NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2820
Corresponding Measures:
Measure Title: Pediatric Computed Tomography (CT) Radiation Dose
Measure Steward: University of California, San Francisco
sp.02. Brief Description of Measure: 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.
1b.01. Developer Rationale: Radiologists and other physicians who perform CT are frequently not aware of the doses they use. Further, they do not know that the doses that are routinely used are in the range where they will cause cancer in some patients. They are also not aware that there is tremendous variation in doses for patients who are seen at different hospitals, even if the clinical question is the same. (ACR, DIR, 2014; Hausleiter, JAMA 2009; Keegan, JACR 2014; Miglioretti, JAMA Pediatrics, 2013; Parker, Pediatrics, 2015; Smith-Bindman, Arch Int Med, 2009; Smith-Bindman, JAMA, 2012; Smith-Bindman, JACR 2014, Smith-Bindman BMJ 2019).
Reducing doses among children who receive unnecessarily high doses could have a significant impact on the number of children who would develop cancer. In our JAMA Pediatrics paper (Miglioretti, JAMA IM 2013) using statistical modeling and observed CT doses across a large number of patients evaluated across an integrated HMO where we collected data on consecutive scans, we modeled what would occur if the highest dose patients (those above the 75% among the observed doses) came down to the median dose. In these patients the two dominant indications for imaging was minor trauma and suspected appendicitis. Using the observed exposures, we would expect that due to CT exposures in children age 15 and younger in the US in 2010, 9,820 future cancers would occur. If the highest exposed individuals instead had doses at the median, 44% of these cancers would be prevented. Thus reducing the outlier doses - which are most never justified or needed, could have a substantial impact.
The motivation for this measure is the belief that if radiologists would be provided with explicitly benchmarks and were informed when their doses routinely exceed those benchmarks, that they would be motivated to improve to bring their performance closer to routine practice. Currently, physicians do not know the typical

radiation doses received by their patients. This tool provides the framework for measurement – the first step towards quality improvement.
Creation of a simple standard for collection of radiation dose information would help facilities understand their current practice, would allow understanding changes in practice over time (Keegan, JACR, 2014; Greenwood, RadioGraphics, 2015) would allow comparisons to local and national standards, and would indicate to facilities whether their is a need to improve. There is currently a high level of interest in this area - facilities are being asked by their patients and governing boards to report whether they are performing CT safely - and this measure is an ideal starting point for facilities to assemble this information to answer these questions. If facilities collect dose information, it is the first step towards trying to compete on a measure of safety and to lower the doses they use.
A 2014 study by Keegan et al. observed a reduction in dose metrics after an earlier NQF measure of CT radiation dose (#0739) was introduced. All dose metrics in inpatient and outpatient facilities that had adopted the measure demonstrated a 30% to 50% reduction between 2010 and 2012, the time period over which the measure was implemented.
Cited in this section: American College of Radiology (ACR). Dose Index Registry (DIR). 2014. http://www.acr.org/
sp.12. Numerator Statement: 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.
sp.14. Denominator Statement: 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.
sp.16. Denominator Exclusions: Examinations with missing anatomic area, patient age, or missing dose length product are excluded.
Measure Type: Outcome: Intermediate Clinical Outcome
sp.28. Data Source: Electronic Health Data; Electronic Health Records; Registry Data
sp.07. Level of Analysis: Facility
IF Endorsement Maintenance – Original Endorsement Date: 5/4/2016
Most Recent Endorsement Date: 5/4/2016

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement, endorsed measures are evaluated periodically to ensure that the measure still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a ***structure, process or intermediate outcome*** measure are that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also

should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following description for this measure:

- This is a maintenance intermediate outcome measure at the facility level that assesses the radiation dose (as 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 developer provides a logic model that describes substantial variation in radiation doses used for CT exams, primarily due to differences in how radiologists choose to perform them, meaning, their choice of a specific imaging protocol (for example, a single or multiple phase CT) and the specific technical parameters used such as scan length, milliampere-seconds, and kilovoltage peak.

The developer provides the following evidence for this measure:

- Systematic Review of the evidence specific to this measure? ☒ Yes ☐ No
- Quality, Quantity and Consistency of evidence provided? ☒ Yes ☐ No
- Evidence graded? ☒ Yes ☐ No

Summary of prior review in that was endorsed in 2016:

- This measure was originally reviewed by the Pediatrics Standing Committee in 2013 but was not recommended for endorsement due to concerns about the potential impact of the measure. The measure returned for review in 2016 with data showing that tracking radiation doses was associated with a change in behavior.
- The Standing Committee agreed the new data demonstrate the measure should lead organizations to address the issue of high doses for children if their doses are higher than national benchmarks, and it should give facilities a framework for setting their dose levels.
- Some Standing Committee members questioned whether non-pediatric radiologists could properly read lower dose scans, which are "noisier." However, radiologists on the Standing Committee explained a lower dose for children would produce an image of the same quality that occurs for an adult at the higher dose.

Changes to evidence from last review

☐ The developer attests that there have been no changes in the evidence since the measure was last evaluated.

☒ The developer provided updated evidence for this measure:

- Since the last review, the developer presented data from two published systematic reviews/meta-analyses providing evidence of increased excess cancer risk from low-dose ionizing radiation in the radiation dose ranges typically used in CT imaging.
    - The first systematic review/meta-analysis focused on early life exposer and contained 21 observational studies, including 11 case-control studies and 10 cohort studies. All studies were assessed to be of good quality, with Newcastle-Ottawa Scale (NOS) scores ranging from 7 to 9.
        - The evidence was graded using the NOS for studies of radiation exposure in children. The NOS assesses the quality of non-randomized studies, using 8 items grouped into 3 domains (i.e., selection, comparability/confounding, and outcome/exposure assessment), with 9 being the best possible score. NOS scores of 6 to 9 equate with "good quality" in the Agency for Healthcare Research and Quality (AHRQ) standards for observational studies. Good quality is the highest possible rating on the AHRQ scale.

- This review focused specifically on children exposed to radiation from medical imaging and found that risks of leukemia and brain tumor increase over background risk by 27 and 9 percent, respectively.

- The second systematic review/meta-analysis included 26 studies, which supported the conclusion that there are significant excess cancer risks from low-dose ionizing radiation.
- Additional references and review of peer-reviewed research were provided by the developer to contribute to the already large body of epidemiologic and biologic evidence supporting the premise that exposure to ionizing radiation from medical imaging increases a person's risk of developing cancer.

Question for the Committee:

- *How strong is the updated evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*

Guidance from the Evidence Algorithm

Intermediate Outcome measure based on systematic review (Box 3) -> QQC presented (Box 4) -> Quantity: High; Quality: Moderate; Consistency: High (Box 5) -> High (Box 5a) -> High

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- In scoring approach one (individual strata), poor performance is classified by whether a facility's median radiation dose is greater than or equal to the benchmark 75th percentile in any single stratum. The proportion of facilities in the UCSF International CT Dose Registry that had a median dose above threshold in at least one stratum was 49 percent (52/106).
- In scoring approach two (all strata combined), poor performance is classified when 50 percent or more of a facility's exams across all strata exceed the relevant stratum-specific benchmark 75th percentile. In the UCSF International CT Dose Registry, 114,443 exams derive from 106 hospitals with at least 10 observed exams. Performance data were as follows:
  - Mean measure score: 26 percent
  - Standard deviation: 16 percent
  - Interquartile range: 18 percent (16% - 34%)

Disparities

- The developer presented data from the UCSF International CT Dose Registry on the proportion of exams across all strata that exceed the relevant stratum-specific benchmark 75th percentile, stratified by sex and age. The data show that males and the youngest children (less than one year old) have slightly higher out-of-range values.
- The developer states that, to the extent they have been studied in both children and adults, social factors are not predictive of radiation dose for CT exams; however, some studies have shown that patients living in poverty are at higher risk for exposure to multiple scans over time and increased cumulative diagnostic imaging radiation.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

1a. Evidence

- [Standing Committee feedback]

1b. Gap in Care/Opportunity for Improvement and Disparities

- [Standing Committee feedback]

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: Alex Sox-Harris, Christie Teigland, Jack Needleman, Sean O'Brien, Jeff Geppert, Larry Glance, Marybeth Farquhar, Sherrie Kaplan, Terri Warholak, Sam Simon, Paul Kurlansky, Eric Weinhandl
Combined Methods Panel Review

- The SMP Passed on Reliability with a score of: H-5; M-4; L-0; I-1
- The SMP Passed on Validity with a score of: H-1; M-7; L-1; I-1

2a. Reliability: Specifications and Testing

For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

For maintenance measures – less emphasis if no new testing data provided.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

Specifications:

- Measure specifications are clear and have changed since last review.
- The developer has made changes to several aspects of the measure including anatomic areas, dose metrics, phantom standardization, and measure scoring:
    - Anatomic areas: the developer has narrowed the measure focus. "Head" has been subdivided into "skull" and "brain" categories. In all, the measure focuses on skull, brain, and abdomen and pelvis.
    - Dose metrics: prior measure specifications collected Volumetric CT Dose Index (CTDIvol) and Size Specific Dose Estimate (SSDE) as well. The measure now solely collects dose length product (DLP).
    - Phantom standardization: skull and brain exams must now be reported using the 16-cm reference phantom. Abdomen exams must be reported using the 32-cm reference phantom.

- Measure scoring: the developer now offers two approaches to scoring. Approach 1 is a binary scoring system of median radiation doses within age-anatomic region strata. Approach 2 measures the overall proportion of high dose exams across all strata.
    - Approach 1: If 50 percent or more of exams in a stratum have DLP values at or below the benchmark for that stratum, then the dose distribution within that stratum will be considered acceptable. If greater than 50percent of exams have DLP values above the benchmark, then the dose distribution within that stratum will be considered poor.
    - Approach 2: Hospitals/Facilities with greater than 50 percent of their CT doses exceeding the appropriate 75th percentile benchmark (for that type of patient and anatomic area) will be considered to have an excessive proportion of high dose exams.
- Numerator Statement:
    - 2022 submission: 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.
    - 2016 submission: Radiation Dose metrics among consecutive patients, who have undergone CT of the head, chest, abdomen/pelvis, or chest/abdomen/pelvis. The metrics are 1) mean dose as measured using DLP, CTDIvol, and SSDE: within age strata. And 2) the proportion of exams with doses greater than the 75th percentile of the benchmark you are comparing with for the same anatomic area strata.
- Denominator Statement:
    - 2022 submission: 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.
    - 2016 submission: Consecutive sample of CTs conducted in the head, chest, abdomen/pelvis and chest/abdomen/pelvis. No examinations should be excluded
- Denominator Exclusions:
    - 2022 submission: Examinations with missing anatomic area, patient age, or missing dose length product are excluded.
    - 2016 submission: CT examinations conducted in anatomic areas not included above (such as CTs of the extremities or lumbar spine) or that combine several areas (head and chest) should not be included. In children, these four included categories will reflect approximately 80 percent of CT scans. Examinations performed as part of diagnostic procedures – such as biopsy procedures – should not be included. CT examinations performed as part of surgical planning or radiation therapy should not be included. Examinations that are considered "limited abdomen" or "limited pelvis" studies should be included in the abdomen and pelvis category. Any examinations that include any parts of the abdomen and or pelvis should count in the abdomen/pelvis category.

Reliability Testing:

- Reliability testing conducted at the accountable entity level:
    - The data source for updated testing was the UCSF international CT Dose Registry (2016-2021), representing 23,319 pediatric CT exams per year on average.

- Hospitals may be included in both the (1) anatomic area-age strata calculations and (2) overall facility median dose calculation as long as they meet minimum sample size requirements for at least one of the 15 anatomic area-age strata.
- Developers use sampling with replacement of CT exams within each anatomic area and age group for each hospital with 1,000 repetitions. Within each anatomic area-age group, hospitals are split into 11 subsets based on decile distribution of sample sizes. Agreement and Cohen's Kappa (between the simulated classifications and the "true class") are calculated.
- Agreement consistently exceeds 90 percent and Cohen's Kappa consistently exceed 0.81 for a sample size in the range of 8 to 11 within an anatomic area-age stratum.
- Developers assessed reliability for scoring within anatomic area-age strata but not for the overall performance based on median radiation dose overall.

- Reliability testing was conducted at the patient/encounter level:
  - In their 2016 submission, the developer conducted data element validation of CT radiation dose metrics (DLP and CTDIvol values) using both manual and electronic abstraction and found Kappa values greater than 95 percent, indicating highly reliable data.

SMP Summary:

- Generally, reviewers found the results support reliability.
- One reviewer commented that the testing methods used for measure score reliability (evaluating classification and use of Cohen's Kappa) were better suited for validity testing than reliability testing.
- Another reviewer commented that the bootstrap approach used could over-estimate reliability and offered recommendations about how to overcome this potential limitation.
- One reviewer noted potential typos/errors in the equations with implications for the calculated sample sizes. The developer has clarified this typo in their response.
  - The developer noted that each reporting entity has a continuous score of between 0 and 100 percent that reflects the overall proportion of exams that are above the benchmarks of the 75th percentile; those that have more than twice the expected number of high dose exams (i.e., those with more than 50 percent) are considered excessive. However, the raw percentage on the continuous range from 0 to 100 percent allows for an understanding of where a facility lies on the spectrum. The developer also noted that the measure is not designed to identify outliers and acknowledged this as a point of interest while also noting the challenge of small sample sizes when approaching this level of detail, which would threaten the reliability. The SMP did not re-vote on the measure.

Questions for the Committee regarding reliability:

- *The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity: Validity testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

For maintenance measures – less emphasis if no new testing data provided.

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Validity Testing

- Validity testing was conducted at the Accountable Entity level:
    - Studies are cited by the developer that demonstrate a relationship between specific organizational structures and processes of care that serve as facilitators/barriers to dose optimization using metrics similar, but not identical, to the measure under consideration.
    - The developer presented information on the relationship between radiation dose levels and organizational factors/care processes associated with high quality care as assessed through a process of care survey taken by 90 imaging centers. There were 14 questions listed within the domains of A) Practice systems, B) Practice readiness C) Change strategies and D) Miscellaneous, including the number of protocols and active quality improvement (QI) projects. The survey found that 32-90 percent of respondents strongly agreed with each question and up to 38 percent of high dose exams were reduced.
    - They also led a randomized controlled trial (RCT) examining the impact of education feedback on radiation doses used in CT imaging, to determine the impact of multicomponent educational feedback on radiation doses used in CT imaging. Hospitals and imaging facilities were provided with median and 75th percentile in dose distribution for head and abdomen CT scans which is highly similar, but not identical, to the specifications of the measure under consideration.
        - The study found that there was a 23-58 percent reduction in the proportion of high-dose exams across anatomic areas based on organ dose, with no observed change in image quality.
        - The developer also reported that 6.8 percent of median head doses in children were reduced ($p<0.004$) and there were fewer abdomen doses in children.
        - The developer concluded that organizations have lower and more optimized doses for CT if they follow identified practices to improve quality in imaging care and put systems in place to do so.
        - The developer also concluded that measure-related feedback provided to hospitals contributes to meaningful radiation dose reduction.
- Validity testing conducted at the Patient/Encounter level:
    - Anatomic area: the developer offers data validating assignment of CT anatomic category in adults, which they indicate is a more complicated assignment than in children. They state that they include this analysis in their validity testing because it demonstrates the validity of assignment of CT exams to specific categories.
        - The developer used an algorithm that assigns categories using CPT and ICD-10-CM codes and used criterion validity to compare agreement between assigned category and a gold standard method based on expert review of the complete medical record.
        - Based on 978 CT exams, sensitivity was 0.86 and specificity was 0.96
    - Radiation dose: The developer notes that DLP is a standardized data element generated by more than 99 percent of all CT machines and that it is well-validated. The developer relies on published work and testing within the UCSF International CT Dose Registry.
        - DLP was reported and within the plausible range for 99.6 percent of CT exams.
    - The developer concluded that anatomic area can be assigned with a high degree of accuracy and that DLP, as a measure of radiation dose is standardized, virtually universally available, well-validated, and used broadly to reflect the radiation dose delivered to the patient.

Exclusions

- The measure does not use exclusions.

Risk-Adjustment

- The measure is not risk-adjusted. The developer stratifies the measure by three anatomic areas including head (low), head (routine), and abdomen. The measure is also stratified by five age groups, including less than one year old, one to four years old, five to nine years old, 10-14 years old, and 15-17 years old.
    - The developer does not adjust for clinical indication/protocol, patient size, or) social risk as they state that the radiation doses used for CT in children (as in adults) primarily varies by anatomic area. The developer states that to the extent they have been studied, social factors including race/ethnicity and socioeconomic status are not predictive of radiation dose for CT exams
    - In their 2016 submission, the developer also clarifies that the measure is not adjusted for patient size because the relative change in dose based on great variation in patient size is still very small compared to other factors (such as physician and facility preferences). They also state it is not important to stratify for clinical indication or protocol because anatomic area provides more accurate information of the dose need, rather than specific indication or protocol.

Meaningful Differences

- The developer states that it is clinically meaningful to detect entities (hospitals) whose median DLP is more than 0.5 standard deviations greater than the threshold (the 75th percentile of the UCSF International CT Dose Registry within a combination of anatomic area and age stratum).
    - They found that the majority of hospitals providing pediatric care can be expected to have sufficient sample size in one year (for skull and abdomen and pelvis, minimum 9 CT exams; for brain exams, minimum 25 exams) to detect if median DLP is more than 0.5 standard deviations about the threshold, with 80 percent power.
- The developer states it is also clinically meaningful to assess the proportion of individual exams within a hospital which exceed the 75th percentile benchmarks for their respective anatomic areas and age groups. (entities whose prevalence of “high dose" exams is at least 50 percent, which is twice the expected prevalence of exams above the 75th percentile benchmark), for classification as hospitals with higher than expected prevalence of “high dose” exams.
    - They found the majority of hospitals providing pediatric care can be expected to have sufficient sample size in one year (minimum 23 exams) to detect a prevalence of high dose exams greater than 50 percent, with 80 percent power.

Missing Data

- Among pediatric examinations in the UCSF International CT Dose Registry, age is missing for 0.05 percent; anatomic area is missing for 2 percent; and radiation dose is missing for <1 percent. (In this registry, CT category is reported in a separate data element). The developer states that no statistical tests were conducted to assess the impact of missingness because the rates were so small.

Comparability

- There is only one set of specifications for this measure.

