produce required images; thus, there is a single CPT®-determined category for each of these body regions. In five other body regions (head, chest, cardiac, abdomen, and combined head and neck), clinical indications do affect the optimal radiation dose, thus these regions were sub-divided based on ICD-10-CM/CPT® defined clinical indications into low, routine, or high radiation dose categories. The "combined head and neck" category was divided into routine and high dose. The approach to determining low, routine, or high radiation doses within these categories was informed by: 1) a review of the published literature; 2) consultation with radiologists with specialty expertise; 3) input from a Technical Expert Panel; and 4) empirical evaluation of about 4.5 million consecutive CT exams from 161 imaging facilities that contribute to the UCSF International CT Dose Registry (January 1, 2016 to December 31, 2019). The categories had face validity as assessed by the Technical Expert Panel, and a manuscript describing this work is under resubmission review in Radiology. The strategy in creating the logic to assign exams to CT categories was to identify indications that were exceptions to the routine radiation dose category, rather than to identify every indication for scanning within the routine category. For example, lung cancer screening is the only defined indication for low-dose chest CT, and evaluation for suspected aortic rupture or dissection (or, more generally, a patient in acute shock) is the only defined indication for high-dose chest CT, leaving all other chest CTs in the routine-dose category. As in this example, all strata were constructed to mimic clinical decision-making regarding the most appropriate imaging protocol and its associated radiation dose range. The logic and code table for assigning body regions and indications to CT categories is provided in sp.11.

Size-adjusted radiation dose and global noise are assessed against thresholds specific to the CT category, as described further below. However, the measure score is binary (in-range or out-of-range), and the total number/proportion of out-of-range exams is summed for a reportable entity without need for separate stratified calculation or reporting. The measure is not weighted by the stratum, but rather, every CT exam contributes equally to overall score. An entity that performs CT exams within only a few strata has its exams judged against the thresholds for the exams that it performs. TYPE SCORE Rate/proportion Better quality = Lower score

COPYRIGHT / DISCLAIMER

that uses or reports performance measures and Alara Imaging, Inc. has no liability to anyone who relies on such measures or specifications.

The Measure can be reproduced and distributed, without modification, for noncommercial purposes (e.g., use by healthcare providers in connection with their practices). Commercial use is defined as the sale, licensing, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain. All commercial uses or requests for modification must be approved by Alara Imaging, Inc. and are subject to a license at the discretion of Alara Imaging, Inc. The Measure is not a clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications.

Alara Imaging, Inc., the University of California San Francisco, and its members and users shall not be responsible for any use or accuracy of the Measure or any code contained within the Measure. THE MEASURE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. Alara Imaging, Inc. disclaims all liability for use or accuracy of any third-party code contained in the specifications. CPT® contained in the Measure specifications is copyright 2004-2021 American Medical Association. LOINC® is copyright 2004-2021 Regenstrief Institute, Inc.

NQF #3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Facility Level)

STEWARD

Alara Imaging

DESCRIPTION

This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient and hospital outpatient care settings are eligible.

TYPE

Outcome: Intermediate Clinical Outcome

DATA SOURCE

Electronic Health Data, Electronic Health Records The measure derives standardized data elements from structured fields within the EHR and the radiology electronic clinical data systems including the Radiology Information System (RIS) and the Picture Archiving and Communication System (PACS). Primary imaging data stored in structured fields in the radiology electronic clinical data systems have been historically inaccessible using the existing eCQM framework. Thus, the eCQM cannot consume CT images and Radiation Dose Structured Reports (RDSR,

which contain the radiation dose) in their original DICOM formats. These primary data, listed below, must be processed to create "calculated" data elements that can then be ingested by the eCQM. The measure developers have created software (available to all users to install locally by agreement, or made accessible through a web interface) to access and process primary data elements from these electronic systems to calculate variables that the eCQM uses to calculate the measure score.

The following primary data elements, their sources, and how they are used in the measure, are illustrated in Table sp-2 below. The steps for how these data elements are accessed, ingested, and processed by the eCQM are described in sp.22.

1. Diagnostic Study, Performed: Categorized CT Exams. All diagnostic CT exams performed during the measurement period, including the type of exam performed (derived from procedure (CPT®) codes associated with the exam bill) and the reason for study (derived from diagnosis (ICD-10-CM) codes associated with the exam order and with the exam bill). A validated algorithm uses combinations of diagnosis and procedure codes to generate the CT Dose and Image Quality Category ("CT category") that specifies the radiation dose and image quality thresholds for each CT exam. (CPT Copyright 2017 American Medical Association. All rights reserved. CPT® is a registered trademark of the American Medical Association.)
2. Diagnostic Study, performed: CT Studies with Radiation Dose Result. Radiation dose is derived from the Radiation Dose Structured Report (RDSR), a DICOM structured element generated by the CT machine for every exam, giving the total radiation dose delivered by the exam (measured as dose length product, mGy-cm). This is used to generate Calculated CT Size-Adjusted Dose ("size-adjusted radiation dose").
3. Diagnostic Study, performed: CT Studies with Image Quality Result. CT image pixel data are generated by the CT machine for every CT exam and stored as DICOM structured data. They are used to measure patient size (measured as diameter on mid-scan axial or coronal images, in mm), which is used in generating the final data element Calculated CT Size-Adjusted Dose. They are also used to generate the final data element Calculated CT Global Noise ("global noise," measured in Hounsfield units).
4. Birth date, to confirm the patient is 18 years of age or older.
5. Supplemental data elements: payer, race, ethnicity, and sex.

Table sp-2. Primary data elements are accessed and combined to generate final data elements. "Radiology Electronic Clinical Data Systems" are the core information systems for data storage and practice management that are nearly universal in radiology practices, including the Picture Archiving and Communication System (PACS) and Radiology Information System (RIS).

Data source
Primary Accessed Data Element
Primary Accessed Data Element
Code System
Calculated
Data Element
Calculated Data Element
Code System
Calculated Data Element

Description
Electronic Health Record (EHR),
or
Radiology Electronic Clinical Data Systems (non-EHR)
Diagnostic Study, performed: CT Studies ICD-10-CM CT Dose and Image Quality Category LOINC Reflects the type of exam performed based on body region and clinical indication. Each CT category has a specific set of dose and image quality thresholds.
CPT®
Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies Result attribute: Radiation Dose Structured Report (RDSR) DICOM Calculated CT Size-Adjusted Dose LOINC Reflects the total radiation dose received during CT, risk-adjusted by patient size. The size-adjusted radiation dose thresholds vary by the CT category.
Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies Result attribute: Image Pixel Data DICOM
Radiology Electronic Clinical Data Systems (non-EHR) Diagnostic Study Performed: CT Studies Result attribute: Image Pixel Data DICOM Calculated CT Global Noise LOINC Reflects the image quality (represented by global noise) of the CT. The global noise thresholds vary by the CT category. The measure adjusts global noise measurement by slice thickness. Electronic Health Record (EHR) Birth Date LOINC Birth Date LOINC MM-DD-YYYY, to confirm the patient is eligible.

LEVEL

Facility

SETTING

Inpatient/Hospital, Outpatient Services

NUMERATOR STATEMENT

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NUMERATOR DETAILS

The numerator represents the total number of out-of-range (i.e. failed) exams.
Through this application, these LOINC variable names will be shortened for brevity, as follows:
Calculated CT Size-Adjusted Dose = size-adjusted radiation dose
Calculated CT Global Noise = global noise
CT Dose and Image Quality Category = CT category
Definitions
Size-adjusted radiation dose reflects the total radiation dose delivered during a CT, risk-adjusted for patient size. The total radiation dose is recorded for each CT exam using the standardized metric of dose length product (ACR–AAPM–SPR: Practice parameter, European Commission, Radiation Protection No. 185, ICRP Publication 135, Kanal 2017, Smith-Bindman 2019). The patient size is defined as the effective diameter of the anatomic area scanned in millimeters, computed on the mid-slice of the scan. Where axial images are available showing the entire

anatomic area, the patient size is computed as the average effective patient diameter on the axial image (Cheng 2013). If axial images showing the entire anatomic area are unavailable, the effective diameter is computed on the coronal localizer image (Christianson 2012). The dose length product is adjusted for patient size using log-transformed linear regression models. The size-adjusted radiation dose value is compared with thresholds that vary by the CT category.
Global noise reflects the image quality of the CT exam. Noise is the most widely used measure of CT image quality. (Catalano 2007, Christianson 2012, Malkus 2017, Schindera 2009, Smith 2008, Szczykutowicz 2017, Szczykutowicz 2021, Willemink 2014) Noise represents differences in the appearance of homogenous areas of tissue that is not a result of inherent tissue composition, but rather of the quality due to imaging technique. In general, image noise in CT reflects the number of x-ray photons hitting the detector, and this will be influenced by the x-ray tube voltage and tube current, as well as patient factors such as the patient's body habitus, the body region being evaluated, and other scanning parameters such as the slice thickness. Different clinical questions require different values of noise, yet in general, the greater the noise, the worse the image quality and the poorer the diagnostic accuracy, although this is not a simple linear relationship. Diagnostic accuracy may be acceptable
for a large range of noise values, but unacceptable only at a high value. Noise can be quantified in CT images by positioning standard elliptical regions of interest in a known density structure (e.g. water, air, soft tissue) and measuring the standard deviation of the measured values in Hounsfield units. (Catalano 2007). Noise as defined in this measure is calculated on every CT image within a scan (a single irradiating event), and the global noise value for each scan is the mean value across all images. For CT exams that have multiple scans (for example a scan without contrast, followed by a scan with contrast, followed by a delayed scan), the exam is assigned the "best" global noise value across all scans, i.e. the highest quality scan. The global noise value for each scan is also standardized to a 3 mm slice thickness. (Alshipli 2017) The global noise value is compared with thresholds that vary by the CT category.
Details needed to calculate the numerator
To calculate the numerator, the size-adjusted radiation dose and global noise for each CT exam are compared against the following evidence-based thresholds specific to the CT Category (Table sp-1). If a CT exam has a size-adjusted radiation dose and/or global noise value exceeding these thresholds, the exam is considered out-of-range (i.e. "failed") and is counted in the numerator.
Table sp-1. Size-adjusted radiation dose and global noise thresholds by CT category.

CT Category	Size-Adjusted Radiation Dose THRESHOLD (Dose length product, mGy-cm)	Global Noise THRESHOLD (Hounsfield units)
Abdomen and Pelvis Low Dose	598	64

Abdomen and Pelvis Routine Dose
644
29
Abdomen and Pelvis High Dose
1260
29
Cardiac Low Dose
93
55
Cardiac Routine Dose
576
32
Chest Low Dose
377
55
Chest Routine Dose
377
49
Cardiac High Dose or Chest High Dose
1282
49
Head Low Dose
582

115
Head Routine Dose
1025
115
Head High Dose
1832
115
Extremity
320
73
Neck or Cervical Spine
1260
25
Thoracic or Lumbar Spine
1260
25
Simultaneous Chest and Abdomen and Pelvis
1637
29

Simultaneous Thoracic and Lumbar Spine
2520
25
Simultaneous Head and Neck Routine Dose
2285
25
Simultaneous Head and Neck High Dose
3092
25
References

ACR–AAPM–SPR: Practice parameter for diagnostic reference levels and achievable doses in medical x-ray imaging. Revised 2018.Alshipli M and Kabir NA, Effect of slice thickness on image noise and diagnostic content of single-source-dual energy computed tomography 2017 J. Phys.: Conf. Ser. 851 012005

Catalano C, Francone M, Ascarelli A, Mangia M, Iacucci I, Passariello R. Optimizing radiation dose and image quality. Eur Radiol. 2007;17 Suppl 6:F26-32.

Cheng PM. Automated estimation of abdominal effective diameter for body size normalization of CT dose. J Digit Imaging. 2013 Jun;26(3):406-11.

Christianson O, Li X, Frush D, Samei E. Automated size-specific CT dose monitoring program: assessing variability in CT dose. Med Phys. 2012 Nov;39(11):7131-9.

Christianson O, Winslow J, Frush DP, Samei E. Automated Technique to Measure Noise in Clinical CT Examinations. AJR Am J Roentgenol. 2015 Jul;205(1):W93-9.

European Commission, Radiation Protection No. 185, European guidelines on diagnostic reference levels for paediatric imaging. 2018.

ICRP, 2017. Diagnostic reference levels in medical imaging. ICRP Publication 135. Ann. ICRP 46(1).

Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-133.

Malkus A, Szczykutowicz TP. A method to extract image noise level from patient images in CT. Med Phys. 2017 Jun;44(6):2173-2184. doi: 10.1002/mp.12240. Epub 2017 Apr 25.

Schindera ST, Nelson RC, Yoshizumi T, et al. Effect of automatic tube current modulation on radiation dose and image quality for low tube voltage multidetector row CT angiography: phantom study. Acad Radiol. 2009;16(8):997-1002.

Smith AB, Dillon WP, Lau BC, et al. Radiation dose reduction strategy for CT protocols: successful implementation in neuroradiology section. Radiology. 2008;247(2):499-506.

Smith-Bindman R, Wang Y, Chu P, et al. International variation in radiation dose for computed tomography examinations: prospective cohort study. BMJ. 2019;364:k4931.

Szczykutowicz TP, DuPlissis A, Pickhardt PJ. Variation in CT Number and Image Noise Uniformity According to Patient Positioning in MDCT. AJR Am J Roentgenol. 2017 May;208(5):1064-1072. doi: 10.2214/AJR.16.17215. Epub 2017 Mar 7.

Szczykutowicz TP, Nett B, Cherkezyan L, et al. Protocol Optimization Considerations for Implementing Deep Learning CT Reconstruction. AJR American journal of roentgenology. 2021;216(6):1668-1677.

Willemink MJ, Takx RA, de Jong PA, et al. Computed tomography radiation dose reduction: effect of different iterative reconstruction algorithms on image quality. J Comput Assist Tomogr. 2014;38(6):815-823.

NUMERATOR DETAILS

The numerator represents the total number of out-of-range (i.e. failed) exams.
Through this application, these LOINC variable names will be shortened for brevity, as follows:
Calculated CT Size-Adjusted Dose = size-adjusted radiation dose
Calculated CT Global Noise = global noise
CT Dose and Image Quality Category = CT category
Definitions
Size-adjusted radiation dose reflects the total radiation dose delivered during a CT, risk-adjusted for patient size. The total radiation dose is recorded for each CT exam using the standardized metric of dose length product (ACR–AAPM–SPR: Practice parameter, European Commission, Radiation Protection No. 185, ICRP Publication 135, Kanal 2017, Smith-Bindman 2019). The patient size is defined as the effective diameter of the anatomic area scanned in millimeters, computed on the mid-slice of the scan. Where axial images are available showing the entire anatomic area, the patient size is computed as the average effective patient diameter on the axial image (Cheng 2013). If axial images showing the entire anatomic area are unavailable, the effective diameter is computed on the coronal localizer image (Christianson 2012). The dose length product is adjusted for patient size using log-transformed linear regression models. The size-adjusted radiation dose value is compared with thresholds that vary by the CT category.
Global noise reflects the image quality of the CT exam. Noise is the most widely used measure of CT image quality. (Catalano 2007, Christianson 2012, Malkus 2017, Schindera 2009, Smith 2008, Szczykutowicz 2017, Szczykutowicz 2021, Willemink 2014) Noise represents differences in the appearance of homogenous areas of tissue that is not a result of inherent tissue composition, but rather of the quality due to imaging technique. In general, image noise in CT reflects the number of x-ray photons hitting the detector, and this will be influenced by the x-ray tube

voltage and tube current, as well as patient factors such as the patient's body habitus, the body region being evaluated, and other scanning parameters such as the slice thickness. Different clinical questions require different values of noise, yet in general, the greater the noise, the worse the image quality and the poorer the diagnostic accuracy, although this is not a simple linear relationship. Diagnostic accuracy may be acceptable for a large range of noise values, but unacceptable only at a high value. Noise can be quantified in CT images by positioning standard elliptical regions of interest in a known density structure (e.g. water, air, soft tissue) and measuring the standard deviation of the measured values in Hounsfield units. (Catalano 2007). Noise as defined in this measure is calculated on every CT image within a scan (a single irradiating event), and the global noise value for each scan is the mean value across all images. For CT exams that have multiple scans (for example a scan without contrast, followed by a scan with contrast, followed by a delayed scan), the exam is assigned the "best" global noise value across all scans, i.e. the highest quality scan. The global noise value for each scan is also standardized to a 3 mm slice thickness. (Alshipli 2017) The global noise value is compared with thresholds that vary by the CT category.

Details needed to calculate the numerator

To calculate the numerator, the size-adjusted radiation dose and global noise for each CT exam are compared against the following evidence-based thresholds specific to the CT Category (Table sp-1). If a CT exam has a size-adjusted radiation dose and/or global noise value exceeding these thresholds, the exam is considered out-of-range (i.e. "failed") and is counted in the numerator.

Table sp-1. Size-adjusted radiation dose and global noise thresholds by CT category.

CT Category	Size-Adjusted Radiation Dose THRESHOLD (Dose length product, mGy-cm)	Global Noise THRESHOLD (Hounsfield units)
Abdomen and Pelvis Low Dose	598	64
Abdomen and Pelvis Routine Dose	644	29
Abdomen and Pelvis High Dose	1260	29
Cardiac Low Dose	93	55
Cardiac Routine Dose	576	

32
Chest Low Dose
377
55
Chest Routine Dose
377
49
Cardiac High Dose or Chest High Dose
1282
49
Head Low Dose
582
115
Head Routine Dose
1025
115
Head High Dose
1832
115
Extremity
320
73
Neck or Cervical Spine
1260
25
Thoracic or Lumbar Spine
1260
25
Simultaneous Chest and Abdomen and Pelvis
1637
29
Simultaneous Thoracic and Lumbar Spine
2520
25
Simultaneous Head and Neck Routine Dose
2285
25
Simultaneous Head and Neck High Dose
3092
25
References
ACR–AAPM–SPR: Practice parameter for diagnostic reference levels and achievable doses in medical x-ray imaging. Revised 2018.Alshipli M and Kabir NA, Effect of slice thickness on image

noise and diagnostic content of single-source-dual energy computed tomography 2017 J. Phys.: Conf. Ser. 851 012005

Catalano C, Francone M, Ascarelli A, Mangia M, Iacucci I, Passariello R. Optimizing radiation dose and image quality. Eur Radiol. 2007;17 Suppl 6:F26-32.

Cheng PM. Automated estimation of abdominal effective diameter for body size normalization of CT dose. J Digit Imaging. 2013 Jun;26(3):406-11.

Christianson O, Li X, Frush D, Samei E. Automated size-specific CT dose monitoring program: assessing variability in CT dose. Med Phys. 2012 Nov;39(11):7131-9.

Christianson O, Winslow J, Frush DP, Samei E. Automated Technique to Measure Noise in Clinical CT Examinations. AJR Am J Roentgenol. 2015 Jul;205(1):W93-9.

European Commission, Radiation Protection No. 185, European guidelines on diagnostic reference levels for paediatric imaging. 2018.

ICRP, 2017. Diagnostic reference levels in medical imaging. ICRP Publication 135. Ann. ICRP 46(1).

Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-133.

Malkus A, Szczykutowicz TP. A method to extract image noise level from patient images in CT. Med Phys. 2017 Jun;44(6):2173-2184. doi: 10.1002/mp.12240. Epub 2017 Apr 25.

Schindera ST, Nelson RC, Yoshizumi T, et al. Effect of automatic tube current modulation on radiation dose and image quality for low tube voltage multidetector row CT angiography: phantom study. Acad Radiol. 2009;16(8):997-1002.

Smith AB, Dillon WP, Lau BC, et al. Radiation dose reduction strategy for CT protocols: successful Implementation in neuroradiology section. Radiology. 2008;247(2):499-506.

Smith-Bindman R, Wang Y, Chu P, et al. International variation in radiation dose for computed tomography examinations: prospective cohort study. BMJ. 2019;364:k4931.

Szczykutowicz TP, DuPlissis A, Pickhardt PJ. Variation in CT Number and Image Noise Uniformity According to Patient Positioning in MDCT. AJR Am J Roentgenol. 2017 May;208(5):1064-1072. doi: 10.2214/AJR.16.17215. Epub 2017 Mar 7.

Szczykutowicz TP, Nett B, Cherkezyan L, et al. Protocol Optimization Considerations for Implementing Deep Learning CT Reconstruction. AJR American journal of roentgenology. 2021;216(6):1668-1677.

Willemink MJ, Takx RA, de Jong PA, et al. Computed tomography radiation dose reduction: effect of different iterative reconstruction algorithms on image quality. J Comput Assist Tomogr. 2014;38(6):815-823.

DENOMINATOR STATEMENT

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

DENOMINATOR DETAILS

Target population

The target population includes all diagnostic CT exams of specified anatomic sites performed on adults during the measurement period.

On a practical level, to be included, the exam must have an assigned CT category and must have a size-adjusted radiation dose value and a global noise value (meaning the relevant CT data must be available to allow calculation of patient size and image quality.)

CT exams performed in conjunction with nuclear medicine (such as SPECT and PET-CT), biopsies, procedures related to an intervention, assessments of bone mineral density, where the body region is not specified, or where no primary images were obtained, are not included as they are not diagnostic CT.

Definitions

CT Dose and Image Quality Category (short term: “CT category”): reflects the type of exam performed based on the body region and the clinical indication for the exam. Each CT category has a specific set of radiation dose and global noise thresholds. The categories are:

1. Abdomen and Pelvis Low Dose
2. Abdomen and Pelvis Routine Dose
3. Abdomen and Pelvis High Dose
4. Cardiac Low Dose
5. Cardiac Routine Dose
6. Chest Low Dose
7. Chest Routine Dose
8. Cardiac High Dose or Chest High Dose
9. Head Low Dose
10. Head Routine Dose
11. Head High Dose
12. Extremity
13. Neck or Cervical Spine
14. Thoracic or Lumbar Spine
15. Simultaneous Chest and Abdomen and Pelvis
16. Simultaneous Thoracic and Lumbar Spine
17. Simultaneous Head and Neck Routine Dose
18. Simultaneous Head and Neck High Dose

Time period for data collection

One calendar year, although shorter periods can be used for high-volume entities

Codes

LOINC codes representing the data elements required for this measure are published in the Value Set Authority Center (VSAC). They are attached in section sp.11. The data elements themselves and data sources are described in section sp.29.

EXCLUSIONS

Denominator exclusions are CT exams that simultaneously include multiple body regions outside of four commonly encountered multiple region groupings (specified as LOINC code 96914-7, CT Dose and Image Quality Category, Full Body). Denominator exclusions are also CT exams with missing patient age, missing size-adjusted radiation dose, or missing global noise. These are technical exclusions ("missing data") from the initial population. Technical exclusions will be flagged, corrected whenever possible, and tracked at the level of the accountable entity.

EXCLUSION DETAILS

Exclusions

CT exams that cannot be placed into a CT category because they are simultaneous include exams of multiple body regions outside of four commonly encountered multiple region groupings are excluded. The four commonly encountered multiple region groupings are: (1) Simultaneous Chest and Abdomen and Pelvis; (2) Simultaneous Thoracic and Lumbar Spine; (3) Simultaneous Head and Neck Routine Dose; and (4) Simultaneous Head and Neck High Dose. Simultaneous exams of the abdomen and lower extremity are already included as a subset of exams included as part of the "Abdomen and Pelvis High Dose" category. Chest and cardiac are not considered separate body regions for purposes of determining whether the exam contains multiple body regions.

Technical exclusions

CT exams missing any of the four data elements required to calculate measure score are considered technical exclusions: CT category; size-adjusted radiation dose; global noise; birth date.

RISK ADJUSTMENT

Statistical risk model with risk factors (specify number of risk factors), stratification by risk category (specify number of categories) the means by which a ct examination is determined to be “out-of-range" with respect to radiation dose is measured by observing whether its patient size-adjusted radiation dose exceeds a pre-determined evidence-based threshold. the value of this size-adjusted radiation dose is calculated with the following equation for any given exam:

d[a] = d[r] * exp(-(d-d[k]) * β[k])

where...

d[a]is the size-adjusted radiation dose of the exam

d[r]is the radiation dose of the exam, without adjustment

d is the diameter of the anatomic area being examined

d[k] is the “expected diameter” of the ct category associated with the exam. this “expected diameter” is equal to the median diameter of all exams associated with the ct category in the ucsf international ct dose registry containing 6.5 million exams from 161 institutions.

β[k] is the “size-adjustment coefficient” of the ct category associated with the exam. this “size-adjustment coefficient” is the slope parameter of a collection of log-transformed linear regression models fit using the ucsf registry. a total of 18 models were fit, each using data from one of the ct dose and image quality categories. the models are parametrized such that, in the kth model and associated dataset, for the jth observation, from the ith hospital, we define:

log({d[r]}[ij]) = {β[0]}[k] + β[k] * d[ij] + {z[i]}[k] + ε[ij]

where d[r] and d are respectively the radiation dose without adjustment and diameter of the anatomic area being examined, β[0]is an intercept term, z is a random effect indicating variation due to the hospital at which the exam was performed, and ε is the residual variation. we restrict the value of β[k]to be greater than 0; when it is less than 0, it is set to 0 and no adjustment is performed. for the estimated values of β[k] across ct categories (strata), please see 2b.30 below.

the intended interpretation of d[a]is the “expected radiation dose of the exam if the diameter of the anatomic area being examined were equal to the population-level median.”

STRATIFICATION

The only stratification variable is the CT category, which is constructed using International Classification of Diseases, Revision, Clinical Modification (ICD-10-CM) diagnosis codes and CPT® (Current Procedural Terminology) procedure codes from the billing entity’s claim (or other mapped fields in the electronic health record) CT categories were constructed to reflect various body regions and different clinical indications for imaging, since different amounts of radiation and image quality are needed to create images sufficient for diagnosis depending on these factors. The framework for creating these categories took an image-quality informed approach, which first relied on categorizing CT exams into 10 body regions. In five of these regions (extremities, neck [including cervical spine], thoraco-lumbar spine [reflecting either thoracic spine or lumbar spine], combined chest-abdomen, and combined thoraco-lumbar spine [reflecting both thoracic and lumber spine]), clinical indications for scanning do not play a substantial role in altering the amount of radiation needed to produce required images; thus, there is a single CPT®-determined category for each of these body regions. In five other body regions (head, chest, cardiac, abdomen, and combined head and neck), clinical indications do affect the optimal radiation dose, thus these regions were sub-divided based on ICD-10-CM/CPT® defined clinical indications into low, routine, or high radiation dose categories. The

"combined head and neck" category was divided into routine and high dose. The approach to determining low, routine, or high radiation doses within these categories was informed by: 1) a review of the published literature; 2) consultation with radiologists with specialty expertise; 3) input from a Technical Expert Panel; and 4) empirical evaluation of about 4.5 million consecutive CT exams from 161 imaging facilities that contribute to the UCSF International CT Dose Registry (January 1, 2016 to December 31, 2019).

ALGORITHM

At a high level, the following steps occur for each CT exam assessed during the reporting period for the reporting entity:

1. The CT exam is assigned to a CT category using diagnosis (ICD-10-CM) and procedure (CPT®) codes.
2. The patient's size is calculated from DICOM (pixel) data included with the CT exam.
3. The size-adjusted radiation dose is calculated from DICOM data, including the Radiation Dose Structured Report (RDSR) and image pixel data, stored with the CT exam.
4. The global noise is calculated from DICOM (pixel) data stored with the CT exam.
5. The size-adjusted radiation dose and global noise are compared with allowable thresholds, and if either (or both) exceed the allowable thresholds, the CT exam is considered out-of-range (failed).
6. The measure score for the reporting entity is calculated as the proportion of out-of-range CT exams for the reporting entity.

As described in section sp.29, the measure derives standardized data elements from structured fields within the EHR and the radiology electronic clinical data systems including the Radiology Information System (RIS) and the Picture Archiving and Communication System (PACS).

In its existing framework, the eCQM cannot consume primary imaging data in its original format and thus cannot access the requisite data for measure calculation. UCSF and Alara Imaging, Inc. have developed software to access and process primary data elements from the electronic systems to calculate the three variables required by the measure – CT category, size-adjusted radiation dose, and global noise – which can then be ingested by the eCQM for calculating the measure score. The calculation of these variables is broadly described as "pre-processing."

This approach was tested across diverse EHR and PACS platforms. The software is installed at imaging facilities or hospitals within the firewall and functions as an edge device, drawing in data from the specified sources and calculating the variables that can be ingested by the eCQM in a manner that minimizes burden. The software can be fully integrated locally into existing data flows using QDM or FHIR or can be available as a web interface for organizations that do not desire a fully integrated solution.

Consecutive, diagnostic CT exams over one calendar year will be evaluated by the eCQM. These exams may be submitted prospectively in real-time or batch-submitted retrospectively (daily, weekly,monthly). The following steps take place to ingest and calculate the measure score on consecutive CT exams:

Ingestion – Edge Device

1. Radiology electronic clinical data systems record and store information related to medical imaging studies. EHRs record and store information related to the patient and medical imaging encounters.

2. Radiology electronic clinical data systems are configured to automatically forward relevant CT studies with included RDSR reports via DICOM protocols to the edge device. Once the CT study is

forwarded to the edge device, the edge device queries the EHR via FHIR or direct API calls for additional information that is then linked to the related exam.

Ingestion – Web Interface

3. For sites not using the integrated edge device, information can be exported from the EHR and radiology electronic clinical data systems via custom reports such as FHIR resources, CCDA documents, and DICOM studies. Relevant information can then be uploaded by sites through a web application for measure calculation. This service will be provided at cost, or free, to minimize burden on providers.

Calculation

1. CT category: The software categorizes the CT exam based on anatomic area (determined by the procedure (CPT®) codes on the exam claims data) and clinical indication (based on the diagnosis (ICD-10-CM) codes associated with the exam order).

2. Size-adjusted radiation dose: The software calculates patient size from image pixel data and receives radiation dose from the Radiation Dose Structured Report (RDSR). The software uses these variables to perform risk adjustment of radiation dose based on patient size. The output of this process is size-adjusted radiation dose.

3. Global noise: The software measures noise in pixel data on CT images. Noise varies by slice thickness, with thinner image slices having higher noise; thus, global noise is adjusted by slice thickness.

4. Software assesses the information for each CT exam for eligibility based on initial population assessment criteria and missing data. Missing data are flagged for the reporting entity and recovered when possible.

5. Remaining CT exams undergo pre-processing on the edge device software or web application, in which the three data elements needed for measure calculation are generated from primary data elements.

6. The eCQM receives all data elements.

7. The eCQM removes denominator exclusions (simultaneous CT exams of multiple body regions outside of four commonly encountered multiple region groupings).

8. For each individual CT exam, the eCQM compares size-adjusted radiation dose and global noise against allowable thresholds specific to the CT category. Exams exceeding dose or noise thresholds are considered failures (out-of-range).

9. The eCQM scores each CT exam in range (pass) or out-of-range (fail). The sum of all out-of-range exams constitutes the numerator for the measure at the patient or population level.

10. An overall measure score (i.e. proportion of CT exams that are out-of-range relative to all evaluated exams) is calculated and can be queried/aggregated at the level of the hospital.

For sites that wish to use existing EHR vendors for eCQM computation and submission, primary data elements are sent via the edge device or downloaded via the web interface for ingestion and storage by site EHRs either as a FHIR observation resource, or if FHIR is unavailable, through an integration with an EHR via API.

The measure score can be reported to CMS by the existing EHR vendor, or if preferred, the measure steward is also able to compute and submit measure results to CMS on behalf of sites. Either way, reporting will follow established CMS implementation guidelines.

Feedback will be provided to the hospital on the proportion of scans that are out-of-range and the reason these scans are out-of-range to encourage performance improvement.

COPYRIGHT / DISCLAIMER

that uses or reports performance measures and Alara Imaging, Inc. has no liability to anyone who relies on such measures or specifications.

The Measure can be reproduced and distributed, without modification, for noncommercial purposes (e.g., use by healthcare providers in connection with their practices). Commercial use is defined as the sale, licensing, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain. All commercial uses or requests for modification must be approved by Alara Imaging, Inc. and are subject to a license at the discretion of Alara Imaging, Inc. The Measure is not a clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications.

Alara Imaging, Inc, the University of California San Francisco, and its members and users shall not be responsible for any use or accuracy of the Measure or any code contained within the Measure. THE MEASURE AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. Alara Imaging, Inc. disclaims all liability for use or accuracy of any third-party code contained in the specifications. CPT® contained in the Measure specifications is copyright 2004-2021 American Medical Association. LOINC® is copyright 2004-2021 Regenstrief Institute, Inc.

Appendix E: Related and Competing Measures

Comparison of NQF #3636 and NQF #0431

Steward/Developer

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Centers for Disease Control and Prevention

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Centers for Disease Control and Prevention

Description

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

This quarterly measure identifies the average percentage of healthcare personnel (HCP) who have ever received a primary COVID-19 vaccination course among the total number of HCP who regularly work in the facility.

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Percentage of healthcare personnel (HCP) who received the influenza vaccination.

Numerator

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

The cumulative number of HCP in the denominator population, who have received a complete vaccination course against COVID-19 administered at the healthcare facility; or reported in writing (paper or electronic) or provided documentation that a complete vaccination course against COVID-19 was received elsewhere.