SMP Summary:

- Overall, reviewers found the testing methods acceptable.
- One reviewer noted that patient/encounter validation reported sensitivity and specificity but did not report Kappa statistic (chance-adjusted agreement) and considered this an important limitation.
- Reviewers did not note any concerns related to exclusions or missing data.
- Reviewers accepted developer's justification for not risk-adjusting the measure.
- One reviewer had concerns with how the measure is scored and classifications are interpreted: "The problem with using the median (or 50% above) is that a high proportion of scans (e.g., 20%) could be near lethal doses but the median could be below the 75th percentile. In other words, sites with median >75 percentile almost certainly have room for improvement but very bad sites can be missed."
- The measure was discussed by SMP because of a member's concern about how the measure is scored and the implications for the measure's validity, specifically, the measure's ability to distinguish sites with poor or excessive radiation from those that do not. The measure has a reference distribution of radiation dose, and sites with a median radiation dose greater than the 75th percentile of the reference distribution are considered poor/excessive. The implication of this is that sites with zero to 49 percent of scans below the 75th percentile are considered acceptable, which is a large range, and up to 49 percent of scans could be very high in the reference distribution. In addition, the mean radiation dose in acceptable sites could be higher than those in poor sites. Another question was also raised about whether a poor site with a median in the 76th percentile is fundamentally different from an acceptable site with a median in the 74th percentile.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

2a. Reliability

- [Standing Committee feedback]

2b. Validity

- [Standing Committee feedback]

2b2-2b6. Potential threats to validity

- [Standing Committee feedback]

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in a combination of electronic sources.
- Data come from information that is coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims).
- Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry).
- There are no plans to create an eCQM.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

3. Feasibility

- [Standing Committee feedback]

Criterion 4: Use and Usability

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported?	☒	Yes	☐	No		
Current use in an accountability program?	☒	Yes	☐	No	☐	UNCLEAR
Planned use in an accountability program?	☐	Yes	☐	No	☒	NA

Accountability program details

- This measure is currently used by the Leapfrog Group. The results are publicly reported at https://ratings.leapfroggroup.org as part of their Hospital and Surgery Center Ratings.
- The developer aims to submit this measure as an eCQM that assesses both radiation dose and image quality for the next round of maintenance review.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer notes that Leapfrog has received feedback on the measure via their Help Desk. The main reported issues are using age instead of size and rewarding low doses in scoring without indication of image quality.
    - Leapfrog recently added a series of fact-finding questions to the Survey to help account for image quality – particularly for those reporting very low doses.
- Based on feedback submitted to the Leapfrog Group and UCSF research, the developer updated the specifications to standardize the use of specific phantoms for head and body exams (this had been a source of confusion and variability for reporting entities). This was based on comments submitted to Leapfrog and on UCSF research.
- Also based on feedback, the developer subdivided the previously specified "head" anatomic area into two new anatomic areas: "skull" representing low radiation dose indications that are focused on imaging of the facial and skull bones, sinus, temporal bone, and for assessment of patency of a ventricular shunt; and "brain" representing all other head indications which have the brain as the primary focus rather than the bony skeleton. This decision was primarily based on analysis of data from the UCSF International CT Dose Registry and recent publications from the American College of Radiology Dose Registry demonstrating meaningful differences in radiation doses used for skull versus brain CT scans. However, it was also frequently noted by reporting entities that a single "head" category did not adequately reflect necessary variation by clinical indication.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The developer reports data on changes over time is not yet available but expects to have this data after this year.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- Given the relationship of radiation dose and image noise, there is theoretical concern that dose reduction could result in deteriorated image quality, which would reduce the diagnostic utility of CT images and could harm patients by requiring repeated scanning (thus doubling the dose). While this measure does not directly assess image quality, the developer states that reporting entities have not informed the developer of any negative impact on image quality.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

4a. Use

- [Standing Committee feedback]

4a. Usability

- [Standing Committee feedback]

Criterion 5: Related and Competing Measures

Related measures:

- 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 (for CT Abdomen-pelvis with contrast/single phase scan, CT Chest without contrast/single).

Harmonization

- NQF #3621 compares performance to their own size-specific DRLs (diagnostic reference levels), which is the 75th percentile. (As opposed to ADs, which is the 50th percentile). The benchmarks are given in the Kanal at al paper published in 2017, which is specific to adults. These benchmarks were developed using exams in adults:
  - The smallest head diameter given is 12-14 cm.
  - Smallest chest is 21-25 cm.
  - Smallest abdomen is 21-25 cm.
- The developer does not see how they would apply these benchmarks to a baby or young child.
- The developer also notes that NQF #3621 has one category of "head and brain" while #2820 separated this into (1) skull and (2) brain as the doses vary between these categories, which reinforces that the measures are related but not competing.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- [Standing Committee feedback]

Public and NQF Member Comments (Submitted as of Month Day, Year)

Member Expression of Support

- Of the X NQF members who have submitted a expression of support, X expressed "support" and X expressed "do not support" for the measure.

Comments

[Insert MIMS pre-evaluation comments export]

Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

1. **Have measure specifications changed since the last review? ☒ Yes ☐ No**
2. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☐ No**
3. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**
   - **Reviewer 6:** None.
   - **Reviewer 7:** No concerns.
   - **Reviewer 9:** No concerns.
   - **Reviewer 10:** None

RELIABILITY: TESTING

4. **Did the developer conduct new reliability testing? ☒ Yes ☐ No**

   4a. If no, summarize the Standing Committee's previous feedback:

   - N/A

   4b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:

   - In the 2016 measure submission, the developer conducted data element validation of DLP values using both manual and electronic abstraction with Kappa values >95 percent.
     - Data source for testing: UCSF international CT Dose Registry (2012-2014) representing 115,000 pediatric CT exams per year on average.
     - Reliability testing was done at the level of data elements using several metrics reflecting CT dose indices, including DLP, CTDIvol, and SSDE.
     - DLP and CTDI are calculated automatically by all current CT scanners, without variability. Reliability of CT radiation dose metric abstraction (DLP and CTDIvol) was tested through both manual and automated data abstraction, both yielding identical results, perfect Kappa statistics.
     - SSDE is a calculated variable that is automatically calculated by dose monitoring programs. Errors from manual calculation were not tested.
     - The developer noted nearly 99 percent of facilities should be able to report on this measure automatically, since any scanner built in the last 10 years reports on the data needed.
     - The Kappas for the reliability testing were high (greater than 95%), but on a limited number of sites.
     - Empirical testing was performed at the performance measure score. The developer indicated a study was conducted comparing each of the dose metrics with measures of absorbed dose among a sample of 10,000 CT examinations showed a "high correlation," >90 percent.

- The 2022 submission reliability testing was conducted at the accountable entity level:
    - The data source for the updated testing was the UCSF international CT Dose Registry (2016-2021), representing 23,319 pediatric CT exams per year on average.
    - The developers used sampling with replacement of CT exams within each anatomic area and age group for each hospital with 1,000 repetitions. Within each anatomic area-age group, hospitals were split into 11 subsets based on decile distribution of sample sizes. Agreement and Cohen's Kappa (between the simulated classifications and the "true class") were calculated.
    - Agreement consistently exceeded 90% and Cohen's Kappa consistently exceed 0.81 for a sample size in the range of 8 to 11 within an anatomic area-age stratum.
    - Developers assessed reliability for scoring within anatomic area-age strata but not for the overall performance based on median radiation dose overall.
- During the 2016 measure evaluation, the Standing Committee had questions about the specifications and the process of collecting the data for this measure, all of which were adequately addressed by the developer. The developer explained that consecutive exams should be used, and that the measure does not include certain procedures (such as radiological oncology). The developer also noted that while there is variability in dose depending on clinical indications, this variability is dwarfed by the variability resulting from institutional preference. (For example, for some clinical questions, one facility will use a single-phase setting while another will use a multiple-phase setting, which results in twice as much radiation exposure.) In addition, the developer noted that this measure only requires that a facility meet the average benchmark, not that every patient be at or below the benchmark.

5. **Reliability testing level: ☒ Accountable-Entity Level ☒ Patient/Encounter Level ☒ Neither**
6. **Reliability testing was conducted with the data source and level of analysis indicated for this measure:**

   ☒ Yes ☐ No
7. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ Yes ☐ No
8. **Assess the method(s) used for reliability testing:**

- **Reviewer 1:** To assess the relationship between the sample size of a hospital and its reliability, developer split the hospitals, within anatomic area and age group, in the UCSF International Dose Registry into 11 subsets, with boundaries defined by the deciles of the distribution of sample sizes. Within each subset, two metrics were calculated:
- The Classification Rate – This is defined within a hospital subset as the prevalence of agreement between a randomly-selected hospital within the subset and its 1000 simulated versions.
- The Cohen's Kappa – This is defined for a single collection of simulated hospitals (out of a total of 1000 collections) as the ratio (p0-pe)/(1-pe), where p0 is the prevalence of agreement between the simulated hospitals and their "true" counterparts, and pe is the hypothetical prevalence of agreement by random chance. This is defined within a hospital subset as the expected Cohen's Kappa between a randomly-selected collection of simulated hospitals and their "true" counterparts.
- Lastly, the recommended sample size associated with a classification rate of at least 90% and a Cohen's Kappa of at least 0.81, both traits associated with "near-perfect" agreement between repeated simulations.
- Reliability testing was performed for scoring approach 1 (classifying a facility's performance as "acceptable" or "poor" based on whether its median radiation dose is below, or greater than or equal to, the benchmark 75th percentile per anatomic area and age strata).

- **Reviewer 2:** reliability testing was performed for scoring approach 1 (classifying a facility's performance as "acceptable" or "poor" based on whether its median radiation dose is below, or greater than or equal to, the benchmark 75th percentile per anatomic area and age strata) using bootstrap methods. Scoring approach 2 was not tested but I'm unclear how the scoring approaches differ.
- **Reviewer 3:** Data element: audit of 1000 records compared to reported data Entity: Classification rate agreement from 1000 bootstrapped samples. Cohen's Kappa for 1000 bootstrapped samples. Analysis of minimum sample sizes needed to achieve acceptable level of performance on these metrics.
- **Reviewer 5:** acceptable
- **Reviewer 6:** Score-level reliability was tested using a bootstrap approach and then assessed by estimating (1) the classification rate – prevalence of agreement between randomly selected hospital with the subset and its 10000 simulated versions; and (2) Cohen's Kappa – defined as the prevalence of agreement between the simulated hospitals and their "true" compartments.
- **Reviewer 7:** Simulated hospitals used. Used structured fields in EMRs to collect data elements.
- **Reviewer 9:** No concerns
- **Reviewer 10:** The developer assessed the reliability of hospital classification mechanism using the UCSF International Dose Registry. Given a combination of hospital, anatomic area (skull, brain, abdomen), and age group, the developer recorded whether this hospital would be deemed poor (hospital median DLP greater than the population 75th percentile threshold) or acceptable (hospital median DLP less than or equal to the population 75th percentile threshold) within this anatomic age and age group. The developer considered this the "true class" of this hospital, within this anatomic area and age group. The developers looked at classification rate and Cohen's Kappa to assess reliability This appears to more of a validity assessment than reliability.
- **Reviewer 11:** Test retest using simulation x 1000 of sampled data compared with the reported value

9. **Assess the results of reliability testing**
   - **Reviewer 1:** On average, classification rates remained high, with even low-volume hospital-stratum combinations having a mean classification rate ~90%. However, the mean classification rate is unstable for low sample sizes, and does not consistently exceed 90% until a sample size to 8 to 11 is observed. Similarly, the Cohen's Kappa consistently exceeds 0.81 once the observed sample size reaches 8 to 11.
   - **Reviewer 2:** Good stability except at very low volume.
   - **Reviewer 3:** Target rates for Classification rate analysis of 90%, Cohen Kappa of 0.81 used to examine reliability for different sample sizes, by strata defined by site and age. Minimum sample sizes by strata identified.
   - **Reviewer 5:** acceptable
   - **Reviewer 6:** Cohen's Kappa ranged between 0.84 an 0.89 depending on sample size.
   - **Reviewer 7:** For those hospitals that have a minimum sample size for each strata, it seems acceptable.
   - **Reviewer 9:** No concerns
   - **Reviewer 10:** Unconvinced that evaluating classification and use of cohen's kappa is appropriate for reliability testing.
   - **Reviewer 11:** kappa 0.8-0.9
10. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? **NOTE:** If multiple methods used, at least one must be appropriate.

    ☒ **Yes** ☒ **No** ☐ **Not applicable**
11. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

☒ **Yes** ☐ **No** ☒ **Not applicable** (patient/encounter level testing was not performed)

12. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☒ **High** (NOTE: Can be HIGH only if accountable-entity level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☒ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

13. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

- **Reviewer 1:** Reliability testing results strong as noted above.
- **Reviewer 2:** Good stability of classification
- **Reviewer 3:** Data element: Low missing data and high concordance between chart review and reported documentation. Entity: High standards chosen for Classification Rate and Cohen's Kappa. Minimum sample size selected to achieve these rates.
- **Reviewer 4:** The developers used bootstrap re-sampling to estimate the probability that a hospital would receive the same classification result (acceptable versus poor) if the measure were to be re-estimated in a different random sample of N patients from the same underlying DLP distribution. A possible issue with the bootstrap procedure is that it can over-estimate reliability when applied to hospitals with very few eligible patients. For example, if a hospital has 2 eligible patients, say patients A and B, and both patients received radiation doses below the benchmark, then every possible bootstrap sample of size 2 drawn from this sample of patients (A+A, A+B, or B+B) will consist of 2 patients who both have radiation doses below the benchmark, and the hospital classification will appear to have perfect consistency i.e., the hospital will always be classified as acceptable. A similar result would occur if both patients at a hospital received unacceptable dosage levels. In that case, the hospital will always be classified as poor. In reality, the probability of receiving a dose level below the benchmark at this hospital is not 0% or 100% but something in between. In general, when hospital-specific probabilities are estimated by their corresponding observed proportions and many of the hospital-specific sample sizes are small then the amount of true signal variation between hospitals will tend to be exaggerated. In light of this, it seems possible that some of the reliability results in Figure 2a.11 and table 2a.11 could be over-estimated. An alternative approach to estimating reliability would be to assume that the proportion of patients receiving a dose below the benchmark arises from a binomial distribution with a probability parameter that is specific to each hospital. One could first use a hierarchical model (e.g., beta-binomial) to estimate the underlying distribution of true hospital-specific probabilities and then use simulations to determine what the level of consistency would be if a population of hospitals and a sample of patients within each hospital were randomly generated from this model.
- I was able to reproduce the same size calculation on page 48 for brain exams (n=25) but not for skull/abdomen/pelvis (n=9). I presume there's a typo in the last line of the equations on page 48 and that 0.3 should be 0.5. After making that change, I get n=25 for brain and n=15 for skull/abdomen/pelvis.
- After making a slight change to the formula on page 49, I was able to reproduce the sample size calculation for detecting a 2-fold increase in the probability of receiving a dosage above the benchmark with 80% power. I think the calculated sample size (n=23) may be based on a one-sided alpha=0.05 not alpha=0.025. And I think either choice of alpha is acceptable.
- **Reviewer 6:** Cohen’s Kappa ranged between 0.84 and 0.89 depending on sample size.

- **Reviewer 10:** See above-- comparison with gold standard is a validity, not reliability evaluation. Reliability not assessed.
- **Reviewer 11:** Data on dose is generated by the equipment. Reliability of the score for the number of tests required to qualify for the metric is high

VALIDITY: TESTING

14. **Did the developer conduct new reliability testing? ☒ Yes ☐ No**

14a. If no, summarize the Standing Committee's previous feedback:

- N/A

14b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:

- In the 2016 measure submission, the developer conducted data element validation of DLP values using both manual and electronic abstraction with Kappa values >95%.
- In the 2022 measure submission, validity testing was conducted at the Accountable Entity level:
    - Developer assessed the relationship between radiation dose levels and organizational factors/care processes associated with high quality care. They also led an RCT examining the impact of education feedback on radiation doses used in CT imaging.
    - Studies are cited by the developer which demonstrate a relationship between specific organizational structures and processes of care that serve as facilitators/barriers to dose optimization using metrics similar, but not identical, to the measure under consideration.
- In the 2022 measure submission, validity testing was also conducted at the Patient/Encounter level:
    - The developer offers data validating assignment of CT anatomic category in adults, which they indicate is a more complicated assignment than in children. The developer compares an algorithm that assigns categories using CPT and ICD-10-CM codes against review of the complete medical record.
    - Based on 978 CT exams, sensitivity was 0.86 and specificity was 0.96
    - Radiation dose (dose length product - DLP). The developer notes that DLP is a standardized data element and well-validated, relying on published work and testing within the UCSF International CT Dose Registry.
    - DLP was reported within the plausible range for 99.6% of exams.

15. **Validity testing level (check all that apply):**

☒ Accountable-Entity Level ☒ Patient or Encounter-Level ☐ Both

NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

16. **If patient/encounter level validity testing was provided, was the method described and appropriate for assessing the accuracy of ALL critical data elements? NOTE:** Data element validation from the literature is acceptable.

☒ **Yes**

☒ **No**

☒ **Not applicable** (patient/encounter level testing was not performed)

17. **Method of establishing validity at the accountable-entity level:**

☒ **Face validity**

☒ **Empirical validity testing at the accountable-entity level**

☒ **N/A (accountable-entity level testing not conducted)**

18. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    ☒ **Yes**

    ☐ **No**

    ☒ **Not applicable** (accountable-entity level testing was not performed)

19. **Assess the method(s) for establishing validity**

- **Reviewer 1:** Patient/encounter-level (data element) validity: Anatomic area: In developing the framework to assign CT exams to CT categories, developer validated the CT categories as being representative of groupings that require different radiation dose and image quality in part by demonstrating significantly greater differences in radiation doses between categories than within categories within the body regions that are subdivided into low, routine, and high dose (N = 4.5 million exams). Developer validated the method of assigning CT category based on CPT® and ICD-10-CM codes against gold standard chart review.
- Empirical validity testing: the developers have not assessed the relationship of this measure score and other related outcome measures because they indicate there are no validated outcome measures related to radiation dose or medical imaging in children.
- **Reviewer 2:** Good methods for both scan and entity-level empirical testing.
- **reviewer 3:** Test of sensitivity and specificity for identification of appropriate cases.
- **Reviewer 5:** acceptable
- **Reviewer 6:** The MD assessed the association between facility score and quality improvement efforts. This is not an ideal approach. However, the alternative, which is to assess the association between dose and outcomes (cancer) is not feasible.
- **Reviewer 7:** Appropriate
- **Reviewer 9:** No concerns
- **Reviewer 10:** The developer used criterion validity for anatomic area to compare agreement between the CT category assigned using this method versus a gold standard method based on expert review of the complete medical record (including notes from the visit when the exam was ordered, information provided as free text with the test order, and information included in the final, dictated radiology report) for a sample of 10,000 CT exams from UCSF Health System. The developer validated the radiation dose data element, by relying on published work and tested the availability of clinically plausible and non-missing values for radiation dose in the UCSF International CT Dose Registry.
- **Reviewer 11:** sites interested in improving patient safety generally agreed that measuring radiation does was important and those sites interested in improvement actually demonstrated improvement in the score

20. **Assess the results(s) for establishing validity**

- **Reviewer 1:** The results of data element validity, weighted by the distribution of CT categories in the UCSF International CT Dose Registry, were: sensitivity = 0.86 and specificity = 0.96 (n=978 CT exams). Anatomic area can be assigned with a high degree of accuracy. Dose length product (DLP), as a measure of radiation dose is standardized, virtually universally available, and well validated and used broadly to reflect the radiation dose delivered to the patient.
- Empirical validity testing: Process of care surveys were received from 90 imaging centers (90%), and 182,415 abdominal CT scans were collected during the study period. The association between process factors and dose found that after adjusting for patient age, gender, and size, quality improvement strategies were association with mean abdominal CT radiation dose and the odds of a high-dose (> 75th percentile in dose) examination. Completed univariate analyses identified strategies and systems

that were significantly associated with lower average doses or lower frequency of high doses for abdominal CT examinations. Organizations that follow identified practices to improve quality in imaging care, and put systems in place to do so, have lower and more optimized doses for CT.

- **Reviewer 2:** The empirical validity testing had a strong methods (linking processes and context factors to the odds of high dose scan) and good results.
- **Reviewer 3:** Case selection: Specificity high. Sensitivity acceptable. There is reliance on data on dose as provided by machines.
- **Reviewer 5:** acceptable
- **Reviewer 6:** Validity testing at the encounter level is acceptable. Hence, results of score-level empirical validity testing are not necessary.
- **Reviewer 7:** Conclusions about quality can be made if enough sample size, which seems to be a non-issue for most hospitals.
- **Reviewer 9:** No concerns
- **Reviewer 10:** Developer validated method of assigning CT category based on CPT® and ICD-10-CM codes against gold standard chart review. The developer did not look at chance-adjusted (Kappa) agreement but rather sensitivity = 0.86 and specificity = 0.96 (n=978 CT exams). This is not sufficient for validity assessment-- simple agreement and SN/SP is helpful but not the full story-- need to understand chance-adjusted agreement. Radiation Dose: In the UCSF Registry, in children, DLP was reported and within plausible range for 99.6% of CT examinations.
- **Reviewer 11:** Because the radiation dose is generated by the equipment no testing of this data element was performed. However testing of validity of assignment to appropriate classification was tested against direct data abstraction and found to high level of accuracy. As noted survey of centers demonstrated high degree of agreement of the importance of measuring radiation dose and centers with active program to improve radiation dose demonstrated improvement in the score.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

21. **Please describe any concerns you have with measure exclusions.**
    - **Reviewer 3:** No concerns
    - **Reviewer 6:** none
    - **Reviewer 7:** Missing data exclusions.
    - **Reviewer 9:** No concerns
    - **Reviewer 10:** None.
    - **Reviewer 11:** Sponsors have trimmed the radiation parameters reported as well as the categories ad provided adequate rationale for doing so
22. **Risk Adjustment**

    22a. **Risk-adjustment method**

    ☒ None (only answer Question 20b and 20e) ☐ Statistical model ☒ Stratification

    ☐ Other method assessing risk factors (please specify)

    22b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

    ☒ Yes ☐ No ☒ Not applicable

    22c. **Social risk adjustment:**

    22c.1 Are social risk factors included in risk model? ☐ Yes ☒ No ☒ Not applicable

    22c.2 Conceptual rationale for social risk factors included? ☒ Yes ☐ No

22c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☒ No

22d.**Risk adjustment summary:**

22d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

22d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

22d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

22d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☐ No

22d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No

22e**. Assess the risk-adjustment approach**

- **Reviewer 3:** Justification provided for not risk adjusting. Good discussion of concerns about higher doses due to old machines at hospitals serving high socially disadvantaged populations, and judgment not necessary to consider in design.
- **Reviewer 6:** Yes, there is justification for no risk adjustment.
- **Reviewer 7:** Stratification based upon current standards of radiology.
- **Reviewer 9:** No concerns
- **Reviewer 10:** acceptable for intermediate-outcome
- **Reviewer 11:** Yes--radiation protocols are institutionally determined and not generally varied by patient factors other than age and possibly size

23. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**
    - **Reviewer 3:** None
    - **Reviewer 6:** none
    - **Reviewer 7:** No concerns.
    - **Reviewer 9:** No concerns
    - **Reviewer 10:** None
24. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    - **Reviewer 3:** Multiple manufacturers. Theoretical concern about differences in calibration of dose reporting across manufacturers.
    - **Reviewer 6:** NA
    - **Reviewer 7:** N/A
    - **Reviewer 10:** None
25. **Please describe any concerns you have regarding missing data.**
    - **Reviewer 3:** None. Missing data reported as low.
    - **Reviewer 6:** none
    - **Reviewer 7:** No concerns.
    - **Reviewer 9:** No concerns
    - **Reviewer 10:** None.
    - **Reviewer 11:** missing data are likely to be minimal as data on the numerator are generated from the equipment with which the test is performed, number of studies is determined from CPT or ICD billing codes and standards of comparison are published and available

For cost/resource use measures ONLY:

If not cost/resource use measure, please skip to question 28.

26. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

27. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

28. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☒ **High** (NOTE: Can be HIGH only if accountable-entity level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity level testing has NOT been conducted)

    ☒ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

    ☒ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)

29. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

- **Reviewer 1:** See interpretation of results section above.
- **Reviewer 2:** The measure is calculated and transformed in two ways (e.g., proportion above 75 percentile (if above 50% then considered excessive), median above 75 percentile (if so, considered poor). (Mathematically, I'm unclear how these are different). The problem with using the median (or 50% above) is that a high proportion of scans (e.g., 20%) could be near lethal doses but the median could be below the 75th percentile. In other words, sites with median >75 percentile almost certainly have room for improvement but very bad sites can be missed. This seems like a very serious problem for the way the measure is scored, and how classifications are interpreted.
- **Reviewer 3:** Measure needs vetting of experts on Standing Committee but construction appears well documented and appropriate.
- **Reviewer 4:** The developers provided a detailed discussion of the considerations for deciding not to adjust for patient weight and other case mix factors. The developers provided a compelling argument that any residual bias from a lack of weight adjustment is small in relation to the amount of true signal variation.
- This measure focuses on radiation but I assume that an overall assessment of CCTA quality would also need to account for image quality. Are there other existing measures that do this? If not, is there any risk of unintended consequences from measuring one but not the other?
- **Reviewer 10:** Kappa (chance-adjusted) agreement not provided and not all data elements empirically tested.
- **Reviewer 11:** Although it is the intention of the sponsors that hospitals report all of their data, there may not be any way of knowing if they are reporting only a selection from their data

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

30. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

31. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**
    - N/A

ADDITIONAL RECOMMENDATIONS

32. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**
    - None

Developer Submission

Criteria 1: Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria

1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

[Response Begins]
Yes
[Yes Please Explain]
Two systematic reviews and meta-analyses were published in recent years providing evidence of increased excess cancer risk from low-dose ionizing radiation in the radiation dose ranges typically used in CT imaging. (Hauptmann 2020, Abalo 2021) One of these reviews (Abalo 2021) focused specifically on children exposed to radiation from medical imaging and found – for a CT exam delivering 10 mGy to the red bone marrow (the average bone marrow exposure from one CT in a child) – the risk of leukemia increases by about 27% over the background risk. For a CT exam delivering 10 mGy to the brain, the risk of brain tumor increases by about 9% over the background risk.
The following evidence section has been updated with these reviews other newer references to peer-reviewed research, contributing to a large body of epidemiologic and biologic evidence of that exposure to ionizing radiation from medical imaging increases a person's risk of developing cancer.

[Response Ends]

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:
2021 Submission:
Updated evidence information here.
2018 Submission:
Evidence from the previous submission here.

1a.01. Provide a logic model.

Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

[Response Begins]
There is substantial variation in the radiation doses used for CT exams (Kanal 2017, Kanal 2021, Smith-Bindman 2009) which is primarily due to differences in how radiologists choose to perform them – in other words, their choice of a specific imaging protocol (for example, a single or multiple phase CT) and the specific technical parameters used such as scan length, milliampere-seconds, and kilovoltage peak. (Smith-Bindman 2019) More than patient or CT machine characteristics, this subjective protocol selection is the single greatest predictor of radiation dose. (Smith-Bindman 2019) However, there are few benchmarks currently available to guide practice. In practice, patients are often assigned to a protocol that uses a higher radiation dose than the underlying indication warrants: for example multiphase scans are frequently used, whereas they are rarely needed. The proposed measure directly assesses radiation dose among children in different age groups, and this is the first step in the process towards using the lowest dose possible. In this framework of assessing the doses used in groups of patients who underwent the three types of scans included in the measure, the

measure assesses both the earlier step of protocol selection and the later step of radiation dose setting given the protocol selected.
There is also substantial evidence (discussed later in this application) that radiation doses used for CT are carcinogenic, and that the risk of cancer is directly proportional to the doses used. Therefore, risks would be directly reduced by reducing doses. However, it is not feasible to identify the incidence of cancer associated with the physician's imaging decisions and resultant patient doses because of the potentially long lag between exposure and cancer onset. As highlighted in this application, cancer risks continue to be elevated for over 50 years after exposure. However, the cancer risk will be directly related to the radiation dose used, which is known at the time of the exam. Thus, the radiation dose for each CT exam is an intermediate outcome that can be used as a surrogate for (future) cancer risk.

[Response Ends]

1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.

A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data.
[Response Begins]
Other systematic review and grading of the body of evidence (e.g., Cochrane Collaboration, AHRQ Evidence Practice Center)
[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)
Group 1 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]
Early life ionizing radiation exposure and cancer risks: systematic review and meta-analysis.
Abalo KD, Rage E, Leuraud K, Richardson DB, Le Pointe HD, Laurier D, Bernier MO.
Pediatr Radiol. 2021 Jan;51(1):45-56. doi: 10.1007/s00247-020-04803-0.
https://link.springer.com/article/10.1007/s00247-020-04803-0
[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]
"CT exposure in childhood appears to be associated with increased risk of cancer (leukemia and brain tumors) while no significant association was observed with diagnostic radiographs." Although the benefits of diagnostic radiation examinations may outweigh the risks associated with the doses delivered by these procedures (benefits were not evaluated in the studied patients), the results of this analysis justify continued efforts to optimize doses to patients.
[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]
Newcastle-Ottawa Scale (NOS) for studies of radiation exposure in children = 7 to 9

The NOS assesses the quality of non-randomized studies, using 8 items grouped into 3 domains (I.e., selection, comparability/confounding, and outcome/exposure assessment), with 9 being the best possible score. NOS scores of 6 to 9 equate with "good quality" in the Agency for Healthcare Research and Quality (AHRQ) standards for observational studies. Good quality is the highest possible rating on the AHRQ scale.

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]
The DerSimonian and Laird random-effect model was used to estimate the overall effect size to account for within- and between-study heterogeneities. The authors reported moderate heterogeneity (I^2 = 60%, p=0.03) among 6 studies of the risk of leukemia following childhood CT exposures, but no substantial alteration of the aggregate excess relative risk (ERR) with exclusion of individual studies from the meta-analysis (with one exception, where exclusion of a Dutch study led to a higher pooled ERR). There was small heterogeneity (I^2 = 32%) among 5 studies reporting on the risk of brain tumors following childhood CT exposures.

Publication and selection bias were assessed and tested using the Egger test. Some evidence of publication bias was reported (p=0.03) in the leukemia analysis, suggesting that studies of small size with negative results were less often published, but this seemed "not to be a major limitation of our analysis as demonstrated by statistical tests." There was no evidence of publication or selection bias in the brain cancer analysis (p=0.16).

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]
N/A – there is no direct recommendation
[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]
N/A
[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]
21 observational studies, including 11 case-control studies and 10 cohort studies, were included in the systematic review. All studies were assessed to be of good quality, with NOS scores ranging from 7 to 9. (Additional included studies looked at prenatal exposure, but the findings discussed below relate only to childhood exposure).
[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]
This review assesses only the risk associated with radiation exposure from medical imaging, not the benefit.
[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

The authors report pooled excessive relative risk (ERR) per unit (Gray, Gy) of exposure for leukemia and brain tumors. ERR is the most commonly reported measure in this domain. Overall, the pooled analysis included over 11 million subjects including 437 cases of leukemia and 478 brain tumor cases. The authors observed a significant increased risk for leukemia (ERR_{pooled}=26.9 Gy^{-1}, 95% CI: 2.7–57.1), which represents an increase of 2.69% per mGy of dose over the background risk of leukemia. The pooled ERR for brain tumors was also significantly increased (ERR_{pooled}=9.1 Gy^{-1}, 95% CI: 5.2–13.1), which represents an increase of 0.91% per mGy of dose over the background risk of brain tumors. In other words, for a CT exam delivering 10 mGy to the red bone marrow, the risk of leukemia increases by about 27% over the background risk, holding all other factors constant. In 2017, this was the average bone marrow exposure from one CT in a child, and just slightly above the average bone marrow dose for an abdomen CT in an adult. For a CT exam delivering 10 mGy to the brain, the risk of brain tumor increases by about 9% over the background risk, holding all other factors constant.
[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]
N/A – the systematic review is from 2021.
[Response Ends]

Group 2 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]
1a.03) Provide the title, author, date, citation (including page number) and URL for the systematic review.
Epidemiological Studies of Low-Dose Ionizing Radiation and Cancer: Summary Bias Assessment and Meta-Analysis.
Michael Hauptmann, Robert D. Daniels, Elisabeth Cardis, Harry M. Cullings, Gerald Kendall, Dominique Laurier, Martha S. Linet, Mark P. Little, Jay H. Lubin, Dale L. Preston, David B. Richardson, Daniel O. Stram, Isabelle Thierry-Chef, Mary K. Schubauer-Berigan, Ethel S. Gilbert, Amy Berrington de Gonzalez
J Natl Cancer Inst Monogr (2020) 2020(56): lgaa010
https://academic.oup.com/jncimono/article/2020/56/188/5869934vv

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]
This systematic review and meta-analysis concludes that “new epidemiological studies directly support excess cancer risks from low-dose ionizing radiation,” in the radiation dose range used in CT imaging. “Furthermore, the magnitude of the cancer risks from these low-dose radiation exposures was statistically compatible with the radiation dose-related cancer risks of the atomic bomb survivors.”
[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]
The included studies were assessed for bias in the following ways:
1. To identify bias in dose estimates, the authors “assessed the strengths and weaknesses of dosimetry systems with respect to the directness, complexity, and completeness of the dosimetry, the dosimetric uncertainty, and the validity of dose estimates.”
2. In assessing the evidence for confounding and selection bias, they “summarized methods to control confounding and assessed the likelihood of uncontrolled confounding as well as its direction.”

3. They "reviewed the possible impact of differential outcome ascertainment across radiation dose levels, and considered loss to follow-up, under- or over ascertainment of cancer outcomes, misclassification of outcomes, and changing classifications over time."
4. They then "performed a summary of the assessments of different biases for each study and considered both the direction of the observed effect and the direction of the bias."
Of 26 eligible studies, 3 had known or suspected bias in dose estimates that could bias the risk estimate away from the null, and 1 study was likely biased toward the null. Various sources of confounding and selection bias were identified, but the authors could not "draw a definitive conclusion on the impact of bias adjustment with the available data." Four studies "may have had cancer ascertainment possibly differential by radiation exposure"; three of these were likely biased away from the null, and one was likely biased toward the null.

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]
In performing the meta-analysis of excess relative risk (ERR), they tested for homogeneity and variance due to heterogeneity (by computing Cochran's Q and the I^2 statistic, respectively.) Heterogeneity was very low for all analyses after excluding one study that contributed significant heterogeneity.
[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]
N/A – there is no direct recommendation
[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]
N/A
[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]
There were 26 eligible human studies on low-dose radiation exposure and cancer risk. Of 22 studies on solid cancer risk, 4 positive studies with potential positive bias were excluded. Of 25 studies on leukemia risk, 5 positive studies with potential positive bias were excluded. Following these exclusions, the authors were able to exclude bias as the cause of the positive associations between low-dose ionizing radiation and elevated cancer risk.
[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]
This review assesses only the risk associated with radiation exposure from medical imaging, not the benefit.
[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

For solid cancers, after excluding 4 positive studies with potential positive bias, 12 of 18 studies reported positive excess relative risks (ERR) per unit of dose. For leukemia, 17 of 20 studies were positive. For both meta-analyses, the authors rejected the null hypothesis that the median ERR per unit of radiation dose equals zero. For adulthood exposure, the meta-ERR at 100 mGy was 0.029 (95% CI = 0.011 to 0.047) for solid cancers and 0.16 (95% CI = 0.07 to 0.25) for leukemia. For childhood exposure, the meta-ERR at 100 mGy for leukemia was 2.84 (95% CI = 0.37 to 5.32). The authors concluded that the majority of studies reported positive risk estimates and that these data directly support excess cancer risks from low-dose ionizing radiation.
[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]
This systematic review was published in 2020; the developers are not aware of any newer studies that have changed the conclusion from this systematic review.
[Response Ends]

1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

[Response Begins]
[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.

[Response Begins]
There is extensive epidemiological and biological evidence that suggests exposure to radiation in the same range as that routinely delivered by CT (10-100 milli-Sieverts, mSv) increases a person's risk of developing cancer (Board of Radiation Effects 2006, Pearce 2012, Pierce 2000, Preston 2007, Brenner 2003, Hong 2019). **It was estimated in 2009 that 2% of cancers diagnosed annually are the result of CT**; in 2019 that would amount to 36,000 cancers diagnosed each year due to the use of CT. (Berrington de Gonzalez 2009, NCI Cancer Statistics).
The relationship between exposure to radiation and cancer has been shown across a large epidemiological literature, including numerous case control studies, cohort studies including the follow up of individuals exposed to radiation from the atomic bombs, and in recent years, cohort studies showing a direct association between CT imaging and cancer risk. For example, Pearce showed that **among 178,604 children exposed to CT radiation between 1985-2002 and followed through 2008, bone marrow and brain organ doses in the range of 30-50 mGy tripled the risk of leukemia and brain cancer within 10 years.** (Pearce 2012) Far from uncommon, these absorbed radiation doses are frequently delivered by CT imaging. (Miglioretti 2013, Stewart 2021) In the longest follow-up study of survivors of the Hiroshima and Nagasaki atomic bombings (where the median dose to survivors was 40 mSv, in the same range as a single CT exam), the survivors remain at significantly elevated risk for every cancer type through all years of follow up. (Sadakane 2019, Brenner 2020, Sakata 2019, Sugiyama 2020) Overall, more than 10% of cancers in this population are attributed to the radiation exposure.
There have been several systematic reviews, summarized above, assessing the relationship between diagnostic medical radiation exposure and cancer. Abalo et al. (2021) performed a literature search of five electronic databases covering publications from 2000 to 2019 on the relationship between medical radiation exposure in children up to age 21 and cancer. Pooled excess relative risk (ERR) was reported, representing the excess of leukemia and brain tumor risk per unit (Gray, Gy) of organ dose – this metric reflects the proportional increase in risk over the background rate of cancer (in the absence of exposure), per unit of dose. The authors observed a significantly increased risk for leukemia (ERR_{pooled}=26.9 Gy−1, 95% CI: 2.7–57.1), which represents an increase of 2.69% per mGy of dose over the background risk of leukemia. The pooled ERR for brain tumors was also significantly increased (ERR_{pooled}=9.1 Gy−1, 95% CI: 5.2–13.1), which represents an increase of 0.91% per mGy of dose over the background risk of brain tumors.
Dr. Amy Berrington De Gonzalez, Chief of Radiation Epidemiology at the National Cancer Institute, was the senior author of a second systematic review and meta-analysis of studies evaluating the association between radiation exposure and cancer. (Hauptmann 2020) The authors identified 26 studies which: 1) reported a mean dose of less than 100 mGy (corresponding to exposures used in medical imaging); 2) individualized dose estimates, risk estimates, and confidence intervals (CI) for the dose-response relationship; and 3) were published between 2006-2017. They systematically assessed

the potential for bias from each primary study and performed a meta-analysis to quantify the ERR and to assess consistency across studies for all solid cancers and leukemia. For adulthood exposure, the meta-ERR at 100 mGy was 0.029 (95% CI: 0.011 to 0.047) for solid cancers and 0.16 (95% CI: 0.07 to 0.25) for leukemia. For childhood exposure, the meta-ERR at 100 mGy for leukemia was 2.84 (95% CI: 0.37 to 5.32). The authors concluded that **the majority of studies reported positive risk estimates and that these data directly support excess cancer risks from low-dose ionizing radiation.** Furthermore, the magnitude of the cancer risks from these low-dose radiation exposures was statistically compatible with the radiation dose-related cancer risks of atomic bomb survivors.

A number of cohort studies are being conducted as part of the EPI-CT study: a European pooled epidemiological study to quantify the risk of radiation-induced cancer from pediatric CT (Bernier, 2019). The full results are forthcoming, but 4 contributing country-specific portions of the cohort have been published and show positive associations between CT and cancer incidence (Table 1a-1):

(1) **The British study reported a positive dose-response relationship between radiation dose and leukemia and CNS tumors in children and young adults.** (Pearce 2012, Berrington 2016)

(2) **The German study reported a significantly increased incidence of all cancer and lymphoma in exposed children** compared with the general population. (Krille 2015)

(3) **The French and the German cohorts reported a dose-related increase for CNS tumors**. (Journy 2015, Journy 2016, Krille 2015)

(4) **The Dutch study reported a dose-response relationship for CNS tumors.** (Meulepas 2016, Meulepas 2019)

Table 1a-1. Results from EPI CT National Cohort (Bernier 2019).