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

HCP in the denominator population who received an influenza vaccination administered at the healthcare facility.

Denominator

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

The number of healthcare personnel (HCP) eligible to work in the healthcare facility for at least one day during the one-week data collection reporting period, excluding persons with contraindications/exclusions to COVID-19 vaccination.

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Number of HCP who are working in the healthcare facility for at least 1 working day between October 1 and March 31 of the following year, regardless of clinical responsibility or patient contact.

Measure Type

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

New

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Maintenance

Data Source

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Varies by facility. Mix of EHR and Paper Sources.

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Mix of EHR and Paper Sources.

Target Population

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

HCP

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

HCP

Care Setting

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Post Acute Care Facility

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Long-term Care Hospital, Post Acute/Long Term Care Facility

Level of Analysis

NQF # 3636 QUARTERLY REPORTING OF COVID-19 VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Facility Level

NQF #0431 INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

Facility Level

Comparison of NQF #3633e and NQF #2820

Steward/Developer

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

University of California, San Francisco; Alara Imaging

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

University of California, San Francisco

Description

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

Radiation dose is measured as the dose-length product for every diagnostic brain, skull, and abdomen and pelvis CT scan performed by a reporting facility on any child less than 18 years of age during the reporting period of 12 months. The dose associated with each scan is evaluated as “high” or “acceptable,” relative to the 75th percentile benchmark for that type of scan and age of patient. Median doses are calculated at the facility level for each type of scan and age of patient stratum, and then compared with the same 75th percentile benchmark. The overall proportion of high dose exams is calculated including all CT scans.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Radiation dose is measured as the dose-length product for every diagnostic brain, skull, and abdomen and pelvis CT scan performed by a reporting facility on any child less than 18 years of age during the reporting period of 12 months. The dose associated with each scan is evaluated as “high” or “acceptable,” relative to the 75th percentile benchmark for that type of scan and age of patient. Median doses are calculated at the facility level for each type of scan and age of patient stratum, and then compared with the same 75th percentile benchmark. The overall proportion of high dose exams is calculated including all CT scans.

Numerator

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUTE IMAGE QUALITY FOR DIAGNSOTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

The number of diagnostic CT scans within an eligible anatomic region (i.e., brain, skull, abdomen and pelvis) and age stratum for which the radiation dose (measured in dose-length product, DLP) exceeds the 75th percentile benchmark for that type of scan and age of patient.

Denominator

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

The denominator is the total number of diagnostic CT scans within an eligible anatomic region and age stratum (infant (<1 year); small child (1-4); medium child (5-9); large child (10-14) and adolescent (15-17)) that were performed during the reporting period. These totals are summed to generate the total number of diagnostic CT scans within all eligible anatomic regions and age strata.

Measure Type

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICAN LEVEL)

New

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Maintenance

Data Source

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

EHR

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

EHR

Target Population

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

Adults (Age > 18)

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Children (Age < 18)

Care Setting

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

Ambulatory Care, Inpatient Care/Hospital, Outpatient Services

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Inpatient Care/Hospital, Outpatient Services

Level of Analysis

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMGRAPHY (CT) IN ADULTS (CLINICAN LEVEL)

Clinican/Indivual Level

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Facility Level

Comparison of NQF #3633e and NQF #3621

Steward/Developer

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

University of California, San Francisco; Alara Imaging

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

American College of Radiology

Description

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

Radiation dose is measured as the dose-length product for every diagnostic brain, skull, and abdomen and pelvis CT scan performed by a reporting facility on any child less than 18 years of age during the reporting period of 12 months. The dose associated with each scan is evaluated as "high" or "acceptable," relative to the 75th percentile benchmark for that type of scan and age of patient. Median doses are calculated at the facility level for each type of scan and age of patient stratum, and then compared with the same 75th percentile benchmark. The overall proportion of high dose exams is calculated including all CT scans.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Weighted average of 3 CT Exam Types: Overall Percent of CT exams for which Dose Length Product is at or below the size-specific diagnostic reference level (for CT Abdomen-pelvis with contrast/single phase scan, CT Chest without contrast/single phase scan and CT Head/Brain without contrast/single phase scan).

Numerator

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUTE IMAGE QUALITY FOR DIAGNSOTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Number of CT Abdomen-Pelvis exams with contrast (single phase scan), CT Chest exams without contrast (single phase scan), and CT Head/Brain exams without contrast (single phase scan) for which Dose Length Product is at or below the size-specific exam-specific diagnostic reference level.

Denominator

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Number of CT Abdomen-pelvis exams with contrast (single phase scans), CT Chest exams without contrast (single phase scans), and CT Head/Brain (single phase scans).

Measure Type

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICAN LEVEL)

New

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

New

Data Source

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

EHR

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Registry Data

Target Population

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

Adults (Age > 18)

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

All patients regardless of age

Care Setting

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN LEVEL)

Ambulatory Care, Inpatient Care/Hospital, Outpatient Services

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Emergency Department and Services, Inpatient/Hospital, Other, Outpatient Services, Dialysis Facility

Level of Analysis

NQF #3633E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMGRAPHY (CT) IN ADULTS (CLINICAN LEVEL)

Clinican/Indivual Level

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Clinician: Group/Practice, Facility

Comparison of NQF #3662e and NQF #2820

Steward/Developer

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

University of California, San Francisco; Alara Imaging

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

University of California, San Francisco

Description

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Radiation dose is measured as the dose-length product for every diagnostic brain, skull, and abdomen and pelvis CT scan performed by a reporting facility on any child less than 18 years of age during the reporting period of 12 months. The dose associated with each scan is evaluated as “high” or “acceptable,” relative to the 75th percentile benchmark for that type of scan and age of patient. Median doses are calculated at the facility level for each type of scan and age of patient stratum, and then compared with the same 75th percentile benchmark. The overall proportion of high dose exams is calculated including all CT scans.

Numerator

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

The number of diagnostic CT scans within an eligible anatomic region (i.e., brain, skull, abdomen and pelvis) and age stratum for which the radiation dose (measured in dose-length product, DLP) exceeds the 75th percentile benchmark for that type of scan and age of patient.

Denominator

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

The denominator is the total number of diagnostic CT scans within an eligible anatomic region and age stratum (infant (<1 year); small child (1-4); medium child (5-9); large child (10-14) and adolescent (15-17)) that were performed during the reporting period. These totals are summed to generate the total number of diagnostic CT scans within all eligible anatomic regions and age strata.

Measure Type

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

New

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Maintenance

Data Source

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

EHR

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

EHR

Target Population

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

Adults (Age >= 18)

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Children (Age < 18)

Care Setting

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

Ambulatory Care, Inpatient Care/Hospital, Outpatient Services

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Inpatient Care/Hospital, Outpatient Services

Level of Analysis

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

Clinician: Group/Practice Level

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Facility Level

Comparison of NQF #3662e and NQF #3621

Steward/Developer

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

University of California, San Francisco; Alara Imaging

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

American College of Radiology

Description

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient, outpatient and ambulatory care settings are eligible.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Weighted average of 3 CT Exam Types: Overall Percent of CT exams for which Dose Length Product is at or below the size-specific diagnostic reference level (for CT Abdomen-pelvis with contrast/single phase scan, CT Chest without contrast/single phase scan and CT Head/Brain without contrast/single phase scan).

Numerator

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Number of CT Abdomen-Pelvis exams with contrast (single phase scan), CT Chest exams without contrast (single phase scan), and CT Head/Brain exams without contrast (single phase scan) for which Dose Length Product is at or below the size-specific exam-specific diagnostic reference level.

Denominator

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Number of CT Abdomen-pelvis exams with contrast (single phase scans), CT Chest exams without contrast (single phase scans), and CT Head/Brain (single phase scans).

Measure Type

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

New

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

New

Data Source

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

EHR

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Registry Data

Target Population

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

Adults (Age >= 18)

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

All patients regardless of age

Care Setting

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

Ambulatory Care, Inpatient Care/Hospital, Outpatient Services

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Emergency Department and Services, Inpatient/Hospital, Other, Outpatient Services, Dialysis Facility

Level of Analysis

NQF #3662E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (CLINICIAN GROUP LEVEL)

Clinician: Group/Practice Level

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Clinician: Group/Practice, Facility

Comparison of NQF #3663e and NQF #2820

Steward/Developer

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

University of California, San Francisco; Alara Imaging

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

University of California, San Francisco

Description

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient and hospital outpatient care settings are eligible.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Radiation dose is measured as the dose-length product for every diagnostic brain, skull, and abdomen and pelvis CT scan performed by a reporting facility on any child less than 18 years of age during the reporting period of 12 months. The dose associated with each scan is evaluated as “high” or “acceptable,” relative to the 75th percentile benchmark for that type of scan and age of patient. Median doses are calculated at the facility level for each type of scan and age of patient stratum, and then compared with the same 75th percentile benchmark. The overall proportion of high dose exams is calculated including all CT scans.

Numerator

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

The number of diagnostic CT scans within an eligible anatomic region (i.e., brain, skull, abdomen and pelvis) and age stratum for which the radiation dose (measured in dose-length product, DLP) exceeds the 75th percentile benchmark for that type of scan and age of patient.

Denominator

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

The denominator is the total number of diagnostic CT scans within an eligible anatomic region and age stratum (infant (<1 year); small child (1-4); medium child (5-9); large child (10-14) and adolescent (15-17)) that were performed during the reporting period. These totals are summed to generate the total number of diagnostic CT scans within all eligible anatomic regions and age strata.

Measure Type

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

New

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Maintenance

Data Source

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

EHR

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

EHR

Target Population

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

Adults (Age > = 18)

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Children (Age < 18)

Care Setting

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

Inpatient Care/Hospital, Outpatient Services

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Inpatient Care/Hospital, Outpatient Services

Level of Analysis

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

Facility Level

NQF #2820 PEDIATRIC COMPUTED TOMOGRAPHY (CT) RADIATION DOSE

Facility Level

Comparison of NQF #3663e and NQF #3621

Steward/Developer

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

University of California, San Francisco; Alara Imaging

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

American College of Radiology

Description

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

This electronic clinical quality measure (eCQM) provides a standardized method for monitoring the performance of diagnostic CT to discourage unnecessarily high radiation doses, a risk factor for cancer, while preserving image quality. It is expressed as a percentage of eligible CT exams that are out-of-range based on having either excessive radiation dose or inadequate image quality, relative to evidence-based thresholds based on the clinical indication for the exam. All diagnostic CT exams of specified anatomic sites performed in inpatient and hospital outpatient care settings are eligible.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Weighted average of 3 CT Exam Types: Overall Percent of CT exams for which Dose Length Product is at or below the size-specific diagnostic reference level (for CT Abdomen-pelvis with contrast/single phase scan, CT Chest without contrast/single phase scan and CT Head/Brain without contrast/single phase scan).

Numerator

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

Diagnostic CT exams that have a size-adjusted radiation dose value greater than the threshold specific to the CT category (reflecting the body region imaged and the radiation dose and image quality required for that exam given the reason for the exam), or a global noise value greater than a threshold specific to the CT Category.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Number of CT Abdomen-Pelvis exams with contrast (single phase scan), CT Chest exams without contrast (single phase scan), and CT Head/Brain exams without contrast (single phase scan) for which Dose Length Product is at or below the size-specific exam-specific diagnostic reference level.

Denominator

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

All diagnostic CT exams performed on adults (aged 18 years and older) during the measurement period of one year that have an assigned CT category, a size-adjusted radiation dose value, and a global noise value.

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Number of CT Abdomen-pelvis exams with contrast (single phase scans), CT Chest exams without contrast (single phase scans), and CT Head/Brain (single phase scans).

Measure Type

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

New

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

New

Data Source

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

EHR

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Registry Data

Target Population

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

Adults (Age > = 18)

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

All patients regardless of age

Care Setting

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

Inpatient Care/Hospital, Outpatient Services

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Emergency Department and Services, Inpatient/Hospital, Other, Outpatient Services, Dialysis Facility

Level of Analysis

NQF #3663E EXCESSIVE RADIATION DOSE OR INADEQUATE IMAGE QUALITY FOR DIAGNOSTIC COMPUTED TOMOGRAPHY (CT) IN ADULTS (FACILITY LEVEL)

Facility Level

NQF #3621 COMPOSITE WEIGHTED AVERAGE FOR 3 CT EXAM TYPES: OVERALL PERCENT OF CT EXAMS FOR WHICH DOSE LENGTH PRODUCT IS AT OR BELOW THE SIZE SPECIFIC DIAGNOSTIC REFERENCE LEVEL

Clinician: Group/Practice, Facility

Appendix F: Pre-Evaluation Comments

Comments received as of January 16, 2022.

NQF #3636 Quarterly Reporting of COVID-19 Vaccination Coverage Among Healthcare Personnel

Comment 1 by: Andrew Geller, Centers for Disease Control and Prevention

Since the original measure information submission, several systematic reviews have been published on the effectiveness of COVID-19 vaccination in reducing COVID-19 infections among those vaccinated. These systematic reviews provide evidence demonstrating that healthcare personnel (HCP) vaccination will reduce infections among HCP. Reductions in HCP infections not only protect HCP themselves but importantly also decrease disruptions of care of patients. Both reductions of HCP infections and decreased disruptions of care are key outcomes expected from increasing HCP vaccination coverage. Three systematic reviews are provided below. There currently are no systematic reviews of COVID-19 vaccine effectiveness among nursing home HCP. A search of the academic literature database MEDLINE (via PubMed) to identify individual studies was conducted and found two US studies demonstrating COVID-19 vaccine effectiveness among HCP at nursing homes/long-term care facilities. These are presented to supplement the measure submission and serve to further support the evidence base for this proposed measure.

Evidence of COVID-19 Vaccine Effectiveness - Systematic Reviews

(1) Harder T, Koch J, Vygen-Bonnet S, et al. Efficacy and effectiveness of COVID-19 vaccines against SARS-CoV-2 infection: interim results of a living systematic review, 1 January to 14 May 2021. *Euro Surveill* 2021 Jul;26(28):2100563. https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2021.26.28.2100563

Key Conclusion: "Results of this living systematic review imply that COVID-19 vaccines are highly effective in preventing SARS-CoV-2 infections, including those which are asymptomatic. From a public health perspective, it can be concluded that fully vaccinated persons might in some instances still become PCR-positive for SARS-CoV-2 but only play a minor role in the transmission of SARS-CoV-2."

Quantity and quality of studies: 30 studies including: 2 Case-control studies, 8 Cohort studies, 3 Matched case-control studies, 2 Prospective cohort studies, 2 Randomized controlled trials, 9 Retrospective cohort studies.

Estimates of benefit: 24 studies reported single-dose efficacy/effectiveness, with most estimates between 60%-70% (range, 16.9%-91.2%). 17 studies reported vaccine effectiveness after the second dose, with most estimates of VE 80%-90% (range, 61.7%-98.6%).

(2) Kow CS, Hasan S. Real-world effectiveness of BNT162b2 mRNA vaccine: a meta-analysis of large observational studies. *Inflammopharmacology* 2021 Aug;29(4):1075-1090. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8266992/

Key Conclusion: "The meta-analysis revealed significant protective effect against RT-PCR confirmed COVID-19 ≥ 14 days after the first dose, with vaccine effectiveness of 53% (95%

confidence interval 32–68%), and ≥ 7 days after the second dose, with vaccine effectiveness of 95% (95% confidence interval: 96–97%)."

Quantity and quality of studies: 19 studies included: 1 Case-control study, 4 Prospective cohort studies, 1 Prospective database review, 1 Prospective multicenter study, 3 Retrospective cohort studies, 1 Retrospective study (other), 7 Retrospective database review studies, 2 Retrospective case-control study.

Estimates of benefit:

8 studies reported hazard ratio (HR) of significant protective effect against RT-PCR confirmed COVID-19 ≥14 days after 1st dose and 5 studies reported significant protective effect by incidence rate ratio (IRR) of reduced RT-PCR confirmed COVID-19 ≥14 days after 1st dose. This resulted in an overall pooled estimate for VE of 53% (95% CI, 32%-68%).

6 studies reported hazard ratio (HR) of significant protective effect against RT-PCR confirmed COVID-19 ≥21 days after 1st dose, and 3 studies reported significant protective effect by incidence rate ratio (IRR) of reduced RT-PCR confirmed COVID-19 ≥21 days after 1st dose. This resulted in an overall pooled estimate for VE of 59% (95% CI, 53%-64%).

3 studies reported hazard ratio (HR) of significant protective effect against RT-PCR confirmed COVID-19 ≥7 days after 2nd dose, 5 studies reported significant protective effect by incidence rate ratio (IRR) of reduced RT-PCR confirmed COVID-19 ≥7 days after 2nd dose, and 3 studies presented odds ratio (OR) as effect measure, showing reduced odds of infection ≥7 days after 2nd dose. This resulted in pooled VE of ≥81%.

3 studies presented HR ≥14 days after the 2nd dose, and 3 IRR ≥14 days after the 2nd dose. This resulted in a pooled vaccine estimate of 96% (95% CI, 95%-97%).

(3) Harder T, Külper-Schiek W, Reda S., et al. Effectiveness of COVID-19 vaccines against SARS-CoV-2 infection with the Delta (B.1.617.2) variant: second interim results of a living systematic review and meta-analysis, 1 January to 25 August 2021. *Euro Surveill* 2021 Oct;26(41):2100920. https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2021.26.41.2100920

Key Conclusion: "Current evidence shows that COVID-19 vaccines licensed in the EU are moderately to highly effective in preventing SARS-CoV-2 infections with the Delta variant, while effectiveness against severe courses of COVID-19 remains high."

Quantity and quality of studies: 17 studies included: 6 Cohort studies, 2 Screening method according to Farrington studies, 2 Serial cross-section design studies, 7 Test-negative design studies.

Estimates of benefit: Most studies reported VE >50%.

For prevention of any infection (n=16 studies), pooled VE estimate was 66.9% (95% confidence interval (CI): 58.4–73.6; I2 = 95.1%) across all studies.

For prevention of asymptomatic infection (n=2), VE estimates ranged between 35.9% and 80.2%, with pooled VE estimate across studies 63.1% (95% CI, 40.9–76.9; I2 = 93%).

For prevention of symptomatic infection (n=9), VE estimates ranged between 56% and 87.9%, with pooled VE estimate 75.7% (95% CI: 69.3–80.8; I2 = 91.9%).

Studies of COVID-19 Vaccine Effectiveness in Nursing Home and Long-term Care Staff

Two studies specific to nursing home and long-term care staff vaccine effectiveness serve to complement the conclusions of the above systematic reviews.

(1) Mor V, Gutman R, Yang X, et al. Short-term impact of nursing home SARS-CoV-2 vaccinations on new infections, hospitalizations, and deaths. *J Am Geriatr Soc* 2021;69(8):2063-2069.

This study found a protective effect each week post-vaccination of NH staff, according to calculated staff infection incident rate ratios; for example, IRR 0.85 (15% fewer COVID-19 infections) among staff 3 weeks post-vaccine clinic, and IRR 0.51 (49% fewer infections) 6 weeks post vaccination clinic.

(2) Linsenmeyer K, Charness M, O'Brien W, et al. Vaccination Status and the Detection of SARS-CoV-2 Infection in Health Care Personnel Under Surveillance in Long-term Residential Facilities. *JAMA Netw Open* 2021;4(11):e2134229.

A protective effect of COVID-19 vaccination was observed in this active surveillance study of long-term care health care personnel, which found fewer SARS-CoV-2 infections among vaccinated versus unvaccinated staff at each time period evaluated.

NQF #3633e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level)

Comment 1 by: Karen Orozco, American College of Radiology
On behalf of Karen Campos, American College of Radiology

The American College of Radiology, representing more than 40,000 radiologists, radiation oncologists, medical physicists, and nuclear medicine physicians, appreciates the opportunity to submit comment on NQF #3633e, #3662e and #3663e: Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level, Clinician Group Level and Facility level, respectively). **The ACR does not support the endorsement of NQF #3633e, #3662e, and #3663e. General Comments**. Protocol selection appropriate for a clinical indication is an important component of radiation dose management along with radiation dose optimization. Each component needs to be addressed as a separate quality action. The specific aspect(s) of performance to be improved is not intuitive due to the multiple components to the measures (size-adjusted dose, image quality, clinical indication). It is premature to measure performance on excessive radiation dose based on thresholds by clinical indication for an exam until the level of standardization and availability of national benchmarks is further along as discussed below. It is true that the most accurate way to address appropriate and safe use of multi-phase studies is to measure both the clinical indication of an exam and the radiation dose output (dose indices per exam) and look at the two separately or distinctly together. **However, these measures conflate**

the appropriateness of protocol for the clinical indication and radiation dose optimization, could be improved by adjusting protocols or by focusing on appropriateness of the ordered exam. Therefore, improvement may be limited.[1]. Dose optimization results in a quality action for facilities to adjust their protocols and is a responsibility of the team as a whole – physicists, technologists, and physicians who oversee the team at the facility. Protocol selection addresses the appropriateness of the exam for the clinical indication and other factors such as patient time on the scanner and optimal radiation dose. There are challenges with the implementation of an indications-based measure. Indications for exams do not have standardized language that could be used to track them. Most health and IT systems capture ICD-10 coding for reimbursement, but typically not enough standardized information to characterize the patient's condition. As a result, the clinical reason for performing an imaging exam is often extremely limited in the exam order. Electronic Health Records (EHRs) are notoriously incomplete with this type of information and interoperability issues exist with other software systems that might contain such information. ***A *validated* method for determining classification of studies using high-dose versus routine protocols appropriate to the indication must be incorporated into such a measure; these three measures include specifications which have not been validated.*** Please refer to the validity section below for more details. **NQF #3633e, #3662e, and #3663e deviate from international standards, like diagnostic reference levels, and lack peer-reviewed, broadly accepted consensus on global noise. For these measures, global noise is defined solely by the measure developer. Endorsing this method may encourage facilities to accept a narrow view of image quality. The ACR requests the developer further clarify the global noise table used in calculating the numerator.** The benchmark source is not transparent, and its applicability is unclear. For example, Table sp-1, Size-adjusted radiation dose and global noise thresholds by CT category, has the same global noise threshold for several CT categories, such as head low dose, head routine dose, and head high dose. Is it intentional that the same global noise threshold should be applied to both low and high dose head CTs? If the image noise thresholds are the same, the size-adjusted radiation dose thresholds should be the same, unless the scan length is remarkably different between the 3 CT categories. Additionally, current CT scanners display dose values based on either a 16 cm or 32 cm phantom for a neck scan, which must be carefully accounted for in measure performance calculations. **There is little to no acknowledgement of limitations.** These measures have multiple limitations, including the lack of widespread acceptance and implementation, and the issues with the method of measuring global noise. The developer states their company can provide the service of quantifying the measure at a cost; this should also be included as a potential limitation. The measure developer does provide specifications for other entities to implement the measure, but the burden of implementation may be significant. Finally, the author cites publications from their group to justify the benchmarks, but they have not been vetted through a broader consensus process. **The ACR strongly encourages the Patient Safety Standing Committee to re-vote on the scientific acceptability of these measures based on the following concerns.**

Validity/Feasibility. These eCQMs require multiple variables that may be captured in software systems external to electronic health records (EHRs), such as dictation systems housing radiology reports or DICOM standard-based systems, such as CT device software. Data element validity testing should demonstrate that the testing sites were able to integrate and validate the variables used to construct the data elements used by the eCQM in addition to the usual validation of the eCQM's electronic output against the medical record review. **We are uncertain that this validation has been completed. Therefore, this submission does not demonstrate the measure can be reproduced in a reliable and valid manner by practices or facilities across multiple settings.** For example, for CT category (or other elements deriving/collecting data using custom natural language processing (NLP) tools), the developer used NLP for obtaining data such as reason

for study or protocol name used in the calculation of this variable. The submission does not provide information on the NLP results' reliability and validity. Because **this comparison of the NLP-derived data against a medical record review was only completed in a sample from one site (UCSF Health System), there is uncertainty whether the results are generalizable across EHRs or other databases**. These measures rely on custom made NLP trained and validated on a small group of pilot sites; it is not clear whether this type of NLP would work outside these sites nor how sites would get access to use this custom NLP tool. Testing information does not demonstrate adequate validation of this critical data element. Additionally, **sufficient evidence should demonstrate that the definitions/variables used are valid and do not rely on one study or use in a single system, such as what is provided to support the thresholds of "out of range" performance values.**While the process to determine these thresholds is detailed, we do not believe that a Technical Expert Panel (TEP) conclusion in the absence of independent data validation is sufficient. M**ultiple unstructured variables are required to construct the data elements for the numerator, denominator, and exclusions. Assessments of the feasibility of the integration of these unstructured data into the measure calculations would be useful to ensure that the underlying data can, in fact, be integrated if practices and facilities that choose not to use the edge device.** For example, the level of effort required to integrate the Binning algorithm for the CT categories and ensure that the results are reproducible and valid remains unclear. The ACR is concerned with the selection bias for the accountable entity-level (measure score) validity. **Assessing measure score face validity through the TEP that created these measures lessens the extent of credibility for these results**. Although the TEP is knowledgeable and represents a variety of stakeholders, there is a vested interest in ensuring these measures are available for use. **Most importantly, as one of the TEP members noted in the survey, the performance score from these measures does not clearly indicate what corrective action needs to be taken by the clinician, clinician group, and/ or the facility to improve performance**. **Usability** While implementing these measures as specified may not impose a substantial burden on clinicians, **it may necessitate substantial organizational effort to access and process the data elements required to calculate the measure score**.The measure steward states that their software is available on a non-commercial basis to calculate this measure, and that other vendors may also develop their own software to implement the measure specifications using the information included in this submission. Will the measure steward review other vendors' software to ensure comparable calculation methods? Measure stewards frequently make specifications available "as is" without warranty, leaving it to the implementer to appropriately update any software or tools as measure specifications are changed. But the complexity of these measure specifications may warrant greater oversight. External vendor software will need to be maintained and updated to ensure the software's accuracy and reflect any changes in specifications and coding. **For all the reasons stated above, the ACR does not support the endorsement of these three measures.** We thank the NQF staff for their transparent endorsement process. **Reference: 1. '**Mahesh M. Benchmarking CT Radiation Doses Based on Clinical Indications: Is Subjective Image Quality Enough?Radiology. 2021 Nov 9:212624. doi: 10.1148/radiol.2021212624. Online ahead of print. PMID: 34751622

Comment 2 by: Angela Keyser, American Association of Physicists in Medicine

What is AAPM:

The American Association of Physicists in Medicine (AAPM) is the primary scientific and professional organization of physics in radiology and radiation oncology in the United States. The mission of AAPM is advancing medicine through excellence in the science, education and professional practice of medical physics; a broad-based scientific and professional discipline which encompasses physical principles with applications in biology and medicine. With 9717 members in

94 countries, AAPM supports the Medical Physics community with a focus on advancing patient care through education, improving safety and efficacy of radiation oncology and medical imaging procedures through research, education and the maintenance of professional standards. AAPM has a staff of 33 and an annual budget of $10.7M, and is located at 1631 Prince Street, Alexandria, VA 22314.

AAPM comments on the proposed measures:

AAPM does not support the endorsement of NQF #3633e, #3662e, and #3663e.
This application proposes electronic clinical quality measures (eCQM) that monitor CT performance to discourage unnecessarily high radiation dose while maintaining adequate image quality. The proposed metrics require CT Category (i.e., the CT exam type), the size adjusted radiation dose [the patient's dose length product (DLP) adjusted by patient size], and the global noise (associated with the variance of the voxel values in CT images). The two reported measures are the percentage of eligible CT cases in a particular category deemed to be "out-of-range" compared to defined thresholds with respect to the size-adjusted radiation dose or the global noise in a set time period. While efforts to enhance consistency of CT practice are noble and include initiatives by AAPM and others worldwide, the proposal has significant limitations that impact its scientific and practical value and overall likelihood of clinical acceptance. These limitations include improper representation of image quality, improper estimation of radiation risk, and substantial oversimplified representation of implementation in practice, including not addressing the challenges of implementation. The authors indicate that their company (Alara Imaging, Inc.) can provide the service of quantifying the measures at a cost. A steward of measures requires an extensive track record for scientific and technical expertise and policy making that represents a broad consensus of the community. These important elements should be carefully reviewed within this application. One cited reference supports the proposed measure, however, this cited article has an accompanied editorial that highlights the limitations of the proposed approach [Mahesh M. Benchmarking CT Radiation Doses Based on Clinical Indications: Is Subjective Image Quality Enough? Radiology. 2021 Nov 9:212624. doi: 10.1148/radiol.2021212624. Online ahead of print. PMID: 34751622]. The editorial and stated limitations are not addressed in the proposal.

The AAPM agrees that effort needs to be continually placed on ensuring diagnostic quality CT imaging, optimizing CT dose, and achieving consistency across facilities, considering differing technologies and practices. The non-profit entities of the AAPM, the American College of Radiology (ACR), and Image Wisely and Image Gently Alliances have spent decades towards this goal and continue to do so through many initiatives. Among them, the non-profit ACR CT Dose Index Registry (DIR; https://www.acr.org/Practice-Management-Quality-Informatics/Registries/Dose-Index-Registry, established in 2011) has the significant stature of implementing a dose registry that enables facilities to compare dose indices nationally, to ensure the highest quality imaging with lowest possible dose. The ACR CT DIR implementation incorporates the expert, consensus opinions of the medical imaging community. ACR dose optimization measure recently endorsed by NQF provides a further valuable measure to manage imaging radiation dose (https://www.qualityforum.org/QPS/3621). The imaging community's valuable clinical benchmarks greatly benefit from consensus decisions based on sound scientific and technical review and discourse. The proposal herein should be carefully reviewed for any additional contributions or advantages it would provide to our existing robust consensus measures and resources, such as available with the ACR.

After a detailed review of the measures by multiple expert members of the AAPM, we have concluded that the **AAPM does not support the endorsement of NQF #3633e, #3662e, and #3663e.** This position stems from eight major concerns about the proposed measures:

1) Unscientific characterization of CT scan risk: The proposal is based on estimation approaches that are not reflective of the consensus of the scientific community and do not acknowledge the uncertainties of the estimates. A NQF measure focused on radiation risk should uphold scientific objectivity, integrity, and responsibility not evident in the presentation and assessment of radiation risk in this proposal.

2) Inactionability of the measures to enable targeted change to improve practice: It is not evident how the proposed measures can be practically used to improve imaging practice and exactly how a facility can do to achieve compliance, given the wide varieties of factors and technologies involved.

3) Inadequate addressing of the complexity of CT categorization: The proposal does not address the magnitude of this challenge nor has suggested means to overcome it given that current standards are even lacking in uniform characterization of protocols. Inaccurate classification of data can lead to significant and misleading errors.

4) Inadequate assessment of noise: Noise in a CT image can be influenced by a variety of factors including justified differences in CT technologies including new reconstruction methods that dramatically alter noise. Further, noise does not have a singular value in a CT exam. A "global noise" ignores this diversity and can misrepresent the quality of an exam.

5) Inadequate assessment of image quality: Image quality is affected by a myriad of factors including resolution and contrast, as well as the intended purpose of the exam. A singular representation of image quality via global noise overly simplifies this space and can lead to gross misrepresentation of image quality and thus mis-service to patient care.

6) Flawed assumption on dose reduction vs dose optimization: The application focuses primarily on radiation dose reduction as oppose to right-sizing the dose for the best care of the patient. Individualization and optimization of care and safety should be the goal not minimization. This approach can lead to some patients getting under exposed, leading to missed diagnosis, while others may be over-dosed for their exact need and condition.

7) Inadequate accuracy in patient size estimation: Assessing a patient size is not a trivial task, stemming from significant variability in the differences in the habitus of different patients, coupled with the existential challenge that there is no single metric capturing the size of a patient of varying diameter at different cross-sectional locations. Algorithms are continuously evolving and no evidence is provided that the company can do this task with sufficient accuracy.

8) Limited expertise and track record of the company: The company is a new (2020) company with no experience of having previously performed a project of such wide scope, scientifically or technically. There is no scientific track record on CT technology, size estimation, or image quality assessment for the company to be considered a steward of measures on which there is a lack of expertise, publication, and scientific history.

These concerns are detailed specially in our complete review submitted via email to patientsafety@qualityforum.org, along with selected specific observations on the proposal on January 19, 2022.

The AAPM recognizes that this topic is complex, including scientific, technical and clinical components. We welcome the opportunity for greater in-depth discussion on meaningful measures of quality imaging practice.

Respectfully submitted,

Comment 3 by: Bradley Delman, Mount Sinai Health System

I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. As an implementation testing partner, I coordinated Mount Sinai Health System's inclusion in the test. To summarize, after installing the data collection software, we routed CT imaging data from PACS and sent order and billing data from various electronic systems to the software. We also worked with UCSF and our CT vendors to ensure the Radiation Dose Structured Report (RDSR) was being saved for each exam sent to PACS. As we discussed in our interview with UCSF, this work fell on the PACS team and IT colleagues, without requiring effort from clinicians above my initial planning and coordination. Besides some technical challenges, which were all resolved, we faced few barriers to successful implementation and had very little missing data. In total we submitted 11,588 scans, representing just over 3 weeks of CT data from our health system. Based on our experience, the participation in the proposed quality measure is feasible. However, I suspect that spirited engagement from PACS, RIS and/or EHR vendors would greatly enhance participation and timely provision of data. We have also been satisfied with the feedback we've received from Alara Imaging on our measure performance, which brought to our attention areas of high radiation dose. This feedback has identified individual exams as well as imaging protocols that contribute high radiation dose. Although we have been a dose-conscious department, the feedback highlighted areas of variability in both routine and size-adjusted datasets. Furthermore, we learned which protocols and classes of studies fell within and beyond expected range for dose, and how dose can vary between scanners for protocols with the same name. We also learned that some types of studies may need to be renamed or reclassified for appropriate grouping of results. A quality measure the quantifies dose while ensuring preservation of imaging quality can help mitigate the use of excessive radiation doses used in CT. I support the work of the measure developers to improve patient safety and CT quality.