Outcome by country	Cases	Risk estimates	Risk (IC 95%)	*
CNS tumor risk according to the brain dose	*	*	*	*
UK[a] (Pearce *et al.,* 2012)	135[b]	ERR per mGy	0.023	(0.0I0, 0.049)
UK[a] (Berrington *et al.,* 2016)	122[b] without PF	ERR per mGy	0.019	(0.008, 0.043)
France (Journy *et al.,* 2015)	22	ERR per mGy	0.022	(-0.016, 0.06 1)
The Netherlands (Meulcpas *et al.,* 2018)	84	ERR per mGy	0.0086	(0.0020, 0.022)
Germany (Krille *er al.,*2015)	7	HR per mGy	1.008	(1.00, 1.01)
France (Journy *et al.,*2016)	15 without PF	HR per I0 mGy	1.07	(0.99, 1.10)
*	7 with PF	HR per 10 mGy	0.8	(04 5, 1.06)
UK[a] (Pearce *et al.,*2012)	135 [b]	RR[50-74 mGy] vs <*5* mGy	2.82	(1.34, 6.03)
Leukemia risk according to RBM dose	*	*	*	*
UK[a] (Pearce *et al.,* 2012)	74	ERR per mGy (RBM dose)	0.036	(0.005, 0.120)
France (Journy *et al.,*2015)	17	ERR per mGy	0.057	(-0.079, 0.193)
The Netherlands (Meulepas *et al.,*2018)	44	ERR per mGy	0.0004	(-0.0012, 0.016)
UK[a] (Berrington *et al.,* 2016)	70 without PF	ERR per mGy	0.037	(0.005, 0.126)
France (Journy *et al.,* 2016)	12 without PF	HR per 10 mGy	1.16	(0.77, 1.27)
France (Journy *et al.,*2016)	*5* with PF	HR per 10 mGy	0.57	(0.06, 1.32)
Germany (Krille *et al.,*2015)	17	HR per mGy	1.009	(0.98, 1.04)
UK (Pearce *et al.,*2012)	74	RR [>30 mGy] vs <*5* mGy	3.18	(1.46, 6.94)
Lymphoma risk according to RBM dose	``````````````	*	*	*
France (Journy *et al.,*2015)	19	ERR per mGy	0.018	(-0.068, 0.104)
UK[a] (Berrington *et al.,* 2017)	65[c]	RR [>20] vs <*5* mGy	0.92	(0.22, 2.94)

CNS, central nervous system; PF, predisposing factor; RBM, red bone marrow; ERR, excess relative risk; RR, relative risk; HR, hazard ratio; mGy, milligray.

* Cell intentionally left empty

a Follow-up period until 2005 only.

b Exclusion period *5* years instead of 2 years.

c Hodgkin lymphoma only.

Lastly, the ongoing Life Span Study (LSS) of atomic bomb survivors in Hiroshima and Nagasaki, Japan, provides quantitative estimates of cancer risks associated with exposure to radiation and is a major source of human data used for risk assessment in establishing radiation safety standards. Although this is not a systematic review, it is the gold standard, epidemiological study of radiation in the same dose range as encountered with CT. The most recent publications describe solid cancer incidence in the LSS cohort through 2009. (Brenner 2020, Grant 2017, Sadakane 2019, Sakata 2019, Sugiyama 2020) The eligible cohort included 105,444 subjects who were alive and had no known history of cancer at the start of follow-up. The follow-up period was 1958-2009, providing 3,079,484 person-years of follow-up. Cases were identified by linkage with population-based Hiroshima and Nagasaki Cancer Registries. Poisson regression methods were used to elucidate the nature of the radiation-associated risks per Gy of weighted absorbed organ doses using both excess relative risk (ERR) and excess absolute risk (EAR) models adjusted for smoking and other covariates**. These analyses demonstrate that solid cancer risks remain elevated more than 60 years after exposure and that approximately 10% of cancers in**

the cohort are due to the radiation. Studies by type of tumor confirm the strong association between radiation exposure and particular cancer types such as CNS tumors (Braganza, 2012 and Brenner, 2020), upper gastrointestinal tract tumors (Sakata, 2019) and liver and pancreas tumors (Sadakane, 2019) and colon tumors (Sugiyama, 2020)
There is also increasing understanding of the mechanisms involved in carcinogenesis. In a prospective evaluation of 67 adults undergoing cardiac CT, patients underwent extensive blood work just prior to and following the exam to look for cellular processes implicated in carcinogenesis. (Nguyen, 2015) Immunohistochemistry and full gene sequencing were performed, and diverse markers of DNA damage, repair, and cell death were evaluated. The average exposure from a single CT exam was 30 mSv (similar to the Hiroshima and Nagasaki exposures), and there was a three-fold increase in markers of DNA damage and cell death. These changes were seen at doses of 7 mSv and greater, and these changes persisted for at least a month.
Despite the known risks of CT, its use has grown substantially over the last few decades (Harvey L Neiman 2017), with 91.4 million CT exams performed in the United States in 2019 (IMV 2020), including 428 exams per 1000 patients aged 65 years and older (Smith-Bindman 2019). The radiation doses used for CT exams are frequently far higher than needed for diagnosis and have been shown to vary up to 200-fold across facilities for patients imaged for the same clinical reason. (Smith-Bindman 2009, Smith-Bindman 2015, Smith-Bindman 2019, Miglioretti 2013, Demb 2017). For example, the American College of Radiology reported that CT exams to assess kidney stones had an average dose of 10 mSv, while the optimum dose is 2-4 mSv. (Lukasiewicz, 2014) In a prospective randomized trial of different imaging strategies for patients with suspected kidney stones, 5% of patients received an appropriate dose of 4 mSv or less. (Smith-Bindman, 2014)
Evidence of the association between medical imaging and cancer risk has been reviewed by many professional societies and government, quality, and oversight organizations, which have all identified CT radiation dose reduction as a safety imperative and issued guidelines asking radiologists to track, optimize, and lower CT radiation doses. These organizations include: the American College of Radiology (Kanal 2017); the Radiology Society of North America (Hricak 2010); The Society of Interventional Radiology (Stecker 2009); The Society of Cardiovascular CT (Halliburton 2011); Cardiovascular Imaging Societies (Writing Committee 2018); Image Wisely (a joint initiative of the American College of Radiology, Radiological Society of North America, American Society of Radiological Technologists, and American Association of Physicists in Medicine); and the FDA (US Food and Drug Administration 2019).

[Response Ends]

1a.15. Detail the process used to identify the evidence.

[Response Begins]
The evidence was obtained through comprehensive searches or PubMed, Embase, and Web of Science from inception to August 2021. Each search consisted of Medical Imaging, Cancer and Epidemiology concept blocks with additional search terms including Computed Tomography and CT. References of all publications were searched to identify additional publications. Additionally, there are a small number of investigators who lead studies in this area (such as Dr. Amy Berrington De Gonzales, Chief of Radiation Epidemiology at the NCI and Dr. Alina Brenner at the Radiation Effects Research Foundation) whose names were added to searches.
[Response Ends]

1a.16. Provide the citation(s) for the evidence.

[Response Begins]

1. Abalo KD, Rage E, Leuraud K, Richardson DB, Le Pointe HD, Laurier D, Bernier MO. Early life ionizing radiation exposure and cancer risks: systematic review and meta-analysis. Pediatr Radiol 2021;51(1):45-56. doi: 10.1007/s00247-020-04803-0
2. Bernier MO, Baysson H, Pearce MS, Moissonnier M, Cardis E, Hauptmann M, Struelens L, Dabin J, Johansen C, Journy N, Laurier D, Blettner M, Le Cornet L, Pokora R, Gradowska P, Meulepas JM, Kjaerheim K, Istad T, Olerud H, Sovik A, Bosch de Basea M, Thierry-Chef I, Kaijser M, Nordenskjold A, Berrington de Gonzalez A, Harbron RW, Kesminiene A. Cohort Profile: the EPI-CT study: a European pooled epidemiological study to quantify the risk of radiation-induced cancer from paediatric CT. Int J Epidemiol 2019;48(2):379-381g. doi: 10.1093/ije/dyy231

3. Berrington de Gonzalez A, Mahesh M, Kim KP, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med. 2009;169(22):2071-2077.
4. Berrington de Gonzalez A, Salotti JA, McHugh K, Little MP, Harbron RW, Lee C, Ntowe E, Braganza MZ, Parker L, Rajaraman P, Stiller C, Stewart DR, Craft AW, Pearce MS. Relationship between paediatric CT scans and subsequent risk of leukaemia and brain tumours: assessment of the impact of underlying conditions. Br J Cancer 2016;114(4):388-394. doi: 10.1038/bjc.2015.415
5. Berrington de Gonzalez A, Daniels RD, Cardis E, et al. Epidemiological Studies of Low-Dose Ionizing Radiation and Cancer: Rationale and Framework for the Monograph and Overview of Eligible Studies. J Natl Cancer Inst Monogr. 2020;2020(56):97-113.
6. Board of Radiation Effects Research Division on Earth and Life Sciences National Research Council of the National Academies. Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2, Washington, D.C.: The National Academies Press; 2006.
7. Braganza MZ, Kitahara CM, Berrington de Gonzalez A, Inskip PD, Johnson KJ, Rajaraman P. Ionizing radiation and the risk of brain and central nervous system tumors: a systematic review. Neuro Oncol 2012;14(11):1316-1324. doi: 10.1093/neuonc/nos208
8. Brenner AV, Sugiyama H, Preston DL, Sakata R, French B, Sadakane A, Cahoon EK, Utada M, Mabuchi K, Ozasa K. Radiation risk of central nervous system tumors in the Life Span Study of atomic bomb survivors, 1958-2009. Eur J Epidemiol 2020;35(6):591-600. doi: 10.1007/s10654-019-00599-y
9. 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.
10. Demb J, Chu P, Nelson T, et al. Optimizing Radiation Doses for Computed Tomography Across Institutions: Dose Auditing and Best Practices. JAMA Intern Med. 2017;177(6):810-81
11. Grant EJ, Brenner A, Sugiyama H, Sakata R, Sadakane A, Utada M, Cahoon EK, Milder CM, Soda M, Cullings HM, Preston DL, Mabuchi K, Ozasa K. Solid Cancer Incidence among the Life Span Study of Atomic Bomb Survivors: 1958-2009. Radiation research 2017;187(5):513-537. doi: 10.1667/RR14492.1
12. Halliburton SS, Abbara S, Chen MY, et al. SCCT guidelines on radiation dose and dose-optimization strategies in cardiovascular CT. J Cardiovasc Comput Tomogr. 2011;5(4):198-224.
13. Harvey L Neiman Health Policy Institute. Harvey L Neiman Health Policy Institute. Medicare Part B Total Computed Tomography Procedures. 2017;

https://www.neimanhpi.org/data_series/medicare-part-b-total-computed-tomography-procedures/#/graph/2017/2017/true , November 12, 2019.

- Hauptmann M, Daniels RD, Cardis E, et al. Epidemiological Studies of Low-Dose Ionizing Radiation and Cancer: Summary Bias Assessment and Meta-Analysis. J Natl Cancer Inst Monogr. 2020;2020(56):188-200.
- Hong JY, Han K, Jung JH, Kim JS. Association of Exposure to Diagnostic Low-Dose Ionizing Radiation with Risk of Cancer Among Youths in South Korea. JAMA Netw Open. 2019;2(9):e1910584.
- Hricak H, Brenner DJ, Adelstein SJ, et al. Managing Radiation Use in Medical Imaging: A Multifaceted Challenge. Radiology. 2010.
- Image Wisely.

https://www.imagewisely.org/ .

- IMV 2019 CT Market Outlook Report,

https://imvinfo.com/ct-departments-seek-workflow-improvements-to-address-increased-ct-utilization/ .

- Journy N, Rehel JL, Ducou Le Pointe H, Lee C, Brisse H, Chateil JF, Caer-Lorho S, Laurier D, Bernier MO. Are the studies on cancer risk from CT scans biased by indication? Elements of answer from a large-scale cohort study in France. Br J Cancer 2015;112(1):185-193. doi: 10.1038/bjc.2014.526
- Journy N, Roue T, Cardis E, et al. Childhood CT scans and cancer risk: impact of predisposing factors for cancer on the risk estimates. J Radiol Prot. 2016;36(1):N1-7.
- 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.
- Kanal KM, Butler PF, Bhargavan-Chatfield M, et al U.S. Diagnostic Reference levels and achievable doses for 10 Pediatric CT exams Radiology 2021 302:1
- Krille L, Dreger S, Schindel R, et al. Risk of cancer incidence before the age of 15 years after exposure to ionising radiation from computed tomography: results from a German cohort study. Radiation and environmental biophysics. 2015;54(1):1-12.
- Lukasiewicz A, Bhargavan-Chatfield M, Coombs L, Ghita M, Weinreb J, Gunabushanam G, Moore CL. Radiation Dose Index of Renal Colic Protocol CT Studies in the United States: A Report from the American College of Radiology National Radiology Data Registry. Radiology 2014;271(2):445-451. doi: 10.1148/radiol.14131601

- Meulepas JM, Ronckers CM, Smets A, Nievelstein RAJ, Gradowska P, Lee C, Jahnen A, van Straten M, de Wit MY, Zonnenberg B, Klein WM, Merks JH, Visser O, van Leeuwen FE, Hauptmann M. Radiation Exposure From Pediatric CT Scans and Subsequent Cancer Risk in the Netherlands. J Natl Cancer Inst 2019;111(3):256-263. doi: 10.1093/jnci/djy104
- Meulepas JM, Ronckers CM, Merks J, Weijerman ME, Lubin JH, Hauptmann M. Confounding of the association between radiation exposure from CT scans and risk of leukemia and brain tumors by cancer susceptibility syndromes. J Radiol Prot 2016;36(4):953-974. doi: 10.1088/0952-4746/36/4/953
- Miglioretti DL, Johnson E, Williams A, et al. The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk. JAMA Pediatr. 2013;167(8):700-707.
- National Cancer Institute Cancer Statistics. Accessed May 25, 2021. https://www.cancer.gov/about-cancer/understanding/statistics
- Nguyen PK, Lee WH, Li YF, Hong WX, Hu S, Chan C, Liang G, Nguyen I, Ong SG, Churko J, Wang J, Altman RB, Fleischmann D, Wu JC. Assessment of the Radiation Effects of Cardiac CT Angiography Using Protein and Genetic Biomarkers. JACC Cardiovasc Imaging 2015;8(8):873-884. doi: 10.1016/j.jcmg.2015.04.016
- Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. Lancet. 2012;380(9840):499-505.
- Pierce DA, Preston DL. Radiation-related cancer risks at low doses among atomic bomb survivors. Radiation research. 2000;154(2):178-186.
- Preston DL, Ron E, Tokuoka S, et al. Solid cancer incidence in atomic bomb survivors: 1958-1998. Radiation research. 2007;168(1):1-64.
- Sadakane A, French B, Brenner AV, Preston DL, Sugiyama H, Grant EJ, Sakata R, Utada M, Cahoon EK, Mabuchi K, Ozasa K. Radiation and Risk of Liver, Biliary Tract, and Pancreatic Cancers among Atomic Bomb Survivors in Hiroshima and Nagasaki: 1958-2009. Radiation research 2019;192(3):299-310. doi: 10.1667/RR15341.1
- Sakata R, Preston DL, Brenner AV, Sugiyama H, Grant EJ, Rajaraman P, Sadakane A, Utada M, French B, Cahoon EK, Mabuchi K, Ozasa K. Radiation-Related Risk of Cancers of the Upper Digestive Tract among Japanese Atomic Bomb Survivors. Radiation research 2019;192(3):331-344. doi: 10.1667/RR15386.1
- Smith-Bindman R, Lipson J, Marcus R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med. 2009;169(22):2078-2086.
- Smith-Bindman R, Aubin C, Bailitz J, Bengiamin RN, Camargo CA, Jr., Corbo J, Dean AJ, Goldstein RB, Griffey RT, Jay GD, Kang TL, Kriesel DR, Ma OJ, Mallin M, Manson W, Melnikow J, Miglioretti DL, Miller SK, Mills LD, Miner JR, Moghadassi M, Noble VE, Press GM, Stoller ML, Valencia VE, Wang J, Wang RC, Cummings SR. Ultrasonography versus Computed Tomography for Suspected Nephrolithiasis. The New England Journal of Medicine. 2014;371(12):1100-1110. doi: 10.1056/NEJMoa1404446.
- Smith-Bindman R, Kwan ML, Marlow EC, et al. Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000-2016. JAMA. 2019;322(9):843-856.
- Smith-Bindman R, Moghadassi M, Wilson N, et al. Radiation Doses in Consecutive CT Examinations from Five University of California Medical Centers. Radiology. 2015;277(1):134-141.
- 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.
- Stecker MS, Balter S, Towbin RB, et al. Guidelines for patient radiation dose management. J Vasc Interv Radiol. 2009;20(7 Suppl):S263-273.
- Stewart C, Smith-Bindman R. It Is Time to Inform Patients of Medical Imaging Risks. JAMA Netw Open. 2021 Oct 1;4(10):e2129681. doi: 10.1001/jamanetworkopen.2021.29681.
- Sugiyama H, Misumi M, Brenner A, Grant EJ, Sakata R, Sadakane A, Utada M, Preston DL, Mabuchi K, Ozasa K. Radiation risk of incident colorectal cancer by anatomical site among atomic bomb survivors: 1958-2009. Int J Cancer 2020;146(3):635-645. doi: 10.1002/ijc.32275
- U.S. Food and Drug Administration. FDA White Paper: Initiative to Reduce Unnecessary Radiation Exposure from Medical Imaging. 2019. https://www.fda.gov/radiation-emitting-products/initiative-reduce-unnecessary-radiation-exposure-medical-imaging/white-paper-initiative-reduce-unnecessary-radiation-exposure-medical-imaging .
- Writing Committee M, Hirshfeld JW, Jr., Ferrari VA, et al. 2018 ACC/HRS/NASCI/SCAI/SCCT Expert Consensus Document on Optimal Use of Ionizing Radiation in Cardiovascular Imaging-Best Practices for Safety and Effectiveness, Part 2: Radiological Equipment Operation, Dose-Sparing Methodologies, Patient and Medical Personnel Protection. J Am Coll Cardiol. 2018.