Comment 4 by: Daniel Hirsch

I write in support of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. They are important proposals that would markedly reduce unnecessary radiation exposures in medicine, and the cancers induced therefrom, while providing the same yield of diagnostic information. Many, many lives could thus be saved were the proposals adopted. I have spent much of my professional career attempting to reduce the risks to public health from ionizing radiation associated with nuclear waste, reactor accidents, nuclear weapons tests, uranium mining and milling, and radioactively contaminated sites involved in the production of nuclear weapons and other nuclear activities. It is with some alarm that I have viewed in recent years the extraordinary increase in public exposures to ionizing radiation associated with the remarkable escalation of exposures in medicine, largely due to ever-more frequent CT scans, resulting in doses from medical procedures now dwarfing exposures from the nuclear activities that have so long concerned me. The proposals made by UCSF would help reign in unnecessarily high radiation doses from these medical procedures while still producing the diagnostic information needed by physicians for their patients. The important revelation in the studies cited in the proposals is that the doses currently received by patients in these procedures are frequently very much higher—often ten times higher—than necessary. One can get the same medical benefit from the procedure at one tenth the cancer risk. The proposals indicate that many thousands of

unnecessary radiation-induced cancers could be avoided were CT scans kept to the minimum level necessary to produce the required image. This seems quite correct. The National Academies of Sciences, Engineering and Medicine has produced over the years the primary studies on the matter of ionizing radiation and cancer induction. The most recent Biological Effects of Ionizing Radiation study (BEIR VII) estimates a risk of 1.17 cancers per 1000 person-rems of exposure, and concludes, as have all the BEIR studies, that there is no threshold below which there is no risk. All radiation protection agencies (e.g., US EPA) have adopted the BEIR conclusions. Currently, exposures to medical radiation are estimated as averaging about 350 millirem/year per person. Given that degree of exposure, and the current U.S. population, medical radiation would be estimated to produce many millions of cancers over the population's lifetime. Reducing unnecessarily high exposures while still producing the necessary diagnostic image could thus prevent a very large number of cancers and deaths, while, not incidentally, also reducing Medicare expenditures for their treatment. I strongly urge adoption of quality measures that assure CT exposures use the lowest reasonable doses necessary for the procedures. Daniel Hirsch retired Director of the Program on Environmental and Nuclear Policy at University of California at Santa Cruz

Comment 5 by: Dawn Ritzwoller

I am a college student and Environmental Biology (E-bio) major, and I am pediatric cancer survivor. I am writing today in support of this radiation dose quality measure. Beginning ten years ago, and both during and after I finished treatment, I received multiple CTs (to multiple parts of my body) as part of my diagnostic and follow-up care. Not once during this period, did any of my doctors or other, discuss with me the downstream risk of all of the radiation exposure I experienced. It was only years after my treatment ended, and now via classes I have take for my E-bio major, that I am beginning to understand the risk associated with radiation exposure. What is also now clear to me is the importance that providers use the most appropriate (low) dose for the specific diagnostic or follow-up exam. I know that image quality is important for diagnosis, but patients (like me) need the confidence that their doctors and hospitals are using the best and lowest dose possible for the exam that they order. Thank you!

Comment 6 by: Debra Ritzwoller

I am writing in support of this important measure. I am a cancer health services researcher *and* a mother of a pediatric cancer survivor. It is well documented in the literature that there has been a significant secular increase in CT use within and across most patient populations. While CT use, and therefore radiation exposure has increased over time, I know that personally and professionally that excessive radiation dose remains a significant quality issue, and it is one that is often not adequately addressed by researchers and healthcare providers/delivery systems. This quality metric is necessary now, in order to provide the incentives and the resources needed to generate the metrics and the benchmarks that may actually influence practice that may in turn translate into a meaningful reductions in the radiation dose that patients are exposed to. This metric is designed to address the clinical indication associated with the respective exam, rather than just the type of advanced imaging that is performed. The measure is also constructed to ensure that the dose benchmarking does not adversely impact the quality of the metric. Given the noted harms of CT based radiation exposure (e.g USPSTF Lung Cancer Screening "B" recommendation), this measure is addresses a timely and needed quality metric.

Comment 7 by: Ehsan Samei, Duke University – Margolis Center for Health Policy Center

Duke University, Ravin Advanced Imaging Laboratories (Ravin Labs) and Clinical Imaging Physics Group (CIPG), Durham, NC 27710 The Ravin Labs is a 50-member leading translation imaging research laboratory in the country with over 30 years of history. The lab conducts rigorous NIH-funded research with an additional mandate to practice its science through CIPG, an imaging physics group of 15 experts dedicated to quality and safety in the practice of radiology. The group, highly integrated into the clinical domain, has devised and put to practice imaging dose and image quality monitoring systems at the level of individual patients within the Duke University Health System with additional pilot installations at MD Anderson Cancer Center and Stanford University. The group has published extensively on its technology and findings (upward of 500 papers), with over 30 referred publications on dose and quality monitoring alone. The effort has led to significant reduction of patient radiation dose at our facilities and right-sizing it per individual needs of patients. **We do not support the proposed measures.** The rationale is detailed below. **Overall:** While we applaud the effort to introduce new quality measures in the practice of medical imaging, the proposed electronic clinical quality measures (eCQM) are misleading and overly simplistic leading to significant unintended consequences. The limitations stem from the fact that the proposed risk measures are based on CT scanner output and not the actual dose burden to individual patients at the organ level, the quality measure is based on noise alone ignoring the multi-faceted reality of diagnostic quality, and lack of methods that standardize protocols across vast diversity of examinations. There is significant ambiguity in the exact method used for noise and size estimation with no track record or peer review of otherwise black-box methods. This approach will likely produce measures that can be orders of magnitude off from their actual values, and therefore lack clinical relevance and fidelity. Measures can lead to misleading and erroneous conclusions while also potentially jeopardizing the use and development of better approaches, as inaccurate low-bar measures can prevent accurate ones in the future. But most importantly, the measure can lead to unintended consequences and even harm the patient. For example, an imaging team can take an action that is not in the best interest of a patient, like applying too little dose for some patients such that disease would be missed, a “wasted dose” with no medical benefit and health and cost consequence of a miss. Conversely others might get more radiation than needed as the measures do not account for individual patient needs and tasks. Improving consistency in imaging practice is a laudable goal that needs a proper solution anchored to scientific understanding of radiation risk, image quality need of patients, diversity of practices, and the CT technology. The proposal is lacking on all these four fronts. A solution to inconsistency in images can only be brought forth through a broad consensus of the scientific and practicing communities (including ACR, AAPM, Image Gently, and Image Wisely), CT manufacturers (represented by MITA), standard methods of data categorizations and measures (supported by the medical community), and evidence-based radiation risk and image quality measures at the level of indication and organ where they are actually relevant to the individual patient. A for-profit company with no track record or transparency of its methods cannot be considered a steward of such a space. Below we further detail 12 concerns regarding the proposed measures:

1. **Inadequate attention to image quality:** The measures are heavily dose related, emphasizing this over measures of quality. Dose and minimizing it is important but equally important is image quality as an inadequate image quality would be a dis-service to the patient regardless of the dose. This is explicitly stated in the International Commission of Radiological Protection (ICRP) in Publication n. 135.

2. **Inaccurate assessment of radiation risk:** The measure of size-adjusted radiation risk, adjusting the CT scanner outputs with ‘patient size’ to perform risk estimation is not a standard method nor endorsed by any scientific or professional body. The method is in fact

explicitly **discouraged** by the AAPM Task Group 204. Patient risk can only be assessed with the knowledge of organ doses that is not even mentioned in the application let alone pursued. The proposed method CANNOT be used as surrogate for future cancer risk.

3. **Incomplete/Inaccurate representation of image quality:** The measures include image noise. Yet, noise is just one component of image quality. For example, the noise of an image can be fine but image quality totally inadequate. And conversely noise can be too high but image quality totally adequate. To assess image quality properly, one should include the actual task at hand (eg, detecting a pancreatic cancer vs bowel obstruction vs kidney stone) as well as other equally important facets of quality, like noise texture, resolution, and contrast. These factors have not been even mentioned let alone tackled in this application. Focusing on noise as a singular metric of quality can lead to major mis-representation of the needs of a quality and safe imaging practice.

4. **Neglecting the impact of image rendition:** Critical and relevant to clinical practice, the measure of noise proposed does not take into consideration how differing reconstruction algorithms and parameters affect noise (up to 200%). Without considering this influence, a measure of noise as proposed is irrelevant and misleading.

5. **Subjectivity:** The measures are anchored to subjective perception by radiologists as how they "like" the images. There is in fact no evidence provided that the measures can lead to an improvement in diagnostic accuracy. In fact, it might lead to a degradation.

6. **Lack of integrating dose and quality:** There is no indication as to how image quality is linked to radiation dose and at what level; or instance, how they propose to manage multiple reconstructions of the same exposure event.

7. **Not addressing the multiplicity of exam components**: A CT exam often includes multiple phases (series) each of which has a noise and radiation dose of its own. Averaging noise across series is meaningless. The measures do not recognize or account for this multiplicity and diversity.

8. **Under-recognizing the diversity of exams:** The measures do not address the notable diversity of exam nomenclature across institutions and practices. This is a significant component of any dose or quality monitoring system. Without a standard for CT protocols, which cannot be devised by a for-profit company without consensus of manufacturers and users, the data can be mislabeled and mishandled leading to major errors in the results and subsequent negative effect on mis-dosing and mis-diagnosing patients.

9. **Inaccurate assessment of patient size:** The measure of size proposed is calibrated to earlier work and publication from our group at Duke University for academic purposes. That early method they have embraced has had major errors (upward of 300% is certain applications) that have been corrected in subsequent versions that have not been shared. Without essential newer refinements to assure fidelity, the company cannot be a responsive steward of the measure that it has had no expertise to advance or maintain.

10. **Inaccurate assessment of noise:** The measure of noise proposed references earlier work and publication from our group at Duke University. That early method exhibited errors, corrected in subsequent versions that have not been shared. Without essential newer refinements, the company cannot be a responsive steward of the measure that it has had no expertise to advance or maintain.

11. **Lack of guidance toward compliance:** To us it is difficult to defend (1) measuring imaging practices based on ambiguous and questionably-relevant metrics promoted to

represent the actual safety or quality of CT practice, and (2) not offering any guidance as to how a practitioner responsible for "outlier" examinations can bring their practice to the proposed definition of compliance. Together, these can easily create signification confusion and potential disruption in the imaging practice

12. **Lack of support from manufacturers:** Having worked in dose and image quality monitoring for over a decade, academic centers of excellence, including ourselves, have a close connection with major CT manufacturers including MITA, Medical Imaging Technology Alliance, which comprises all CT manufactures. Our discussions regarding this measure lead us to believe that there will be little support from scanner manufacturers for a non-transparent and unpredictable product that lacks maturity from a private for-profit entity. There are substantial differences in image processing, detector efficiency, and such across scanners that will have significant bearing on the CT image. The proposed measure does not account for such important nuances, leading to erroneous results.

Comment 8 by: J. Leonard Lichtenfeld

I am pleased to provide this comment in support of NQF quality measures 3633e, 3662e and 3663e. These comments reflect my personal opinion and not any other organization with which I may be affiliated. CT scans have assumed a primary role in the evaluation and diagnosis of many medical conditions, and are very commonly performed procedures. Less appreciated by the public and many professionals (including non-radiology physicians) is the variation in image quality and dose that has been recognized for many years by researchers who have evaluated these factors. As such, there can be substantial variation in CT scan dose and quality, even within the same institution. As a patient, this consideration has figured prominently in my own decisions as to whether or not to proceed with serial CT scans for follow-up of medical conditions. These measures have been carefully crafted to create an effective and validated method to monitor CT image and quality based on indications for the studies and in consideration of individual patient-related variables. As such, they provide a useful and meaningful way to offer our patients and the public the assurance that the scans they are receiving meet reasonable safety and professional standards--which is not routinely available otherwise. These quality measures will meaningfully improve the ability of physicians and health systems alike to monitor the equipment utilized for these studies in a manner that minimizes interference with the typical workflow of a radiology center (or other center) where such studies are performed and will provide a significant and substantial increase in the quality of scans while reducing dose variability that can occur because of machine settings/performance or patient characteristics. Cumulative radiation dose should decline as a result of implementing these measures. At the very least, there will be assurance that the right dose is used for the right scan in the right patient. As a physician and patient advocate for many years, I offer my support for these measures for the reasons stated. And as someone who served as an advisor for this measure, I will add that I was impressed by the exceptional commitment of the developers and their colleagues to provide a meaningful, validated and effective quality measure as they created new processes to measure CT dose and quality, always with an eye towards making this measure acceptable to the professional and consumer communities. (Disclosures: As noted, I was an advisor during the development of this measure and received compensation for those services. I have also served on the NQF Cancer Committee without compensation. I have no other relevant conflicts.)

Comment 9 by: James Anthony Seibert, University of California, Davis Medical Center

January 27, 2022 To: National Quality Forum Dear NQF Standing Committee, I am writing to lend support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco, where I have served on the Technical Expert Panel and have been a long-time collaborator for similar projects between UCSF and UC Davis. I led the implementation of measure testing at my institution, University of California Davis Health, which required local installation of the software, configuring connections to the PACS, extracting CPT and ICD-10 data from the EHR, and supervising the aggregation and transfer of all this data to the UCSF software. Most of this work was completed by our PACS administrator and did not impact the work of our clinicians at any time. One challenge we encountered was that transfer of data from PACS to the software was slow; we believe this was due to capacity limitations of our PACS relative to the query-retrieve process. Nevertheless, we set up auto-transfers of the data over nights and weekends so as not to impact the operation of our PACS during our busiest clinical hours. Besides this issue, the testing was completed successfully with minimal missing data. Based on our experience, the proposed quality measure is highly feasible, and will, in my opinion, be able to appropriately identify CT exams that are significantly above diagnostic reference level (DRL) doses(*), as well as inadequate CT exams with insufficient dose, for specific diagnosis indications versus radiation dose versus image quality. There are certainly many parameters and issues that can potentially confound such CT quality measures, particularly with the assessment of corresponding image quality, but significant advances in developing robust algorithms to recognize such confounding factors have largely mitigated such concerns. I believe this quality measure can significantly reduce the use of excessive high radiation dose as well as inadequate, sub-optimal low dose used for clinical CT studies, by identifying outliers and thereby increasing the awareness and importance of CT protocol optimization. I support the work to improve patient safety and CT quality as described in these measures. Sincerely, J. Anthony Seibert, PhD, FAAPM, FACR, FSIIM, FIOMP Professor Emeritus, Department of Radiology UC Davis Health (*) Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology 284(1), 120-133, 2017. Disclosure: I have served on the Technical Expert Panel for this effort and have received some minor compensation for participation (honoraria) but have no other relevant conflicts. The opinions expressed here are my own.

Comment 10 by: Kenneth Wang

I am pleased to provide my support for the proposed CT quality measures 3633e, 3662e and 3663e developed by the University of California, San Francisco. I have been a practicing radiologist in the Veterans Affairs (VA) system for more than ten years, during which time I have led efforts in CT dose optimization across the VA Maryland Health Care System. I also serve in a number of volunteer roles within the Radiological Society of North America (RSNA) and the American College of Radiology (ACR), leading efforts in informatics, standards, interoperability and registries. However, this letter reflects my personal opinion, and not necessarily those of any organization with which I am affiliated. I have also served as a member of the Technical Expert Panel (TEP) advising on the formulation of these proposed quality measures, since the inception of this project.

The impetus for this work rests on fundamental principles which are widely accepted. Namely, that CT constitutes an important source of radiation dose to patients, and that CT imaging presents an opportunity for dose reduction, but that it is of paramount importance to maintain the diagnostic quality of the imaging obtained. The proposed measures have been developed using a scientific approach incorporating extensive testing and validation, as well as expert consensus, while

maintaining a focus on practicality. This has been all the more impressive given the complex nature of the technical factors involved, such as CT exam types, size-adjusted dose, and diagnostic image quality. By leveraging extensive data, including but not limited to data in the UCSF International CT Dose Registry, data obtained from practicing radiologists on image quality, and feedback from testing facilities, the measures strike a practical balance intended to identify opportunities for CT dose reduction while maintaining a floor for diagnostic quality (which was rarely violated in measure testing).
As such, these measures represent an important step beyond simple dose reduction. I also believe that these measures will provide actionable feedback, especially given the many different techniques now available on modern CT scanners for dose adjustment.
As a radiologist, I know there will never be universal agreement on subjective assessments such as image quality. However, the proposed measures take a balanced approach, informed by extensive testing and validation, which serves a very practical and important quality objective. For these reasons, I support the adoption of these measures.

Comment 11 by: Krishna Nallamshetty, Radiology Partners

I would like to submit a comment in support of this measure. I am a practicing radiologist for the past 15 years and serve as the Associate Chief Medical Officer of Radiology Partners, the largest medical imaging practice in the United States. I am the chair of our national Patient Safety Committee. We have seen tremendous growth in medical imaging that requires radiation, specifically computed tomography (CT). The public awareness of the potential long-term effects of ionizing radiation has become mainstream and as a result, a primary objective of the American College of Radiology and other governing bodies. The objective focuses on reducing radiation exposure as much as possible without compromising the diagnostic information that is obtained
We have recognized that there is large variability in how CT scans are acquired all over the country. Techniques and radiation exposure is extremely varied but yet appropriate clinical diagnosis are made. This measure evaluates radiation dose for every patient who undergoes CT *based on the clinical indication for imaging* rather than solely on the type of examination that is performed. It ensures patients receive the most appropriate CT acquisition protocol and level of radiation for their individual condition. The measure also assesses image noise, safeguarding image quality against potential effects of dose reduction, and is the first quality measure to do so.
The measure would have a large, positive impact on patients and protect them from unnecessary over-exposure of radiation without compromising the diagnostic value of medical imaging. It would be the first time a measure addresses both radiation and image quality.

Comment 12 by: Maribel Escobar

Submitting on behalf of ARA's CMO, Dr. John Kish: January 25, 2022 Dear NQF Standing Committee, I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. As an implementation testing partner, my institution, ARA Diagnostic Imaging, was required to install the data collection software, route CT data from PACS and order and billing data from various electronic systems to the software, and oversee the migration of data. We also worked with UCSF and our CT vendors to

ensure the Radiation Dose Structured Report (RDSR) was being saved from each exam in the PACS. As we discussed in an interview with UCSF, this work fell on the PACS team and IT colleagues and did not require effort from clinicians. Besides some technical hiccups, which were all resolved, we faced few barriers to successful implementation and had very little missing data. Based on our experience, the proposed quality measure is highly feasible. We have also been satisfied with the feedback we have received from Alara Imaging on our measure performance, which brought to our attention some areas of opportunity to decrease radiation dose. The feedback provided by Alara Imaging has taken the burden of researching problem areas away from my institution, by identifying specific exams, imaging protocols and even specific CT units that contribute to high radiation dose and need improvement. We have plans to address each accordingly. Given our positive experience, my organization is moving towards a commercial relationship with Alara to continue to submit data, receive feedback, and strive to optimize our CT doses. I earnestly believe this quality measure can help mitigate the use of excessive radiation doses used in CT. I support these measure developments in order to improve patient safety and CT quality. Sincerely, John Kish, MD Chief Medical Officer

Comment 13 by: Mary White

I am writing in support of this CT radiation dose safety measure. As a cancer epidemiologist, I recognize that excessive exposure to medical radiation increases cancer risk. And I understand that this measure will be valuable for protecting patients from unnecessarily high levels of radiation from CT imaging. The measure is designed to evaluate radiation dose for every patient based on the clinical indication for imaging. The measure also assesses image noise, ensuring adequate image quality despite the reduction in radiation dose. This measure fills an important quality void and has the potential to substantially reduce the contribution of CT scans to the incidence of cancer in the population.

Comment 14 by: Matthew Nielsen

I am writing in support of this important measure. The utilization of CT imaging in the United States has dramatically increased over recent decades, providing numerous benefits to patients and clinicians in the management of countless medical conditions. There has also been increasing recognition of the potential for unintended harms due to potentially avoidable variation in radiation in radiation dose for many patients. Evidence from research and quality improvement efforts demonstrates the potential to mitigate these harms with a feedback loop and benchmarking to radiologists and staff. This measure provides needed resources to disseminate these early successes, preserving the benefit of advanced imaging with CT while providing a means for healthcare facilities and clinicians to improve the safety of the studies they provide patients. The design of this measure importantly takes into account the indication for the study as the framework for dose benchmarking, with balancing measures of image quality to assure that efforts to reduce dose do not come at the expense of diagnostic quality. Given the increased recognition from patients and providers of the potential harms of imaging-associated radiation, this measure fills a timely and important gap in the current measurement portfolio.

Comment 15 by: Pavlina Pike, Huntsville Hospital

I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. I am a Medical Physicist and

Radiation Safety Officer at Huntsville Hospital and led the testing of UCSF's quality measure at my health system, which involved installing the data collection software, routing CT data from PACS and order and billing data from various electronic systems to the software, and overseeing the migration of data. We came onboard late in the testing period, leaving a tight window of time to collect the data prior to UCSF's submission deadlines. I am proud of my PACS and IT colleagues for pulling together so efficiently and completing the work rapidly with very little missing data. The work in no way impacted our physicians or clinical workflows. We faced few barriers to implementation, and based on our experience, the proposed quality measure is highly feasible.

We have also been satisfied with the feedback we've received from Alara Imaging on our measure performance, which brought to our attention areas of high radiation dose. Our exams were compared to thresholds established based on input from 125 radiologists and 50,000 CT examinations from other facilities. The analysis includes comparisons of the performance of different model CT scanners, exams, protocols, patient size, facility, etc. The feedback from the Alara software is helpful and actionable as we are able to identify what changes will have the greatest impact on patient dose and make the appropriate changes. In addition it provides suggestions for billing inconsistencies which was very helpful to our administration.

I earnestly believe this quality measure can help mitigate the use of excessive radiation doses used in CT. I support the work of the measure developers to improve patient safety and CT quality.

Comment 16 by: Robert Gould, University of California, San Francisco Medical Center

I am writing as a physician who has worked for decades as a leader in Physicians for Social Responsibility, as well as the International Physicians for the Prevention of Nuclear War toward eliminating nuclear weapons, cognizant of the public health dangers of radiation initially derived from studies of victims of the twin atomic bombings in Japan. Informed by the central tenet of physician practice to "at first do no harm," I strongly support CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. While my long experience as a practicing pathologist has made me understand at a profound level how diagnostic radiation is a critical tool in medical practice, it has also underscored to me the often-overlooked risks of carcinogenesis that must always be balanced against the benefits of various radiological procedures. Over time, research has documented that many radiological procedures are medically unnecessary when information that is desired can be obtained by other means than exposing a patient to ionizing radiation; it is also unwarranted when employed as a "hedge" against possibility of malpractice litigation. In addition, when radiological imaging is indeed required and justifiable, it is not uncommon, where standards are not uniformly applied in practice, for radiation exposures to exceed what would be required for achieving images satisfactory for diagnostic purposes. As such, the lack of attention to standardizing, and minimizing exposures inevitably results in the induction of significant numbers of unnecessary cancers that would not occur if lower doses were employed to achieve adequate imaging. I believe that CT quality measures 3633e, 3662e, and 3663e would be important steps to assuring that physicians can obtain the information necessary from diagnostic imaging while minimizing the number of unnecessary cancers induced by the procedures.

Comment 17 by: Suz Schrandt

As a patient advocate with significant experience navigating the healthcare system--including repeated exposures to a variety of diagnostic imaging studies--I submit these comments in endorsement of this measure. The measure takes into account different contexts and parameters for a given patient and his or her unique benefit/risk profile. At a more foundational level, the

measure calls into focus the significant variation in practices in CT imaging that can expose patients to unnecessary and/or unsafe levels of radiation, a risk many patients are not even aware of. The wide-spread use of this measure could standardize imaging practices and should the measure be adopted, I strongly encourage a robust dissemination plan to inform patients and families of its existence. Our ability to access safe and effective care should not be left to change; measures such as this help to close key gaps in our system.

Comment 18 by: Melissa Danforth, The Leapfrog Group

Founded in 2000 by large employers and other purchasers, The Leapfrog Group is a national nonprofit organization driving a movement for giant leaps forward in the quality and safety of American health care. The flagship Leapfrog Hospital Survey collects and transparently reports hospital performance, empowering purchasers to find the highest-value care and giving consumers the lifesaving information they need to make informed decisions. For the past several year's Leapfrog has been collecting and publicly reporting hospital performance on an NQF-endorsed Pediatric CT Radiation Dose (NQF 2820) measure. The new Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level) fills a critical gap in evaluating radiation dose for adult patients who undergo CT. Additionally, because the measure is based on the clinical indication for imaging – rather than on the type of examination the radiologist chose to perform – it can help ensure patients receive the right type of CT and amount of radiation for their individual condition, which is a primary concern of Leapfrog and our purchaser and employer membership. The measure also assesses image noise, safeguarding image quality against potential effects of dose reduction, and is the first quality measure to do so. Leapfrog strongly supports this measure.

Comment 19 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

Comment Part 1:

We thank the American College of Radiology for their comments from 1/19/22 but wish to address several factual inaccuracies in the comments. (Reponse PART 1) **Comment:** *Indications for exams do not have standardized language that could be used to track them. Most health and IT systems capture…coding for reimbursement, but typically not enough... As a result, the clinical reason for performing an imaging exam is often extremely limited in the exam order... A validated method for determining classification of studies .. must be incorporated into such a measure.* **Response:** This statement indicates that the commenter does not understand how clinical indication is determined in the proposed measure. It does not rely on the clinical reason for performing an imaging exam in the exam order. As described in Specifications, sp-11, clinical indication for imaging is determined using an algorithm that combines procedure (CPT®) and diagnosis (ICD-10-CM) codes associated with the clinical visit when the test was ordered, information provided as part of the order, and information on the final bill. The codes are available in the radiology electronic systems and/or the EHR or billing systems. The goal in creating the CT categorization decision rules was to identify exams that are exceptions to the routine dose category (i.e. either high or low dose). The approach of assigning CT exams to the various CT categories in an automated fashion using an algorithm was developed using over 4.5 million CT exams in the UCSF International CT Dose Registry. We confirmed that the CT categories were representative of groupings that require different radiation dose and image quality (Smith-Bindman 2021). The algorithm was validated using over 10,000 patient records from UCSF Health. The CT category assignment determined by the algorithm was

compared with a "gold standard" chart review, as described in Validity sections 2b.02 and 2b.03. Since we did not have access to complete medical records at testing sites, we developed a second referent standard that determined CT category based on natural language processing of DICOM data and the full radiology report. This second referent standard was found to be accurate compared to the gold standard chart review of the same sample of UCSF Health exams (sensitivity = 0.92, specificity = 0.97; see 2b.02). When the algorithm was deployed at testing sites, the correct classification rate of CT category assignment was on average 92% across clinician groups and hospitals and 95% in individual clinicians (see 2b.03). Knowing that the algorithm was developed using data from a single health system, we performed detailed investigation of the categorization results at testing sites – comparing the assigned CT category against full radiology reports – for the purpose of improving the algorithm, which we did. **Reference:** Smith-Bindman R, Yu S, Wang Y, et al. An Image Quality-informed Framework for CT Characterization. Radiology. 2021 Nov 9:210591. **Comment:** *The developer states their company can provide the service of quantifying the measure at a cost; this should also be included as a potential limitation. The measure developer does provide specifications for other entities to implement the measure, but the burden of implementation may be significant.* **Response:** This is inaccurate. As stated in Feasibility, 3.07, there are no fees for users submitting their eCQM data to CMS programs. The eCQM can be run and the measure score calculated by any EHR vendor or hospital and reporting entities can partner with any commercial partner capable of developing reporting software using the eCQM specifications. The measure steward's software to ingest this data and calculate the measure is freely available. Alara Imaging has created an edge device that can assemble data from different electronic sources (e.g. EHR, RIS [Radiology Information Systems], PACS [Picture Archiving and Communication Systems], and billing) to calculate the CT category, size-adjusted dose, and image noise that can then be consumed by the eCQM. If practices want to calculate these variables without using the Alara edge device, they may access a free online portal to calculate these variables and provide them to any entity implementing the measure. A prototype of this software was deployed at 8 testing sites (7 hospital systems and 1 ambulatory imaging network). Sites were asked to install the software, configure local connections to PACS, EHR, and other electronic systems as needed, and oversee the transfer of data to it from these sources. Burden was found to be no more or less onerous than the effort required by participation in other eCQMs or national registries, such as the ACR Dose Index Registry (Feasibility, 3.06). **Comment:** *For CT category … the developer used NLP for obtaining data such as reason for study or protocol name used in the calculation of this variable. The submission does not provide information on the NLP results' reliability and validity… or how sites would get access to use this custom NLP tool.* **Response:** This is incorrect; the measure does not use NLP. As described in the submission and above, it uses an algorithm that combines CPT® and ICD-10-CM codes to categorize CT exams. NLP was deployed as a method to validate the CT categorization determined by the algorithm at testing sites, where we did not have access to medical records. The sensitivity and specificity of this NLP referent standard are given above. **Comment:** *Multiple unstructured variables are required to construct the data elements for the numerator, denominator, and exclusions…* **Response:** This is incorrect; the measure does not use unstructured data. All data elements used to calculate the measure come from structured variables listed in the feasibility scorecards and in Specifications, Table sp-2: CPT® and ICD-10-CM codes; dose length product stored in the DICOM data; and patient diameter and image noise calculated on imaging data. The measure would not have met the requirements of an eCQM had it relied on unstructured data. **Comment:** *Protocol selection appropriate for a clinical indication is an important component of radiation dose management along with radiation dose optimization. Each component needs to be addressed as a separate quality action. The specific aspect(s) of performance to be improved is not intuitive due to the multiple components to the measures… It is true that the most accurate way to address appropriate and safe use of multi-phase studies is to measure both the clinical indication*

of an exam and the radiation dose output… However, these measures conflate the appropriateness of protocol for the clinical indication and radiation dose optimization… a facility may not be able to determine if its performance could be improved by adjusting protocols or by focusing on appropriateness of the ordered exam. **Response:** We agree that selecting an appropriate CT protocol and limiting radiation dose given the selected protocol are separate quality actions, but the commenter misses the crucial point that intermediate outcome measures typically reflect multiple opportunities for improvement. By analogy, we recognize systolic blood pressure control and glycosylated hemoglobin control as intermediate outcome measures for patients with hypertension and diabetes, respectively, even though there are many potential ways to manage these conditions. The fact that these intermediate outcomes can be improved by diet, exercise, medications, or combined approaches does not invalidate glycosylated hemoglobin or blood pressure control as quality measures. Similarly, the fact that our measure would be responsive to multiple, interrelated process steps is a key strength that will improve its value for reducing radiation exposure at the population level. Further, reporting entities will be provided with feedback for each CT exam, including its assigned CT category, radiation dose, size-adjusted radiation dose, and image noise, allowing recipients to identify the causes of performance gaps. Reporting entities will be able to assess if they are systematically assigning patients to the wrong protocol, or if they are choosing protocol settings that are inappropriate with respect to radiation dose or image noise. The actionability of the feedback is noted in the other letters written in support of the measure. To further demonstrate the potential of this measure, we conducted a randomized controlled trial in 100 hospitals and outpatient radiology practices to study the impact of providing detailed audit feedback, similar to what will be provided as part of the feedback on this measure. We found that this intervention resulted in significant reductions in radiation dose and dose variation with no impact to image quality, described in Usability, 4b.01. (Smith-Bindman, 2020) **Reference:** Smith-Bindman R, Chu P, Wang Y, et al. Comparison of the Effectiveness of Single-Component and Multicomponent Interventions for Reducing Radiation Doses in Patients Undergoing Computed Tomography: A Randomized Clinical Trial. JAMA Intern Med. 2020 May 1;180(5):666-675.