[Response Ends]

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]
Radiologists and other physicians who perform CT are frequently not aware of the doses they use. Further, they do not know that the doses that are routinely used are in the range where they will cause cancer in some patients. They are also not aware that there is tremendous variation in doses for patients who are seen at different hospitals, even if the clinical question is the same. (ACR, DIR, 2014; Hausleiter, JAMA 2009; Keegan, JACR 2014; Miglioretti, JAMA Pediatrics, 2013; Parker, Pediatrics, 2015; Smith-Bindman, Arch Int Med, 2009; Smith-Bindman, JAMA, 2012; Smith-Bindman, JACR 2014, Smith-Bindman BMJ 2019).
Reducing doses among children who receive unnecessarily high doses could have a significant impact on the number of children who would develop cancer. In our JAMA Pediatrics paper (Miglioretti, JAMA IM 2013) using statistical modeling and observed CT doses across a large number of patients evaluated across an integrated HMO where we collected data on consecutive scans, we modeled what would occur if the highest dose patients (those above the 75% among the observed doses) came down to the median dose. In these patients the two dominant indications for imaging was minor trauma and suspected appendicitis. Using the observed exposures, we would expect that due to CT exposures in children age 15 and younger in the US in 2010, 9,820 future cancers would occur. If the highest exposed individuals instead had doses at the median, 44% of these cancers would be prevented. Thus reducing the outlier doses - which are most never justified or needed, could have a substantial impact.
The motivation for this measure is the belief that if radiologists would be provided with explicity benchmarks and were informed when their doses routinely exceed those benchmarks, that they would be motivated to improve to bring their performance closer to routine practice. Currently, physicians do not know the typical radiation doses received by their patients. This tool provides the framework for measurement – the first step towards quality improvement.
Creation of a simple standard for collection of radiation dose information would help facilities understand their current practice, would allow understanding changes in practice over time (Keegan, JACR, 2014; Greenwood, RadioGraphics, 2015) would allow comparisons to local and national standards, and would indicate to facilities whether their is a need to improve. There is currently a high level of interest in this area - facilities are being asked by their patients and governing boards to report whether they are performing CT safely - and this measure is an ideal starting point for facilities to assemble this information to answer these questions. If facilities collect dose information, it is the first step towards trying to compete on a measure of safety and to lower the doses they use.
A 2014 study by Keegan et al. observed a reduction in dose metrics after an earlier NQF measure of CT radiation dose (#0739) was introduced. All dose metrics in inpatient and outpatient facilities that had adopted the measure demonstrated a 30% to 50% reduction between 2010 and 2012, the time period over which the measure was implemented.
Cited in this section:
American College of Radiology (ACR). Dose Index Registry (DIR). 2014. http://www.acr.org/

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]
There are two sources of data for performance scores. These entities are described further in section 2a.05.
Scoring approach 1, at the individual strata level: **poor performance is classified by whether a facility's median radiation dose is greater than or equal to the benchmark 75th percentile in any single stratum.**
The measure has been in use by the Leapfrog Group and implemented at 1,447 hospitals, including 24 free-standing pediatric hospitals. The hospitals are diverse with regard to community vs. academic, urban vs. nonurban care settings, and geographic location – representing all 50 states and the District of Columbia. The measure is reported at the hospital

(facility) level. Data were collected over a 1 year period: 1/1/19 through 12/31/2019; 1/1/20 through 12/31/20; or 7/1/20 through 6/30/21. As the Leapfrog Group does not collect data at the individual patient level, we are unable to state how many patients were included. The proportion of hospitals overall that had a median dose above threshold in at least one stratum was 28% (375/1340).
The proportion of facilities in the UCSF International CT Dose Registry that had a median dose above threshold in at least one stratum was 49% (52/106).
Scoring approach 2, all strata combined: **poor performance is classified when 50% or more of a facility's exams across all strata exceed the relevant stratum-specific benchmark 75th percentile**
As the Leapfrog Group does not collect individual patient-level data, we cannot provide performance scores using this approach.
In the UCSF International CT Dose Registry, 114,443 exams derive from 106 hospitals with at least 10 observed exams.
Performance data is as follows:
Mean measure score: 26%
Standard deviation: 16%
Minimum score = 0%
Maximum score = 100%
Interquartile range: 18 % (16 %- 34%)
Measure scores by percentile:

- 10th = 9%
- 20th = 13%
- 30th = 17%
- 40th = 19%
- 50th = 22%
- 60th = 27%
- 70th = 30%
- 80th = 38%
- 90th = 48%

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]
/media/ACR/Documents/PDF/QualitySafety/NRDR/DIR/DIR%20Measures.pdf
Registry designed to showcase measures for certain CT procedure types.
Calvert C, Strauss KJ, Mooney DP. Variation in computed tomography radiation dose in community hospitals. Journal of pediatric surgery. Jun 2012;47(6):1167-1169.
"Radiation exposure is a concern among those who evaluate injured children…This study identified a thirty-times range of radiation dosage for CT scans performed across 40 different hospitals." – Conclusion statement from Abstract
Demb J, manuscript under preparation. CT Radiation Dose Standardization Across the University of California Medical Centers Using Audits to Optimize Dose. 2015.
Following an in-person meeting regarding CT radiation dose, radiologists, technologists and medical physicists from University of California medical centers strategized how to best optimize dosing practices at their sites, which were then analyzed for effectiveness and success after implementation.
Dorfman AL, Fazel R, Einstein AJ, et al. Use of Medical Imaging Procedures With Ionizing Radiation in Children: A Population-Based Study. Arch Pediatr Adolesc Med. Jan 3 2011.
" Exposure to ionizing radiation from medical diagnostic imaging procedures may occur frequently among children. Efforts to optimize and ensure appropriate use of these procedures in the pediatric population should be encouraged." – Conclusion statement from Abstract
Duncan J, Street M, Strother M, et al. Optimizing Radiation Use During Fluoroscopic Procedures: A Quality and Safety Improvement Project. J Am Coll Radiol. 2013;10:847-853

"A systematic approach to improving radiation use during procedures led to a substantial and sustained reduction in risk with no reduction in benefits. Data were readily captured by both manual and automated processes." – Conclusion statement from Abstract
Einstein AJ, Henzlova MJ, Rajagopalan S. Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. JAMA 2007;298:317-23.
"… estimates derived from our simulation models suggest that use of 64-slice CTCA is associated with a nonnegligible LAR (lifetime attributable risk) of cancer. This risk varies markedly and is considerably greater for women, younger patients..."– Conclusion statement from Abstract
Greenwood T, Lopez-Costa R, Rhoades P, et al. CT Dose Optimization in Pediatric Radiology: A Multiyear Effort to Preserve the Benefits of Imaging While Reducing the Risks. RadioGraphics. Jan 2015;35(5):1539-1554
"This systematic approach involving education, streamlining access to magnetic resonance imaging and ultrasonography, auditing with comparison with benchmarks, applying modern CT technology, and revising CT protocols has led to a more than twofold reduction in CT radiation exposure between 2005 and 2012…" – Conclusion statement from Abstract
Hausleiter, J., T. Meyer, et al. Estimated radiation dose associated with cardiac CT angiography. JAMA 301(5): 500-7. 2009
"Median doses of CCTA differ significantly between study sites and CT systems. Effective strategies to reduce radiation dose are available but some strategies are not frequently used. The comparable diagnostic image quality may support an increased use of dose-saving strategies in adequately selected patients."– Conclusion statement from Abstract
Keegan J, Miglioretti DL, Gould R, Donnelly LF, Wilson ND, Smith-Bindman R. Radiation Dose Metrics in CT: Assessing Dose Using the National Quality Forum CT Patient Safety Measure. Journal of the American College of Radiology: JACR; 11(3):309-315.
http://download.journals.elsevierhealth.com/pdfs/journals/1546-1440/PIIS1546144013006625.pdf. Mar 2014
Looking at dose metrics as per compliance with the previously endorsed #0739 NQF measure results in reasonably timed acquisition of CT doses, and seeing such doses resulted in 30-50% dose reduction.
Kumar K, manuscript under preparation. Radiation Dose Benchmarks in Children.
This paper will describe dose metrics among 29,000 children within age strata <1, 1-4 years, 5-9 years, 10-14 years, and 15-19 years. 2015.
Miglioretti D, Johnson E, Vanneman N, Smith-Bindman R, al e. Use of Computed Tomography and Associated Radiation Exposure and Leukemia Risk in Children and Young Adults across Seven Integrated Healthcare Systems from 1994 – 2010. JAMA Pediatrics Published online June 10, 2013 joli:101001/jamapediatrics2013311. 2013.
Radiation-induced cancers in children could be dramatically reduced if the highest quartile of CT radiation doses were lowered.
Miglioretti DL, YX Zhang, E Johnson, N Vanneman, R Smith-Bindman. Personalized Technologist Dose Audit Feedback for Reducing Patient Radiation Exposure from Computed Tomography. Journal of the American College of Radiology: JACR 2014.
"Personalized audit feedback and education can change technologists´ attitudes about, and awareness of, radiation and can lower patient radiation exposure from CT imaging." – Conclusion statement from Abstract
Nationwide Evaluation of X-ray Trends: NEXT 2005-2006. This presentation was given by David Spelic, physicist with the Food and Drug Administration (FDA), to the 39th Conference of Radiation Control Program Directors (CRCPD) annual meeting, held in Spokane Washington, May 21-24, 2007.
Morin, R. L. (2006). CT dosimetry--an enigma surrounded by a conundrum. J Am Coll Radiol 3(8): 630.
An explanation of the difficulties surrounding CT dosing and estimations of its harmful effects.
Morin, R. L. (2006). What are the national radiation doses? J Am Coll Radiol 3(12): 956.
An explanation of why benchmarks or national measures are so difficult to set (related to the article listed above).
Parker M, Shah S, Hall M, et al. Computed Tomography and Shifts to Alternate Imaging Modalities in Hospitalized Children. Pediatrics. 2015-0995.
"For the 10 most common All-Patient Refined Diagnosis Related Groups (APR-DRGs) for which children received CT in 2004, a decrease in CT utilization was found in 2012. Alternative imaging modalities for 8 of the diagnoses were used." – Conclusion statement from Abstract
Smith-Bindman R, Lipson J, Marcus R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med 2009;169:2078-86.
" Radiation doses from commonly performed diagnostic CT examinations are higher and more variable than generally quoted, highlighting the need for greater standardization across institutions." – Conclusion statement from Abstract
Smith-Bindman R. Is computed tomography safe? N Engl J Med 2010;363:1-4.
An explanation of the harmful effects of CT overdose, and why its diagnostic purposes are often misused.
Smith-Bindman R. Environmental causes of breast cancer and radiation from medical imaging: findings from the Institute of Medicine report. Arch Intern Med 2012;172:1023-7.

" The IOM´s conclusion of a causal relation between radiation exposure and cancer is consistent with a large and varied literature showing that exposure to radiation in the same range as used for computed tomography will increase the risk of cancer." – Conclusion statement from Abstract
Smith-Bindman R, Miglioretti DL, Johnson E, et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. JAMA 2012;307:2400-9.
" Within integrated health care systems, there was a large increase in the rate of advanced diagnostic imaging and associated radiation exposure between 1996 and 2010." – Conclusion statement from Abstract
Smith-Bindman R, Moghadassi M, Wilson N, et al. Radiation Doses in Consecutive CT Examinations from Five University of California Centers. Radiology 2015:277: 134–141
"These summary dose data provide a starting point for institutional evaluation of CT radiation doses." – Conclusion statement from Abstract
Wilson N. CT Radiation Dose Standardization Across the Five University of California Medical Centers. ARRS: Annual Toronto Meeting presentation. April 19-24, 2015
Understanding the reasons for variation in commonly performed CT procedures, and figuring out how to standardize them.

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]
Table 1b.04 Proportion of exams exceeding the benchmarks by sex and age.
The following data are from the UCSF International CT Dose Registry and reflect the facility's proportion of exams across all strata that exceed the relevant stratum-specific benchmark 75th percentile (i.e. scoring approach 2). There are slightly higher proportions above benchmark in males and in the <1 year old age group. The measure steward will continue to monitor these rates as the measure rolls out, and the proposed measure may have a role in reducing disparities.

	Female	**Male**	**Age <1**	**Age 1-4**	**Age 5-9**	**Age 10-14**	**Age 15-17**
Mean Measure Score	0.22	0.29	0.36	0.25	0.25	0.27	0.24
Standard Deviation	0.16	0.20	0.28	0.19	0.19	0.18	0.17
Minimum Score	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum Score	0.86	1.00	1.00	0.74	0.75	0.82	0.77
First Quartile	0.10	0.15	0.09	0.09	0.10	0.14	0.12
Third Quartile	0.32	0.41	0.56	0.40	0.34	0.39	0.32
10th Percentile	0.05	0.10	0.04	0.04	0.05	0.08	0.07
20th Percentile	0.09	0.14	0.07	0.07	0.09	0.12	0.11
30th Percentile	0.15	0.17	0.12	0.13	0.12	0.15	0.14
40th Percentile	0.17	0.21	0.25	0.17	0.19	0.19	0.17
50th Percentile	0.20	0.25	0.35	0.20	0.20	0.21	0.20
60th Percentile	0.23	0.29	0.39	0.26	0.25	0.27	0.23
70th Percentile	0.28	0.38	0.49	0.32	0.31	0.35	0.30
80th Percentile	0.35	0.45	0.64	0.45	0.39	0.43	0.37
90th Percentile	0.47	0.52	0.76	0.53	0.54	0.54	0.47

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

To the extent they have been studied in both children and adults, social factors including sex, race/ethnicity, and socioeconomic status are not predictive of radiation dose for CT exams. (Strauchler 2012, Freeman 2012, Hou 2014, Messenger 2015). However, as described in the studies led by Strauchler and Freeman, 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. Thus, it is particularly important to ensure that the doses used for CT in these individuals are not excessive, because vulnerable patients are at greatest risk of chronic disease and more likely to be exposed to many irradiating exams.

References

1. Freeman K, Strauchler D, Miller TS. Impact of socioeconomic status on ionizing radiation exposure from medical imaging in children. J Am Coll Radiol. 2012 Nov;9(11):799-807. doi: 10.1016/j.jacr.2012.06.005. PMID: 23122347.
2. Hou, J.K., Malaty, H.M. & Thirumurthi, S. Radiation Exposure from Diagnostic Imaging Studies Among Patients with Inflammatory Bowel Disease in a Safety-Net Health-Care System. Dig Dis Sci 59, 546–553 (2014). https://doi.org/10.1007/s10620-013-2852-1

- Messenger B, Li D, Nasir K, Carr JJ, Blankstein R, Budoff MJ. Coronary calcium scans and radiation exposure in the multi-ethnic study of atherosclerosis. Int J Cardiovasc Imaging. 2016 Mar;32(3):525-9. doi: 10.1007/s10554-015-0799-3. Epub 2015 Oct 29. PMID: 26515964.
- Strauchler D, Freeman K, Miller TS. The impact of socioeconomic status and comorbid medical conditions on ionizing radiation exposure from diagnostic medical imaging in adults. J Am Coll Radiol. 2012 Jan;9(1):58-63. doi: 10.1016/j.jacr.2011.07.009. PMID: 22221637.

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

[Response Begins]
Yes
[Yes Please Explain]
Changes to specifications are explained in spma.02 below.

[Response Ends]

spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous NQF CDP review.

[Response Begins]
1. Anatomic areas
We have subdivided the previously specified "head" anatomic area into two new anatomic areas: "skull" representing low radiation dose indications that are focused on imaging of the facial and skull bones, sinus, temporal bone, and for assessment of patency of a ventricular shunt; and "brain" representing all other head indications which have the brain as the primary focus rather than the bony skeleton. This decision was made based on analysis of data from the UCSF International CT Dose Registry (Smith-Bindman 2021, Chu 2021) and recent publications from the American College of Radiology Dose Registry describing CT radiation doses in children. (Kanal 2021) These sources all demonstrate meaningful differences in radiation doses used for skull versus brain CT, necessitating two separate head categories.
To simplify the measure calculation and reporting, we now focus on only the three most common CT categories in children (skull, brain, and abdomen and pelvis), which together account for over 80% of CT scans in children (Chu 2021, Kanal 2021). We have dropped from the measure less frequent anatomic areas, which each contribute fewer than 10% of CT exams performed in children. (The Leapfrog Group found that it was not feasible to report data on these less frequent anatomic categories.)

2. Dose metrics
This measure focuses on reducing radiation dose in CT, an intermediate outcome important to cancer prevention. As radiation dose is known to be directly related and proportional to future cancer risk (Board of Radiation Effects 2006, Pearce 2012, Pierce 2000, Preston 2007, Brenner 2003, Hong 2019, Berrington de Gonzalez 2009, Hauptmann 2020, Abalo 2021), any reduction in the total radiation imparted to patients would be expected to lead to a proportional reduction in incident cancers in subsequent years.
Dose Length Product (DLP) reflects the total amount of radiation imparted to the patient and is therefore the single most important dose metric with respect to ensuring patients receive the appropriate radiation dose for CT. The DLP is universally reported by CT manufacturers using the Digital Imaging and Communication in Medicine (DICOM) standard. Although several other dose metrics were included in the currently endorsed measure, DLP is the only one used by the Leapfrog Group, which currently has over 1400 hospitals reporting on this measure. To simplify data collection and reporting, we have dropped the two additional dose metrics originally specified in the measure: Volumetric CT Dose Index (CTDIvol) and Size Specific Dose Estimate (SSDE), which reflect the average dose per slice rather than the total dose to the patient. We believe including a single measure of dose, and one that is uniformly available, will reduce the burden of reporting.

3. Phantom standardization

For calculation purposes, dose metrics rely on the use of a specific reference phantom to calibrate the amount of radiation reported for the CT scan. The DLP is only meaningful if you know whether it is being reported using one of two possible phantoms (the smaller 16 cm reference phantom or the larger 32 cm phantom). There tends to be consistency in the use of the phantoms across hospitals, (Chu 2021), but if the atypical phantom is used, the DLP will be inaccurate. We have introduced a new requirement that head (brain and skull) exams be reported using the 16-cm reference phantom and abdomen exams be reported using the 32-cm reference phantom. This standard will facilitate radiation dose comparisons across sites because varying the phantom used introduces distortions in dose reporting: a dose referenced to a 16-cm phantom will be approximately double the same dose referenced to a 32-cm phantom, and vice versa. According to recent research, reference phantom selection is already highly consistent, with 98.0% of head exams referenced to a 16-cm phantom and 94.4% of abdomen exams referenced to a 32-cm phantom. (Chu 2021) Therefore, this requirement will not impose any burden on the great majority of reporting facilities that are already reporting in a consistent fashion.

4. Measure scoring

We have updated the benchmarks for hospitals/imaging facilities to use for comparing their own doses to median and 75th percentile reference benchmark levels that were derived from pediatric exams in the UCSF International CT Dose Registry. These are provided for age-anatomic area strata, as described below.

Hospitals and outpatient imaging facilities will assess whether the DLP for each eligible CT scan performed in children is equal to or above the 75% percentile for that age-anatomic area stratum. The results can be scored separately for each stratum and overall:

- ***Approach 1) comparison of median radiation doses within age-anatomic region strata:*** Hospitals/Facilities with a median radiation dose within an age-anatomic area stratum below the 75th percentile benchmark will be considered to have an acceptable distribution of doses in that stratum. In other words, if 50% or more of exams in a stratum have DLP values at or *below* the benchmark for that stratum, then the dose distribution within that stratum will be considered acceptable. If greater than 50% of exams have DLP values *above* the benchmark, then the dose distribution within that stratum will be considered poor.

- ***Approach 2) overall proportion of high dose exams across all strata:*** Hospitals/Facilities with greater than 50% of their CT doses exceeding the appropriate 75th percentile benchmark (for that type of patient and anatomic area) will be considered to have an excessive proportion of high dose exams.