Comment 20 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

Comment Part 2:

We thank the American College of Radiology for their comments from 1/19/22 but wish to address several factual inaccuracies in the comments. (Response PART 2) **Comment:** *NQF #3633e, #3662e, and #3663e deviate from international standards, like diagnostic reference levels, and lack peer reviewed, broadly accepted consensus on global noise. For these measures, global noise is defined solely by the measure developer. Endorsing this method may encourage facilities to accept a narrow view of image quality.* **Response:** The ACR correctly notes that we have defined an approach to measuring noise. We did so only after testing and comparing multiple approaches described in peer-reviewed literature and validating noise measurements against radiologists' assessment of image adequacy for diagnosis. Image quality is a much less common problem than excessive use of radiation in CT imaging. While there may be other reasons to study CT image quality, our interest was simply to ensure that CT image quality did not erode as an unintended consequence of lowering radiation doses. There is no reason to believe that endorsing this measure will encourage facilities to "accept a narrow view of image quality" because radiologists have a requirement for adequate images to perform their work. They have no desire or motivation to alter their standards of what constitutes an adequate image. Radiologists do not want to read inadequate images and routinely request that such images be repeated or complemented by other imaging

modalities. **Comment:** *The ACR requests the developer further clarify the global noise table used in calculating the numerator... For example, Table sp-1 has the same global noise threshold for several CT categories, such as head low dose, head routinedose, and head high dose... If the image noise thresholds are the same, the size-adjusted radiation dose thresholds should be the same.* **Response:** We tested various published methods for measuring image noise and opted for a modified version of the method proposed by Malkus in 2017. The approach for setting the thresholds for image quality and radiation dose were based on the referent standard of radiologists' satisfaction with image quality. This did not always result in the relationship the ACR has suggested. For example, radiologists might want a minimum level of image quality for all head CT categories whereas the upper dose threshold might vary across the three head categories reflecting the different clinical indications comprising each group. Radiologists in our image quality study graded the majority of head exams as having acceptable image quality, even those at the lower dose range, meaning the minimum noise threshold is similar for all three categories. **Reference:** Malkus A, Szczykutowicz TP. A method to extract image noise level from patient images in CT. Med Phys. 2017 Jun;44(6):2173-2184. **Comment:** *Additionally, current CT scanners display dose values based oneither a 16 cm or 32 cm phantom for a neck scan, which must be carefully accounted for in measure performance calculations.* **Response:** As the ACR correctly notes, CT scanners display dose values based on a 16 cm or 32 cm phantom. If comparisons are made across reporting entities it is important that they use the same phantom, as this impacts the scanner reported DLP. The manufacturers are highly consistent in their use of phantoms for different body regions. In a study of 106,837 pediatric patients (a population where potential variation in phantom choice would most likely occur), 100% of CT exams in the neck are referenced to the 32 cm phantom, and it is thus unnecessary to account for phantom selection (Chu 2021). **Reference:** Chu PW, Yu S, Wang Y, et al. Reference phantom selection in pediatric computed tomography using data from a large, multicenter registry. Pediatr Radiol. 2021 Dec 6. **Comment:** *These eCQMs require multiple variables that may be captured in software systems external to electronic health records (EHRs), such as dictation systems housing radiology reports or DICOM standard-based systems, such as CT device software. Data element validity testing should demonstrate that the testing sites were able to integrate and validate the variables used to construct the data elements used by the eCQM in addition to the usual validation of the eCQM's electronic output against the medical record review. We are uncertain that this validation has been completed. Therefore, this submission does not demonstrate the measure can be reproduced in a reliable and valid manner by practices or facilities across multiple settings.* **Response:** This comment is entirely erroneous. No data are pulled from dictation systems or CT device software. The measure derives and uses codified and specified data from DICOM standard based systems, such as PACS, and EHR and billing claims. Our data element validity testing did demonstrate that 8 testing sites, reflecting 16 hospitals and 13 outpatient imaging facilities, were able to integrate, collect, and report the variables used to construct the data elements ingested by the eCQM. The letters of support from these testing sites independently confirm their ability to assemble the required data across diverse practice types and settings. **Comment:** *The ACR is concerned with the selection bias for the accountable entity-level… validity. Assessing measure score face validity through the TEP that created these measures lessens the extent of credibility for these results. Although the TEP is knowledgeable and represents a variety of stakeholders, there is a vested interest in ensuring these measures are available for use.* **Response:** All of the TEP members and their affiliations are identified in our submission materials (2b.02). Conflicts of interest were reviewed at each meeting and included with meeting minutes in a publicly available website (https://ctqualitymeasure.ucsf.edu/). The TEP members all voluntarily provided public service by joining the TEP. None of our TEP members has any "vested interest" in the outcome of the NQF endorsement process other than the ACR which served as a single member of the TEP. None of our TEP members is employed by the developer organization (UCSF)

or its funder (CMS), nor has any financial interest in the company that is offering technical support for software implementation (Alara Imaging). To be clear, these measures were developed by an academic radiology, quality improvement, and analytics team based at UCSF and supported by CMS, NIH and PCORI. The TEP was organized and tasked to provide broad multidisciplinary input to this team. Their endorsement of the validity of the measures is highly credible, as it reflects the fact that their advice was heeded at every stage of the development and testing process. Our TEP process followed the CMS Blueprint as well as NQF guidance, and 16/17 members agreed that that implementation of the measure will lead to a reduction in average CT radiation dose while maintaining adequate CT image quality if adopted (reported in 2b.03).

Comment 21 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

We thank the American Association of Physicists in Medicine for their perspectives but wish to address several factual inaccuracies: **Comment 1:** *Unscientific characterization of CT scan risk: The proposal is based on estimation approaches that are not reflective of the consensus of the scientific community* **Response:** The measure is not focused on radiation risk and does not calculate nor report radiation risk. The measure evaluates dose length product (DLP), and specifically whether size-adjusted DLP exceeds thresholds specific to CT category. DLP is the radiation dose measure most directly under the control of providers, determined by selected parameters. Further, DLP is universally reported by CT manufacturers. It is thus the ideal measurement to use when assessing the quality of CT exams. The TEP, which included the ACR, radiologists and a medical physicist, unanimously supported the radiation dose measure used and agreed is a relevant metric of quality for CT imaging (2b.03).There is also considerable precedent for using DLP to evaluate radiation dose in CT. The American College of Radiology has used DLP to set benchmarks [Kanal 2017] and to measure dose in its own NQF-endorsed quality measure #3621. **Reference:** Kanal KM et al. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-133. **Comment 2:** *Inactionability of the measures to enable targeted change to improve practice: It is not evident how the proposed measures can be practically used* **Response:** Reporting entities will be provided with specific feedback for each CT scan on its assigned CT category, radiation dose, size-adjusted radiation dose, and image noise, allowing recipients to identify causes of performance gaps and make targeted changes to improve quality. Comments in support of the measure from the testing sites describe how useful the information provided was to allow them to understand and improve their practice. As described in our submission, we found in a randomized controlled trial in 100 imaging facilities that providing detailed audit feedback on radiation doses, similar to what will be provided as part of the feedback on this measure, resulted in significant reductions in radiation dose with no impact on satisfaction with image quality (see Usability, 4b.01). (Smith-Bindman, 2020) **Reference:** Smith-Bindman R et al. Comparison of the Effectiveness of Single-Component and Multicomponent Interventions for Reducing Radiation Doses in Patients Undergoing Computed Tomography: A Randomized Clinical Trial. JAMA Intern Med. 2020 May 1;180(5):666-675. **Comment 3:** *Inadequate addressing of the complexity of CT categorization* **Response:** A detailed response to this question was provided in our response to the ACR. In short, the approach of assigning CT examinations to the different CT categories as specified in the measure was first developed using records from over 4.5 million CT exams in the UCSF International CT Dose Registry (Smith-Bindman, 2021). We then developed an approach for determining the clinical indication for imaging using an algorithm that combines procedure (CPT®) and diagnosis codes (ICD-10-CM) provided in Specifications, sp-11. This algorithm was developed using detailed review of over 10,000 patient records from UCSF Health. We validated the CT category assignment using the algorithm against “gold standard” chart review, as described in

Validity sections 2b.02 and 2b.03. When the algorithm was deployed at our testing sites the correct classification rate of the assignment of CT exams to CT category was on average 92% across clinician groups and hospitals and 95% in individual clinicians. **Comment 4:** *Inadequate assessment of noise: Noise in a CT image can be influenced by a variety of factors.* **Comment 5:** *Inadequate assessment of image quality: Image quality is affected by a myriad of factors* **Response:** The primary focus of our measure is to assess radiation dose adjusted for body size. The image quality component was included to protect against the unlikely possibility of substantial degradation of image quality as an unintended consequence of dose reduction. Our measure of image quality reflects what radiologists in practice regard as adequate. Others might have an interest in other ratings of image quality for other purposes, but that was not our intent. We tested and found that noise as a measure of image quality was associated with radiologists' satisfaction with the adequacy of CT images. These results were included in the submission (2b.03). **Comment 6:** *Flawed assumption on dose reduction vs dose optimization: The application focuses primarily on radiation dose reduction as opposed to right-sizing the dose.* **Response:** This is incorrect. We created the CT categories based on radiation dose and image quality requirements specific to clinical indications for imaging. Using radiologists' satisfaction with image quality, we established an image quality floor for each category, below which an exam is considered to have inadequate quality, and a radiation dose ceiling, beyond which doses are considered unnecessarily high. The purpose is to allow detailed assessment of each CT exam to ensure the dose is optimal based on the clinical indication for imaging. In current practice, there are no such benchmarks created by clinical indication, making it impossible for providers to know the right dose range for each patient. In our testing data, far more CT exams exceeded the radiation dose ceiling (average = 30%) than failed to meet the image quality requirement (average < 1%) (see section 1b.02). The measure encourages entities to reduce the proportion of exams that may "be overdosed for their exact need and condition" while preserving the minimum image quality. **Comment 7:** *Inadequate accuracy in patient size estimation: Assessing a patient size is not a trivial task, stemming from significant variability in the differences in the habitus of different patients, coupled with the existential challenge that there is no single metric.* **Response:** We agree that measuring patient size is important. Our approach for using mid-scan diameter is highly correlated with patient weight: in separate, NIH-funded research on CT use in children up to age 21 (Kwan 2022), we have shown that diameter in 4,239 children as measured on mid-scan axial images is highly predictive of patient weight, correlation = 0.904, Figure 1.

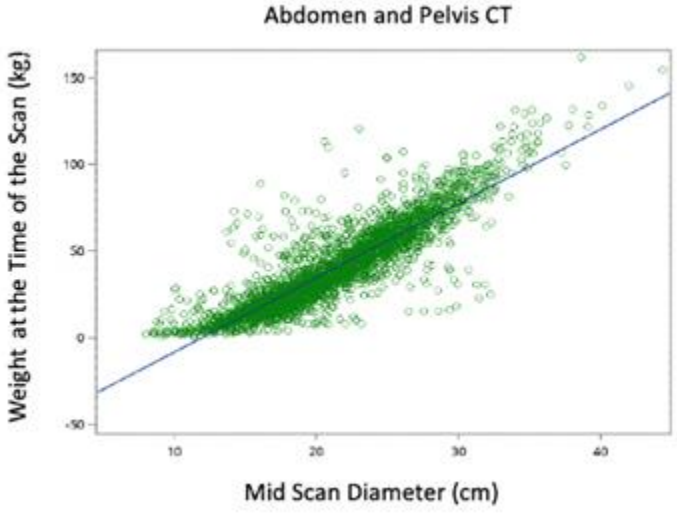


Figure 1. Mid scan patient diameter versus patient weight at the time of the CT, Kwan 2022

For this measure, patient size was measured using CT image pixel data, either on the mid-scan axial image or the coronal scout image when the mid-scan axial image was not available. This approach has been validated using data from UCSF Health, the UCSF Registry, as well as the data assembled

for measure testing. While there may be different ways to measure patient size, and different reasons for measuring patient size, it is a crucial piece of information that must be practically defined to ensure that the types of patients (case mix) at different practices do not bias the number of scans graded as out-of-range. We are adjusting for patient size primarily to ensure that entities that see larger patients are not penalized for doing so. Figure 2a shows the relationship between radiation dose (in DLP) and patient diameter using data from the UCSF Registry for abdomen CT. We chose abdomen CT as this is the category most influenced by patient size, and where patient mix could impact an entity's out-of-range rate. The raw correlation between patient diameter and unadjusted DLP is 0.50, and the marginal R-squared of the log-linear model used for adjustment is 0.15. After size-adjustment, the relationship is nearly removed: Figure 2b shows size-adjusted DLP by patient diameter using the same data; the raw correlation is far lower (-0.09), and the modeled marginal R-squared post-adjustment is 0. This demonstrates adequacy of the approach for adjustment of patient size.

Figure 2a: Unadjusted Dose Length Product vs Patient Diameter

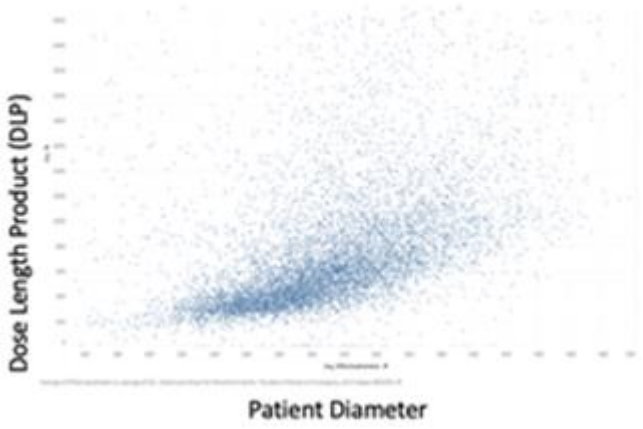
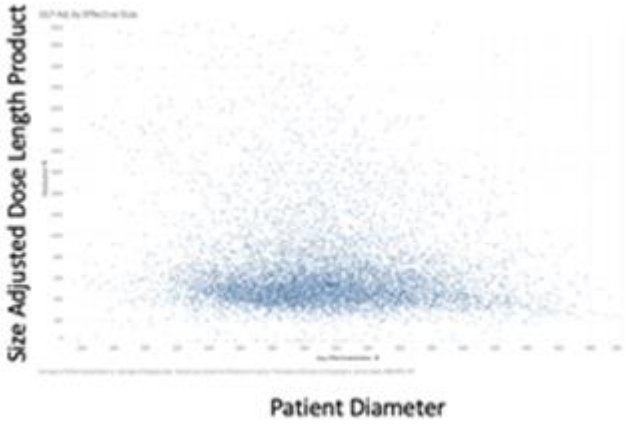


Figure 2b: Size-Adjusted Dose Length Product vs Patient Diameter



The adequacy of size adjustment was shown empirically using data assembled from the testing sites. The proportion of exams with out-of-range rates based on **unadjusted** and **adjusted** DLP are shown in Tables 1a and 1b. Without adjustment, the out-of-range values are strongly associated with patient size; after adjustment this relationship is gone.

Table 1a) Proportion of exams out-of-range on routine dose abdomen exams based on **unadjusted** DLP across the 16 hospitals, shown by decile in patient size. The proportion of out-of-range exams increased with patient size, seen in the table as an increase in dark shading in the lower rows of the table. Among patients in the highest size decile – those in last row– the out-of-range proportions across the 16 hospitals ranged from 93-100%.

Size Decile	Hospitals 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.27	0.22	0.11	0.00	0.27	0.29	0.30	0.14	0.34	0.20	0.24	0.40	0.06	0.17	0.11	0.17
2	0.30	0.00	0.00	0.00	0.08	0.11	0.13	0.29	0.24	0.30	0.04	0.19	0.00	0.07	0.16	0.09
3	0.15	0.06	0.15	0.00	0.17	0.18	0.22	0.75	0.21	0.30	0.17	0.18	0.08	0.18	0.17	0.12
4	0.07	0.17	0.29	0.15	0.25	0.32	0.09	0.82	0.43	0.25	0.07	0.42	0.10	0.17	0.19	0.21
5	0.45	0.15	0.13	0.14	0.28	0.43	0.00	0.93	0.40	0.42	0.19	0.38	0.00	0.14	0.48	0.55
6	0.42	0.20	0.25	0.36	0.55	0.61	0.27	0.96	0.55	0.19	0.31	0.51	0.08	0.46	0.47	0.78
7	0.79	0.47	0.45	0.58	0.70	0.75	0.17	1.00	0.69	0.37	0.26	0.73	0.06	0.71	0.66	0.90
8	0.81	0.37	0.75	0.69	0.67	0.86	0.24	1.00	0.89	0.35	0.58	0.77	0.22	0.80	0.91	0.95
9	0.96	0.85	1.00	0.75	0.88	0.93	0.26	0.93	0.94	0.64	0.78	0.93	0.63	0.90	1.00	1.00
10	1.00	0.96	0.98	0.93	0.97	0.97	0.93	0.94	1.00	0.95	0.98	0.96	0.85	0.95	0.94	1.00

Table 1b) Proportion of exams out-of-range on routine dose abdomen exams, based on **size-adjusted** DLP across the 16 hospitals shown by decile in patient size. High proportion of out-of-range exams are no longer concentrated among the larger patients. Among patients in the highest size decile, out-of-range rates ranged from 11-53%.

Size Decile	Hospitals 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.55	0.61	0.22	0.20	0.42	0.48	0.48	0.61	0.49	0.70	0.37	0.62	0.10	0.37	0.22	0.51
2	0.50	0.25	0.00	0.18	0.19	0.31	0.22	0.71	0.36	0.61	0.15	0.38	0.00	0.08	0.33	0.21
3	0.15	0.18	0.15	0.11	0.37	0.39	0.33	0.96	0.30	0.50	0.26	0.36	0.10	0.23	0.43	0.22
4	0.27	0.17	0.29	0.15	0.35	0.54	0.09	0.97	0.43	0.38	0.11	0.53	0.10	0.18	0.31	0.30
5	0.45	0.15	0.13	0.14	0.26	0.46	0.00	0.93	0.40	0.42	0.19	0.38	0.00	0.19	0.52	0.59
6	0.33	0.10	0.25	0.36	0.39	0.45	0.27	0.96	0.47	0.13	0.31	0.40	0.05	0.34	0.45	0.72
7	0.29	0.18	0.20	0.46	0.50	0.42	0.17	0.90	0.57	0.16	0.17	0.60	0.04	0.50	0.36	0.70
8	0.43	0.05	0.19	0.25	0.54	0.39	0.12	0.70	0.58	0.09	0.35	0.62	0.09	0.59	0.53	0.83
9	0.48	0.26	0.48	0.30	0.45	0.27	0.11	0.19	0.72	0.07	0.18	0.56	0.06	0.62	0.60	0.66
10	0.35	0.27	0.40	0.39	0.38	0.11	0.27	0.11	0.61	0.37	0.29	0.44	0.11	0.27	0.53	0.36

Reference: Marilyn Kwan et al. Quantifying cancer risk from exposures to medical imaging in the Risk of Pediatric and Adolescent Cancer Associated with Medical Imaging (RIC) Study: Research Methods and Cohort Profile. In press, Cancer Causes and Control. **Additional Comment:** *One cited reference supports the proposed measure, however, this cited article has an accompanied editorial that highlights the limitations of the proposed approach [Mahesh M. Benchmarking CT Radiation Doses...Radiology. 2021 Nov 9:212624.]* **Response**: We find it surprising that Dr. Mahesh's editorial was used to criticize the measure. Dr. Mahesh is a board member of American College of Radiology and American Association of Physicists in Medicine, and he was very positive about our image quality-informed framework for assessing radiation dose. He noted the observed, significant differences *between* CT categories versus *within* categories was "an encouraging result for anyone trying to optimize CT studies based on clinical indications." He noted the study was "a good start" on the road to optimizing CT protocols based on image quality. He opined that the CT classification would be more useable and easier to implement if based on current procedural terminology codes. This is precisely what we have done in this measure.

Comment 22 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

We thank Dr. Ehsan Samei for sharing his perspectives on the measure and for collaborating with us early in the measure development process. We wish to address a few inaccuracies and misunderstandings in Dr. Samei's comments. The majority of Dr. Samei's comments focus on image quality and his concern that the measure does not offer a comprehensive assessment of image quality. Our measure is not intended to be a comprehensive assessment of image quality.

Criticizing the proposed measure for what it is not is beyond the scope of what should be considered in assessing the usefulness of what has been submitted. The primary focus of our measure is to assess radiation dose adjusted for body size, and the image quality component provides a means to protect against the unlikely possibility of substantial degradation of image quality as an unintended consequence of dose reduction. The approach for creating thresholds is described in Validity, 2b.02. **Comment:** ***Inaccurate assessment of patient size:*** *The measure of size proposed is calibrated to earlier work and publication from our group at Duke University for academic purposes. That early method they have embraced has had major errors.* **Response**: We are adjusting for patient size primarily to ensure that entities that see larger patients are not penalized for doing so. Although we explored code that Dr. Samei provided early in our initial efforts to measure patient body habitus we found that it was inadequate, particularly for some CT categories, and we have not relied upon it. We developed our own approach for measuring size using CT image pixel data from the mid-scan axial image or the coronal scout image when the mid-scan axial image was not available. Our approach of measuring size was shown to be highly correlated with patient weight (correlation = 0.904) in a large study in children described in our response to the AAPM. For this measure, the measurement of size was validated using data from UCSF Health, the UCSF Registry, as well as the data assembled for measure testing. The adequacy of the approach we have adopted for size adjustment is described in the initial application and the response to the comments by the AAPM. **Comment:** ***Inaccurate assessment of noise:*** *The measure of noise proposed references earlier work and publication from our group at Duke University. That early method exhibited errors, corrected in subsequent versions that have not been shared…* **Response:** Dr. Samei's approach and code for measuring image quality were explored in the process of developing our measure but were not included in the final measure specifications. Any errors in his approach are not relevant to the measure. **Comment:** ***Inaccurate assessment of radiation risk:*** *The measure of size-adjusted radiation risk, adjusting the CT scanner outputs with 'patient size' to perform risk estimation is not a standard method nor endorsed by any scientific or professional body… Patient risk can only be assessed with the knowledge of organ doses that is not even mentioned in the application let alone pursued. The proposed method CANNOT be used as surrogate for future cancer risk.* **Response:** The measure does not calculate or report radiation risk. The measure evaluates radiation *dose* (measured in dose length product, DLP), and whether size-adjusted DLP exceeds thresholds specific to CT category. The empirical validity of the risk-adjustment approach based on patient size is described in the application (section 2b.26 – 2b.31) and in our response to the comments by the AAPM. The approach of evaluating CT safety by comparing machine output (whether DLP or CTDIvol) against benchmarks is widely accepted in the radiology field. (Kanal 2017) In contrast, organ dose has no standard definition, is not reported by the manufacturers, is not available in a structured format, would be time intensive to calculate in clinical settings and most importantly has limited actionability as this is not under the direct control of technologists or physicians. Organ doses may be useful for counseling patients or in the context of epidemiological studies, but we do not believe it has a role as a metric for CT quality measurement. **Reference:** Kanal KM, Butler PF, Sengupta D, et al. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-1 ***Comment:*** ***Subjectivity:*** *The measures are anchored to subjective perception by radiologists as how they "like" the images. There is in fact no evidence provided that the measures can lead to an improvement in diagnostic accuracy. In fact, it might lead to a degradation.* **Response:** The measure is not intended to improve diagnostic accuracy. The purpose of the measure is to establish a radiation dose ceiling to avoid excessive radiation exposure, and an image quality floor to safeguard against unintended deterioration of image quality. There is precedent for using radiologist satisfaction with image quality to set or validate noise targets, including work by Dr. Samei. (Cheng 2019, IAEA 2009) This also reflects clinical practice: radiologists subjectively assess images and regularly ask for scans to

be repeated when they are not adequate. As described in the response to ACR comments, Radiologists do not want to read inadequate images and routinely request that such images be repeated or complemented by other imaging modalities. Radiologist's subjective assessment provides a practical way to ensure the image quality is not degraded through efforts to optimize the radiation doses. **References:** Cheng Y, Abadi E, Smith TB, Ria F, Meyer M, Marin D, Samei E. Validation of algorithmic CT image quality metrics with preferences of radiologists. Med Phys. 2019 Nov;46(11):4837-4846. doi: 10.1002/mp.13795. Epub 2019 Sep 20. International Atomic Energy Agency (IAEA), Dose Reduction in CT while Maintaining Diagnostic Confidence: A Feasibility/Demonstration Study, TECDOC Series, 2009.

NQF #3662e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level)

Comment 1 by: Karen Orozco, American College of Radiology
On behalf of Karen Campos, American College of Radiology

The American College of Radiology, representing more than 40,000 radiologists, radiation oncologists, medical physicists, and nuclear medicine physicians, appreciates the opportunity to submit comment on NQF #3633e, #3662e and #3663e: Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level, Clinician Group Level and Facility level, respectively). **The ACR does not support the endorsement of NQF #3633e, #3662e, and #3663e.* *General Comments** Protocol selection appropriate for a clinical indication is an important component of radiation dose management along with radiation dose optimization. Each component needs to be addressed as a separate quality action. The specific aspect(s) of performance to be improved is not intuitive due to the multiple components to the measures (size-adjusted dose, image quality, clinical indication). It is premature to measure performance on excessive radiation dose based on thresholds by clinical indication for an exam until the level of standardization and availability of national benchmarks is further along as discussed below. It is true that the most accurate way to address appropriate and safe use of multi-phase studies is to measure both the clinical indication of an exam and the radiation dose output (dose indices per exam) and look at the two separately or distinctly together. **However, these measures conflate the appropriateness of protocol for the clinical indication and radiation dose optimization, disregarding applicability, from which a facility may not be able to determine if its performance could be improved by adjusting protocols or by focusing on appropriateness of the ordered exam. Therefore, improvement may be limited.[1]** Dose optimization results in a quality action for facilities to adjust their protocols and is a responsibility of the team as a whole – physicists, technologists, and physicians who oversee the team at the facility. Protocol selection addresses the appropriateness of the exam for the clinical indication and other factors such as patient time on the scanner and optimal radiation dose. There are challenges with the implementation of an indications-based measure. Indications for exams do not have standardized language that could be used to track them. Most health and IT systems capture ICD-10 coding for reimbursement, but typically not enough standardized information to characterize the patient's condition. As a result, the clinical reason for performing an imaging exam is often extremely limited in the exam order. Electronic Health Records (EHRs) are notoriously incomplete with this type of information and interoperability issues exist with other software systems that might contain such information. **A *validated* method for determining classification of studies using high-dose versus routine protocols appropriate to the indication must be incorporated into such a measure; these three measures include specifications which have not been validated.** Please refer to the validity

section below for more details. **NQF #3633e, #3662e, and #3663e deviate from international standards, like diagnostic reference levels, and lack peer-reviewed, broadly accepted consensus on global noise. For these measures, global noise is defined solely by the measure developer. Endorsing this method may encourage facilities to accept a narrow view of image quality. The ACR requests the developer further clarify the global noise table used in calculating the numerator.*** The benchmark source is not transparent, and its applicability is unclear. For example, Table sp-1, Size-adjusted radiation dose and global noise thresholds by CT category, has the same global noise threshold for several CT categories, such as head low dose, head routine dose, and head high dose. Is it intentional that the same global noise threshold should be applied to both low and high dose head CTs? If the image noise thresholds are the same, the size-adjusted radiation dose thresholds should be the same, unless the scan length is remarkably different between the 3 CT categories. Additionally, current CT scanners display dose values based on either a 16 cm or 32 cm phantom for a neck scan, which must be carefully accounted for in measure performance calculations. **There is little to no acknowledgement of limitations.** These measures have multiple limitations, including the lack of widespread acceptance and implementation, and the issues with the method of measuring global noise. The developer states their company can provide the service of quantifying the measure at a cost; this should also be included as a potential limitation. The measure developer does provide specifications for other entities to implement the measure, but the burden of implementation may be significant. Finally, the author cites publications from their group to justify the benchmarks, but they have not been vetted through a broader consensus process. **The ACR strongly encourages the Patient Safety Standing Committee to re-vote on the scientific acceptability of these measures based on the following concerns.**
Validity/Feasibility.These eCQMs require multiple variables that may be captured in software systems external to electronic health records (EHRs), such as dictation systems housing radiology reports or DICOM standard-based systems, such as CT device software. Data element validity testing should demonstrate that the testing sites were able to integrate and validate the variables used to construct the data elements used by the eCQM in addition to the usual validation of the eCQM's electronic output against the medical record review. **We are uncertain that this validation has been completed. Therefore, this submission does not demonstrate the measure can be reproduced in a reliable and valid manner by practices or facilities across multiple settings.** For example, for CT category (or other elements deriving/collecting data using custom natural language processing (NLP) tools), the developer used NLP for obtaining data such as reason for study or protocol name used in the calculation of this variable. The submission does not provide information on the NLP results' reliability and validity. Because **this comparison of the NLP-derived data against a medical record review was only completed in a sample from one site (UCSF Health System), there is uncertainty whether the results are generalizable across EHRs or other databases**. These measures rely on custom made NLP trained and validated on a small group of pilot sites; it is not clear whether this type of NLP would work outside these sites nor how sites would get access to use this custom NLP tool. Testing information does not demonstrate adequate validation of this critical data element. Additionally, **sufficient evidence should demonstrate that the definitions/variables used are valid and do not rely on one study or use in a single system, such as what is provided to support the thresholds of "out of range" performance values.** While the process to determine these thresholds is detailed, we do not believe that a Technical Expert Panel (TEP) conclusion in the absence of independent data validation is sufficient.
M**ultiple unstructured variables are required to construct the data elements for the numerator, denominator, and exclusions. Assessments of the feasibility of the integration of these unstructured data into the measure calculations would be useful to ensure that the underlying data can, in fact, be integrated if practices and facilities that choose not to use the edge device.** For example, the level of effort required to integrate the Binning algorithm for the CT categories

and ensure that the results are reproducible and valid remains unclear. The ACR is concerned with the selection bias for the accountable entity-level (measure score) validity. **Assessing measure score face validity through the TEP that created these measures lessens the extent of credibility for these results**. Although the TEP is knowledgeable and represents a variety of stakeholders, there is a vested interest in ensuring these measures are available for use. **Most importantly, as one of the TEP members noted in the survey, the performance score from these measures does not clearly indicate what corrective action needs to be taken by the clinician, clinician group, and/or the facility to improve performance**. **Usability** While implementing these measures as specified may not impose a substantial burden on clinicians, **it may necessitate substantial organizational effort to access and process the data elements required to calculate the measure score**. The measure steward states that their software is available on a non-commercial basis to calculate this measure, and that other vendors may also develop their own software to implement the measure specifications using the information included in this submission. Will the measure steward review other vendors' software to ensure comparable calculation methods? Measure stewards frequently make specifications available "as is" without warranty, leaving it to the implementer to appropriately update any software or tools as measure specifications are changed. But the complexity of these measure specifications may warrant greater oversight. External vendor software will need to be maintained and updated to ensure the software's accuracy and reflect any changes in specifications and coding. **For all the reasons stated above, the ACR does not support the endorsement of these three measures.** We thank the NQF staff for their transparent endorsement process. **Reference: 1. '**Mahesh M. Benchmarking CT Radiation Doses Based on Clinical Indications: Is Subjective Image Quality Enough?Radiology. 2021 Nov 9:212624. doi: 10.1148/radiol.2021212624. Online ahead of print. PMID: 34751622

Comment 2 by: Angela Keyser, American Association of Physicists in Medicine

What is AAPM:

The American Association of Physicists in Medicine (AAPM) is the primary scientific and professional organization of physics in radiology and radiation oncology in the United States. The mission of AAPM is advancing medicine through excellence in the science, education and professional practice of medical physics; a broad-based scientific and professional discipline which encompasses physical principles with applications in biology and medicine. With 9717 members in 94 countries, AAPM supports the Medical Physics community with a focus on advancing patient care through education, improving safety and efficacy of radiation oncology and medical imaging procedures through research, education and the maintenance of professional standards. AAPM has a staff of 33 and an annual budget of $10.7M, and is located at 1631 Prince Street, Alexandria, VA 22314.

AAPM comments on the proposed measures:

AAPM does not support the endorsement of NQF #3633e, #3662e, and #3663e.
This application proposes electronic clinical quality measures (eCQM) that monitor CT performance to discourage unnecessarily high radiation dose while maintaining adequate image quality. The proposed metrics require CT Category (i.e., the CT exam type), the size adjusted radiation dose [the patient's dose length product (DLP) adjusted by patient size], and the global noise (associated with the variance of the voxel values in CT images). The two reported measures are the percentage of eligible CT cases in a particular category deemed to be "out-of-range" compared to defined thresholds with respect to the size-adjusted radiation dose or the global noise in a set time period. While efforts to enhance consistency of CT practice are noble and include initiatives by AAPM and others worldwide, the proposal has significant limitations that impact its scientific and practical

value and overall likelihood of clinical acceptance. These limitations include improper representation of image quality, improper estimation of radiation risk, and substantial oversimplified representation of implementation in practice, including not addressing the challenges of implementation. The authors indicate that their company (Alara Imaging, Inc.) can provide the service of quantifying the measures at a cost. A steward of measures requires an extensive track record for scientific and technical expertise and policy making that represents a broad consensus of the community. These important elements should be carefully reviewed within this application. One cited reference supports the proposed measure, however, this cited article has an accompanied editorial that highlights the limitations of the proposed approach [Mahesh M. Benchmarking CT Radiation Doses Based on Clinical Indications: Is Subjective Image Quality Enough? Radiology. 2021 Nov 9:212624. doi: 10.1148/radiol.2021212624. Online ahead of print. PMID: 34751622]. The editorial and stated limitations are not addressed in the proposal.

The AAPM agrees that effort needs to be continually placed on ensuring diagnostic quality CT imaging, optimizing CT dose, and achieving consistency across facilities, considering differing technologies and practices. The non-profit entities of the AAPM, the American College of Radiology (ACR), and Image Wisely and Image Gently Alliances have spent decades towards this goal and continue to do so through many initiatives. Among them, the non-profit ACR CT Dose Index Registry (DIR; https://www.acr.org/Practice-Management-Quality-Informatics/Registries/Dose-Index-Registry, established in 2011) has the significant stature of implementing a dose registry that enables facilities to compare dose indices nationally, to ensure the highest quality imaging with lowest possible dose. The ACR CT DIR implementation incorporates the expert, consensus opinions of the medical imaging community. ACR dose optimization measure recently endorsed by NQF provides a further valuable measure to manage imaging radiation dose (https://www.qualityforum.org/QPS/3621). The imaging community's valuable clinical benchmarks greatly benefit from consensus decisions based on sound scientific and technical review and discourse. The proposal herein should be carefully reviewed for any additional contributions or advantages it would provide to our existing robust consensus measures and resources, such as available with the ACR.

After a detailed review of the measures by multiple expert members of the AAPM, we have concluded that the **AAPM does not support the endorsement of NQF #3633e, #3662e, and #3663e.** This position stems from eight major concerns about the proposed measures:

1) Unscientific characterization of CT scan risk: The proposal is based on estimation approaches that are not reflective of the consensus of the scientific community and do not acknowledge the uncertainties of the estimates. A NQF measure focused on radiation risk should uphold scientific objectivity, integrity, and responsibility not evident in the presentation and assessment of radiation risk in this proposal.

2) Inactionability of the measures to enable targeted change to improve practice: It is not evident how the proposed measures can be practically used to improve imaging practice and exactly how a facility can do to achieve compliance, given the wide varieties of factors and technologies involved.

3) Inadequate addressing of the complexity of CT categorization: The proposal does not address the magnitude of this challenge nor has suggested means to overcome it given that current standards are even lacking in uniform characterization of protocols. Inaccurate classification of data can lead to significant and misleading errors.

4) Inadequate assessment of noise: Noise in a CT image can be influenced by a variety of factors including justified differences in CT technologies including new reconstruction

methods that dramatically alter noise. Further, noise does not have a singular value in a CT exam. A "global noise" ignores this diversity and can misrepresent the quality of an exam.

5) Inadequate assessment of image quality: Image quality is affected by a myriad of factors including resolution and contrast, as well as the intended purpose of the exam. A singular representation of image quality via global noise overly simplifies this space and can lead to gross misrepresentation of image quality and thus mis-service to patient care.

6) Flawed assumption on dose reduction vs dose optimization: The application focuses primarily on radiation dose reduction as oppose to right-sizing the dose for the best care of the patient. Individualization and optimization of care and safety should be the goal not minimization. This approach can lead to some patients getting under exposed, leading to missed diagnosis, while others may be over-dosed for their exact need and condition.

7) Inadequate accuracy in patient size estimation: Assessing a patient size is not a trivial task, stemming from significant variability in the differences in the habitus of different patients, coupled with the existential challenge that there is no single metric capturing the size of a patient of varying diameter at different cross-sectional locations. Algorithms are continuously evolving and no evidence is provided that the company can do this task with sufficient accuracy.

8) Limited expertise and track record of the company: The company is a new (2020) company with no experience of having previously performed a project of such wide scope, scientifically or technically. There is no scientific track record on CT technology, size estimation, or image quality assessment for the company to be considered a steward of measures on which there is a lack of expertise, publication, and scientific history.

These concerns are detailed specially in our complete review submitted via email to patientsafety@qualityforum.org, along with selected specific observations on the proposal on January 19, 2022.

The AAPM recognizes that this topic is complex, including scientific, technical and clinical components. We welcome the opportunity for greater in-depth discussion on meaningful measures of quality imaging practice.

Respectfully submitted,
American Association of Physicists in Medicine (AAPM)
January 19, 2022

Comment 3 by: Bradley Delman, Mount Sinai Health System

I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. As an implementation testing partner, I coordinated Mount Sinai Health System's inclusion in the test. To summarize, after installing the data collection software, we routed CT imaging data from PACS and sent order and billing data from various electronic systems to the software. We also worked with UCSF and our CT vendors to ensure the Radiation Dose Structured Report (RDSR) was being saved for each exam sent to PACS. As we discussed in our interview with UCSF, this work fell on the PACS team and IT colleagues, without requiring effort from clinicians above my initial planning and coordination. Besides some technical challenges, which were all resolved, we faced few barriers to successful implementation and had very little missing data. In total we submitted 11,588 scans, representing just over 3 weeks of CT data from our health system. Based on our experience, the participation in the proposed quality measure is feasible. However, I suspect that spirited engagement from PACS, RIS and/or EHR vendors would greatly enhance participation and timely provision of data. We

have also been satisfied with the feedback we've received from Alara Imaging on our measure performance, which brought to our attention areas of high radiation dose. This feedback has identified individual exams as well as imaging protocols that contribute high radiation dose. Although we have been a dose-conscious department, the feedback highlighted areas of variability in both routine and size-adjusted datasets. Furthermore, we learned which protocols and classes of studies fell within and beyond expected range for dose, and how dose can vary between scanners for protocols with the same name. We also learned that some types of studies may need to be renamed or reclassified for appropriate grouping of results. A quality measure the quantifies dose while ensuring preservation of imaging quality can help mitigate the use of excessive radiation doses used in CT. I support the work of the measure developers to improve patient safety and CT quality.

Comment 4 by: Daniel Hirsch

I write in support of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. They are important proposals that would markedly reduce unnecessary radiation exposures in medicine, and the cancers induced therefrom, while providing the same yield of diagnostic information. Many, many lives could thus be saved were the proposals adopted. I have spent much of my professional career attempting to reduce the risks to public health from ionizing radiation associated with nuclear waste, reactor accidents, nuclear weapons tests, uranium mining and milling, and radioactively contaminated sites involved in the production of nuclear weapons and other nuclear activities. It is with some alarm that I have viewed in recent years the extraordinary increase in public exposures to ionizing radiation associated with the remarkable escalation of exposures in medicine, largely due to ever-more frequent CT scans, resulting in doses from medical procedures now dwarfing exposures from the nuclear activities that have so long concerned me. The proposals made by UCSF would help reign in unnecessarily high radiation doses from these medical procedures while still producing the diagnostic information needed by physicians for their patients. The important revelation in the studies cited in the proposals is that the doses currently received by patients in these procedures are frequently very much higher—often ten times higher—than necessary. One can get the same medical benefit from the procedure at one tenth the cancer risk. The proposals indicate that many thousands of unnecessary radiation-induced cancers could be avoided were CT scans kept to the minimum level necessary to produce the required image. This seems quite correct. The National Academies of Sciences, Engineering and Medicine has produced over the years the primary studies on the matter of ionizing radiation and cancer induction. The most recent Biological Effects of Ionizing Radiation study (BEIR VII) estimates a risk of 1.17 cancers per 1000 person-rems of exposure, and concludes, as have all the BEIR studies, that there is no threshold below which there is no risk. All radiation protection agencies (e.g., US EPA) have adopted the BEIR conclusions. Currently, exposures to medical radiation are estimated as averaging about 350 millirem/year per person. Given that degree of exposure, and the current U.S. population, medical radiation would be estimated to produce many millions of cancers over the population's lifetime. Reducing unnecessarily high exposures while still producing the necessary diagnostic image could thus prevent a very large number of cancers and deaths, while, not incidentally, also reducing Medicare expenditures for their treatment. I strongly urge adoption of quality measures that assure CT exposures use the lowest reasonable doses necessary for the procedures. Daniel Hirsch retired Director of the Program on Environmental and Nuclear Policy at University of California at Santa Cruz

Comment 5 by: Dawn Ritzwoller

I am a college student and Environmental Biology (E-bio) major, and I am pediatric cancer survivor. I am writing today in support of this radiation dose quality measure. Beginning ten years ago, and both during and after I finished treatment, I received multiple CTs (to multiple parts of my body) as part of my diagnostic and follow-up care. Not once during this period, did any of my doctors or other, discuss with me the downstream risk of all of the radiation exposure I experienced. It was only years after my treatment ended, and now via classes I have take for my E-bio major, that I am beginning to understand the risk associated with radiation exposure. What is also now clear to me is the importance that providers use the most appropriate (low) dose for the specific diagnostic or follow-up exam. I know that image quality is important for diagnosis, but patients (like me) need the confidence that their doctors and hospitals are using the best and lowest dose possible for the exam that they order. Thank you!

Comment 6 by: Debra Ritzwoller

I am writing in support of this important measure. I am a cancer health services researcher *and* a mother of a pediatric cancer survivor. It is well documented in the literature that there has been a significant secular increase in CT use within and across most patient populations. While CT use, and therefore radiation exposure has increased over time, I know that personally and professionally that excessive radiation dose remains a significant quality issue, and it is one that is often not adequately addressed by researchers and healthcare providers/delivery systems. This quality metric is necessary now, in order to provide the incentives and the resources needed to generate the metrics and the benchmarks that may actually influence practice that may in turn translate into a meaningful reductions in the radiation dose that patients are exposed to. This metric is designed to address the clinical indication associated with the respective exam, rather than just the type of advanced imaging that is performed. The measure is also constructed to ensure that the dose benchmarking does not adversely impact the quality of the metric. Given the noted harms of CT based radiation exposure (e.g USPSTF Lung Cancer Screening "B" recommendation), this measure is addresses a timely and needed quality metric.

Comment 7 by: Ehsan Samei, Duke University – Margolis Center for Health Policy Center

Duke University, Ravin Advanced Imaging Laboratories (Ravin Labs) and Clinical Imaging Physics Group (CIPG), Durham, NC 27710 The Ravin Labs is a 50-member leading translation imaging research laboratory in the country with over 30 years of history. The lab conducts rigorous NIH-funded research with an additional mandate to practice its science through CIPG, an imaging physics group of 15 experts dedicated to quality and safety in the practice of radiology. The group, highly integrated into the clinical domain, has devised and put to practice imaging dose and image quality monitoring systems at the level of individual patients within the Duke University Health System with additional pilot installations at MD Anderson Cancer Center and Stanford University. The group has published extensively on its technology and findings (upward of 500 papers), with over 30 referred publications on dose and quality monitoring alone. The effort has led to significant reduction of patient radiation dose at our facilities and right-sizing it per individual needs of patients. **We do not support the proposed measures.** The rationale is detailed below. **Overall:** While we applaud the effort to introduce new quality measures in the practice of medical imaging, the proposed electronic clinical quality measures (eCQM) are misleading and overly simplistic leading to significant unintended consequences. The limitations stem from the fact that the proposed risk measures are based on CT scanner output and not the actual dose burden to

individual patients at the organ level, the quality measure is based on noise alone ignoring the multi-faceted reality of diagnostic quality, and lack of methods that standardize protocols across vast diversity of examinations. There is significant ambiguity in the exact method used for noise and size estimation with no track record or peer review of otherwise black-box methods. This approach will likely produce measures that can be orders of magnitude off from their actual values, and therefore lack clinical relevance and fidelity. Measures can lead to misleading and erroneous conclusions while also potentially jeopardizing the use and development of better approaches, as inaccurate low-bar measures can prevent accurate ones in the future. But most importantly, the measure can lead to unintended consequences and even harm the patient. For example, an imaging team can take an action that is not in the best interest of a patient, like applying too little dose for some patients such that disease would be missed, a "wasted dose" with no medical benefit and health and cost consequence of a miss. Conversely others might get more radiation than needed as the measures do not account for individual patient needs and tasks. Improving consistency in imaging practice is a laudable goal that needs a proper solution anchored to scientific understanding of radiation risk, image quality need of patients, diversity of practices, and the CT technology. The proposal is lacking on all these four fronts. A solution to inconsistency in images can only be brought forth through a broad consensus of the scientific and practicing communities (including ACR, AAPM, Image Gently, and Image Wisely), CT manufacturers (represented by MITA), standard methods of data categorizations and measures (supported by the medical community), and evidence-based radiation risk and image quality measures at the level of indication and organ where they are actually relevant to the individual patient. A for-profit company with no track record or transparency of its methods cannot be considered a steward of such a space. Below we further detail 12 concerns regarding the proposed measures:

1. **Inadequate attention to image quality:** The measures are heavily dose related, emphasizing this over measures of quality. Dose and minimizing it is important but equally important is image quality as an inadequate image quality would be a dis-service to the patient regardless of the dose. This is explicitly stated in the International Commission of Radiological Protection (ICRP) in Publication n. 135.

2. **Inaccurate assessment of radiation risk:** The measure of size-adjusted radiation risk, adjusting the CT scanner outputs with 'patient size' to perform risk estimation is not a standard method nor endorsed by any scientific or professional body. The method is in fact explicitly **discouraged** by the AAPM Task Group 204. Patient risk can only be assessed with the knowledge of organ doses that is not even mentioned in the application let alone pursued. The proposed method CANNOT be used as surrogate for future cancer risk.

3. **Incomplete/Inaccurate representation of image quality:** The measures include image noise. Yet, noise is just one component of image quality. For example, the noise of an image can be fine but image quality totally inadequate. And conversely noise can be too high but image quality totally adequate. To assess image quality properly, one should include the actual task at hand (eg, detecting a pancreatic cancer vs bowel obstruction vs kidney stone) as well as other equally important facets of quality, like noise texture, resolution, and contrast. These factors have not been even mentioned let alone tackled in this application. Focusing on noise as a singular metric of quality can lead to major mis-representation of the needs of a quality and safe imaging practice.

4. **Neglecting the impact of image rendition:** Critical and relevant to clinical practice, the measure of noise proposed does not take into consideration how differing reconstruction algorithms and parameters affect noise (up to 200%). Without considering this influence, a measure of noise as proposed is irrelevant and misleading.

5. **Subjectivity:** The measures are anchored to subjective perception by radiologists as how they "like" the images. There is in fact no evidence provided that the measures can lead to an improvement in diagnostic accuracy. In fact, it might lead to a degradation.

6. **Lack of integrating dose and quality:** There is no indication as to how image quality is linked to radiation dose and at what level; or instance, how they propose to manage multiple reconstructions of the same exposure event.

7. **Not addressing the multiplicity of exam components**: A CT exam often includes multiple phases (series) each of which has a noise and radiation dose of its own. Averaging noise across series is meaningless. The measures do not recognize or account for this multiplicity and diversity.

8. **Under-recognizing the diversity of exams:** The measures do not address the notable diversity of exam nomenclature across institutions and practices. This is a significant component of any dose or quality monitoring system. Without a standard for CT protocols, which cannot be devised by a for-profit company without consensus of manufacturers and users, the data can be mislabeled and mishandled leading to major errors in the results and subsequent negative effect on mis-dosing and mis-diagnosing patients.

9. **Inaccurate assessment of patient size:** The measure of size proposed is calibrated to earlier work and publication from our group at Duke University for academic purposes. That early method they have embraced has had major errors (upward of 300% is certain applications) that have been corrected in subsequent versions that have not been shared. Without essential newer refinements to assure fidelity, the company cannot be a responsive steward of the measure that it has had no expertise to advance or maintain.

10. **Inaccurate assessment of noise:** The measure of noise proposed references earlier work and publication from our group at Duke University. That early method exhibited errors, corrected in subsequent versions that have not been shared. Without essential newer refinements, the company cannot be a responsive steward of the measure that it has had no expertise to advance or maintain.

11. **Lack of guidance toward compliance:** To us it is difficult to defend (1) measuring imaging practices based on ambiguous and questionably-relevant metrics promoted to represent the actual safety or quality of CT practice, and (2) not offering any guidance as to how a practitioner responsible for "outlier" examinations can bring their practice to the proposed definition of compliance. Together, these can easily create signification confusion and potential disruption in the imaging practice

12. **Lack of support from manufacturers:** Having worked in dose and image quality monitoring for over a decade, academic centers of excellence, including ourselves, have a close connection with major CT manufacturers including MITA, Medical Imaging Technology Alliance, which comprises all CT manufactures. Our discussions regarding this measure lead us to believe that there will be little support from scanner manufacturers for a non-transparent and unpredictable product that lacks maturity from a private for-profit entity. There are substantial differences in image processing, detector efficiency, and such across scanners that will have significant bearing on the CT image. The proposed measure does not account for such important nuances, leading to erroneous results.

Comment 8 by: J. Leonard Lichtenfeld

I am pleased to provide this comment in support of NQF quality measures 3633e, 3662e and 3663e. These comments reflect my personal opinion and not any other organization with which I may be affiliated. CT scans have assumed a primary role in the evaluation and diagnosis of many medical conditions, and are very commonly performed procedures. Less appreciated by the public and many professionals (including non-radiology physicians) is the variation in image quality and dose that has been recognized for many years by researchers who have evaluated these factors. As such, there can be substantial variation in CT scan dose and quality, even within the same institution. As a patient, this consideration has figured prominently in my own decisions as to whether or not to proceed with serial CT scans for follow-up of medical conditions. These measures have been carefully crafted to create an effective and validated method to monitor CT image and quality based on indications for the studies and in consideration of individual patient-related variables. As such, they provide a useful and meaningful way to offer our patients and the public the assurance that the scans they are receiving meet reasonable safety and professional standards--which is not routinely available otherwise. These quality measures will meaningfully improve the ability of physicians and health systems alike to monitor the equipment utilized for these studies in a manner that minimizes interference with the typical workflow of a radiology center (or other center) where such studies are performed and will provide a significant and substantial increase in the quality of scans while reducing dose variability that can occur because of machine settings/performance or patient characteristics. Cumulative radiation dose should decline as a result of implementing these measures. At the very least, there will be assurance that the right dose is used for the right scan in the right patient. As a physician and patient advocate for many years, I offer my support for these measures for the reasons stated. And as someone who served as an advisor for this measure, I will add that I was impressed by the exceptional commitment of the developers and their colleagues to provide a meaningful, validated and effective quality measure as they created new processes to measure CT dose and quality, always with an eye towards making this measure acceptable to the professional and consumer communities. (Disclosures: As noted, I was an advisor during the development of this measure and received compensation for those services. I have also served on the NQF Cancer Committee without compensation. I have no other relevant conflicts.)

Comment 9 by: James Anthony Seibert, University of California, Davis Medical Center

January 27, 2022 To: National Quality Forum Dear NQF Standing Committee, I am writing to lend support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco, where I have served on the Technical Expert Panel and have been a long-time collaborator for similar projects between UCSF and UC Davis. I led the implementation of measure testing at my institution, University of California Davis Health, which required local installation of the software, configuring connections to the PACS, extracting CPT and ICD-10 data from the EHR, and supervising the aggregation and transfer of all this data to the UCSF software. Most of this work was completed by our PACS administrator and did not impact the work of our clinicians at any time. One challenge we encountered was that transfer of data from PACS to the software was slow; we believe this was due to capacity limitations of our PACS relative to the query-retrieve process. Nevertheless, we set up auto-transfers of the data over nights and weekends so as not to impact the operation of our PACS during our busiest clinical hours. Besides this issue, the testing was completed successfully with minimal missing data. Based on our experience, the proposed quality measure is highly feasible, and will, in my opinion, be able to appropriately identify CT exams that are significantly above diagnostic reference level (DRL) doses(*), as well as inadequate CT exams with insufficient dose, for specific diagnosis indications

versus radiation dose versus image quality. There are certainly many parameters and issues that can potentially confound such CT quality measures, particularly with the assessment of corresponding image quality, but significant advances in developing robust algorithms to recognize such confounding factors have largely mitigated such concerns. I believe this quality measure can significantly reduce the use of excessive high radiation dose as well as inadequate, sub-optimal low dose used for clinical CT studies, by identifying outliers and thereby increasing the awareness and importance of CT protocol optimization. I support the work to improve patient safety and CT quality as described in these measures. Sincerely, J. Anthony Seibert, PhD, FAAPM, FACR, FSIIM, FIOMP Professor Emeritus, Department of Radiology UC Davis Health (*) Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology 284(1), 120-133, 2017. Disclosure: I have served on the Technical Expert Panel for this effort and have received some minor compensation for participation (honoraria) but have no other relevant conflicts. The opinions expressed here are my own.

Comment 10 by: Kenneth Wang

I am pleased to provide my support for the proposed CT quality measures 3633e, 3662e and 3663e developed by the University of California, San Francisco. I have been a practicing radiologist in the Veterans Affairs (VA) system for more than ten years, during which time I have led efforts in CT dose optimization across the VA Maryland Health Care System. I also serve in a number of volunteer roles within the Radiological Society of North America (RSNA) and the American College of Radiology (ACR), leading efforts in informatics, standards, interoperability and registries. However, this letter reflects my personal opinion, and not necessarily those of any organization with which I am affiliated. I have also served as a member of the Technical Expert Panel (TEP) advising on the formulation of these proposed quality measures, since the inception of this project.

The impetus for this work rests on fundamental principles which are widely accepted. Namely, that CT constitutes an important source of radiation dose to patients, and that CT imaging presents an opportunity for dose reduction, but that it is of paramount importance to maintain the diagnostic quality of the imaging obtained. The proposed measures have been developed using a scientific approach incorporating extensive testing and validation, as well as expert consensus, while maintaining a focus on practicality. This has been all the more impressive given the complex nature of the technical factors involved, such as CT exam types, size-adjusted dose, and diagnostic image quality. By leveraging extensive data, including but not limited to data in the UCSF International CT Dose Registry, data obtained from practicing radiologists on image quality, and feedback from testing facilities, the measures strike a practical balance intended to identify opportunities for CT dose reduction while maintaining a floor for diagnostic quality (which was rarely violated in measure testing).

As such, these measures represent an important step beyond simple dose reduction. I also believe that these measures will provide actionable feedback, especially given the many different techniques now available on modern CT scanners for dose adjustment.

As a radiologist, I know there will never be universal agreement on subjective assessments such as image quality. However, the proposed measures take a balanced approach, informed by extensive testing and validation, which serves a very practical and important quality objective. For these reasons, I support the adoption of these measures.

Comment 11 by: Krishna Nallamshetty, Radiology Partners

I would like to submit a comment in support of this measure. I am a practicing radiologist for the past 15 years and serve as the Associate Chief Medical Officer of Radiology Partners, the largest medical imaging practice in the United States. I am the chair of our national Patient Safety Committee. We have seen tremendous growth in medical imaging that requires radiation, specifically computed tomography (CT). The public awareness of the potential long-term effects of ionizing radiation has become mainstream and as a result, a primary objective of the American College of Radiology and other governing bodies. The objective focuses on reducing radiation exposure as much as possible without compromising the diagnostic information that is obtained

We have recognized that there is large variability in how CT scans are acquired all over the country. Techniques and radiation exposure is extremely varied but yet appropriate clinical diagnosis are made. This measure evaluates radiation dose for every patient who undergoes CT *based on the clinical indication for imaging* rather than solely on the type of examination that is performed. It ensures patients receive the most appropriate CT acquisition protocol and level of radiation for their individual condition. The measure also assesses image noise, safeguarding image quality against potential effects of dose reduction, and is the first quality measure to do so.

The measure would have a large, positive impact on patients and protect them from unnecessary over-exposure of radiation without compromising the diagnostic value of medical imaging. It would be the first time a measure addresses both radiation and image quality.

Comment 12 by: Maribel Escobar

Submitting on behalf of ARA's CMO, Dr. John Kish. January 25, 2022 Dear NQF Standing Committee, I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. As an implementation testing partner, my institution, ARA Diagnostic Imaging, was required to install the data collection software, route CT data from PACS and order and billing data from various electronic systems to the software, and oversee the migration of data. We also worked with UCSF and our CT vendors to ensure the Radiation Dose Structured Report (RDSR) was being saved from each exam in the PACS. As we discussed in an interview with UCSF, this work fell on the PACS team and IT colleagues and did not require effort from clinicians. Besides some technical hiccups, which were all resolved, we faced few barriers to successful implementation and had very little missing data. Based on our experience, the proposed quality measure is highly feasible. We have also been satisfied with the feedback we have received from Alara Imaging on our measure performance, which brought to our attention some areas of opportunity to decrease radiation dose. The feedback provided by Alara Imaging has taken the burden of researching problem areas away from my institution, by identifying specific exams, imaging protocols and even specific CT units that contribute to high radiation dose and need improvement. We have plans to address each accordingly. Given our positive experience, my organization is moving towards a commercial relationship with Alara to continue to submit data, receive feedback, and strive to optimize our CT doses. I earnestly believe this quality measure can help mitigate the use of excessive radiation doses used in CT. I support these measure developments in order to improve patient safety and CT quality. Sincerely, John Kish, MD Chief Medical Officer

Comment 13 by: Mary White

I am writing in support of this CT radiation dose safety measure. As a cancer epidemiologist, I recognize that excessive exposure to medical radiation increases cancer risk. And I understand that this measure will be valuable for protecting patients from unnecessarily high levels of radiation from CT imaging. The measure is designed to evaluate radiation dose for every patient based on the clinical indication for imaging. The measure also assesses image noise, ensuring adequate image quality despite the reduction in radiation dose. This measure fills an important quality void and has the potential to substantially reduce the contribution of CT scans to the incidence of cancer in the population.

Comment 14 by: Matthew Nielsen

I am writing in support of this important measure. The utilization of CT imaging in the United States has dramatically increased over recent decades, providing numerous benefits to patients and clinicians in the management of countless medical conditions. There has also been increasing recognition of the potential for unintended harms due to potentially avoidable variation in radiation in radiation dose for many patients. Evidence from research and quality improvement efforts demonstrates the potential to mitigate these harms with a feedback loop and benchmarking to radiologists and staff. This measure provides needed resources to disseminate these early successes, preserving the benefit of advanced imaging with CT while providing a means for healthcare facilities and clinicians to improve the safety of the studies they provide patients. The design of this measure importantly takes into account the indication for the study as the framework for dose benchmarking, with balancing measures of image quality to assure that efforts to reduce dose do not come at the expense of diagnostic quality. Given the increased recognition from patients and providers of the potential harms of imaging-associated radiation, this measure fills a timely and important gap in the current measurement portfolio.

Comment 15 by: Pavlina Pike, Huntsville Hospital

I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. I am a Medical Physicist and Radiation Safety Officer at Huntsville Hospital and led the testing of UCSF's quality measure at my health system, which involved installing the data collection software, routing CT data from PACS and order and billing data from various electronic systems to the software, and overseeing the migration of data. We came onboard late in the testing period, leaving a tight window of time to collect the data prior to UCSF's submission deadlines. I am proud of my PACS and IT colleagues for pulling together so efficiently and completing the work rapidly with very little missing data. The work in no way impacted our physicians or clinical workflows. We faced few barriers to implementation, and based on our experience, the proposed quality measure is highly feasible.

We have also been satisfied with the feedback we've received from Alara Imaging on our measure performance, which brought to our attention areas of high radiation dose. Our exams were compared to thresholds established based on input from 125 radiologists and 50,000 CT examinations from other facilities. The analysis includes comparisons of the performance of different model CT scanners, exams, protocols, patient size, facility, etc. The feedback from the Alara software is helpful and actionable as we are able to identify what changes will have the greatest impact on patient dose and make the appropriate changes. In addition it provides suggestions for billing inconsistencies which was very helpful to our administration.

I earnestly believe this quality measure can help mitigate the use of excessive radiation doses used in CT. I support the work of the measure developers to improve patient safety and CT quality.

Comment 16 by: Robert Gould, University of California, San Francisco Medical Center

I am writing as a physician who has worked for decades as a leader in Physicians for Social Responsibility, as well as the International Physicians for the Prevention of Nuclear War toward eliminating nuclear weapons, cognizant of the public health dangers of radiation initially derived from studies of victims of the twin atomic bombings in Japan. Informed by the central tenet of physician practice to "at first do no harm," I strongly support CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. While my long experience as a practicing pathologist has made me understand at a profound level how diagnostic radiation is a critical tool in medical practice, it has also underscored to me the often-overlooked risks of carcinogenesis that must always be balanced against the benefits of various radiological procedures. Over time, research has documented that many radiological procedures are medically unnecessary when information that is desired can be obtained by other means than exposing a patient to ionizing radiation; it is also unwarranted when employed as a "hedge" against possibility of malpractice litigation. In addition, when radiological imaging is indeed required and justifiable, it is not uncommon, where standards are not uniformly applied in practice, for radiation exposures to exceed what would be required for achieving images satisfactory for diagnostic purposes. As such, the lack of attention to standardizing, and minimizing exposures inevitably results in the induction of significant numbers of unnecessary cancers that would not occur if lower doses were employed to achieve adequate imaging. I believe that CT quality measures 3633e, 3662e, and 3663e would be important steps to assuring that physicians can obtain the information necessary from diagnostic imaging while minimizing the number of unnecessary cancers induced by the procedures.

Comment 17 by: Suz Schrandt

As a patient advocate with significant experience navigating the healthcare system--including repeated exposures to a variety of diagnostic imaging studies--I submit these comments in endorsement of this measure. The measure takes into account different contexts and parameters for a given patient and his or her unique benefit/risk profile. At a more foundational level, the measure calls into focus the significant variation in practices in CT imaging that can expose patients to unnecessary and/or unsafe levels of radiation, a risk many patients are not even aware of. The wide-spread use of this measure could standardize imaging practices and should the measure be adopted, I strongly encourage a robust dissemination plan to inform patients and families of its existence. Our ability to access safe and effective care should not be left to change; measures such as this help to close key gaps in our system.

Comment 18 by: Melissa Danforth, The Leapfrog Group

Founded in 2000 by large employers and other purchasers, The Leapfrog Group is a national nonprofit organization driving a movement for giant leaps forward in the quality and safety of American health care. The flagship Leapfrog Hospital Survey collects and transparently reports hospital performance, empowering purchasers to find the highest-value care and giving consumers the lifesaving information they need to make informed decisions. For the past several year's Leapfrog has been collecting and publicly reporting hospital performance on an NQF-endorsed Pediatric CT Radiation Dose (NQF 2820) measure. The new Excessive Radiation Dose or Inadequate

Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level) fills a critical gap in evaluating radiation dose for adult patients who undergo CT. Additionally, because the measure is based on the clinical indication for imaging – rather than on the type of examination the radiologist chose to perform – it can help ensure patients receive the right type of CT and amount of radiation for their individual condition, which is a primary concern of Leapfrog and our purchaser and employer membership. The measure also assesses image noise, safeguarding image quality against potential effects of dose reduction, and is the first quality measure to do so. Leapfrog strongly supports this measure.