Citations

1. Abalo KD, Rage E, Leuraud K, Richardson DB, Le Pointe HD, Laurier D, Bernier MO. Early life ionizing radiation exposure and cancer risks: systematic review and meta-analysis. Pediatr Radiol. 2021 Jan;51(1):45-56. doi: 10.1007/s00247-020-04803-0. Epub 2020 Sep 10. Erratum in: Pediatr Radiol. 2020 Oct 22.
2. Berrington de Gonzalez A, Mahesh M, Kim KP, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med. 2009;169(22):2071-2077.
3. Board of Radiation Effects Research Division on Earth and Life Sciences National Research Council of the National Academies. Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2, Washington, D.C.: The National Academies Press; 2006.
4. 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.
5. Chu PW, Yu S, Wang Y, Seibert JA, Cervantes LF, Kasraie N, Chu CA, Smith-Bindman R. Reference phantom selection in pediatric computed tomography using data from a large, multicenter registry. Pediatr Radiol. 2021 Dec 6. doi: 10.1007/s00247-021-05227-0. Epub ahead of print. PMID: 34866159.
6. Hauptmann M, Daniels RD, Cardis E, Cullings HM, Kendall G, Laurier D, Linet MS, Little MP, Lubin JH, Preston DL, Richardson DB, Stram DO, Thierry-Chef I, Schubauer-Berigan MK, Gilbert ES, Berrington de Gonzalez A. Epidemiological Studies of Low-Dose Ionizing Radiation and Cancer: Summary Bias Assessment and Meta-Analysis. J Natl Cancer Inst Monogr. 2020 Jul 1;2020(56):188-200. doi: 10.1093/jncimonographs/lgaa010.
7. Hong JY, Han K, Jung JH, Kim JS. Association of Exposure to Diagnostic Low-Dose Ionizing Radiation with Risk of Cancer Among Youths in South Korea. JAMA Netw Open. 2019;2(9):e1910584.
8. Kanal KM, Butler PF, Chatfield MB, Wells J, Samei E, Simanowith M, Golden D, Gress DA, Burleson J, Sensakovic WF, Strauss KJ, Frush D. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Pediatric CT Examinations. Radiology. 2022 Jan;302(1):164-174. doi: 10.1148/radiol.2021211241. Epub 2021 Oct 26. Erratum in: Radiology. 2022 Jan;302(1):E6.
9. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. Lancet. 2012;380(9840):499-505.

10. Pierce DA, Preston DL. Radiation-related cancer risks at low doses among atomic bomb survivors. Radiation research. 2000;154(2):178-186.
11. Preston DL, Ron E, Tokuoka S, et al. Solid cancer incidence in atomic bomb survivors: 1958-1998. Radiation research. 2007;168(1):1-64.
12. Smith-Bindman R, Yu S, Wang Y, Kohli MD, Chu P, Chung R, Luong J, Bos D, Stewart C, Bista B, Alejandrez Cisneros A, Delman B, Einstein AJ, Flynn M, Romano P, Seibert JA, Westphalen AC, Bindman A. An Image Quality-informed Framework for CT Characterization. Radiology. 2021 Nov 9:210591. doi: 10.1148/radiol.2021210591. Epub ahead of print.

[Response Ends]

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]
Pediatric Computed Tomography (CT) Radiation Dose
[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]
2022 submission:
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.
Updated reference from 2016 submission:
Demb J, Chu P, Nelson T, Hall D, Seibert A, Lamba R, Boone J, Krishnam M, Cagnon C, Bostani M, Gould R, Miglioretti D, Smith-Bindman R. Optimizing Radiation Doses for Computed Tomography Across Institutions: Dose Auditing and Best Practices. JAMA Intern Med. 2017 Jun 1;177(6):810-817. doi: 10.1001/jamainternmed.2017.0445. PMID: 28395000; PMCID: PMC5818828.
Note, the Kumar reference noted as being "in preparation" in the 2016 submission was never published.
2016 submission:
The measure requires hospitals and output facilities that conduct Computed Tomography (CT) examinations in children to: 1. Review their CT radiation dose metrics, 2. calculate the distribution of the results, and 3.compare their results to benchmarks. This would then imply a fourth step to investigate instances where results exceed a trigger value for underlying cause, such as issues with protocol, tech, equipment, patient, etc.
It is important to review doses of radiation used for CT, as the doses are far higher than conventional radiographs (x-rays), the doses are in the same range known to be carcinogenic (Pearce, Lancet, 2012; Ozasa, Radiation Research, 2012), and the higher the doses, the greater the risk of subsequent cancer (Miglioretti, JAMA Pediatrics, 2013) Thus the goal of the measure is to provide a framework where facilities can easily assess their doses, compare them to benchmarks, and take corrective action to lower their doses if they exceed threshold values, as per specifications in benchmarks.
The measure calls for assessment of doses for the most frequently conducted CT examination types, and compare these doses to published benchmarks. The measure calls for the assessment of radiation doses within four anatomic areas (CT's of the head, chest, abdomen/pelvis and combined chest/abdomen/pelvis.) The measure provides a simple framework for how facilities can assess their dose, compare their doses to published benchmarks (Smith-Bindman, Radiology, 2015) and identify opportunities to improve if their doses are higher than the benchmarks. For example, If a hospital finds their doses are higher than published benchmarks, they can review the processes and procedures they use for performance of

CT in children and take corrective action, and follow published guidelines for how to lower doses (such as “child sizing” the doses, reducing multiple phase scans, and reducing scan lengths).
Published benchmarks for radiation dose in children exist (Smith-Bindman, Radiology, 2015) and additional benchmarks are under development and will be published within the year by us. (Kumar, 2015) Other groups have also published benchmarks (Goeske) or in the process of doing so.
Our work and that of others have shown that institutional review of dose metrics as outlined in this measure results in a significant lowering of average and outlier doses. (Demb, 2015; Greenwood, RadioGraphics, 2015; Miglioretti, JAMA Pediatrics, 2013; Keegan, JACR, 2104; Wilson, ARRS, 2015).
This measure is being proposed for diagnostic CT in children, but can also be used for CT in adults, and CT used in conjunction with radiation therapy for cancer. Whenever context the doses are used, the doses should be compared with appropriate benchmarks.
A similar measure (#0739) was previously endorsed by the NQF in 2011. The NQF did not provide ongoing endorsement when the measure was up for renewal in 2015, primarily because there was no evidence that assessing doses as called for in the measure would result in an improvement in outcomes (i.e. patient dose). Since that time, there has been additional research that has shown that assessing doses using the format outlined in the measure does indeed result in lower doses, and thus we are re-submitting a similar although updated measure.
Of note, the surrogate measure we are using for outcomes is radiation dose. The true outcome of interest is the number of cancers that result from imaging. Because of the lag time between exposure to radiation and cancer development (years to decades) it is not feasible to use cancer cases as the outcome of a quality improvement effort. Thus while there is ample evidence that radiation causes cancer (sited below), and evidenced that cancer risk is proportional to dose, there are no direct data that suggest that lowering doses lowers cancer risk. However, we have used mathematical modeling to try to understand the relationship between lowering doses and cancers and estimated that if the top quartile of doses were reduced in children (i.e. the very high doses are brought down the average doses), the number of cancer cases would be reduced by approximately 43%, the equivalent to preventing 4,350 cancer cases / year in the US among children (Miglioretti, JAMA Pediatrics 2013).
Cited in this section:
Demb J, manuscript under preparation. CT Radiation Dose Standardization Across the University of California Medical Centers Using Audits to Optimize Dose. 2015.
Following an in-person meeting regarding CT radiation dose, radiologists, technologists and medical physicists from University of California medical centers strategized how to best optimize dosing practices at their sites, which were then analyzed for effectiveness and success after implementation.
Greenwood T, Lopez-Costa R, Rhoades P, et al. CT Dose Optimization in Pediatric Radiology: A Multiyear Effort to Preserve the Benefits of Imaging While Reducing the Risks. RadioGraphics. Jan 2015;35(5):1539-1554
“This systematic approach involving education, streamlining access to magnetic resonance imaging and ultrasonography, auditing with comparison with benchmarks, applying modern CT technology, and revising CT protocols has led to a more than twofold reduction in CT radiation exposure between 2005 and 2012…” – Conclusion statement from Abstract
Keegan J, Miglioretti DL, Gould R, Donnelly LF, Wilson ND, Smith-Bindman R. Radiation Dose Metrics in CT: Assessing Dose Using the National Quality Forum CT Patient Safety Measure. Journal of the American College of Radiology: JACR; 11(3):309-315.
http://download.journals.elsevierhealth.com/pdfs/journals/1546-1440/PIIS1546144013006625.pdf. Mar 2014
Looking at dose metrics as per compliance with the previously endorsed #0739 NQF measure results in reasonably timed acquisition of CT doses, and seeing such doses resulted in 30-50% dose reduction.
Kumar K, manuscript under preparation. Radiation Dose Benchmarks in Children.
This paper will describe dose metrics among 29,000 children within age strata <1, 1-4 years, 5-9 years, 10-14 years, and 15-19 years. 2015.
Miglioretti D, Johnson E, Vanneman N, Smith-Bindman R, al e. Use of Computed Tomography and Associated Radiation Exposure and Leukemia Risk in Children and Young Adults across Seven Integrated Healthcare Systems from 1994 – 2010. JAMA Pediatrics Published online June 10, 2013 joli:101001/jamapediatrics2013311, 2013.
Radiation-induced cancers in children could be dramatically reduced if the highest quartile of CT radiation doses were lowered.
Miglioretti, YX Zhang, E Johnson, N Vanneman, R Smith-Bindman. Personalized Technologist Dose Audit Feedback for Reducing Patient Radiation Exposure from Computed Tomography. Journal of the American College of Radiology: JACR 2014.
“Personalized audit feedback and education can change technologists´ attitudes about, and awareness of, radiation and can lower patient radiation exposure from CT imaging.” – Conclusion statement from Abstract
Ozasa K, Shimizu Y, Suyama A, et al. Studies of the mortality of atomic bomb survivors, Report 14, 1950-2003: an overview of cancer and noncancer diseases. Radiation Research; 177(3):229-243. Mar 2012

Fourteenth follow-up report on the lifetime health effects from radiation on atomic bomb survivor showing that: 58% of the 86,611 LSS cohort members with DS02 dose estimates have died, 17% more cancer deaths especially among those under age 10 at exposure (58% more deaths).
Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. Lancet;380(9840):499-505. Aug 4 2012
"Use of CT scans in children to deliver cumulative doses of about 50 mGy might almost triple the risk of leukaemia and doses of about 60 mGy might triple the risk of brain cancer… although clinical benefits should outweigh the small absolute risks, radiation doses from CT scans ought to be kept as low as possible" – Conclusion statement from Abstract
Smith-Bindman R, Moghadassi M, Wilson N, et al. Radiation Doses in Consecutive CT Examinations from Five University of California Centers. Radiology 2015:277: 134–141
"These summary dose data provide a starting point for institutional evaluation of CT radiation doses." – Conclusion statement from Abstract
Wilson N. CT Radiation Dose Standardization Across the Five University of California Medical Centers. ARRS: Annual Toronto Meeting presentation. April 19-24, 2015
Understanding the reasons for variation in commonly performed CT procedures, and figuring out how to standardize them.

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Surgery: General*

[Response Begins]
Cancer
Other (specify)
[Other (specify) Please Explain]
Radiology

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]
Safety
[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]
Children (Age < 18)
[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Facility
[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
[Response Begins]
Inpatient/Hospital
Outpatient Services
[Response Ends]

sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".

[Response Begins]
2022 submission:
The current NQF-endorsed measure is reported by the Leapfrog Group. Their web pages with current detailed specifications appear below. Please note that we propose some specification changes (see above) that are not reflected in these web pages:
https://ratings.leapfroggroup.org/measure/hospital/radiation-dose-head-scans
https://ratings.leapfroggroup.org/measure/hospital/radiation-dose-abdomenpelvis-scans
2016 submission:
N/A

[Response Ends]

sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff . Provide descriptors for any codes. Use one file with multiple worksheets, if needed.
[Response Begins]
No data dictionary/code table – all information provided in the submission form
[Response Ends]

sp.12. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).
DO NOT include the rationale for the measure.

[Response Begins]
2022 submission:

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.

2016 submission:

Radiation Dose metrics among consecutive patients, who have undergone CT of the head, chest, abdomen/pelvis, or chest/abdomen/pelvis. The metrics are 1) mean dose as measured using DLP, CTDIvol, and SSDE: within age strata. And 2) the proportion of exams with doses greater than the 75th percentile of the benchmark you are comparing with for the same anatomic area strata (Kumar, 2015; Smith-Bindman, Radiology, 2015; Goske, Radiology, 2013)

The CTDIvol and DLP are directly reported by the scanner using an "industry wide" standardized dose report (DICOM Radiation Dose Structured Report). The data should be assembled for the entire CT examination. If there are several series, the CTDIvol values should be averaged, and the DLP values should be added.

SSDE can be calculated using any dose monitoring software product, or using published multiplier coefficients which are highly valid.

These different metrics are highly correlated, but nonetheless reveal important differences regarding radiology practice and performance and are thus complimentary. However, if a practice only assesses data from a single metric, there is substantial opportunity for data-driven improvement.

CTDIvol reflects the average dose per small scan length. Modern CT scanners directly generate this.

DLP reflects the CTDIvol x scan length, and is directly generated by modern CT scanners.

SSDE is a modified measure of CTDIvol that takes into account the size of the patient scanned and is useful for scaling dose to patient size. Several current radiation tracking software tools directly report SSDE.

Cited in this section

Goske MJ, Strauss KJ, Coombs LP, et al. Diagnostic reference ranges for pediatric abdominal CT. Radiology. Jul 2013;268(1):208-218.

"Calculation of reference doses as a function of BW (body weight) for an individual practice provides a tool to help develop site-specific CT protocols that help manage pediatric patient radiation doses." – Conclusion statement from Abstract

Kumar K, manuscript under preparation. Radiation Dose Benchmarks in Children.

This paper will describe dose metrics among 29,000 children within age strata <1, 1-4 years, 5-9 years, 10-14 years, and 15-19 years. 2015.

Smith-Bindman R, Moghadassi M, Wilson N, et al. Radiation Doses in Consecutive CT Examinations from Five University of California Centers. Radiology 2015:277: 134–141

"These summary dose data provide a starting point for institutional evaluation of CT radiation doses." – Conclusion statement from Abstract

Smith-Bindman R, Miglioretti DL. CTDIvol, DLP, and Effective Dose are excellent measures for use in CT quality improvement. Radiology. Dec 2011;261(3):999; author reply 999-1000.

An explanation as to why these radiation dose metrics are useful in calculating a patient's absorbed doses.

Huda W, Ogden KM, Khorasani MR. Converting dose-length product to effective dose at CT. Radiology. Sep 2008;248(3):995-1003.

"This article describes a method of providing CT users with a practical and reliable estimate of adult patient EDs by using the DLP displayed on the CT console at the end of any given examination." – Conclusion statement from Abstract

[Response Ends]

sp.13. Provide details needed to calculate the numerator.

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

2022 submission:

Calculating the numerator and scoring the measure

The numerator is comprised of the total number of CT exams in the denominator for which the DLP exceeds the 75th percentile benchmark for the specific anatomic and age strata.

There are two ways of scoring the measure:
1) At the individual strata level: A hospital or outpatient imaging facility's performance, by anatomic area and by age group, are classified using the following scale aligning with the Leapfrog Group's implementation:

- Acceptable = the hospital or outpatient imaging facility's median radiation dose is below the 75th percentile for the stratum.
- Poor = the hospital or outpatient imaging facility's median radiation dose is greater or equal to the benchmark 75th percentile.

2) At the overall level, including all strata combined: A hospital or imaging facility's proportion of high dose exams is defined as the percent of examinations, across all strata, that exceed the relevant stratum specific benchmark 75th percentile.

- Performance is classified as poor when the out-of-range rate is more than twice the expected rate, i.e., when 50% or more examinations exceed the 75th percentile.

The overarching goal is to assess whether an individual reporting entity's distribution of CT exams (within strata, and across all strata) on average exceeds the 75th percentile, and to what degree. The measure classifies both (1) median radiation doses exceeding the 75th percentile within a stratum, and (2) a rate of 50% or more of all exams exceeding their respective 75th percentile levels as poor performance.

Reference phantoms

Radiation doses for head exams (skull and brain) must be reported using the 16-cm reference phantom. Radiation doses for abdomen and pelvis exams must be reported using the 32-cm reference phantom.

While reference phantom selection is highly standardized across imaging facilities (Chu 2021), there is a small amount of variation by CT manufacturer in the abdomen and pelvis category for children up to 10 years of age. Abdomen and pelvis doses referenced to a 16-cm phantom will be approximately double the corresponding doses based on the correct 32 cm phantom. (Nelson 2014, Seibert 2014) Hospitals and imaging facilities that report using the less common phantom need to adjust their DLP values prior to reporting. Abdomen and pelvis doses reported using a 16 cm phantom should be halved, and head doses referenced to a 32-cm phantom should be doubled. (Chu 2021) This is a workaround if facilities are unable to report using the standard phantom selection.

Benchmarks

We have generated benchmarks for CT examinations in children for the three CT categories using data on 116,597 pediatric exams from the UCSF International CT Dose Registry, provided in table sp-1. These benchmark data are being drafted for publication. (Bos 2022, in preparation) These categories reflect the indications that led to imaging, rather than decisions made by the radiologist, for example, whether to do single phase or multiple phase examinations. All skull exams, all brain exams and all abdomen and pelvis exams should be included in the skull, brain and abdomen and pelvis categories, whether a single non-contrast phase, a single contrast phase, or a multiphase exam with and without contrast was done for an included patient.

Table sp-13-1. Median and 75th percentile radiation doses, measured in dose length product (DLP), for the 3 anatomic areas and 5 age groups, derived from the UCSF International CT Dose Registry.

Anatomic Area & Age Group	Median DLP (mGy·cm)	75th Percentile DLP (mGy·cm)
Skull	-	-
< 1 year	122	224
1-4 years	181	280
5-9 years	203	307
10-14 years	254	393
15-17 years	296	517
Brain	-	-
< 1 year	223	326
1-4 years	350	486
5-9 years	463	605
10-14 years	599	784
15-17 years	726	967
Abdomen and pelvis	-	-
< 1 year	50	89
1-4 years	76	110
5-9 years	126	197
10-14 years	269	373
15-17 years	353	549

Cells marked with a dash (-) are left intentionally blank

We have used the UCSF Registry to create benchmarks as these are currently the best data to summarize performance for the included anatomic areas and as specified in the measure (e.g., including all skull CT examinations in a single category, all brain CT examinations in a single category and all abdomen and pelvis CT examinations in a single category) and using a single age schema across all anatomic areas simplifying reporting. These benchmarks will be periodically updated and reassessed and we will continue to collaborate with the Leapfrog Group and other users to do so.
The Leapfrog Group, which is the current the primary user of this measure, has developed their own benchmarks based on hospital-reported data, which closely align with the recommended UCSF benchmarks (Table sp-2 for the abdomen and pelvis category). The Leapfrog Group does not currently subdivide head examinations into skull and brain, thus we cannot directly compare those benchmarks.

Table sp-13-2. The Leapfrog Group 75th percentile benchmarks for pediatric abdomen and pelvis examinations and those created from the UCSF International CT Dose Registry

Abdomen and pelvis	*	**75th percentile benchmark UCSF Registry**	**75th percentile benchmark used by the Leapfrog Group**
< 1 year	89	*	73
1-4 years	110	*	110
5-9 years	197	*	176
10-14 years	373	*	394
15-17 years	549	*	565

*cells intentionally left empty

In our 2016 submission, we did not include recommended benchmarks but suggested measure implementers may use any established benchmarks of their choosing; this is why the Leapfrog Group to date has used benchmarks based on their own collected data. We have notified the Leapfrog Group of the proposed changes in specifications (e.g. splitting the head category and updating radiation dose benchmarks), and we plan to work closely with them to ensure the measure is implemented in keeping with the newer specifications. We believe our benchmarks are the right ones to use, but as noted above, we'll continue to work with the Leapfrog Group (and any future users) to periodically reassess and update benchmarks as needed.