Comment 19 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

Comment Part 1:

We thank the American College of Radiology for their comments from 1/19/22 but wish to address several factual inaccuracies in the comments. (Reponse PART 1) **Comment:** *Indications for exams do not have standardized language that could be used to track them. Most health and IT systems capture…coding for reimbursement, but typically not enough... As a result, the clinical reason for performing an imaging exam is often extremely limited in the exam order... A validated method for determining classification of studies .. must be incorporated into such a measure.* **Response:** This statement indicates that the commenter does not understand how clinical indication is determined in the proposed measure. It does not rely on the clinical reason for performing an imaging exam in the exam order. As described in Specifications, sp-11, clinical indication for imaging is determined using an algorithm that combines procedure (CPT®) and diagnosis (ICD-10-CM) codes associated with the clinical visit when the test was ordered, information provided as part of the order, and information on the final bill. The codes are available in the radiology electronic systems and/or the EHR or billing systems. The goal in creating the CT categorization decision rules was to identify exams that are exceptions to the routine dose category (i.e. either high or low dose). The approach of assigning CT exams to the various CT categories in an automated fashion using an algorithm was developed using over 4.5 million CT exams in the UCSF International CT Dose Registry. We confirmed that the CT categories were representative of groupings that require different radiation dose and image quality (Smith-Bindman 2021). The algorithm was validated using over 10,000 patient records from UCSF Health. The CT category assignment determined by the algorithm was compared with a "gold standard" chart review, as described in Validity sections 2b.02 and 2b.03. Since we did not have access to complete medical records at testing sites, we developed a second referent standard that determined CT category based on natural language processing of DICOM data and the full radiology report. This second referent standard was found to be accurate compared to the gold standard chart review of the same sample of UCSF Health exams (sensitivity = 0.92, specificity = 0.97; see 2b.02). When the algorithm was deployed at testing sites, the correct classification rate of CT category assignment was on average 92% across clinician groups and hospitals and 95% in individual clinicians (see 2b.03). Knowing that the algorithm was developed using data from a single health system, we performed detailed investigation of the categorization results at testing sites – comparing the assigned CT category against full radiology reports – for the purpose of improving the algorithm, which we did. **Reference:** Smith-Bindman R, Yu S, Wang Y, et al. An Image Quality-informed Framework for CT Characterization. Radiology. 2021 Nov 9:210591. **Comment:** *The developer states their company can provide the service of quantifying the measure at a cost; this should also be included as a potential limitation. The measure developer does provide specifications for other entities to implement the measure, but the burden of implementation may be significant.* **Response:** This is inaccurate. As stated in Feasibility, 3.07, there are no fees for users submitting their eCQM data to CMS programs. The eCQM can be run and the

measure score calculated by any EHR vendor or hospital and reporting entities can partner with any commercial partner capable of developing reporting software using the eCQM specifications. The measure steward's software to ingest this data and calculate the measure is freely available. Alara Imaging has created an edge device that can assemble data from different electronic sources (e.g. EHR, RIS [Radiology Information Systems], PACS [Picture Archiving and Communication Systems], and billing) to calculate the CT category, size-adjusted dose, and image noise that can then be consumed by the eCQM. If practices want to calculate these variables without using the Alara edge device, they may access a free online portal to calculate these variables and provide them to any entity implementing the measure. A prototype of this software was deployed at 8 testing sites (7 hospital systems and 1 ambulatory imaging network). Sites were asked to install the software, configure local connections to PACS, EHR, and other electronic systems as needed, and oversee the transfer of data to it from these sources. Burden was found to be no more or less onerous than the effort required by participation in other eCQMs or national registries, such as the ACR Dose Index Registry (Feasibility, 3.06). **Comment:** *For CT category … the developer used NLP for obtaining data such as reason for study or protocol name used in the calculation of this variable. The submission does not provide information on the NLP results' reliability and validity… or how sites would get access to use this custom NLP tool.* **Response:** This is incorrect; the measure does not use NLP. As described in the submission and above, it uses an algorithm that combines CPT® and ICD-10-CM codes to categorize CT exams. NLP was deployed as a method to validate the CT categorization determined by the algorithm at testing sites, where we did not have access to medical records. The sensitivity and specificity of this NLP referent standard are given above. **Comment:** *Multiple unstructured variables are required to construct the data elements for the numerator, denominator, and exclusions…* **Response:** This is incorrect; the measure does not use unstructured data. All data elements used to calculate the measure come from structured variables listed in the feasibility scorecards and in Specifications, Table sp-2: CPT® and ICD-10-CM codes; dose length product stored in the DICOM data; and patient diameter and image noise calculated on imaging data. The measure would not have met the requirements of an eCQM had it relied on unstructured data. **Comment***: Protocol selection appropriate for a clinical indication is an important component of radiation dose management along with radiation dose optimization. Each component needs to be addressed as a separate quality action. The specific aspect(s) of performance to be improved is not intuitive due to the multiple components to the measures… It is true that the most accurate way to address appropriate and safe use of multi-phase studies is to measure both the clinical indication of an exam and the radiation dose output… However, these measures conflate the appropriateness of protocol for the clinical indication and radiation dose optimization… a facility may not be able to determine if its performance could be improved by adjusting protocols or by focusing on appropriateness of the ordered exam.* **Response:** We agree that selecting an appropriate CT protocol and limiting radiation dose given the selected protocol are separate quality actions, but the commenter misses the crucial point that intermediate outcome measures typically reflect multiple opportunities for improvement. By analogy, we recognize systolic blood pressure control and glycosylated hemoglobin control as intermediate outcome measures for patients with hypertension and diabetes, respectively, even though there are many potential ways to manage these conditions. The fact that these intermediate outcomes can be improved by diet, exercise, medications, or combined approaches does not invalidate glycosylated hemoglobin or blood pressure control as quality measures. Similarly the fact that our measure would be responsive to multiple, interrelated process steps is a key strength that will improve its value for reducing radiation exposure at the population level. Further, reporting entities will be provided with feedback for each CT exam, including its assigned CT category, radiation dose, size-adjusted radiation dose, and image noise, allowing recipients to identify the causes of performance gaps. Reporting entities will be able to assess if they are systematically assigning patients to the wrong

protocol, or if they are choosing protocol settings that are inappropriate with respect to radiation dose or image noise. The actionability of the feedback is noted in the other letters written in support of the measure. To further demonstrate the potential of this measure, we conducted a randomized controlled trial in 100 hospitals and outpatient radiology practices to study the impact of providing detailed audit feedback, similar to what will be provided as part of the feedback on this measure. We found that this intervention resulted in significant reductions in radiation dose and dose variation with no impact to image quality, described in Usability, 4b.01. (Smith-Bindman, 2020) **Reference:** Smith-Bindman R, Chu P, Wang Y, et al. Comparison of the Effectiveness of Single-Component and Multicomponent Interventions for Reducing Radiation Doses in Patients Undergoing Computed Tomography: A Randomized Clinical Trial. JAMA Intern Med. 2020 May 1;180(5):666-675.

Comment 20 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

Comment Part 2:

We thank the American College of Radiology for their comments from 1/19/22 but wish to address several factual inaccuracies in the comments. (Response PART 2) **Comment:** *NQF #3633e, #3662e, and #3663e deviate from international standards, like diagnostic reference levels, and lack peer reviewed, broadly accepted consensus on global noise. For these measures, global noise is defined solely by the measure developer. Endorsing this method may encourage facilities to accept a narrow view of image quality.* **Response:** The ACR correctly notes that we have defined an approach to measuring noise. We did so only after testing and comparing multiple approaches described in peer-reviewed literature and validating noise measurements against radiologists' assessment of image adequacy for diagnosis. Image quality is a much less common problem than excessive use of radiation in CT imaging. While there may be other reasons to study CT image quality, our interest was simply to ensure that CT image quality did not erode as an unintended consequence of lowering radiation doses. There is no reason to believe that endorsing this measure will encourage facilities to "accept a narrow view of image quality" because radiologists have a requirement for adequate images to perform their work. They have no desire or motivation to alter their standards of what constitutes an adequate image. Radiologists do not want to read inadequate images and routinely request that such images be repeated or complemented by other imaging modalities. **Comment:** *The ACR requests the developer further clarify the global noise table used in calculating the numerator... For example, Table sp-1 has the same global noise threshold for several CT categories, such as head low dose, head routine dose, and head high dose... If the image noise thresholds are the same, the size-adjusted radiation dose thresholds should be the same.* **Response:** We tested various published methods for measuring image noise and opted for a modified version of the method proposed by Malkus in 2017. The approach for setting the thresholds for image quality and radiation dose were based on the referent standard of radiologists' satisfaction with image quality. This did not always result in the relationship the ACR has suggested. For example, radiologists might want a minimum level of image quality for all head CT categories whereas the upper dose threshold might vary across the three head categories reflecting the different clinical indications comprising each group. Radiologists in our image quality study graded the majority of head exams as having acceptable image quality, even those at the lower dose range, meaning the minimum noise threshold is similar for all three categories. **Reference:** Malkus A, Szczykutowicz TP. A method to extract image noise level from patient images in CT. Med Phys. 2017 Jun;44(6):2173-2184. **Comment:** *Additionally, current CT scanners display dose values based oneither a 16 cm or 32 cm phantom for a neck scan, which must be carefully accounted for in measure performance calculations.* **Response:** As the ACR correctly notes, CT scanners display dose values based on a 16

cm or 32 cm phantom. If comparisons are made across reporting entities it is important that they use the same phantom, as this impacts the scanner reported DLP. The manufacturers are highly consistent in their use of phantoms for different body regions. In a study of 106,837 pediatric patients (a population where potential variation in phantom choice would most likely occur), 100% of CT exams in the neck are referenced to the 32 cm phantom, and it is thus unnecessary to account for phantom selection (Chu 2021). **Reference:** Chu PW, Yu S, Wang Y, et al. Reference phantom selection in pediatric computed tomography using data from a large, multicenter registry. Pediatr Radiol. 2021 Dec 6. **Comment:** *These eCQMs require multiple variables that may be captured in software systems external to electronic health records (EHRs), such as dictation systems housing radiology reports or DICOM standard-based systems, such as CT device software. Data element validity testing should demonstrate that the testing sites were able to integrate and validate the variables used to construct the data elements used by the eCQM in addition to the usual validation of the eCQM's electronic output against the medical record review. We are uncertain that this validation has been completed. Therefore, this submission does not demonstrate the measure can be reproduced in a reliable and valid manner by practices or facilities across multiple settings.* **Response:** This comment is entirely erroneous. No data are pulled from dictation systems or CT device software. The measure derives and uses codified and specified data from DICOM standard based systems, such as PACS, and EHR and billing claims. Our data element validity testing did demonstrate that 8 testing sites, reflecting 16 hospitals and 13 outpatient imaging facilities, were able to integrate, collect, and report the variables used to construct the data elements ingested by the eCQM. The letters of support from these testing sites independently confirm their ability to assemble the required data across diverse practice types and settings. **Comment:** *The ACR is concerned with the selection bias for the accountable entity-level… validity. Assessing measure score face validity through the TEP that created these measures lessens the extent of credibility for these results. Although the TEP is knowledgeable and represents a variety of stakeholders, there is a vested interest in ensuring these measures are available for use.* **Response:** All of the TEP members and their affiliations are identified in our submission materials (2b.02). Conflicts of interest were reviewed at each meeting and included with meeting minutes in a publicly available website (https://ctqualitymeasure.ucsf.edu/). The TEP members all voluntarily provided public service by joining the TEP. None of our TEP members has any "vested interest" in the outcome of the NQF endorsement process other than the ACR which served as a single member of the TEP. None of our TEP members is employed by the developer organization (UCSF) or its funder (CMS), nor has any financial interest in the company that is offering technical support for software implementation (Alara Imaging). To be clear, these measures were developed by an academic radiology, quality improvement, and analytics team based at UCSF and supported by CMS, NIH and PCORI. The TEP was organized and tasked to provide broad multidisciplinary input to this team. Their endorsement of the validity of the measures is highly credible, as it reflects the fact that their advice was heeded at every stage of the development and testing process. Our TEP process followed the CMS Blueprint as well as NQF guidance, and 16/17 members agreed that that implementation of the measure will lead to a reduction in average CT radiation dose while maintaining adequate CT image quality if adopted (reported in 2b.03).

Comment 21 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

We thank the American Association of Physicists in Medicine for their perspectives but wish to address several factual inaccuracies: **Comment 1:** *Unscientific characterization of CT scan risk: The proposal is based on estimation approaches that are not reflective of the consensus of the scientific community.* **Response:** The measure is not focused on radiation risk and does not calculate nor

report radiation risk. The measure evaluates dose length product (DLP), and specifically whether size-adjusted DLP exceeds thresholds specific to CT category. DLP is the radiation dose measure most directly under the control of providers, determined by selected parameters. Further, DLP is universally reported by CT manufacturers. It is thus the ideal measurement to use when assessing the quality of CT exams. The TEP, which included the ACR, radiologists and a medical physicist, unanimously supported the radiation dose measure used and agreed is a relevant metric of quality for CT imaging (2b.03). There is also considerable precedent for using DLP to evaluate radiation dose in CT. The American College of Radiology has used DLP to set benchmarks [Kanal 2017] and to measure dose in its own NQF-endorsed quality measure #3621. **Reference:** Kanal KM et al. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-133. **Comment 2:** *Inactionability of the measures to enable targeted change to improve practice: It is not evident how the proposed measures can be practically used* **Response:** Reporting entities will be provided with specific feedback for each CT scan on its assigned CT category, radiation dose, size-adjusted radiation dose, and image noise, allowing recipients to identify causes of performance gaps and make targeted changes to improve quality. Comments in support of the measure from the testing sites describe how useful the information provided was to allow them to understand and improve their practice. As described in our submission, we found in a randomized controlled trial in 100 imaging facilities that providing detailed audit feedback on radiation doses, similar to what will be provided as part of the feedback on this measure, resulted in significant reductions in radiation dose with no impact on satisfaction with image quality (see Usability, 4b.01). (Smith-Bindman, 2020) **Reference:** Smith-Bindman R et al. Comparison of the Effectiveness of Single-Component and Multicomponent Interventions for Reducing Radiation Doses in Patients Undergoing Computed Tomography: A Randomized Clinical Trial. JAMA Intern Med. 2020 May 1;180(5):666-675. **Comment 3:** *Inadequate addressing of the complexity of CT categorization*. **Response:** A detailed response to this question was provided in our response to the ACR. In short, the approach of assigning CT examinations to the different CT categories as specified in the measure was first developed using records from over 4.5 million CT exams in the UCSF International CT Dose Registry (Smith-Bindman, 2021). We then developed an approach for determining the clinical indication for imaging using an algorithm that combines procedure (CPT®) and diagnosis codes (ICD-10-CM) provided in Specifications, sp-11. This algorithm was developed using detailed review of over 10,000 patient records from UCSF Health. We validated the CT category assignment using the algorithm against "gold standard" chart review, as described in Validity sections 2b.02 and 2b.03. When the algorithm was deployed at our testing sites the correct classification rate of the assignment of CT exams to CT category was on average 92% across clinician groups and hospitals and 95% in individual clinicians. **Comment 4:** *Inadequate assessment of noise: Noise in a CT image can be influenced by a variety of factors*. **Comment 5:** *Inadequate assessment of image quality: Image quality is affected by a myriad of factors*. **Response:** The primary focus of our measure is to assess radiation dose adjusted for body size. The image quality component was included to protect against the unlikely possibility of substantial degradation of image quality as an unintended consequence of dose reduction. Our measure of image quality reflects what radiologists in practice regard as adequate. Others might have an interest in other ratings of image quality for other purposes, but that was not our intent. We tested and found that noise as a measure of image quality was associated with radiologists' satisfaction with the adequacy of CT images. These results were included in the submission (2b.03). **Comment 6:** *Flawed assumption on dose reduction vs dose optimization: The application focuses primarily on radiation dose reduction as opposed to right-sizing the dose.* **Response:** This is incorrect. We created the CT categories based on radiation dose and image quality requirements specific to clinical indications for imaging. Using radiologists' satisfaction with image quality, we established an image quality floor for each category, below which an exam is considered to have inadequate quality, and a

radiation dose ceiling, beyond which doses are considered unnecessarily high. The purpose is to allow detailed assessment of each CT exam to ensure the dose is optimal based on the clinical indication for imaging. In current practice, there are no such benchmarks created by clinical indication, making it impossible for providers to know the right dose range for each patient. In our testing data, far more CT exams exceeded the radiation dose ceiling (average = 30%) than failed to meet the image quality requirement (average < 1%) (see section 1b.02). The measure encourages entities to reduce the proportion of exams that may "be overdosed for their exact need and condition" while preserving the minimum image quality. **Comment 7:** *Inadequate accuracy in patient size estimation: Assessing a patient size is not a trivial task, stemming from significant variability in the differences in the habitus of different patients, coupled with the existential challenge that there is no single metric* **Response:** We agree that measuring patient size is important. Our approach for using mid-scan diameter is highly correlated with patient weight: in separate, NIH-funded research on CT use in children up to age 21 (Kwan 2022), we have shown that diameter in 4,239 children as measured on mid-scan axial images is highly predictive of patient weight, correlation = 0.904, Figure 1.

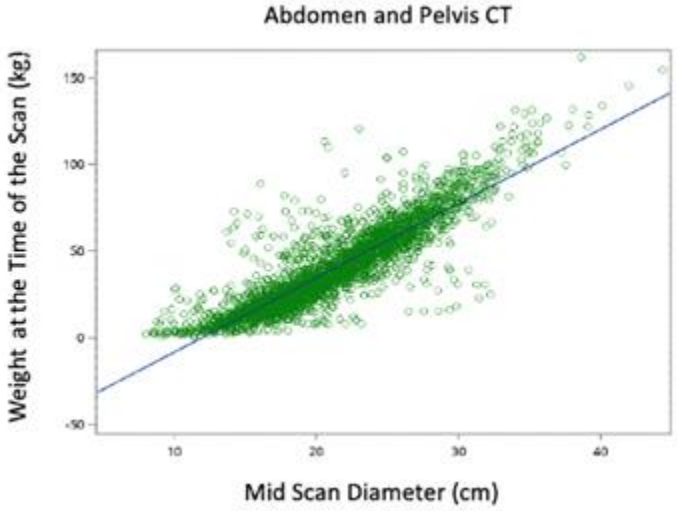


Figure 1. Mid scan patient diameter versus patient weight at the time of the CT, Kwan 2022

For this measure, patient size was measured using CT image pixel data, either on the mid-scan axial image or the coronal scout image when the mid-scan axial image was not available. This approach has been validated using data from UCSF Health, the UCSF Registry, as well as the data assembled for measure testing. While there may be different ways to measure patient size, and different reasons for measuring patient size, it is a crucial piece of information that must be practically defined to ensure that the types of patients (case mix) at different practices do not bias the number of scans graded as out-of-range. We are adjusting for patient size primarily to ensure that entities that see larger patients are not penalized for doing so. Figure 2a shows the relationship between radiation dose (in DLP) and patient diameter using data from the UCSF Registry for abdomen CT. We chose abdomen CT as this is the category most influenced by patient size, and where patient mix could impact an entity's out-of-range rate. The raw correlation between patient diameter and unadjusted DLP is 0.50, and the marginal R-squared of the log-linear model used for adjustment is 0.15. After size-adjustment, the relationship is nearly removed: Figure 2b shows size-adjusted DLP by patient diameter using the same data; the raw correlation is far lower (-0.09), and the modeled marginal R-squared post-adjustment is 0. This demonstrates adequacy of the approach for adjustment of patient size.

Figure 2a: Unadjusted Dose Length Product vs Patient Diameter

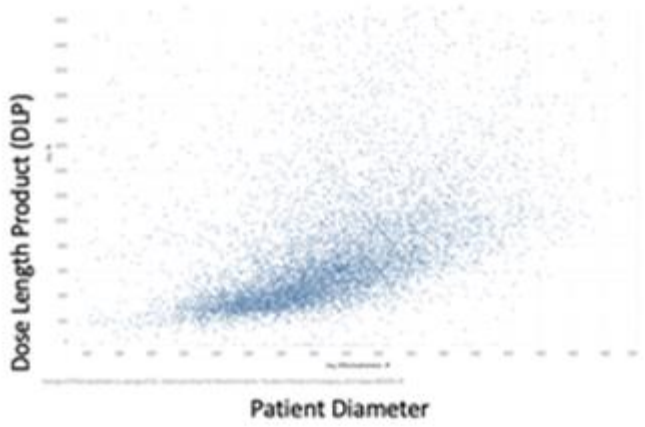
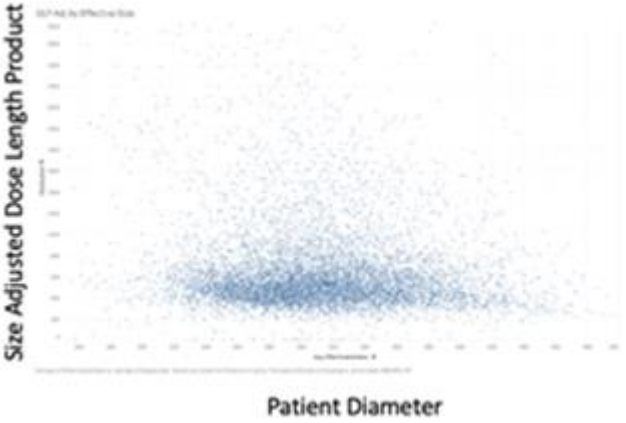


Figure 2b: Size-Adjusted Dose Length Product vs Patient Diameter



The adequacy of size adjustment was shown empirically using data assembled from the testing sites. The proportion of exams with out-of-range rates based on **unadjusted** and **adjusted** DLP are shown in Tables 1a and 1b. Without adjustment, the out-of-range values are strongly associated with patient size; after adjustment this relationship is gone.

Table 1a) Proportion of exams out-of-range on routine dose abdomen exams based on **unadjusted** DLP across the 16 hospitals, shown by decile in patient size. The proportion of out-of-range exams increased with patient size, seen in the table as an increase in dark shading in the lower rows of the table. Among patients in the highest size decile – those in last row– the out-of-range proportions across the 16 hospitals ranged from 93-100%.

Size Decile	Hospitals 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.27	0.22	0.11	0.00	0.27	0.29	0.30	0.14	0.34	0.20	0.24	0.40	0.06	0.17	0.11	0.17
2	0.30	0.00	0.00	0.00	0.08	0.11	0.13	0.29	0.24	0.30	0.04	0.19	0.00	0.07	0.16	0.09
3	0.15	0.06	0.15	0.00	0.17	0.18	0.22	0.75	0.21	0.30	0.17	0.18	0.08	0.18	0.17	0.12
4	0.07	0.17	0.29	0.15	0.25	0.32	0.09	0.82	0.43	0.25	0.07	0.42	0.10	0.17	0.19	0.21
5	0.45	0.15	0.13	0.14	0.28	0.43	0.00	0.93	0.40	0.42	0.19	0.38	0.00	0.14	0.48	0.55
6	0.42	0.20	0.25	0.36	0.55	0.61	0.27	0.96	0.55	0.19	0.31	0.51	0.08	0.46	0.47	0.78
7	0.79	0.47	0.45	0.58	0.70	0.75	0.17	1.00	0.69	0.37	0.26	0.73	0.06	0.71	0.66	0.90
8	0.81	0.37	0.75	0.69	0.67	0.86	0.24	1.00	0.89	0.35	0.58	0.77	0.22	0.80	0.91	0.95
9	0.96	0.85	1.00	0.75	0.88	0.93	0.26	0.93	0.94	0.64	0.78	0.93	0.63	0.90	1.00	1.00
10	1.00	0.96	0.98	0.93	0.97	0.97	0.93	0.94	1.00	0.95	0.98	0.96	0.85	0.95	0.94	1.00

Table 1b) Proportion of exams out-of-range on routine dose abdomen exams, based on **size-adjusted** DLP across the 16 hospitals shown by decile in patient size. High proportion of out-of-range exams are no longer concentrated among the larger patients. Among patients in the highest size decile, out-of-range rates ranged from 11-53%.

Size Decile	Hospitals 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.55	0.61	0.22	0.20	0.42	0.48	0.48	0.61	0.49	0.70	0.37	0.62	0.10	0.37	0.22	0.51
2	0.50	0.25	0.00	0.18	0.19	0.31	0.22	0.71	0.36	0.61	0.15	0.38	0.00	0.08	0.33	0.21
3	0.15	0.18	0.15	0.11	0.37	0.39	0.33	0.96	0.30	0.50	0.26	0.36	0.10	0.23	0.43	0.22
4	0.27	0.17	0.29	0.15	0.35	0.54	0.09	0.97	0.43	0.38	0.11	0.53	0.10	0.18	0.31	0.30
5	0.45	0.15	0.13	0.14	0.26	0.46	0.00	0.93	0.40	0.42	0.19	0.38	0.00	0.19	0.52	0.59
6	0.33	0.10	0.25	0.36	0.39	0.45	0.27	0.96	0.47	0.13	0.31	0.40	0.05	0.34	0.45	0.72
7	0.29	0.18	0.20	0.46	0.50	0.42	0.17	0.90	0.57	0.16	0.17	0.60	0.04	0.50	0.36	0.70
8	0.43	0.05	0.19	0.25	0.54	0.39	0.12	0.70	0.58	0.09	0.35	0.62	0.09	0.59	0.53	0.83
9	0.48	0.26	0.48	0.30	0.45	0.27	0.11	0.19	0.72	0.07	0.18	0.56	0.06	0.62	0.60	0.66
10	0.35	0.27	0.40	0.39	0.38	0.11	0.27	0.11	0.61	0.37	0.29	0.44	0.11	0.27	0.53	0.36

Reference: Marilyn Kwan et al. Quantifying cancer risk from exposures to medical imaging in the Risk of Pediatric and Adolescent Cancer Associated with Medical Imaging (RIC) Study: Research Methods and Cohort Profile. In press, Cancer Causes and Control. **Additional Comment:** *One cited reference supports the proposed measure, however, this cited article has an accompanied editorial that highlights the limitations of the proposed approach [Mahesh M. Benchmarking CT Radiation Doses...Radiology. 2021 Nov 9:212624.]* **Response**: We find it surprising that Dr. Mahesh's editorial was used to criticize the measure. Dr. Mahesh is a board member of American College of Radiology and American Association of Physicists in Medicine, and he was very positive about our image quality-informed framework for assessing radiation dose. He noted the observed, significant differences *between* CT categories versus *within* categories was "an encouraging result for anyone trying to optimize CT studies based on clinical indications." He noted the study was "a good start" on the road to optimizing CT protocols based on image quality. He opined that the CT classification would be more useable and easier to implement if based on current procedural terminology codes. This is precisely what we have done in this measure.

Comment 22 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

We thank Dr. Ehsan Samei for sharing his perspectives on the measure and for collaborating with us early in the measure development process. We wish to address a few inaccuracies and misunderstandings in Dr. Samei's comments. The majority of Dr. Samei's comments focus on image quality and his concern that the measure does not offer a comprehensive assessment of image quality. Our measure is not intended to be a comprehensive assessment of image quality.

Criticizing the proposed measure for what it is not is beyond the scope of what should be considered in assessing the usefulness of what has been submitted. The primary focus of our measure is to assess radiation dose adjusted for body size, and the image quality component provides a means to protect against the unlikely possibility of substantial degradation of image quality as an unintended consequence of dose reduction. The approach for creating thresholds is described in Validity, 2b.02. **Comment:** ***Inaccurate assessment of patient size:*** *The measure of size proposed is calibrated to earlier work and publication from our group at Duke University for academic purposes. That early method they have embraced has had major errors.* **Response**: We are adjusting for patient size primarily to ensure that entities that see larger patients are not penalized for doing so. Although we explored code that Dr. Samei provided early in our initial efforts to measure patient body habitus we found that it was inadequate, particularly for some CT categories, and we have not relied upon it. We developed our own approach for measuring size using CT image pixel data from the mid-scan axial image or the coronal scout image when the mid-scan axial image was not available. Our approach of measuring size was shown to be highly correlated with patient weight (correlation = 0.904) in a large study in children described in our response to the AAPM. For this measure, the measurement of size was validated using data from UCSF Health, the UCSF Registry, as well as the data assembled for measure testing. The adequacy of the approach we have adopted for size adjustment is described in the initial application and the response to the comments by the AAPM. **Comment:** ***Inaccurate assessment of noise:*** *The measure of noise proposed references earlier work and publication from our group at Duke University. That early method exhibited errors, corrected in subsequent versions that have not been shared…* **Response:** Dr. Samei's approach and code for measuring image quality were explored in the process of developing our measure but were not included in the final measure specifications. Any errors in his approach are not relevant to the measure. **Comment:** ***Inaccurate assessment of radiation risk:*** *The measure of size-adjusted radiation risk, adjusting the CT scanner outputs with 'patient size' to perform risk estimation is not a standard method nor endorsed by any scientific or professional body… Patient risk can only be assessed with the knowledge of organ doses that is not even mentioned in the application let alone pursued. The proposed method CANNOT be used as surrogate for future cancer risk.* **Response:** The measure does not calculate or report radiation risk. The measure evaluates radiation *dose* (measured in dose length product, DLP), and whether size-adjusted DLP exceeds thresholds specific to CT category. The empirical validity of the risk-adjustment approach based on patient size is described in the application (section 2b.26 – 2b.31) and in our response to the comments by the AAPM. The approach of evaluating CT safety by comparing machine output (whether DLP or CTDIvol) against benchmarks is widely accepted in the radiology field. (Kanal 2017) In contrast, organ dose has no standard definition, is not reported by the manufacturers, is not available in a structured format, would be time intensive to calculate in clinical settings and most importantly has limited actionability as this is not under the direct control of technologists or physicians. Organ doses may be useful for counseling patients or in the context of epidemiological studies, but we do not believe it has a role as a metric for CT quality measurement. **Reference:** Kanal KM, Butler PF, Sengupta D, et al. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-1 ***Comment:*** ***Subjectivity:*** *The measures are anchored to subjective perception by radiologists as how they "like" the images. There is in fact no evidence provided that the measures can lead to an improvement in diagnostic accuracy. In fact, it might lead to a degradation.* **Response:** The measure is not intended to improve diagnostic accuracy. The purpose of the measure is to establish a radiation dose ceiling to avoid excessive radiation exposure, and an image quality floor to safeguard against unintended deterioration of image quality. There is precedent for using radiologist satisfaction with image quality to set or validate noise targets, including work by Dr. Samei. (Cheng 2019, IAEA 2009) This also reflects clinical practice: radiologists subjectively assess images and regularly ask for scans to

be repeated when they are not adequate. As described in the response to ACR comments, Radiologists do not want to read inadequate images and routinely request that such images be repeated or complemented by other imaging modalities. Radiologist's subjective assessment provides a practical way to ensure the image quality is not degraded through efforts to optimize the radiation doses. **References:** Cheng Y, Abadi E, Smith TB, Ria F, Meyer M, Marin D, Samei E. Validation of algorithmic CT image quality metrics with preferences of radiologists. Med Phys. 2019 Nov;46(11):4837-4846. doi: 10.1002/mp.13795. Epub 2019 Sep 20. International Atomic Energy Agency (IAEA), Dose Reduction in CT while Maintaining Diagnostic Confidence: A Feasibility/Demonstration Study, TECDOC Series, 2009.

NQF #3663e Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Facility Level)

Comment 1 by: Karen Orozco, American College of Radiology
On behalf of Karen Campos, American College of Radiology

The American College of Radiology, representing more than 40,000 radiologists, radiation oncologists, medical physicists, and nuclear medicine physicians, appreciates the opportunity to submit comment on NQF #3633e, #3662e and #3663e: Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Level, Clinician Group Level and Facility level, respectively). **The ACR does not support the endorsement of NQF #3633e, #3662e, and #3663e.General Comments** Protocol selection appropriate for a clinical indication is an important component of radiation dose management along with radiation dose optimization. Each component needs to be addressed as a separate quality action. The specific aspect(s) of performance to be improved is not intuitive due to the multiple components to the measures (size-adjusted dose, image quality, clinical indication). It is premature to measure performance on excessive radiation dose based on thresholds by clinical indication for an exam until the level of standardization and availability of national benchmarks is further along as discussed below. It is true that the most accurate way to address appropriate and safe use of multi-phase studies is to measure both the clinical indication of an exam and the radiation dose output (dose indices per exam) and look at the two separately or distinctly together. **However, these measures conflate the appropriateness of protocol for the clinical indication and radiation dose optimization, disregarding applicability, from which a facility may not be able to determine if its performance could be improved by adjusting protocols or by focusing on appropriateness of the ordered exam. Therefore, improvement may be limited.[1]** Dose optimization results in a quality action for facilities to adjust their protocols and is a responsibility of the team as a whole – physicists, technologists, and physicians who oversee the team at the facility. Protocol selection addresses the appropriateness of the exam for the clinical indication and other factors such as patient time on the scanner and optimal radiation dose. There are challenges with the implementation of an indications-based measure. Indications for exams do not have standardized language that could be used to track them. Most health and IT systems capture ICD-10 coding for reimbursement, but typically not enough standardized information to characterize the patient's condition. As a result, the clinical reason for performing an imaging exam is often extremely limited in the exam order. Electronic Health Records (EHRs) are notoriously incomplete with this type of information and interoperability issues exist with other software systems that might contain such information. **A *validated* method for determining classification of studies using high-dose versus routine protocols appropriate to the indication must be incorporated into such a measure; these three measures include specifications which have not been validated.** Please refer to the validity section below for more details. **NQF #3633e, #3662e, and #3663e deviate from international**

standards, like diagnostic reference levels, and lack peer-reviewed, broadly accepted consensus on global noise. For these measures, global noise is defined solely by the measure developer. Endorsing this method may encourage facilities to accept a narrow view of image quality. The ACR requests the developer further clarify the global noise table used in calculating the numerator. The benchmark source is not transparent, and its applicability is unclear. For example, Table sp-1, Size-adjusted radiation dose and global noise thresholds by CT category, has the same global noise threshold for several CT categories, such as head low dose, head routine dose, and head high dose. Is it intentional that the same global noise threshold should be applied to both low and high dose head CTs? If the image noise thresholds are the same, the size-adjusted radiation dose thresholds should be the same, unless the scan length is remarkably different between the 3 CT categories. Additionally, current CT scanners display dose values based on either a 16 cm or 32 cm phantom for a neck scan, which must be carefully accounted for in measure performance calculations. **There is little to no acknowledgement of limitations.** These measures have multiple limitations, including the lack of widespread acceptance and implementation, and the issues with the method of measuring global noise. The developer states their company can provide the service of quantifying the measure at a cost; this should also be included as a potential limitation. The measure developer does provide specifications for other entities to implement the measure, but the burden of implementation may be significant. Finally, the author cites publications from their group to justify the benchmarks, but they have not been vetted through a broader consensus process. **The ACR strongly encourages the Patient Safety Standing Committee to re-vote on the scientific acceptability of these measures based on the following concerns.**

Validity/Feasibility These eCQMs require multiple variables that may be captured in software systems external to electronic health records (EHRs), such as dictation systems housing radiology reports or DICOM standard-based systems, such as CT device software. Data element validity testing should demonstrate that the testing sites were able to integrate and validate the variables used to construct the data elements used by the eCQM in addition to the usual validation of the eCQM's electronic output against the medical record review. **We are uncertain that this validation has been completed. Therefore, this submission does not demonstrate the measure can be reproduced in a reliable and valid manner by practices or facilities across multiple settings.** For example, for CT category (or other elements deriving/collecting data using custom natural language processing (NLP) tools), the developer used NLP for obtaining data such as reason for study or protocol name used in the calculation of this variable. The submission does not provide information on the NLP results' reliability and validity. Because **this comparison of the NLP-derived data against a medical record review was only completed in a sample from one site (UCSF Health System), there is uncertainty whether the results are generalizable across EHRs or other databases**. These measures rely on custom made NLP trained and validated on a small group of pilot sites; it is not clear whether this type of NLP would work outside these sites nor how sites would get access to use this custom NLP tool. Testing information does not demonstrate adequate validation of this critical data element. Additionally, **sufficient evidence should demonstrate that the definitions/variables used are valid and do not rely on one study or use in a single system, such as what is provided to support the thresholds of "out of range" performance values.** While the process to determine these thresholds is detailed, we do not believe that a Technical Expert Panel (TEP) conclusion in the absence of independent data validation is sufficient.