Citations:

1. Bos, D. Pediatric Radiation Dose Benchmarks from the UCSF International CT Dose Registry. 2022, in preparation.
2. Chu PW, Yu S, Wang Y, Seibert JA, Cervantes LF, Kasraie N, Chu CA, Smith-Bindman R. Reference phantom selection in pediatric computed tomography using data from a large, multicenter registry. Pediatr Radiol. 2021 Dec 6. doi: 10.1007/s00247-021-05227-0. Epub ahead of print. PMID: 34866159.
3. Nelson TR. Practical strategies to reduce pediatric CT radiation dose. J Am Coll Radiol. 2014 Mar;11(3):292-9. doi: 10.1016/j.jacr.2013.10.011. PMID: 24589405.
4. Seibert JA, Boone JM, Wootton-Gorges SL, Lamba R. Dose is not always what it seems: where very misleading values can result from volume CT dose index and dose length product. J Am Coll Radiol. 2014 Mar;11(3):233-7. doi: 10.1016/j.jacr.2013.10.010. PMID: 24589395.

2016 submission:

Radiation dose distribution for the three metrics (CTDIvol, DLP, and SSDE) need to be recorded for a consecutive sample of CT examinations within anatomic area and age stratum. The mean, median, and the percent of examinations above the published 75% percentile needs to be generated.

These data can be extracted from the CT examinations in several ways. These numbers can written down directly from the CT scanner itself at the time of the examination; they can be written down from the PACS (computer terminal where images are reviewed and stored); or can be written down from the medical record if the facility stores these data as part of the medical record (all facilities in California due this based on statutory requirements.) The CT manufacturers have agreed (through MITA, Medical Imaging and Technology Alliance, the professional trade association of imaging manufacturers) to make these data electronically available through export from the CT machines to a local server), and these data can also be collected electronically. A growing number of companies are leveraging the standardized data format to systematically collect dose metrics directly from a facilities imaging infrastructure. This not only improves the accuracy of the data but also markedly reduces the costs of data collection. From the PACS, Radiology Information System, EPIC program if the data are exported there, or using any number of dose monitoring software programs allowing the collection and reporting of these dose data. The easiest way to collect these data is through one of the 6 or so commercial software programs developed for dose tracking, and several free-ware programs that enable directly extracting CT dose information from the PACS. We have published (Keegan, JACR 2014) several examples of techniques for dose extraction that can be completed even by a small facility.

The strata for this measure include:
Anatomic area strata: head, chest, abdomen/pelvis, Chest/abdomen/pelvis
Age strata: infant (<1); small child (1-5); medium child (>5 - 10); large child (>10-15) and adult (>15)
NOTE: The SSDE was developed as a metric for adjusting for size. However, it does not completely adjust for size and analysis within age strata are still needed among children to account for the different doses that are used and should be used for infants to obese children.
Cited in this section:
Keegan J, Miglioretti DL, Gould R, Donnelly LF, Wilson ND, Smith-Bindman R. Radiation Dose Metrics in CT: Assessing Dose Using the National Quality Forum CT Patient Safety Measure. Journal of the American College of Radiology: JACR; 11(3):309-315.
http://download.journals.elsevierhealth.com/pdfs/journals/1546-1440/PIIS1546144013006625.pdf. Mar 2014
Looking at dose metrics as per compliance with the previously endorsed #0739 NQF measure results in reasonably timed acquisition of CT doses, and seeing such doses resulted in 30-50% dose reduction.

[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
2022 submission:
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.
2016 submission:
Consecutive sample of CTs conducted in the head, chest, abdomen/pelvis and chest/abdomen/pelvis. No examinations should be excluded

[Response Ends]

sp.15. Provide details needed to calculate the denominator.

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.
Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]
2022 submission:
Target population
The target population includes all diagnostic CT exams of specified anatomic areas (skull, brain, abdomen and pelvis) performed in children aged 0-17 years during the measurement period. These can be most easily identified using CPT® codes (see below) but can also be identified using specific protocol names available in the Picture Archiving and Information System (PACS) or Radiology Information System (RIS), which correspond to the CPT® codes and descriptions below.
Of note, examinations that are not diagnostic CT are not included. These include: CT examinations performed in conjunction with nuclear medicine (such as SPECT and PET-CT) or as part of diagnostic procedures such as a biopsy or interventional therapeutic procedures; examinations performed as part of surgical planning or radiation therapy; and those where the anatomic area is not specified or where no primary images were obtained. These have different CPT® codes and are not included in the measure.
Table sp-15-1. CPT® codes and descriptions.
Cells marked with a dash (-) are left intentionally blank.

SKULL	**-**
70480	Computed tomography, orbit, sella, or posterior fossa or outer, middle, or inner ear; without [contrast material]
70481	Computed tomography, orbit, sella, or posterior fossa or outer, middle, or inner ear; with[contrast material]
70482	Computed tomography, orbit, sella, or posterior fossa or outer, middle, or inner ear; without [contrast material], followed by contrast material[s] and further sections
70486	Computed tomography, maxillofacial area; without [contrast material]
70487	Computed tomography, maxillofacial area; with [contrast material]
70488	Computed tomography, maxillofacial area; without [contrast material], followed by contrast material[s] and further sections
70450	Computed tomography, head or brain; without [contrast material] in conjunction with an ICD-10-CM code that identifies this as an exam done to evaluate a ventricular shunt (see below). These are rare examination types, and therefore if an entity cannot identify ICD-10-CM codes, these examinations should be included in the brain category.
-	**ICD-10-CM codes and descriptors for ventricular shunt evaluation. These should be identified in combination with CPT® code 70450 to consider a head CT as a skull exam.**
T8501XA	Breakdown (mechanical) of ventricular intracranial (communicating) shunt, initial encounter
T8501XD	Breakdown (mechanical) of ventricular intracranial (communicating) shunt, subsequent encounter
T8501XS	Breakdown (mechanical) of ventricular intracranial (communicating) shunt, sequela
T8502XD	Displacement of ventricular intracranial (communicating) shunt, subsequent encounter
T8502XS	Displacement of ventricular intracranial (communicating) shunt, sequela
T8503XA	Leakage of ventricular intracranial (communicating) shunt, initial encounter
T8503XD	Leakage of ventricular intracranial (communicating) shunt, subsequent encounter
T8503XS	Leakage of ventricular intracranial (communicating) shunt, sequela
T8509XA	Other mechanical complication of ventricular intracranial (communicating) shunt, initial encounter
T8509XD	Other mechanical complication of ventricular intracranial (communicating) shunt, subsequent encounter
T8509XS	Other mechanical complication of ventricular intracranial (communicating) shunt, sequela
T85730A	Infection and inflammatory reaction due to ventricular intracranial (communicating) shunt, initial encounter
T85730D	Infection and inflammatory reaction due to ventricular intracranial (communicating) shunt, subsequent encounter
T85730S	Infection and inflammatory reaction due to ventricular intracranial (communicating) shunt, sequela

Cells marked by a dash (-) are intentionally blank

BRAIN	**-**
CPT® codes	**Descriptions**
70450	Computed tomography, head or brain; without [contrast material]
70460	Computed tomography, head or brain; with contrast material[s]
70470	Computed tomography, head or brain; without [contrast material], followed by contrast material[s] and further sections
0042T	Cerebral perfusion analysis using computed tomography with contrast administration, including post-processing of parametric maps with determination of cerebral blood flow, cerebral blood volume, and mean transit time
70496	Computed tomographic angiography, head, with contrast material[s], including noncontrast images, if performed, and image post-processing

ABDOMEN AND PELVIS	-
CPT® codes	**Descriptions**
72191	Computed tomographic angiography, pelvis, with contrast material[s], including non-contrast images, if performed, and image post-processing
72192	Computed tomography, pelvis without contrast
72193	Computed tomography, pelvis; with [contrast material]
72194	Computed tomography, pelvis; without [contrast material] in one or both body regions, followed by contrast material[s] and further sections in one or both body regions
74150	Computed tomography, abdomen; without [contrast material]
74160	Computed tomography, abdomen; with contrast material[s]
74170	Computed tomography, abdomen; without [contrast material], followed by contrast material[s] and further sections
74174	Computed tomographic angiography, abdomen and pelvis, with contrast material[s], including noncontrast images, if performed, and image post-processing
74175	Computed tomographic angiography, abdomen, with contrast material[s], including noncontrast images, if performed, and image post-processing
74176	CT scan of abdomen and pelvis without contrast
74177	Computed tomography, abdomen and pelvis; with [contrast material]
74178	Computed tomography, abdomen and pelvis; without [contrast material] in one or both body regions, followed by contrast material[s] and further sections in one or both body regions
74261	Computed tomographic colonography, diagnostic, without contrast material
74262	Computed tomographic colonography, diagnostic, with contrast material(s), including non-contrast images if performed
74263	Screening CT scan of large bowel
75635	CT angiography, abdominal aorta with bilateral iliofemoral lower extremity runoff, with contrast material, including noncontrast images, if performed

Cells marked by a dash (-) are intentionally blank

Minimum sample size required

For scoring approach 1 (assessment of the facility's median value for each stratum), 11 exams within each age and anatomic area stratum are required for skull and abdomen and pelvis categories, and 25 exams for the skull category. For scoring approach 2 (assessment of the facility's overall proportion of exams), a total of 23 exams (across all age and anatomic area strata) are required.

The rationale for these sample size requirements is provided in sp.25.

Time Period for Data Collection

One year. The rationale for this time period is that many facilities do not reach the minimal sample size in a shorter duration. Measure implementers may define their own 12-month periods; it does not need to be a calendar year. However, all exams are compared to the same set of benchmarks, and measure score calculation does not take the specific time period into consideration. Thus, measure scores are comparable regardless of the 12-month period selected.

2016 submission:

Consecutive sample of CTs conducted in the head, chest, abdomen/pelvis, chest/abdomen/pelvis

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

2022 submission:

Examinations with missing anatomic area, patient age, or missing dose length product are excluded.

2016 submission:

CT examinations conducted in anatomic areas not included above (such as CTs of the extremities or lumbar spine) or that combine several areas (head and chest) should not be included. In children, these four included categories will reflect approximately 80% of CT scans.
Examinations performed as part of diagnostic procedures – such as biopsy procedures – should not be included. CT examinations performed as part of surgical planning or radiation therapy should not be included.
Examinations that are considered "limited abdomen" or "limited pelvis" studies should be included in the abdomen and pelvis category. Any examinations that include any parts of the abdomen and or pelvis should count in the abdomen/pelvis category.

[Response Ends]

sp.17. Provide details needed to calculate the denominator exclusions.

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]
2022 submission:
Missing data on anatomic area imaged, patient age, or radiation dose should not be included.
2016 submission:
Most abdominal/pelvis CT scans in adult patients include scanning of the abdomen and pelvis as one contiguous area. If examinations are conducted limited to one region, these should also be included, as it is difficult/impossible to define what areas would be considered limited.

[Response Ends]

sp.18. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]
2022 submission:
Anatomic areas stratum
These anatomic areas can be identified using specific CPT® codes or protocol names found in the radiology information systems (such as PACS or RIS) and specified in sp.15 above.
Skull: including all imaging of the facial skeleton, sinus, skull bones, or for the assessment of a ventricular shunt.
Brain: including imaging of the head not specified as part of skull and includes imaging for suspected hemorrhage, trauma, headache, altered mental status, seizures and all other indication for head CT not captured as part of skull imaging. This group should include the very small number of head CTs (<< 1%) that include perfusion angiography. Exams that include both the skull and brain as part of a single evaluation but cannot be separated into the component exams (e.g., performed as part of a single evaluation on the same date and time) should be included with brain imaging.
Abdomen and pelvis: including imaging for all abdomen and/or pelvis CT indications. Examinations that are considered "limited abdomen" or "limited pelvis" studies should be included in the abdomen and pelvis category as there is no reliable way to separate these types of examinations. The scan lengths are not very different between exams codified as abdomen, codified as abdomen and pelvis, or codified as limited pelvis. Thus examinations that include any parts of the abdomen and/or pelvis should count in the abdomen and pelvis category. Multiphase exams of the abdomen and pelvis should be included.
These three anatomic areas were chosen based on being the most common CT examination types conducted in the US, comprising >80% of all CT examinations in children, and because dose varies across these categories. (Chu 2021, Kanal 2021, Smith-Bindman 2021)
Age Strata
Infant (<1 year)

Small child (1-4 years)
Medium child (5-9 years)
Large child (10-14 years)
Adolescent (15-17)
These patient age groups were chosen based on the widespread practice of varying CT machine settings and the resulting radiation dose variation based on patient size or age (age is frequently used as a surrogate for size.) The International Commission on Radiation Protection (ICRP) uses these child size categories, which correspond to available phantoms. (ICRP publications 121 and 135) Other literature has similarly supported these age groupings. (Vassileva 2015).

Citations

1. Chu PW, Yu S, Wang Y, Seibert JA, Cervantes LF, Kasraie N, Chu CA, Smith-Bindman R. Reference phantom selection in pediatric computed tomography using data from a large, multicenter registry. Pediatr Radiol. 2021 Dec 6. doi: 10.1007/s00247-021-05227-0. Epub ahead of print. PMID: 34866159.
2. International Commission on Radiation Protection. ICRP Publication 121: Radiological protection in paediatric diagnostic and interventional radiology. Ann. ICRP 42(2).
3. International Commission on Radiation Protection. ICRP Publication 135: Diagnostic Reference Levels in Medical Imaging. Ann ICRP 2017;46(1):1–144.
4. Kanal KM, Butler PF, Chatfield MB, Wells J, Samei E, Simanowith M, Golden D, Gress DA, Burleson J, Sensakovic WF, Strauss KJ, Frush D. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Pediatric CT Examinations. Radiology. 2022 Jan;302(1):164-174. doi: 10.1148/radiol.2021211241. Epub 2021 Oct 26. Erratum in: Radiology. 2022 Jan;302(1):E6.
5. Vassileva J, Rehani M. Patient grouping for dose surveys and establishment of diagnostic reference levels in paediatric computed tomography. Radiat Prot Dosimetry. 2015 Jul;165(1-4):81-5. doi: 10.1093/rpd/ncv113. Epub 2015 Apr 1. PMID: 25836695.
6. Smith-Bindman R, Yu S, Wang Y, Kohli MD, Chu P, Chung R, Luong J, Bos D, Stewart C, Bista B, Alejandrez Cisneros A, Delman B, Einstein AJ, Flynn M, Romano P, Seibert JA, Westphalen AC, Bindman A. An Image Quality-informed Framework for CT Characterization. Radiology. 2021 Nov 9:210591. doi: 10.1148/radiol.2021210591. Epub ahead of print.

2016 submission:

Anatomic area strata: head, chest, abdomen/pelvis, chest/abdomen/pelvis
These were chosen based on being the most common CT examination types conducted in the US, comprising >80% of all CT scans, and because dose varies by these groups.
Age strata: infant (<1); small child (1-5); medium child (>5 - 10); large child (>10-15) and adult (>15)
These patient age groups were chosen based on the variation of CT settings and resulting radiation dose based on patient size (and age is frequently used as a surrogate for size.) The ICRU (International Commission on Radiation Units and Measurements) uses these child size categories, they correspond to available phantoms, and they are the ones found to be most reliable
Geographic location where studies were done (zip code or state), to facilitate using the data to create geographically specific benchmarks

[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.
[Response Begins]
Stratification by risk category/subgroup (specify number of risk factors)
[Stratification by risk category/subgroup (specify number of risk factors) Please Explain]
Radiation doses primarily vary by age and the body region. The measure is specified in 5 age groups and 3 anatomic areas for a total number of 15 subgroups. There are no other risk factors that need to be taken into account or risk adjusted

[Response Ends]

sp.20. Select the most relevant type of score.

Attachment: If available, please provide a sample report.
[Response Begins]

Rate/proportion
[Response Ends]

sp.21. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score
[Response Begins]
Passing score defines better quality
[Response Ends]

sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]
2022 submission:
1. Each diagnostic CT examination performed within the 12month period is assessed for inclusion based on non-missing anatomic area, patient age, and radiation dose data.
2. Radiation dose (DLP) is recorded for all included exams.
3. The DLP is compared to the benchmark (75^{th} percentile) value for that anatomic area-age specific stratum.
4. The numerator for the measure documents whether the DLP is above the benchmark stratum.
5. The total number of scans above the benchmark is calculated (aggregated) for each anatomic area-age stratum.
6. The total proportion of CT examinations with DLP greater than the corresponding 75^{th} percentile benchmark across all categories is calculated.
7. Performance is classified for each stratum (median) and overall (proportion of high dose exams) according to the scale described in sp.13. If the median is above the 75% percentile benchmark for a stratum, the hospital or facility is considered to have a poor dose distribution in that category. If the overall proportion of high dose exams exceeds 50% then the overall dose distribution is considered poor.
2016 submission:
N/A

[Response Ends]

sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

[Response Begins]
2022 submission:
For assessment of the facility's median value for each stratum and whether it is acceptable or poor
For skull and abdomen and pelvis categories, 11 exams within each age and anatomic area stratum will provide adequate sample size. For the brain category, 25 exams within each age and anatomic area stratum will provide adequate sample size. One year of data should be extracted for reporting.
These minimums were selected to ensure reliability and validity (i.e. "detectability of results of interest") of the measure score. We selected the larger minimum between reliability and validity, displayed in Table sp-25, as the required minimum sample size.
The sample size for brain is higher because there is less variation in doses in this category, and the distribution for this category is much less skewed, and therefore requires more sample size to identify outliers.
For assessment of the overall proportion of exams above the 75% benchmark (e.g., proportion exceeding benchmarks)
A total number of CT exams (across all age and anatomic area strata) of 23 will allow a reliable assessment of the proportion exceeding benchmarks and assessment of whether the proportion is over 50% (i.e. twice the expected rate of 25%).
This minimum was selected to ensure reliability and validity (i.e. "detectability of results of interest") of the measure score.
Table sp-25. Minimum sample sizes by anatomic area.

Method of Testing	Skull	Brain	Abdomen	All Body Regions
Reliability	11	11	11	*
Validity (detectability of results of interest)	9	25	9	23
Minimum Required Sample Size	11	25	11	23

* Cell intentionally left empty

2016 submission:

For child categories, 25 patients within each strata will provide adequate sample size. (One year of data should be extracted if the minimum cannot be met within a shorter time interval).

[Response Ends]

sp.28. Select only the data sources for which the measure is specified.

[Response Begins]

Electronic Health Data
Electronic Health Records
Registry Data

[Response Ends]

sp.29. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

2022 submission:

The measure derives standardized data elements from structured fields stored electronically, including:

1. Type of CT examination (i.e., anatomic area imaged)
2. Radiation dose (DLP) stored electronically in standardized DICOM format
3. Patient age

The data can be extracted either manually or automatically from several sources:

1. Derived directly from the CT scanner at the time of examinations;
2. Derived from the Picture Archiving and Communication System (PACS), which is the electric system where imaging data are stored and reviewed; or the Radiology Information System (RIS)
3. Derived from the electronic health record (EHR), where many facilities – whether by custom or law – store radiation dose information.
4. Derived from widely used commercial radiation dose software programs such as Dose Watch, PACS Health and Radimetrics.

We have also published several techniques for dose extraction that can be completed even by small facilities. (Keegan, JACR 2014)

Citations

1. Keegan J, Miglioretti DL, Gould R, Donnelly LF, Wilson ND, Smith-Bindman R. Radiation Dose Metrics in CT: Assessing Dose Using the National Quality Forum CT Patient Safety Measure. Journal of the American College of Radiology: JACR; 11(3):309-315.