M**ultiple unstructured variables are required to construct the data elements for the numerator, denominator, and exclusions. Assessments of the feasibility of the integration of these unstructured data into the measure calculations would be useful to ensure that the underlying data can, in fact, be integrated if practices and facilities that choose not to use the edge device.** For example, the level of effort required to integrate the Binning algorithm for the CT categories and ensure that the results are reproducible and valid remains unclear. The ACR is concerned with

the selection bias for the accountable entity-level (measure score) validity. **Assessing measure score face validity through the TEP that created these measures lessens the extent of credibility for these results**. Although the TEP is knowledgeable and represents a variety of stakeholders, there is a vested interest in ensuring these measures are available for use. **Most importantly, as one of the TEP members noted in the survey, the performance score from these measures does not clearly indicate what corrective action needs to be taken by the clinician, clinician group, and/or the facility to improve performance**. **Usability** While implementing these measures as specified may not impose a substantial burden on clinicians, **it may necessitate substantial organizational effort to access and process the data elements required to calculate the measure score**. The measure steward states that their software is available on a non-commercial basis to calculate this measure, and that other vendors may also develop their own software to implement the measure specifications using the information included in this submission. Will the measure steward review other vendors' software to ensure comparable calculation methods? Measure stewards frequently make specifications available "as is" without warranty, leaving it to the implementer to appropriately update any software or tools as measure specifications are changed. But the complexity of these measure specifications may warrant greater oversight. External vendor software will need to be maintained and updated to ensure the software's accuracy and reflect any changes in specifications and coding. **For all the reasons stated above, the ACR does not support the endorsement of these three measures.** We thank the NQF staff for their transparent endorsement process. **Reference: 1. '**Mahesh M. Benchmarking CT Radiation Doses Based on Clinical Indications: Is Subjective Image Quality Enough?Radiology. 2021 Nov 9:212624. doi: 10.1148/radiol.2021212624. Online ahead of print. PMID: 34751622

Comment 2 by: Angela Keyser, American Association of Physicists in Medicine

What is AAPM:

The American Association of Physicists in Medicine (AAPM) is the primary scientific and professional organization of physics in radiology and radiation oncology in the United States. The mission of AAPM is advancing medicine through excellence in the science, education and professional practice of medical physics; a broad-based scientific and professional discipline which encompasses physical principles with applications in biology and medicine. With 9717 members in 94 countries, AAPM supports the Medical Physics community with a focus on advancing patient care through education, improving safety and efficacy of radiation oncology and medical imaging procedures through research, education and the maintenance of professional standards. AAPM has a staff of 33 and an annual budget of $10.7M, and is located at 1631 Prince Street, Alexandria, VA 22314.

AAPM comments on the proposed measures:

AAPM does not support the endorsement of NQF #3633e, #3662e, and #3663e.
This application proposes electronic clinical quality measures (eCQM) that monitor CT performance to discourage unnecessarily high radiation dose while maintaining adequate image quality. The proposed metrics require CT Category (i.e., the CT exam type), the size adjusted radiation dose [the patient's dose length product (DLP) adjusted by patient size], and the global noise (associated with the variance of the voxel values in CT images). The two reported measures are the percentage of eligible CT cases in a particular category deemed to be "out-of-range" compared to defined thresholds with respect to the size-adjusted radiation dose or the global noise in a set time period. While efforts to enhance consistency of CT practice are noble and include initiatives by AAPM and others worldwide, the proposal has significant limitations that impact its scientific and practical value and overall likelihood of clinical acceptance. These limitations include improper

representation of image quality, improper estimation of radiation risk, and substantial oversimplified representation of implementation in practice, including not addressing the challenges of implementation. The authors indicate that their company (Alara Imaging, Inc.) can provide the service of quantifying the measures at a cost. A steward of measures requires an extensive track record for scientific and technical expertise and policy making that represents a broad consensus of the community. These important elements should be carefully reviewed within this application. One cited reference supports the proposed measure, however, this cited article has an accompanied editorial that highlights the limitations of the proposed approach [Mahesh M. Benchmarking CT Radiation Doses Based on Clinical Indications: Is Subjective Image Quality Enough? Radiology. 2021 Nov 9:212624. doi: 10.1148/radiol.2021212624. Online ahead of print. PMID: 34751622]. The editorial and stated limitations are not addressed in the proposal.

The AAPM agrees that effort needs to be continually placed on ensuring diagnostic quality CT imaging, optimizing CT dose, and achieving consistency across facilities, considering differing technologies and practices. The non-profit entities of the AAPM, the American College of Radiology (ACR), and Image Wisely and Image Gently Alliances have spent decades towards this goal and continue to do so through many initiatives. Among them, the non-profit ACR CT Dose Index Registry (DIR; https://www.acr.org/Practice-Management-Quality-Informatics/Registries/Dose-Index-Registry, established in 2011) has the significant stature of implementing a dose registry that enables facilities to compare dose indices nationally, to ensure the highest quality imaging with lowest possible dose. The ACR CT DIR implementation incorporates the expert, consensus opinions of the medical imaging community. ACR dose optimization measure recently endorsed by NQF provides a further valuable measure to manage imaging radiation dose (https://www.qualityforum.org/QPS/3621). The imaging community's valuable clinical benchmarks greatly benefit from consensus decisions based on sound scientific and technical review and discourse. The proposal herein should be carefully reviewed for any additional contributions or advantages it would provide to our existing robust consensus measures and resources, such as available with the ACR.

After a detailed review of the measures by multiple expert members of the AAPM, we have concluded that the **AAPM does not support the endorsement of NQF #3633e, #3662e, and #3663e.** This position stems from eight major concerns about the proposed measures:

1) Unscientific characterization of CT scan risk: The proposal is based on estimation approaches that are not reflective of the consensus of the scientific community and do not acknowledge the uncertainties of the estimates. A NQF measure focused on radiation risk should uphold scientific objectivity, integrity, and responsibility not evident in the presentation and assessment of radiation risk in this proposal.

2) Inactionability of the measures to enable targeted change to improve practice: It is not evident how the proposed measures can be practically used to improve imaging practice and exactly how a facility can do to achieve compliance, given the wide varieties of factors and technologies involved.

3) Inadequate addressing of the complexity of CT categorization: The proposal does not address the magnitude of this challenge nor has suggested means to overcome it given that current standards are even lacking in uniform characterization of protocols. Inaccurate classification of data can lead to significant and misleading errors.

4) Inadequate assessment of noise: Noise in a CT image can be influenced by a variety of factors including justified differences in CT technologies including new reconstruction methods that dramatically alter noise. Further, noise does not have a singular value in a CT exam. A "global noise" ignores this diversity and can misrepresent the quality of an exam.

5) Inadequate assessment of image quality: Image quality is affected by a myriad of factors including resolution and contrast, as well as the intended purpose of the exam. A singular representation of image quality via global noise overly simplifies this space and can lead to gross misrepresentation of image quality and thus mis-service to patient care.

6) Flawed assumption on dose reduction vs dose optimization: The application focuses primarily on radiation dose reduction as oppose to right-sizing the dose for the best care of the patient. Individualization and optimization of care and safety should be the goal not minimization. This approach can lead to some patients getting under exposed, leading to missed diagnosis, while others may be over-dosed for their exact need and condition.

7) Inadequate accuracy in patient size estimation: Assessing a patient size is not a trivial task, stemming from significant variability in the differences in the habitus of different patients, coupled with the existential challenge that there is no single metric capturing the size of a patient of varying diameter at different cross-sectional locations. Algorithms are continuously evolving and no evidence is provided that the company can do this task with sufficient accuracy.

8) Limited expertise and track record of the company: The company is a new (2020) company with no experience of having previously performed a project of such wide scope, scientifically or technically. There is no scientific track record on CT technology, size estimation, or image quality assessment for the company to be considered a steward of measures on which there is a lack of expertise, publication, and scientific history.

These concerns are detailed specially in our complete review submitted via email to patientsafety@qualityforum.org, along with selected specific observations on the proposal on January 19, 2022.

The AAPM recognizes that this topic is complex, including scientific, technical and clinical components. We welcome the opportunity for greater in-depth discussion on meaningful measures of quality imaging practice.

Respectfully submitted,
American Association of Physicists in Medicine (AAPM)
January 19, 2022

Comment 3 by: Bradley Delman, Mount Sinai Health System

I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. As an implementation testing partner, I coordinated Mount Sinai Health System's inclusion in the test. To summarize, after installing the data collection software, we routed CT imaging data from PACS and sent order and billing data from various electronic systems to the software. We also worked with UCSF and our CT vendors to ensure the Radiation Dose Structured Report (RDSR) was being saved for each exam sent to PACS. As we discussed in our interview with UCSF, this work fell on the PACS team and IT colleagues, without requiring effort from clinicians above my initial planning and coordination. Besides some technical challenges, which were all resolved, we faced few barriers to successful implementation and had very little missing data. In total we submitted 11,588 scans, representing just over 3 weeks of CT data from our health system. Based on our experience, the participation in the proposed quality measure is feasible. However, I suspect that spirited engagement from PACS, RIS and/or EHR vendors would greatly enhance participation and timely provision of data. We have also been satisfied with the feedback we've received from Alara Imaging on our measure performance, which brought to our attention areas of high radiation dose. This feedback has

identified individual exams as well as imaging protocols that contribute high radiation dose. Although we have been a dose-conscious department, the feedback highlighted areas of variability in both routine and size-adjusted datasets. Furthermore, we learned which protocols and classes of studies fell within and beyond expected range for dose, and how dose can vary between scanners for protocols with the same name. We also learned that some types of studies may need to be renamed or reclassified for appropriate grouping of results. A quality measure the quantifies dose while ensuring preservation of imaging quality can help mitigate the use of excessive radiation doses used in CT. I support the work of the measure developers to improve patient safety and CT quality.

Comment 4 by: Daniel Hirsch

I write in support of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. They are important proposals that would markedly reduce unnecessary radiation exposures in medicine, and the cancers induced therefrom, while providing the same yield of diagnostic information. Many, many lives could thus be saved were the proposals adopted. I have spent much of my professional career attempting to reduce the risks to public health from ionizing radiation associated with nuclear waste, reactor accidents, nuclear weapons tests, uranium mining and milling, and radioactively contaminated sites involved in the production of nuclear weapons and other nuclear activities. It is with some alarm that I have viewed in recent years the extraordinary increase in public exposures to ionizing radiation associated with the remarkable escalation of exposures in medicine, largely due to ever-more frequent CT scans, resulting in doses from medical procedures now dwarfing exposures from the nuclear activities that have so long concerned me. The proposals made by UCSF would help reign in unnecessarily high radiation doses from these medical procedures while still producing the diagnostic information needed by physicians for their patients. The important revelation in the studies cited in the proposals is that the doses currently received by patients in these procedures are frequently very much higher—often ten times higher—than necessary. One can get the same medical benefit from the procedure at one tenth the cancer risk. The proposals indicate that many thousands of unnecessary radiation-induced cancers could be avoided were CT scans kept to the minimum level necessary to produce the required image. This seems quite correct. The National Academies of Sciences, Engineering and Medicine has produced over the years the primary studies on the matter of ionizing radiation and cancer induction. The most recent Biological Effects of Ionizing Radiation study (BEIR VII) estimates a risk of 1.17 cancers per 1000 person-rems of exposure, and concludes, as have all the BEIR studies, that there is no threshold below which there is no risk. All radiation protection agencies (e.g., US EPA) have adopted the BEIR conclusions. Currently, exposures to medical radiation are estimated as averaging about 350 millirem/year per person. Given that degree of exposure, and the current U.S. population, medical radiation would be estimated to produce many millions of cancers over the population's lifetime. Reducing unnecessarily high exposures while still producing the necessary diagnostic image could thus prevent a very large number of cancers and deaths, while, not incidentally, also reducing Medicare expenditures for their treatment. I strongly urge adoption of quality measures that assure CT exposures use the lowest reasonable doses necessary for the procedures. Daniel Hirsch retired Director of the Program on Environmental and Nuclear Policy at University of California at Santa Cruz

Comment 5 by: Dawn Ritzwoller

I am a college student and Environmental Biology (E-bio) major, and I am pediatric cancer survivor. I am writing today in support of this radiation dose quality measure. Beginning ten years ago, and

both during and after I finished treatment, I received multiple CTs (to multiple parts of my body) as part of my diagnostic and follow-up care. Not once during this period, did any of my doctors or other, discuss with me the downstream risk of all of the radiation exposure I experienced. It was only years after my treatment ended, and now via classes I have take for my E-bio major, that I am beginning to understand the risk associated with radiation exposure. What is also now clear to me is the importance that providers use the most appropriate (low) dose for the specific diagnostic or follow-up exam. I know that image quality is important for diagnosis, but patients (like me) need the confidence that their doctors and hospitals are using the best and lowest dose possible for the exam that they order. Thank you!

Comment 6 by: Debra Ritzwoller

I am writing in support of this important measure. I am a cancer health services researcher *and* a mother of a pediatric cancer survivor. It is well documented in the literature that there has been a significant secular increase in CT use within and across most patient populations. While CT use, and therefore radiation exposure has increased over time, I know that personally and professionally that excessive radiation dose remains a significant quality issue, and it is one that is often not adequately addressed by researchers and healthcare providers/delivery systems. This quality metric is necessary now, in order to provide the incentives and the resources needed to generate the metrics and the benchmarks that may actually influence practice that may in turn translate into a meaningful reductions in the radiation dose that patients are exposed to. This metric is designed to address the clinical indication associated with the respective exam, rather than just the type of advanced imaging that is performed. The measure is also constructed to ensure that the dose benchmarking does not adversely impact the quality of the metric. Given the noted harms of CT based radiation exposure (e.g USPSTF Lung Cancer Screening "B" recommendation), this measure is addresses a timely and needed quality metric.

Comment 7 by: Ehsan Samei, Duke University, Margolis Center for Health Policy Center

Duke University, Ravin Advanced Imaging Laboratories (Ravin Labs) and Clinical Imaging Physics Group (CIPG), Durham, NC 27710 The Ravin Labs is a 50-member leading translation imaging research laboratory in the country with over 30 years of history. The lab conducts rigorous NIH-funded research with an additional mandate to practice its science through CIPG, an imaging physics group of 15 experts dedicated to quality and safety in the practice of radiology. The group, highly integrated into the clinical domain, has devised and put to practice imaging dose and image quality monitoring systems at the level of individual patients within the Duke University Health System with additional pilot installations at MD Anderson Cancer Center and Stanford University. The group has published extensively on its technology and findings (upward of 500 papers), with over 30 referred publications on dose and quality monitoring alone. The effort has led to significant reduction of patient radiation dose at our facilities and right-sizing it per individual needs of patients. **We do not support the proposed measures.** The rationale is detailed below. **Overall:** While we applaud the effort to introduce new quality measures in the practice of medical imaging, the proposed electronic clinical quality measures (eCQM) are misleading and overly simplistic leading to significant unintended consequences. The limitations stem from the fact that the proposed risk measures are based on CT scanner output and not the actual dose burden to individual patients at the organ level, the quality measure is based on noise alone ignoring the multi-faceted reality of diagnostic quality, and lack of methods that standardize protocols across vast diversity of examinations. There is significant ambiguity in the exact method used for noise and size estimation with no track record or peer review of otherwise black-box methods. This

approach will likely produce measures that can be orders of magnitude off from their actual values, and therefore lack clinical relevance and fidelity. Measures can lead to misleading and erroneous conclusions while also potentially jeopardizing the use and development of better approaches, as inaccurate low-bar measures can prevent accurate ones in the future. But most importantly, the measure can lead to unintended consequences and even harm the patient. For example, an imaging team can take an action that is not in the best interest of a patient, like applying too little dose for some patients such that disease would be missed, a "wasted dose" with no medical benefit and health and cost consequence of a miss. Conversely others might get more radiation than needed as the measures do not account for individual patient needs and tasks. Improving consistency in imaging practice is a laudable goal that needs a proper solution anchored to scientific understanding of radiation risk, image quality need of patients, diversity of practices, and the CT technology. The proposal is lacking on all these four fronts. A solution to inconsistency in images can only be brought forth through a broad consensus of the scientific and practicing communities (including ACR, AAPM, Image Gently, and Image Wisely), CT manufacturers (represented by MITA), standard methods of data categorizations and measures (supported by the medical community), and evidence-based radiation risk and image quality measures at the level of indication and organ where they are actually relevant to the individual patient. A for-profit company with no track record or transparency of its methods cannot be considered a steward of such a space. Below we further detail 12 concerns regarding the proposed measures:

1. **Inadequate attention to image quality:** The measures are heavily dose related, emphasizing this over measures of quality. Dose and minimizing it is important but equally important is image quality as an inadequate image quality would be a dis-service to the patient regardless of the dose. This is explicitly stated in the International Commission of Radiological Protection (ICRP) in Publication n. 135.

2. **Inaccurate assessment of radiation risk:** The measure of size-adjusted radiation risk, adjusting the CT scanner outputs with 'patient size' to perform risk estimation is not a standard method nor endorsed by any scientific or professional body. The method is in fact explicitly **discouraged** by the AAPM Task Group 204. Patient risk can only be assessed with the knowledge of organ doses that is not even mentioned in the application let alone pursued. The proposed method CANNOT be used as surrogate for future cancer risk.

3. **Incomplete/Inaccurate representation of image quality:** The measures include image noise. Yet, noise is just one component of image quality. For example, the noise of an image can be fine but image quality totally inadequate. And conversely noise can be too high but image quality totally adequate. To assess image quality properly, one should include the actual task at hand (eg, detecting a pancreatic cancer vs bowel obstruction vs kidney stone) as well as other equally important facets of quality, like noise texture, resolution, and contrast. These factors have not been even mentioned let alone tackled in this application. Focusing on noise as a singular metric of quality can lead to major mis-representation of the needs of a quality and safe imaging practice.

4. **Neglecting the impact of image rendition:** Critical and relevant to clinical practice, the measure of noise proposed does not take into consideration how differing reconstruction algorithms and parameters affect noise (up to 200%). Without considering this influence, a measure of noise as proposed is irrelevant and misleading.

5. **Subjectivity:** The measures are anchored to subjective perception by radiologists as how they "like" the images. There is in fact no evidence provided that the measures can lead to an improvement in diagnostic accuracy. In fact, it might lead to a degradation.

6. **Lack of integrating dose and quality:** There is no indication as to how image quality is linked to radiation dose and at what level; or instance, how they propose to manage multiple reconstructions of the same exposure event.

7. **Not addressing the multiplicity of exam components**: A CT exam often includes multiple phases (series) each of which has a noise and radiation dose of its own. Averaging noise across series is meaningless. The measures do not recognize or account for this multiplicity and diversity.

8. **Under-recognizing the diversity of exams:** The measures do not address the notable diversity of exam nomenclature across institutions and practices. This is a significant component of any dose or quality monitoring system. Without a standard for CT protocols, which cannot be devised by a for-profit company without consensus of manufacturers and users, the data can be mislabeled and mishandled leading to major errors in the results and subsequent negative effect on mis-dosing and mis-diagnosing patients.

9. **Inaccurate assessment of patient size:** The measure of size proposed is calibrated to earlier work and publication from our group at Duke University for academic purposes. That early method they have embraced has had major errors (upward of 300% is certain applications) that have been corrected in subsequent versions that have not been shared. Without essential newer refinements to assure fidelity, the company cannot be a responsive steward of the measure that it has had no expertise to advance or maintain.

10. **Inaccurate assessment of noise:** The measure of noise proposed references earlier work and publication from our group at Duke University. That early method exhibited errors, corrected in subsequent versions that have not been shared. Without essential newer refinements, the company cannot be a responsive steward of the measure that it has had no expertise to advance or maintain.

11. **Lack of guidance toward compliance:** To us it is difficult to defend (1) measuring imaging practices based on ambiguous and questionably-relevant metrics promoted to represent the actual safety or quality of CT practice, and (2) not offering any guidance as to how a practitioner responsible for "outlier" examinations can bring their practice to the proposed definition of compliance. Together, these can easily create signification confusion and potential disruption in the imaging practice

12. **Lack of support from manufacturers:** Having worked in dose and image quality monitoring for over a decade, academic centers of excellence, including ourselves, have a close connection with major CT manufacturers including MITA, Medical Imaging Technology Alliance, which comprises all CT manufactures. Our discussions regarding this measure lead us to believe that there will be little support from scanner manufacturers for a non-transparent and unpredictable product that lacks maturity from a private for-profit entity. There are substantial differences in image processing, detector efficiency, and such across scanners that will have significant bearing on the CT image. The proposed measure does not account for such important nuances, leading to erroneous results.

Comment 8 by: J. Leonard Lichtenfeld

I am pleased to provide this comment in support of NQF quality measures 3633e, 3662e and 3663e. These comments reflect my personal opinion and not any other organization with which I may be affiliated. CT scans have assumed a primary role in the evaluation and diagnosis of many medical conditions, and are very commonly performed procedures. Less appreciated by the public

and many professionals (including non-radiology physicians) is the variation in image quality and dose that has been recognized for many years by researchers who have evaluated these factors. As such, there can be substantial variation in CT scan dose and quality, even within the same institution. As a patient, this consideration has figured prominently in my own decisions as to whether or not to proceed with serial CT scans for follow-up of medical conditions. These measures have been carefully crafted to create an effective and validated method to monitor CT image and quality based on indications for the studies and in consideration of individual patient-related variables. As such, they provide a useful and meaningful way to offer our patients and the public the assurance that the scans they are receiving meet reasonable safety and professional standards--which is not routinely available otherwise. These quality measures will meaningfully improve the ability of physicians and heatlh systems alike to monitor the equipment utilized for these studies in a manner that minimizes interference with the tyipical workflow of a radiology center (or other center) where such studies are performed and will provide a significant and substantial increase in the quality of scans while reducing dose variability that can occur because of machine settings/performance or patient characteristics. Cumulative radiation dose should decline as a result of implementing these measures. At the very least, there will be assurance that the right dose is used for the right scan in the right patient. As a physician and patient advocate for many years, I offer my support for these measures for the reasons stated. And as someone who served as an advisor for this measure, I will add that I was impressed by the exceptional commitment of the developers and their colleagues to provide a meaningful, validated and effective quality measure as they created new processes to measure CT dose and quality, always with an eye towards making this measure acceptable to the professional and consumer communities. (Disclosures: As noted, I was an advisor during the development of this measure and received compensation for those services. I have also served on the NQF Cancer Committee without compensation. I have no other relevant conflicts.)

Comment 9 by: James Anthony Seibert, University of California, Davis Medical Center

January 27, 2022 To: National Quality Forum Dear NQF Standing Committee, I am writing to lend support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco, where I have served on the Technical Expert Panel and have been a long-time collaborator for similar projects between UCSF and UC Davis. I led the implementation of measure testing at my institution, University of California Davis Health, which required local installation of the software, configuring connections to the PACS, extracting CPT and ICD-10 data from the EHR, and supervising the aggregation and transfer of all this data to the UCSF software. Most of this work was completed by our PACS administrator and did not impact the work of our clinicians at any time. One challenge we encountered was that transfer of data from PACS to the software was slow; we believe this was due to capacity limitations of our PACS relative to the query-retrieve process. Nevertheless, we set up auto-transfers of the data over nights and weekends so as not to impact the operation of our PACS during our busiest clinical hours. Besides this issue, the testing was completed successfully with minimal missing data. Based on our experience, the proposed quality measure is highly feasible, and will, in my opinion, be able to appropriately identify CT exams that are significantly above diagnostic reference level (DRL) doses(*), as well as inadequate CT exams with insufficient dose, for specific diagnosis indications versus radiation dose versus image quality. There are certainly many parameters and issues that can potentially confound such CT quality measures, particularly with the assessment of corresponding image quality, but significant advances in developing robust algorithms to recognize such confounding factors have largely mitigated such concerns. I believe this quality measure can significantly reduce the use of excessive high radiation dose as well as inadequate, sub-optimal low

dose used for clinical CT studies, by identifying outliers and thereby increasing the awareness and importance of CT protocol optimization. I support the work to improve patient safety and CT quality as described in these measures. Sincerely, J. Anthony Seibert, PhD, FAAPM, FACR, FSIIM, FIOMP Professor Emeritus, Department of Radiology UC Davis Health (*) Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology 284(1), 120-133, 2017. Disclosure: I have served on the Technical Expert Panel for this effort and have received some minor compensation for participation (honoraria) but have no other relevant conflicts. The opinions expressed here are my own.

Comment 10 by: Kenneth Wang

I am pleased to provide my support for the proposed CT quality measures 3633e, 3662e and 3663e developed by the University of California, San Francisco. I have been a practicing radiologist in the Veterans Affairs (VA) system for more than ten years, during which time I have led efforts in CT dose optimization across the VA Maryland Health Care System. I also serve in a number of volunteer roles within the Radiological Society of North America (RSNA) and the American College of Radiology (ACR), leading efforts in informatics, standards, interoperability and registries. However, this letter reflects my personal opinion, and not necessarily those of any organization with which I am affiliated. I have also served as a member of the Technical Expert Panel (TEP) advising on the formulation of these proposed quality measures, since the inception of this project.

The impetus for this work rests on fundamental principles which are widely accepted. Namely, that CT constitutes an important source of radiation dose to patients, and that CT imaging presents an opportunity for dose reduction, but that it is of paramount importance to maintain the diagnostic quality of the imaging obtained. The proposed measures have been developed using a scientific approach incorporating extensive testing and validation, as well as expert consensus, while maintaining a focus on practicality. This has been all the more impressive given the complex nature of the technical factors involved, such as CT exam types, size-adjusted dose, and diagnostic image quality. By leveraging extensive data, including but not limited to data in the UCSF International CT Dose Registry, data obtained from practicing radiologists on image quality, and feedback from testing facilities, the measures strike a practical balance intended to identify opportunities for CT dose reduction while maintaining a floor for diagnostic quality (which was rarely violated in measure testing).

As such, these measures represent an important step beyond simple dose reduction. I also believe that these measures will provide actionable feedback, especially given the many different techniques now available on modern CT scanners for dose adjustment.

As a radiologist, I know there will never be universal agreement on subjective assessments such as image quality. However, the proposed measures take a balanced approach, informed by extensive testing and validation, which serves a very practical and important quality objective. For these reasons, I support the adoption of these measures.

Comment 11 by: Krishna Nallamshetty, Radiology Partners

I would like to submit a comment in support of this measure. I am a practicing radiologist for the past 15 years and serve as the Associate Chief Medical Officer of Radiology Partners, the largest medical imaging practice in the United States. I am the chair of our national Patient Safety Committee. We have seen tremendous growth in medical imaging that requires radiation, specifically computed tomography (CT). The public awareness of the potential long-term effects of

ionizing radiation has become mainstream and as a result, a primary objective of the American College of Radiology and other governing bodies. The objective focuses on reducing radiation exposure as much as possible without compromising the diagnostic information that is obtained We have recognized that there is large variability in how CT scans are acquired all over the country. Techniques and radiation exposure is extremely varied but yet appropriate clinical diagnosis are made. This measure evaluates radiation dose for every patient who undergoes CT *based on the clinical indication for imaging* rather than solely on the type of examination that is performed. It ensures patients receive the most appropriate CT acquisition protocol and level of radiation for their individual condition. The measure also assesses image noise, safeguarding image quality against potential effects of dose reduction, and is the first quality measure to do so.
The measure would have a large, positive impact on patients and protect them from unnecessary over-exposure of radiation without compromising the diagnostic value of medical imaging. It would be the first time a measure addresses both radiation and image quality.

Comment 12 by: Maribel Escobar

Submitting on behalf of ARA's CMO, Dr. John Kish: January 25, 2022 Dear NQF Standing Committee, I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. As an implementation testing partner, my institution, ARA Diagnostic Imaging, was required to install the data collection software, route CT data from PACS and order and billing data from various electronic systems to the software, and oversee the migration of data. We also worked with UCSF and our CT vendors to ensure the Radiation Dose Structured Report (RDSR) was being saved from each exam in the PACS. As we discussed in an interview with UCSF, this work fell on the PACS team and IT colleagues and did not require effort from clinicians. Besides some technical hiccups, which were all resolved, we faced few barriers to successful implementation and had very little missing data. Based on our experience, the proposed quality measure is highly feasible. We have also been satisfied with the feedback we have received from Alara Imaging on our measure performance, which brought to our attention some areas of opportunity to decrease radiation dose. The feedback provided by Alara Imaging has taken the burden of researching problem areas away from my institution, by identifying specific exams, imaging protocols and even specific CT units that contribute to high radiation dose and need improvement. We have plans to address each accordingly. Given our positive experience, my organization is moving towards a commercial relationship with Alara to continue to submit data, receive feedback, and strive to optimize our CT doses. I earnestly believe this quality measure can help mitigate the use of excessive radiation doses used in CT. I support these measure developments in order to improve patient safety and CT quality. Sincerely, John Kish, MD Chief Medical Officer

Comment 13 by: Mary White

I am writing in support of this CT radiation dose safety measure. As a cancer epidemiologist, I recognize that excessive exposure to medical radiation increases cancer risk. And I understand that this measure will be valuable for protecting patients from unnecessarily high levels of radiation from CT imaging. The measure is designed to evaluate radiation dose for every patient based on the clinical indication for imaging. The measure also assesses image noise, ensuring adequate image quality despite the reduction in radiation dose. This measure fills an important quality void and has the potential to substantially reduce the contribution of CT scans to the incidence of cancer in the population.

Comment 14 by: Matthew Nielsen

I am writing in support of this important measure. The utilization of CT imaging int the United States has dramatically increased over recent decades, providing numerous benefits to patients and clinicians in the management of countless medical conditions. There has also been increasing recognition of the potential for unintended harms due to potentially avoidable variation in radiation in radiation dose for many patients. Evidence from research and quality improvement efforts demonstrates the potential to mitigate these harms witha feedback loop and benchmarking to radiologists and staff. This measure provides needed resources to disseminate these early successes, preserving the benefit of advanced imaging with CT while providing a means for healthcare facilities and clinicians to improve the safety of the studies they provide patients. The design of this measure importantly takes into account the indication for the study as the framework for dose benchmarking, with balancing measures of image quality to assure that efforts to reduce dose do not come at the expense of diagnostic quality. Given the increased recognition from patients and providers of the potential harms of imaging-associated radiation, this measure fills a timely and important gap in the current measurement portfolio.

Comment 15 by: Pavlina Pike, Huntsville Hospital

I am writing to lend my support for the endorsement of CT quality measures 3633e, 3662e, and 3663e developed by the University of California, San Francisco. I am a Medical Physicist and Radiation Safety Officer at Huntsville Hospital and led the testing of UCSF's quality measure at my health system, which involved installing the data collection software, routing CT data from PACS and order and billing data from various electronic systems to the software, and overseeing the migration of data. We came onboard late in the testing period, leaving a tight window of time to collect the data prior to UCSF's submission deadlines. I am proud of my PACS and IT colleagues for pulling together so efficiently and completing the work rapidly with very little missing data. The work in no way impacted our physicians or clinical workflows. We faced few barriers to implementation, and based on our experience, the proposed quality measure is highly feasible.

We have also been satisfied with the feedback we've received from Alara Imaging on our measure performance, which brought to our attention areas of high radiation dose. Our exams were compared to thresholds established based on input from 125 radiologists and 50,000 CT examinations from other facilities. The analysis includes comparisons of the performance of different model CT scanners, exams, protocols, patient size, facility, etc. The feedback from the Alara software is helpful and actionable as we are able to identify what changes will have the greatest impact on patient dose and make the appropriate changes. In addition it provides suggestions for billing inconsistencies which was very helpful to our administration.

I earnestly believe this quality measure can help mitigate the use of excessive radiation doses used in CT. I support the work of the measure developers to improve patient safety and CT quality.

Comment 16 by: Robert Gould, University of California, San Francisco Medical Center

I am writing as a physician who has worked for decades as a leader in Physicians for Social Responsibility, as well as the International Physicians for the Prevention of Nuclear War toward eliminating nuclear weapons, cognizant of the public health dangers of radiation initially derived from studies of victims of the twin atomic bombings in Japan. Informed by the central tenet of physician practice to "at first do no harm," I strongly support CT quality measures 3633e, 3662e,

and 3663e developed by the University of California, San Francisco. While my long experience as a practicing pathologist has made me understand at a profound level how diagnostic radiation is a critical tool in medical practice, it has also underscored to me the often-overlooked risks of carcinogenesis that must always be balanced against the benefits of various radiological procedures. Over time, research has documented that many radiological procedures are medically unnecessary when information that is desired can be obtained by other means than exposing a patient to ionizing radiation; it is also unwarranted when employed as a "hedge" against possibility of malpractice litigation. In addition, when radiological imaging is indeed required and justifiable, it is not uncommon, where standards are not uniformly applied in practice, for radiation exposures to exceed what would be required for achieving images satisfactory for diagnostic purposes. As such, the lack of attention to standardizing, and minimizing exposures inevitably results in the induction of significant numbers of unnecessary cancers that would not occur if lower doses were employed to achieve adequate imaging. I believe that CT quality measures 3633e, 3662e, and 3663e would be important steps to assuring that physicians can obtain the information necessary from diagnostic imaging while minimizing the number of unnecessary cancers induced by the procedures.

Comment 17 by: Suz Schrandt

As a patient advocate with significant experience navigating the healthcare system--including repeated exposures to a variety of diagnostic imaging studies--I submit these comments in endorsement of this measure. The measure takes into account different contexts and parameters for a given patient and his or her unique benefit/risk profile. At a more foundational level, the measure calls into focus the significant variation in practices in CT imaging that can expose patients to unnecessary and/or unsafe levels of radiation, a risk many patients are not even aware of. The wide-spread use of this measure could standardize imaging practices and should the measure be adopted, I strongly encourage a robust dissemination plan to inform patients and families of its existence. Our ability to access safe and effective care should not be left to change; measures such as this help to close key gaps in our system.

Comment 18 by: Melissa Danforth, The Leapfrog Group

Founded in 2000 by large employers and other purchasers, The Leapfrog Group is a national nonprofit organization driving a movement for giant leaps forward in the quality and safety of American health care. The flagship Leapfrog Hospital Survey collects and transparently reports hospital performance, empowering purchasers to find the highest-value care and giving consumers the lifesaving information they need to make informed decisions. For the past several year's Leapfrog has been collecting and publicly reporting hospital performance on an NQF-endorsed Pediatric CT Radiation Dose (NQF 2820) measure. The new Excessive Radiation Dose or Inadequate Image Quality for Diagnostic Computed Tomography (CT) in Adults (Clinician Group Level) fills a critical gap in evaluating radiation dose for adult patients who undergo CT. Additionally, because the measure is based on the clinical indication for imaging – rather than on the type of examination the radiologist chose to perform – it can help ensure patients receive the right type of CT and amount of radiation for their individual condition, which is a primary concern of Leapfrog and our purchaser and employer membership. The measure also assesses image noise, safeguarding image quality against potential effects of dose reduction, and is the first quality measure to do so. Leapfrog strongly supports this measure.

Comment 19 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

Comment Part 1:

We thank the American College of Radiology for their comments from 1/19/22 but wish to address several factual inaccuracies in the comments. (Response PART 1) **Comment:** *Indications for exams do not have standardized language that could be used to track them. Most health and IT systems capture…coding for reimbursement, but typically not enough... As a result, the clinical reason for performing an imaging exam is often extremely limited in the exam order... A validated method for determining classification of studies .. must be incorporated into such a measure.* **Response:** This statement indicates that the commenter does not understand how clinical indication is determined in the proposed measure. It does not rely on the clinical reason for performing an imaging exam in the exam order. As described in Specifications, sp-11, clinical indication for imaging is determined using an algorithm that combines procedure (CPT®) and diagnosis (ICD-10-CM) codes associated with the clinical visit when the test was ordered, information provided as part of the order, and information on the final bill. The codes are available in the radiology electronic systems and/or the EHR or billing systems. The goal in creating the CT categorization decision rules was to identify exams that are exceptions to the routine dose category (i.e. either high or low dose). The approach of assigning CT exams to the various CT categories in an automated fashion using an algorithm was developed using over 4.5 million CT exams in the UCSF International CT Dose Registry. We confirmed that the CT categories were representative of groupings that require different radiation dose and image quality (Smith-Bindman 2021). The algorithm was validated using over 10,000 patient records from UCSF Health. The CT category assignment determined by the algorithm was compared with a "gold standard" chart review, as described in Validity sections 2b.02 and 2b.03. Since we did not have access to complete medical records at testing sites, we developed a second referent standard that determined CT category based on natural language processing of DICOM data and the full radiology report. This second referent standard was found to be accurate compared to the gold standard chart review of the same sample of UCSF Health exams (sensitivity = 0.92, specificity = 0.97; see 2b.02). When the algorithm was deployed at testing sites, the correct classification rate of CT category assignment was on average 92% across clinician groups and hospitals and 95% in individual clinicians (see 2b.03). Knowing that the algorithm was developed using data from a single health system, we performed detailed investigation of the categorization results at testing sites – comparing the assigned CT category against full radiology reports – for the purpose of improving the algorithm, which we did. **Reference:** Smith-Bindman R, Yu S, Wang Y, et al. An Image Quality-informed Framework for CT Characterization. Radiology. 2021 Nov 9:210591. **Comment:** *The developer states their company can provide the service of quantifying the measure at a cost; this should also be included as a potential limitation. The measure developer does provide specifications for other entities to implement the measure, but the burden of implementation may be significant.* **Response:** This is inaccurate. As stated in Feasibility, 3.07, there are no fees for users submitting their eCQM data to CMS programs. The eCQM can be run and the measure score calculated by any EHR vendor or hospital and reporting entities can partner with any commercial partner capable of developing reporting software using the eCQM specifications. The measure steward's software to ingest this data and calculate the measure is freely available. Alara Imaging has created an edge device that can assemble data from different electronic sources (e.g. EHR, RIS [Radiology Information Systems], PACS [Picture Archiving and Communication Systems], and billing) to calculate the CT category, size-adjusted dose, and image noise that can then be consumed by the eCQM. If practices want to calculate these variables without using the Alara edge device, they may access a free online portal to calculate these variables and provide them to any entity implementing the measure. A prototype of this software was deployed at 8 testing sites (7 hospital systems and 1 ambulatory imaging network). Sites were asked to install the software,

configure local connections to PACS, EHR, and other electronic systems as needed, and oversee the transfer of data to it from these sources. Burden was found to be no more or less onerous than the effort required by participation in other eCQMs or national registries, such as the ACR Dose Index Registry (Feasibility, 3.06). **Comment:** *For CT category … the developer used NLP for obtaining data such as reason for study or protocol name used in the calculation of this variable. The submission does not provide information on the NLP results' reliability and validity… or how sites would get access to use this custom NLP tool.* **Response:** This is incorrect; the measure does not use NLP. As described in the submission and above, it uses an algorithm that combines CPT® and ICD-10-CM codes to categorize CT exams. NLP was deployed as a method to validate the CT categorization determined by the algorithm at testing sites, where we did not have access to medical records. The sensitivity and specificity of this NLP referent standard are given above. **Comment:** *Multiple unstructured variables are required to construct the data elements for the numerator, denominator, and exclusions…* **Response:** This is incorrect; the measure does not use unstructured data. All data elements used to calculate the measure come from structured variables listed in the feasibility scorecards and in Specifications, Table sp-2: CPT® and ICD-10-CM codes; dose length product stored in the DICOM data; and patient diameter and image noise calculated on imaging data. The measure would not have met the requirements of an eCQM had it relied on unstructured data. **Comment:** *Protocol selection appropriate for a clinical indication is an important component of radiation dose management along with radiation dose optimization. Each component needs to be addressed as a separate quality action. The specific aspect(s) of performance to be improved is not intuitive due to the multiple components to the measures… It is true that the most accurate way to address appropriate and safe use of multi-phase studies is to measure both the clinical indication of an exam and the radiation dose output… However, these measures conflate the appropriateness of protocol for the clinical indication and radiation dose optimization… a facility may not be able to determine if its performance could be improved by adjusting protocols or by focusing on appropriateness of the ordered exam.* **Response:** We agree that selecting an appropriate CT protocol and limiting radiation dose given the selected protocol are separate quality actions, but the commenter misses the crucial point that intermediate outcome measures typically reflect multiple opportunities for improvement. By analogy, we recognize systolic blood pressure control and glycosylated hemoglobin control as intermediate outcome measures for patients with hypertension and diabetes, respectively, even though there are many potential ways to manage these conditions. The fact that these intermediate outcomes can be improved by diet, exercise, medications, or combined approaches does not invalidate glycosylated hemoglobin or blood pressure control as quality measures. Similarly the fact that our measure would be responsive to multiple, interrelated process steps is a key strength that will improve its value for reducing radiation exposure at the population level. Further, reporting entities will be provided with feedback for each CT exam, including its assigned CT category, radiation dose, size-adjusted radiation dose, and image noise, allowing recipients to identify the causes of performance gaps. Reporting entities will be able to assess if they are systematically assigning patients to the wrong protocol, or if they are choosing protocol settings that are inappropriate with respect to radiation dose or image noise. The actionability of the feedback is noted in the other letters written in support of the measure. To further demonstrate the potential of this measure, we conducted a randomized controlled trial in 100 hospitals and outpatient radiology practices to study the impact of providing detailed audit feedback, similar to what will be provided as part of the feedback on this measure. We found that this intervention resulted in significant reductions in radiation dose and dose variation with no impact to image quality, described in Usability, 4b.01. (Smith-Bindman, 2020) **Reference:** Smith-Bindman R, Chu P, Wang Y, et al. Comparison of the Effectiveness of Single-Component and Multicomponent Interventions for Reducing Radiation Doses in Patients

Undergoing Computed Tomography: A Randomized Clinical Trial. JAMA Intern Med. 2020 May 1;180(5):666-675.

Comment 20 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

Comment Part 2:

We thank the American College of Radiology for their comments from 1/19/22 but wish to address several factual inaccuracies in the comments. (Response PART 2) **Comment:** *NQF #3633e, #3662e, and #3663e deviate from international standards, like diagnostic reference levels, and lack peer reviewed, broadly accepted consensus on global noise. For these measures, global noise is defined solely by the measure developer. Endorsing this method may encourage facilities to accept a narrow view of image quality.* **Response:** The ACR correctly notes that we have defined an approach to measuring noise. We did so only after testing and comparing multiple approaches described in peer-reviewed literature and validating noise measurements against radiologists' assessment of image adequacy for diagnosis. Image quality is a much less common problem than excessive use of radiation in CT imaging. While there may be other reasons to study CT image quality, our interest was simply to ensure that CT image quality did not erode as an unintended consequence of lowering radiation doses. There is no reason to believe that endorsing this measure will encourage facilities to "accept a narrow view of image quality" because radiologists have a requirement for adequate images to perform their work. They have no desire or motivation to alter their standards of what constitutes an adequate image. Radiologists do not want to read inadequate images and routinely request that such images be repeated or complemented by other imaging modalities. **Comment:** *The ACR requests the developer further clarify the global noise table used in calculating the numerator... For example, Table sp-1 has the same global noise threshold for several CT categories, such as head low dose, head routine dose, and head high dose... If the image noise thresholds are the same, the size-adjusted radiation dose thresholds should be the same.* **Response:** We tested various published methods for measuring image noise and opted for a modified version of the method proposed by Malkus in 2017. The approach for setting the thresholds for image quality and radiation dose were based on the referent standard of radiologists' satisfaction with image quality. This did not always result in the relationship the ACR has suggested. For example, radiologists might want a minimum level of image quality for all head CT categories whereas the upper dose threshold might vary across the three head categories reflecting the different clinical indications comprising each group. Radiologists in our image quality study graded the majority of head exams as having acceptable image quality, even those at the lower dose range, meaning the minimum noise threshold is similar for all three categories. **Reference:** Malkus A, Szczykutowicz TP. A method to extract image noise level from patient images in CT. Med Phys. 2017 Jun;44(6):2173-2184. **Comment:** *Additionally, current CT scanners display dose values based on either a 16 cm or 32 cm phantom for a neck scan, which must be carefully accounted for in measure performance calculations.* **Response:** As the ACR correctly notes, CT scanners display dose values based on a 16 cm or 32 cm phantom. If comparisons are made across reporting entities it is important that they use the same phantom, as this impacts the scanner reported DLP. The manufacturers are highly consistent in their use of phantoms for different body regions. In a study of 106,837 pediatric patients (a population where potential variation in phantom choice would most likely occur), 100% of CT exams in the neck are referenced to the 32 cm phantom, and it is thus unnecessary to account for phantom selection (Chu 2021). **Reference:** Chu PW, Yu S, Wang Y, et al. Reference phantom selection in pediatric computed tomography using data from a large, multicenter registry. Pediatr Radiol. 2021 Dec 6. **Comment:** *These eCQMs require multiple variables that may be captured in software systems external to electronic health records (EHRs), such as dictation systems*

housing radiology reports or DICOM standard-based systems, such as CT device software. Data element validity testing should demonstrate that the testing sites were able to integrate and validate the variables used to construct the data elements used by the eCQM in addition to the usual validation of the eCQM's electronic output against the medical record review. We are uncertain that this validation has been completed. Therefore, this submission does not demonstrate the measure can be reproduced in a reliable and valid manner by practices or facilities across multiple settings. **Response:** This comment is entirely erroneous. No data are pulled from dictation systems or CT device software. The measure derives and uses codified and specified data from DICOM standard based systems, such as PACS, and EHR and billing claims. Our data element validity testing did demonstrate that 8 testing sites, reflecting 16 hospitals and 13 outpatient imaging facilities, were able to integrate, collect, and report the variables used to construct the data elements ingested by the eCQM. The letters of support from these testing sites independently confirm their ability to assemble the required data across diverse practice types and settings. **Comment:** *The ACR is concerned with the selection bias for the accountable entity-level… validity. Assessing measure score face validity through the TEP that created these measures lessens the extent of credibility for these results. Although the TEP is knowledgeable and represents a variety of stakeholders, there is a vested interest in ensuring these measures are available for use.* **Response:** All of the TEP members and their affiliations are identified in our submission materials (2b.02). Conflicts of interest were reviewed at each meeting and included with meeting minutes in a publicly available website (https://ctqualitymeasure.ucsf.edu/). The TEP members all voluntarily provided public service by joining the TEP. None of our TEP members has any "vested interest" in the outcome of the NQF endorsement process other than the ACR which served as a single member of the TEP. None of our TEP members is employed by the developer organization (UCSF) or its funder (CMS), nor has any financial interest in the company that is offering technical support for software implementation (Alara Imaging). To be clear, these measures were developed by an academic radiology, quality improvement, and analytics team based at UCSF and supported by CMS, NIH and PCORI. The TEP was organized and tasked to provide broad multidisciplinary input to this team. Their endorsement of the validity of the measures is highly credible, as it reflects the fact that their advice was heeded at every stage of the development and testing process. Our TEP process followed the CMS Blueprint as well as NQF guidance, and 16/17 members agreed that that implementation of the measure will lead to a reduction in average CT radiation dose while maintaining adequate CT image quality if adopted (reported in 2b.03).

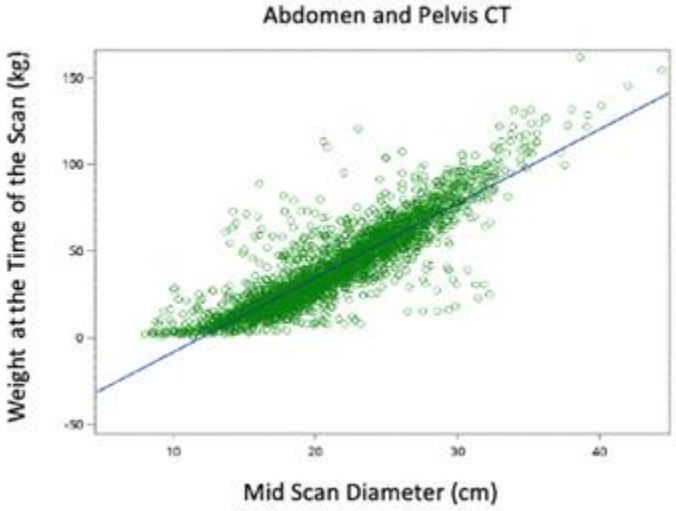
Comment 21 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

We thank the American Association of Physicists in Medicine for their perspectives but wish to address several factual inaccuracies: **Comment 1:** *Unscientific characterization of CT scan risk: The proposal is based on estimation approaches that are not reflective of the consensus of the scientific community* **Response:** The measure is not focused on radiation risk and does not calculate nor report radiation risk. The measure evaluates dose length product (DLP), and specifically whether size-adjusted DLP exceeds thresholds specific to CT category. DLP is the radiation dose measure most directly under the control of providers, determined by selected parameters. Further, DLP is universally reported by CT manufacturers. It is thus the ideal measurement to use when assessing the quality of CT exams. The TEP, which included the ACR, radiologists and a medical physicist, unanimously supported the radiation dose measure used and agreed is a relevant metric of quality for CT imaging (2b.03).There is also considerable precedent for using DLP to evaluate radiation dose in CT. The American College of Radiology has used DLP to set benchmarks [Kanal 2017] and to measure dose in its own NQF-endorsed quality measure #3621. **Reference:** Kanal KM et al. U.S.

Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-133. **Comment 2:** *Inactionability of the measures to enable targeted change to improve practice: It is not evident how the proposed measures can be practically used* **Response:** Reporting entities will be provided with specific feedback for each CT scan on its assigned CT category, radiation dose, size-adjusted radiation dose, and image noise, allowing recipients to identify causes of performance gaps and make targeted changes to improve quality. Comments in support of the measure from the testing sites describe how useful the information provided was to allow them to understand and improve their practice. As described in our submission, we found in a randomized controlled trial in 100 imaging facilities that providing detailed audit feedback on radiation doses, similar to what will be provided as part of the feedback on this measure, resulted in significant reductions in radiation dose with no impact on satisfaction with image quality (see Usability, 4b.01). (Smith-Bindman, 2020) **Reference:** Smith-Bindman R et al. Comparison of the Effectiveness of Single-Component and Multicomponent Interventions for Reducing Radiation Doses in Patients Undergoing Computed Tomography: A Randomized Clinical Trial. JAMA Intern Med. 2020 May 1;180(5):666-675. **Comment 3:** *Inadequate addressing of the complexity of CT categorization* **Response:** A detailed response to this question was provided in our response to the ACR. In short, the approach of assigning CT examinations to the different CT categories as specified in the measure was first developed using records from over 4.5 million CT exams in the UCSF International CT Dose Registry (Smith-Bindman, 2021). We then developed an approach for determining the clinical indication for imaging using an algorithm that combines procedure (CPT®) and diagnosis codes (ICD-10-CM) provided in Specifications, sp-11. This algorithm was developed using detailed review of over 10,000 patient records from UCSF Health. We validated the CT category assignment using the algorithm against "gold standard" chart review, as described in Validity sections 2b.02 and 2b.03. When the algorithm was deployed at our testing sites the correct classification rate of the assignment of CT exams to CT category was on average 92% across clinician groups and hospitals and 95% in individual clinicians. **Comment 4:** *Inadequate assessment of noise: Noise in a CT image can be influenced by a variety of factors* **Comment 5:** *Inadequate assessment of image quality: Image quality is affected by a myriad of factors* **Response:** The primary focus of our measure is to assess radiation dose adjusted for body size. The image quality component was included to protect against the unlikely possibility of substantial degradation of image quality as an unintended consequence of dose reduction. Our measure of image quality reflects what radiologists in practice regard as adequate. Others might have an interest in other ratings of image quality for other purposes, but that was not our intent. We tested and found that noise as a measure of image quality was associated with radiologists' satisfaction with the adequacy of CT images. These results were included in the submission (2b.03). **Comment 6:** *Flawed assumption on dose reduction vs dose optimization: The application focuses primarily on radiation dose reduction as opposed to right-sizing the dose.* **Response:** This is incorrect. We created the CT categories based on radiation dose and image quality requirements specific to clinical indications for imaging. Using radiologists' satisfaction with image quality, we established an image quality floor for each category, below which an exam is considered to have inadequate quality, and a radiation dose ceiling, beyond which doses are considered unnecessarily high. The purpose is to allow detailed assessment of each CT exam to ensure the dose is optimal based on the clinical indication for imaging. In current practice, there are no such benchmarks created by clinical indication, making it impossible for providers to know the right dose range for each patient. In our testing data, far more CT exams exceeded the radiation dose ceiling (average = 30%) than failed to meet the image quality requirement (average < 1%) (see section 1b.02). The measure encourages entities to reduce the proportion of exams that may "be overdosed for their exact need and condition" while preserving the minimum image quality. **Comment 7:** *Inadequate accuracy in patient size estimation: Assessing a patient size is not a trivial task, stemming from*

significant variability in the differences in the habitus of different patients, coupled with the existential challenge that there is no single metric **Response:** We agree that measuring patient size is important. Our approach for using mid-scan diameter is highly correlated with patient weight: in separate, NIH-funded research on CT use in children up to age 21 (Kwan 2022), we have shown that diameter in 4,239 children as measured on mid-scan axial images is highly predictive of patient weight, correlation = 0.904, Figure 1.

Figure 1. Mid scan patient diameter versus patient weight at the time of the CT, Kwan 2022



For this measure, patient size was measured using CT image pixel data, either on the mid-scan axial image or the coronal scout image when the mid-scan axial image was not available. This approach has been validated using data from UCSF Health, the UCSF Registry, as well as the data assembled for measure testing. While there may be different ways to measure patient size, and different reasons for measuring patient size, it is a crucial piece of information that must be practically defined to ensure that the types of patients (case mix) at different practices do not bias the number of scans graded as out-of-range. We are adjusting for patient size primarily to ensure that entities that see larger patients are not penalized for doing so. Figure 2a shows the relationship between radiation dose (in DLP) and patient diameter using data from the UCSF Registry for abdomen CT. We chose abdomen CT as this is the category most influenced by patient size, and where patient mix could impact an entity's out-of-range rate. The raw correlation between patient diameter and unadjusted DLP is 0.50, and the marginal R-squared of the log-linear model used for adjustment is 0.15. After size-adjustment, the relationship is nearly removed: Figure 2b shows size-adjusted DLP by patient diameter using the same data; the raw correlation is far lower (-0.09), and the modeled marginal R-squared post-adjustment is 0. This demonstrates adequacy of the approach for adjustment of patient size.

Figure 2a: Unadjusted Dose Length Product vs Patient Diameter

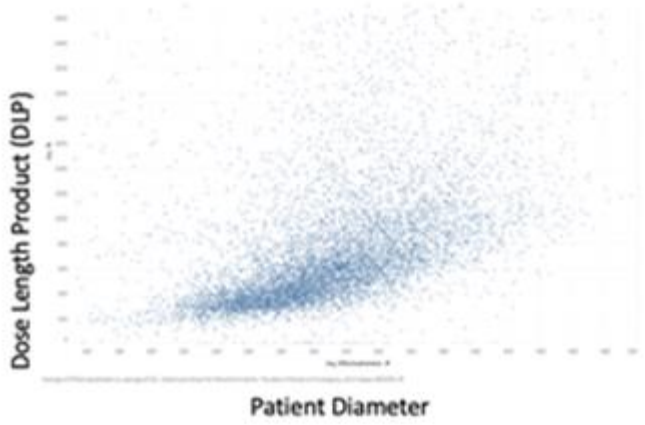
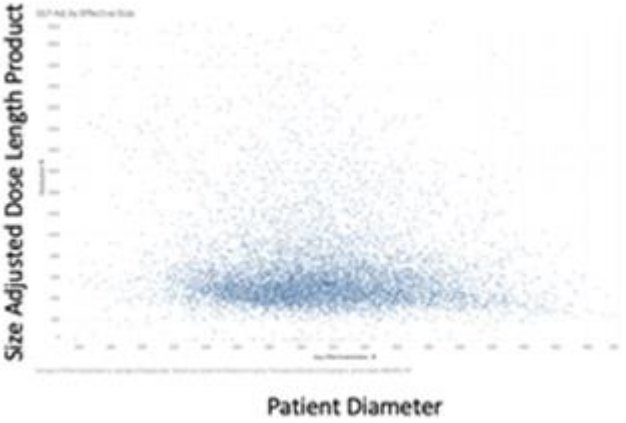


Figure 2b: Size-Adjusted Dose Length Product vs Patient Diameter



The adequacy of size adjustment was shown empirically using data assembled from the testing sites. The proportion of exams with out-of-range rates based on **unadjusted** and **adjusted** DLP are shown in Tables 1a and 1b. Without adjustment, the out-of-range values are strongly associated with patient size; after adjustment this relationship is gone.

Table 1a) Proportion of exams out-of-range on routine dose abdomen exams based on **unadjusted** DLP across the 16 hospitals, shown by decile in patient size. The proportion of out-of-range exams increased with patient size, seen in the table as an increase in dark shading in the lower rows of the table. Among patients in the highest size decile – those in last row– the out-of-range proportions across the 16 hospitals ranged from 93-100%.

Size Decile	Hospitals 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.27	0.22	0.11	0.00	0.27	0.29	0.30	0.14	0.34	0.20	0.24	0.40	0.06	0.17	0.11	0.17
2	0.30	0.00	0.00	0.00	0.08	0.11	0.13	0.29	0.24	0.30	0.04	0.19	0.00	0.07	0.16	0.09
3	0.15	0.06	0.15	0.00	0.17	0.18	0.22	0.75	0.21	0.30	0.17	0.18	0.08	0.18	0.17	0.12
4	0.07	0.17	0.29	0.15	0.25	0.32	0.09	0.82	0.43	0.25	0.07	0.42	0.10	0.17	0.19	0.21
5	0.45	0.15	0.13	0.14	0.28	0.43	0.00	0.93	0.40	0.42	0.19	0.38	0.00	0.14	0.48	0.55
6	0.42	0.20	0.25	0.36	0.55	0.61	0.27	0.96	0.55	0.19	0.31	0.51	0.08	0.46	0.47	0.78
7	0.79	0.47	0.45	0.58	0.70	0.75	0.17	1.00	0.69	0.37	0.26	0.73	0.06	0.71	0.66	0.90
8	0.81	0.37	0.75	0.69	0.67	0.86	0.24	1.00	0.89	0.35	0.58	0.77	0.22	0.80	0.91	0.95
9	0.96	0.85	1.00	0.75	0.88	0.93	0.26	0.93	0.94	0.64	0.78	0.93	0.63	0.90	1.00	1.00
10	1.00	0.96	0.98	0.93	0.97	0.97	0.93	0.94	1.00	0.95	0.98	0.96	0.85	0.95	0.94	1.00

Table 1b) Proportion of exams out-of-range on routine dose abdomen exams, based on **size-adjusted** DLP across the 16 hospitals shown by decile in patient size. High proportion of out-of-range exams are no longer concentrated among the larger patients. Among patients in the highest size decile, out-of-range rates ranged from 11-53%.

Size Decile	Hospitals 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.55	0.61	0.22	0.20	0.42	0.48	0.48	0.61	0.49	0.70	0.37	0.62	0.10	0.37	0.22	0.51
2	0.50	0.25	0.00	0.18	0.19	0.31	0.22	0.71	0.36	0.61	0.15	0.38	0.00	0.08	0.33	0.21
3	0.15	0.18	0.15	0.11	0.37	0.39	0.33	0.96	0.30	0.50	0.26	0.36	0.10	0.23	0.43	0.22
4	0.27	0.17	0.29	0.15	0.35	0.54	0.09	0.97	0.43	0.38	0.11	0.53	0.10	0.18	0.31	0.30
5	0.45	0.15	0.13	0.14	0.26	0.46	0.00	0.93	0.40	0.42	0.19	0.38	0.00	0.19	0.52	0.59
6	0.33	0.10	0.25	0.36	0.39	0.45	0.27	0.96	0.47	0.13	0.31	0.40	0.05	0.34	0.45	0.72
7	0.29	0.18	0.20	0.46	0.50	0.42	0.17	0.90	0.57	0.16	0.17	0.60	0.04	0.50	0.36	0.70
8	0.43	0.05	0.19	0.25	0.54	0.39	0.12	0.70	0.58	0.09	0.35	0.62	0.09	0.59	0.53	0.83
9	0.48	0.26	0.48	0.30	0.45	0.27	0.11	0.19	0.72	0.07	0.18	0.56	0.06	0.62	0.60	0.66
10	0.35	0.27	0.40	0.39	0.38	0.11	0.27	0.11	0.61	0.37	0.29	0.44	0.11	0.27	0.53	0.36

Reference: Marilyn Kwan et al. Quantifying cancer risk from exposures to medical imaging in the Risk of Pediatric and Adolescent Cancer Associated with Medical Imaging (RIC) Study: Research Methods and Cohort Profile. In press, Cancer Causes and Control. **Additional Comment:** *One cited reference supports the proposed measure, however, this cited article has an accompanied editorial that highlights the limitations of the proposed approach [Mahesh M. Benchmarking CT Radiation Doses...Radiology. 2021 Nov 9:212624.]* **Response**: We find it surprising that Dr. Mahesh's editorial was used to criticize the measure. Dr. Mahesh is a board member of American College of Radiology and American Association of Physicists in Medicine, and he was very positive about our image quality-informed framework for assessing radiation dose. He noted the observed, significant differences *between* CT categories versus *within* categories was "an encouraging result for anyone trying to optimize CT studies based on clinical indications." He noted the study was "a good start" on the road to optimizing CT protocols based on image quality. He opined that the CT classification would be more useable and easier to implement if based on current procedural terminology codes. This is precisely what we have done in this measure.

Comment 22 by: Carly Stewart, University of California, San Francisco
On behalf of Rebecca Smith-Bindman, University of California, San Francisco

We thank Dr. Ehsan Samei for sharing his perspectives on the measure and for collaborating with us early in the measure development process. We wish to address a few inaccuracies and misunderstandings in Dr. Samei's comments. The majority of Dr. Samei's comments focus on image quality and his concern that the measure does not offer a comprehensive assessment of image quality. Our measure is not intended to be a comprehensive assessment of image quality.

Criticizing the proposed measure for what it is not is beyond the scope of what should be considered in assessing the usefulness of what has been submitted. The primary focus of our measure is to assess radiation dose adjusted for body size, and the image quality component provides a means to protect against the unlikely possibility of substantial degradation of image quality as an unintended consequence of dose reduction. The approach for creating thresholds is described in Validity, 2b.02. **Comment:** ***Inaccurate assessment of patient size:*** *The measure of size proposed is calibrated to earlier work and publication from our group at Duke University for academic purposes. That early method they have embraced has had major errors.* **Response**: We are adjusting for patient size primarily to ensure that entities that see larger patients are not penalized for doing so. Although we explored code that Dr. Samei provided early in our initial efforts to measure patient body habitus we found that it was inadequate, particularly for some CT categories, and we have not relied upon it. We developed our own approach for measuring size using CT image pixel data from the mid-scan axial image or the coronal scout image when the mid-scan axial image was not available. Our approach of measuring size was shown to be highly correlated with patient weight (correlation = 0.904) in a large study in children described in our response to the AAPM. For this measure, the measurement of size was validated using data from UCSF Health, the UCSF Registry, as well as the data assembled for measure testing. The adequacy of the approach we have adopted for size adjustment is described in the initial application and the response to the comments by the AAPM. **Comment:** ***Inaccurate assessment of noise:*** *The measure of noise proposed references earlier work and publication from our group at Duke University. That early method exhibited errors, corrected in subsequent versions that have not been shared…* **Response:** Dr. Samei's approach and code for measuring image quality were explored in the process of developing our measure but were not included in the final measure specifications. Any errors in his approach are not relevant to the measure. **Comment:** ***Inaccurate assessment of radiation risk:*** *The measure of size-adjusted radiation risk, adjusting the CT scanner outputs with 'patient size' to perform risk estimation is not a standard method nor endorsed by any scientific or professional body… Patient risk can only be assessed with the knowledge of organ doses that is not even mentioned in the application let alone pursued. The proposed method CANNOT be used as surrogate for future cancer risk.* **Response:** The measure does not calculate or report radiation risk. The measure evaluates radiation *dose* (measured in dose length product, DLP), and whether size-adjusted DLP exceeds thresholds specific to CT category. The empirical validity of the risk-adjustment approach based on patient size is described in the application (section 2b.26 – 2b.31) and in our response to the comments by the AAPM. The approach of evaluating CT safety by comparing machine output (whether DLP or CTDIvol) against benchmarks is widely accepted in the radiology field. (Kanal 2017) In contrast, organ dose has no standard definition, is not reported by the manufacturers, is not available in a structured format, would be time intensive to calculate in clinical settings and most importantly has limited actionability as this is not under the direct control of technologists or physicians. Organ doses may be useful for counseling patients or in the context of epidemiological studies, but we do not believe it has a role as a metric for CT quality measurement. **Reference:** Kanal KM, Butler PF, Sengupta D, et al. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations. Radiology. 2017;284(1):120-1 ***Comment:*** ***Subjectivity:*** *The measures are anchored to subjective perception by radiologists as how they "like" the images. There is in fact no evidence provided that the measures can lead to an improvement in diagnostic accuracy. In fact, it might lead to a degradation.* **Response:** The measure is not intended to improve diagnostic accuracy. The purpose of the measure is to establish a radiation dose ceiling to avoid excessive radiation exposure, and an image quality floor to safeguard against unintended deterioration of image quality. There is precedent for using radiologist satisfaction with image quality to set or validate noise targets, including work by Dr. Samei. (Cheng 2019, IAEA 2009) This also reflects clinical practice: radiologists subjectively assess images and regularly ask for scans to

be repeated when they are not adequate. As described in the response to ACR comments, Radiologists do not want to read inadequate images and routinely request that such images be repeated or complemented by other imaging modalities. Radiologist's subjective assessment provides a practical way to ensure the image quality is not degraded through efforts to optimize the radiation doses. **References:** Cheng Y, Abadi E, Smith TB, Ria F, Meyer M, Marin D, Samei E. Validation of algorithmic CT image quality metrics with preferences of radiologists. Med Phys. 2019 Nov;46(11):4837-4846. doi: 10.1002/mp.13795. Epub 2019 Sep 20. International Atomic Energy Agency (IAEA), Dose Reduction in CT while Maintaining Diagnostic Confidence: A Feasibility/Demonstration Study, TECDOC Series, 2009.

National Quality Forum
1099 14th Street NW, Suite 500
Washington, DC 20005
http://www.qualityforum.org