2016 submission:

The data sources will include electronic CT images [captured from the CT console at the time of scanning or harvested from the PACS (Picture Archiving Communication System) - the computerized systems for reviewing and storing imaging data], Radiology Information System, EPIC, printed CT images, or information stored in the medical record. Numerous other software products are now available for capturing these data (Bayer, GE, etc.) and several free ware programs are also available. Of note, the 2012 California law now requires the reporting of several of the dose metrics outlined in this measure in the patient medical record, and as a results, many software companies have provided techniques for collating these data.

[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]
No data collection instrument provided
[Response Ends]

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:
Current Submission:
Updated testing information here.
Previous Submission:
Testing from the previous submission here.

[Response Begins]
Yes
[Response Ends]

2ma.02. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:
Current Submission:
Updated testing information here.
Previous Submission:
Testing from the previous submission here.

[Response Begins]
Yes
[Response Ends]

2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

[Response Begins]
Yes
[Response Ends]

2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]
No additional risk adjustment analysis included
[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

• Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
• All required sections must be completed.
• For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
• If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
• An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
• Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage .
• For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance .

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.
2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.
2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.
2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
2b3. For outcome measures and other measures when indicated (e.g., resource use):
• an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
OR
• rationale/data support no risk adjustment/ stratification.
2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;
OR
there is evidence of overall less-than-optimal performance.
2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.
2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.
(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions
Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
Risk factors that influence outcomes should not be specified as exclusions.
With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:
2021 Submission:
Updated testing information here.
2018 Submission:
Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]
Electronic Health Data
Electronic Health Records
Registry Data
[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]
2022 submission:

The measure was tested in two separate datasets: UCSF International CT Dose Registry and the Leapfrog Group implementation data. Details on these data sources are provided in 2a.05.

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]
2022 submission:
UCSF International CT Dose Registry: 01/01/2016 - 01/01/2021
The Leapfrog Group implementation data: 01/01/2019 – 06/30/2021
2016 submission:
January 1, 2008 – December 31 , 2013

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Facility
[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]
2022 submission:
For the 2022 maintenance review, the measure has been tested in two additional environments named in 2a.02. These include:
UCSF International CT Dose Registry: a database of over 6.5 million CT exams from 161 healthcare facilities across 7 countries (though over 80% of examinations come from United States facilities in urban, suburban, and rural areas). Of these facilities, 143 contributed examinations in children, including 5 pediatric hospitals. Of examinations included in the Registry, 116,597 head and abdomen examinations were performed in children and included in this testing report. These examinations reflect submission of all CT examinations during the period of reporting (e.g. they reflect consecutive examinations).
The Leapfrog Group implementation data: the Leapfrog Group has been actively using this measure in its "Leapfrog Hospital Survey," which collects and publicly reports quality and safety information provided voluntarily by hospitals. As of 12/3/21, 1447 hospitals had reported on this measure from all 50 US states and the District of Columbia, including 24 pediatric hospitals. Together, these sites provided data on 349,278 head and abdomen and pelvis examinations in children, which are included in this testing report.
2016 submission:

The measure has been tested in several settings: Group Health Research Institute, a large integrated Health System in the Pacific Northwest. CT examinations on over 10,000 examinations have been assembled and included in several publications. (Miglioretti, JACR 2014; Miglioretti JAMA Pediatrics 2013)
The measure was tested in a consortium of integrated health care systems (n=6) and data were assembled for over 5000 CT examinations, and were published. (Smith-Bindman JAMA 2012)
The measure was tested across the five University of California Medical Centers, including over 100,000 CT examinations. The data has in part been published (Keegan, JACR 2014) and additional manuscripts were presented at national meetings (RSNA 2012) and are in preparation. A manuscript is in press describing the results of assessment of dose using this measure using data from across the University of California (In press, Radiology) and a second paper is under preparation demonstrating a 10-30% reduction in dose using a before and after design using assessment as specified in this measure.
For all, analyses were done using consecutive sample of CT examinations within anatomic area, age and machine type stratum as specified in this measure, or using a randomly selected subset of examinations and analyzed per measure specifications.
A quality improvement activity assembling data per the NQF specifications was approved by the Board of the American College of Radiology for PQRS credit.

Citations

Miglioretti D, Johnson E, Vanneman N, Smith-Bindman R, al e. Use of Computed Tomography and Associated Radiation Exposure and Leukemia Risk in Children and Young Adults across Seven Integrated Healthcare Systems from 1994 – 2010. JAMA Pediatrics Published online June 10, 2013 joli:101001/jamapediatrics2013311 2013.
Miglioretti DL, Zhang Y, Johnson E, Lee C, Morin RL, Vanneman N, Smith-Bindman R. Personalized technologist dose audit feedback for reducing patient radiation exposure from CT. J Am Coll Radiol. 2014 Mar;11(3):300-8. doi: 10.1016/j.jacr.2013.10.017.
Smith-Bindman R, Lipson J, Marcus R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med 2009;169:2078-86.
Smith-Bindman R. Is computed tomography safe? N Engl J Med 2010;363:1-4.
Smith-Bindman R. Environmental causes of breast cancer and radiation from medical imaging: findings from the Institute of Medicine report. Arch Intern Med 2012;172:1023-7.
Smith-Bindman R, Miglioretti DL, Johnson E, et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. JAMA 2012;307:2400-9.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

Table 2a.06 Total number of CT scans per year in the UCSF International CT Dose Registry and The Leapfrog, and the sex and age distributions of these scans.

2022 submission:

*	*	**Sex (proportion)**		**Age (proportion)**	*	*	*	*
*	Average CT Exams per Year	Female	Male	< 1 year	1-4 years	5-9 years	10-14 years	15-17 years
UCSF International CT Dose Registry	23,319	0.45	0.55	0.08	0.15	0.19	0.27	0.31
The Leapfrog Group implementation data	141,542	NA	NA	0.07	0.14	0.16	0.29	0.34

*Cells marked by an asterisk are intentionally left blank.

2016 submission:
The summary of CT dose has been done in children and adults, and using consecutive scans without exclusion (i.e. scans were not excluded on any individuals) and analyzed within stratum. Because the measure is specified at the institutional level, there is no reason to exclude any individuals. While there are individual patients who will and should have doses above averages, the measure calls for assessment of institutional data, and individual patients will have a small impact, if any, on overall calculations.
The stratum for this measure include:

- Anatomic area stratum: head, chest, ~~abdomen/~~abdomen and pelvis. These anatomic areas reflect approximately 85% of CT examination types in adults, and approximately 75% of CT examination types in children
- Age stratum: infant (<1); small child (1-5); medium child (>5 - 10); large child (>10-15) and adult (>15)
- CT machine (manufacturer, type)

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]
2022 submission:
N/A
2016 submission:
There are no differences in the data or sample used for different aspects of testing.

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]
2022 submission:
These were not available, nor tested. Social factors are not known to affect radiation dose, because technical decisions on how to perform CT are made at the hospital or outpatient imaging facility's level rather than at the individual patient level. See 2b.25 below for additional discussion.
2016 submission:
These were not available, nor tested.

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.
[Response Begins]
Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)
Accountable Entity Level (e.g., signal-to-noise analysis)
[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

2022 submission:

We assessed the reliability of our hospital classification mechanism using the UCSF International Dose Registry.
Given a combination of hospital, anatomic area (skull, brain, abdomen), and age group, we record whether this hospital would be deemed poor (hospital median DLP greater than the population 75th percentile threshold) or acceptable (hospital median DLP less than or equal to the population 75th percentile threshold) within this anatomic age and age group. We consider this the "true class" of this hospital, within this anatomic area and age group.
We then sample, with replacement, a number of exams from this hospital, within this anatomic area and age group, equal to the number of observed exams from this hospital, within this anatomic area and age group. This sampling is repeated 1000 times to simulate 1000 versions of this hospital, within this anatomic area and age group. Each of these 1000 simulated hospitals is also classified, based on its median DLP, as poor or acceptable.
To assess the relationship between the sample size of a hospital and its reliability, we split the hospitals, within anatomic area and age group, in the UCSF International Dose Registry into 11 subsets, with boundaries defined by the deciles of the distribution of sample sizes. Within each subset, we compute two metrics:

1. The Classification Rate – This is defined within a hospital subset as the prevalence of agreement between a randomly-selected hospital within the subset and its 1000 simulated versions.
2. The Cohen's Kappa – This is defined for a single collection of simulated hospitals (out of a total of 1000 collections) as the ratio $(p_0-p_e)/(1-p_e)$, where p_0 is the prevalence of agreement between the simulated hospitals and their "true" counterparts, and p_e is the hypothetical prevalence of agreement by random chance. This is defined within a hospital subset as the expected Cohen's Kappa between a randomly-selected collection of simulated hospitals and their "true" counterparts.

Lastly, we recommend the sample size associated with a classification rate of at least 90% and a Cohen's Kappa of at least 0.81, both traits associated with "near-perfect" agreement between repeated simulations.
Any combination of hospital, anatomic area, and age group whose observed sample size is 1 or 0 are ignored for purposes of this analysis.
Reliability testing was performed for scoring approach 1 (classifying a facility's performance as "acceptable" or "poor" based on whether its median radiation dose is below, or greater than or equal to, the benchmark 75th percentile per anatomic area and age strata). We did not test reliability of scoring approach 2 (classifying a facility's performance on based on its median radiation dose overall). The rationale for this decision is explained in 2a.12.

2016 submission:

The Proposed CT Dose measure calls for the collection of several metrics reflecting CT dose indices including DLP, CTDIvol, and SSDE. CTDI and DLP are calculated automatically by all current CT scanners, without variability. When these data are manually extracted, there is possibility of errors of writing down the values and has been found to be in the ballpark of a 5-10% error related to transcription, not calculation (Keegan JACR). SSDE is a calculated variable, and while dose monitoring programs automatically calculate this variable, sites that choose to calculate this manually will likely introduce errors, although this has not been quantified.
CTDIvol and DLP measures have been widely used for over a decade in several other countries, are used for in a bill that is in effect in California, and I have personal and recent experience collecting these dose Indices across 12 large institutions reflecting dozens of machines and thousands of patients. Reliability of CT radiation dose metric abstraction (DLP and CTDIvol) was tested by our group in several ways. First, manual data abstraction of data recorded from the PACS system was repeated in two large samples (one at Group Health, and one at UCSF) where the data was abstracted by a single observer, yielding highly reliable measures between abstractions (i.e. the measures were concordant, nearly perfect Kappa statistics, with a 5% variation in our analysis.) Second, data were extracted via commercial software product using two different tools for extracting the data from the stored CT files in PACS, and these were reviewed by a medical physicist to ensure the data were correct. This was performed at five separate institutions, and found the electronically captured data was identical to the manual review, perfect Kappa statistics.
SSDE is a relatively new metric that tries to take into account patient size. In our published work (Keegan JACR 2014) it tracks in parallel to the other metrics, but has not undergone formal reliability testing in large cohorts of patients.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]
2022 submission:
The relationship between observed sample size and classification rate does not seem to differ by anatomic are or age group (Figure 2a.11), allowing reliability of our measure to be assessed holistically.

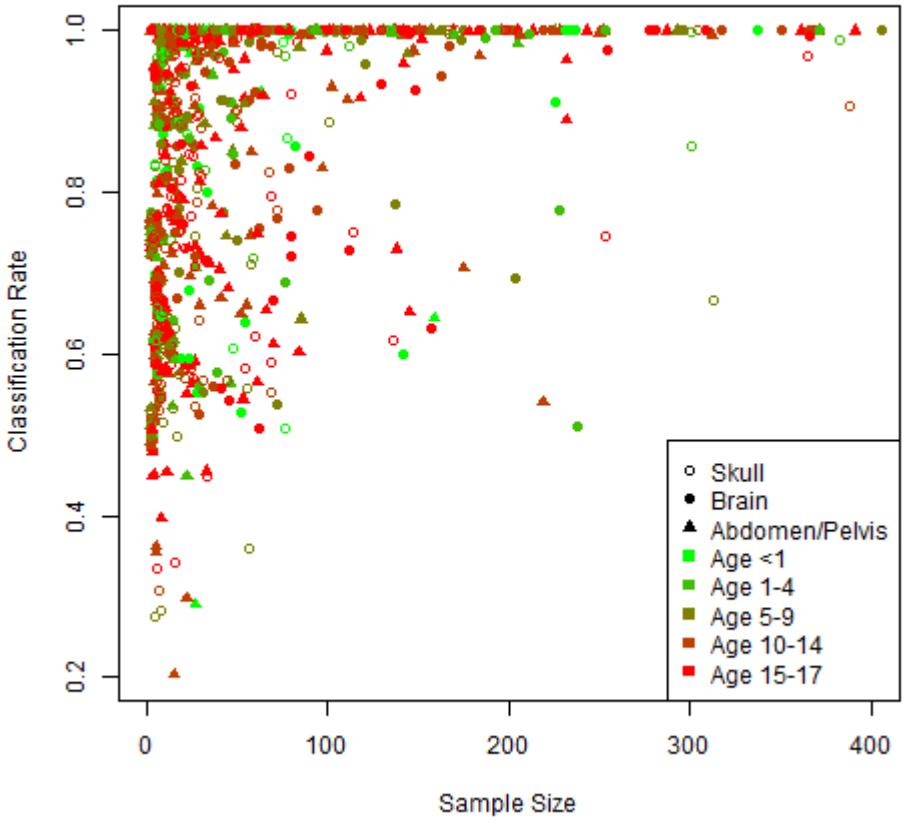


Figure 2a.11 – Scatter plot of observed sample size and classification rate, separated by anatomic area and age group. Though sample sizes observed within the UCSF Registry can be as high as 2084, the majority (95%) are below 400. Thus, the scatter plot is zoomed in for better view of where most data lie.

On average, classification rates remained high, with even low-volume hospital-stratum combinations having a mean classification rate ~90% (Table 2a.11). However, the mean classification rate is unstable for low sample sizes, and does not consistently exceed 90% until a sample size to 8 to 11 is observed.
Similarly, the Cohen's Kappa consistently exceeds 0.81 once the observed sample size reaches 8 to 11 (within a hospital-stratum combination).

Table 2a.11 – Mean classification rate and Cohen's Kappa by sample size.

Observed Sample Size	Mean Classification Rate	Cohen's Kappa
2	0.93	0.84
3	0.88	0.82
4	0.91	0.77

Observed Sample Size	Mean Classification Rate	Cohen's Kappa
5	0.87	0.74
6	0.90	0.73
7	0.90	0.84
8	0.92	0.77
9	0.91	0.89
10	0.94	0.82
11	0.91	0.90
12	0.93	0.87
13-18	0.93	0.83
19-29	0.95	0.87
30-54	0.94	0.84
55-102	0.97	0.88
103-226	0.99	0.89
>226	0.99	0.88

2016 submission:
Highly reliable, Kappas > 95%

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]
2022 submission:
The sample size of 8-11 acquired in section 2a.11 ensures that a facility's performance within an anatomic area and age stratum (i.e. whether the hospital is deemed as "poor") remains the same after a simulated re-sampling 90% of the time (classification rate). 11 was selected as the minimal sample size for reliability, as it happens to also coincide with the minimal sample size required for a Cohen's Kappa of 0.90, which is a measure comparable to classification rate but which makes adjustments to account for the possibility of re-sampling agreeing by chance.
Among the 106 UCSF International Dose Registry hospitals who submitted ten observed pediatric exams within at least one anatomic area and age group, 71% have sufficient annualized sample size to achieve reliability in at least one combination of anatomic area and age group.
Among the 5 dedicated pediatric hospitals in the UCSF International Dose Registry, 100% have sufficient annualized sample size to achieve reliability in at least one combination of anatomic are and age group.
Among the 1340 Leapfrog Group Implementation Data hospitals who submitted ten observed pediatric exams within at least one anatomic area and age group, 72% have sufficient annualized sample size to achieve reliability in at least one combination of anatomic area and age group.
Among the 24 dedicated pediatric hospitals in the Leapfrog Group Implementation Data, 92% have sufficient annualized sample size to achieve reliability in at least one combination of anatomic are and age group.
There are indeed a number of hospitals only able to achieve reliability within a single stratum. This is because hospitals vary tremendously in whether or not they evaluate pediatric patients. Extremely large hospitals are likely to see pediatric patients even if there is a pediatric hospital nearby. Extremely small hospitals in rural areas are also likely to see a high number of pediatric patients because there isn't a dedicated pediatric hospital nearby. For all hospitals in between, their likelihood of seeing pediatric patients will vary by a large number of factors, most importantly the availability of alternative places for patients to get care.
However, this is not expected to affect the assessment of a hospital's prevalence of individual exams over the threshold (i.e. - the "overall" measure score). This is due to a combination of two factors. First is that benchmark values are selected

such that each age and anatomic area stratum is expected to observe 25% of exams over the threshold, and thus every hospital has the same expected prevalence of exams over the threshold (25%) regardless of its distribution of age and anatomic area strata. Second is that the relationship between sample size and classification rate does not seem to differ by age and anatomic area strata, as shown in figure 2a.11; one may conjecture then that the minimal sample size requirements for a hospital's "overall" reliability also would not depend on a hospital's distribution of age and anatomic area strata.
Note, reliability testing was performed for scoring approach 1 (classifying a facility's performance as "acceptable" or "poor" based on whether its median radiation dose is below, or greater than or equal to, the benchmark 75th percentile per anatomic area and age strata). We did not test reliability of scoring approach 2 (classifying a facility's performance on based on its median radiation dose overall). However, considering that the relationship between sample size and classification rate doesn't seem to differ between different age and anatomic area strata (figure 2a.11), one may conjecture that the sample size requirements for reliability at the overall hospital level would be comparable to sample size requirements for reliability within age and anatomic area strata. Importantly, when we define the minimum sample size, it is driven based on validity testing (required larger sample sizes than reliability testing).
2016 submission:
Highly reliable, Kappas > 95%

[Response Ends]

2b.01. Select the level of validity testing that was conducted.

[Response Begins]
Patient or Encounter-Level (data element validity must address ALL critical data elements)
Empirical validity testing
[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]
2022 submission:
Patient/encounter-level (data element) validity
Anatomic area: reporting facilities should assign examinations to one of the three anatomic areas based on the CPT® and ICD-10-CM codes (provided in sp-15) or protocol names that determine how they image. The protocol names are stored in the Picture Archiving and Communication System (PACS) and Radiology Information System (RIS) and specify at a minimum the anatomic region imaged, which can be used to assign CT scans to the correct anatomic area. Thus it is straightforward to assign each CT to one of the included anatomic areas.
In adults, the framework for assigning CT category is more complicated as it requires knowing both the anatomic area and the specific indication for whey the CT was done. The framework of assigning CT exams to the various CT categories in adults – which are distinguished based on the radiation dose and image quality required for underlying clinical indications – was developed using over 4.5 million CT exams in the UCSF International CT Dose Registry; this work has undergone peer review (Smith-Bindman, 2021). This work, while more complicated than in children, demonstrates the validity of assignment of CT exams to specific categories and is thus included.
In developing a related CT quality measure for adults that assigns CT categories by an algorithm that uses CPT® and ICD-10-CM codes, we used criterion validity to compare agreement between the CT category assigned using this method versus a gold standard method based on expert review of the complete medical record (including notes from the visit when the exam was ordered, information provided as free text with the test order, and information included in the final, dictated radiology report) for a sample of 10,000 CT exams from UCSF Health System.
Radiation Dose: The measure uses dose length product (DLP), which gives the total radiation imparted to the patient by the CT machine. This is a standardized data element, generated by virtually (>99%) all CT machines, is well validated and used broadly to reflect the radiation dose delivered to the patient. (Kanal 2017, Smith-Bindman 2019.) Further, DLP is frequently used in benchmarking in the U.S. and internationally (ACR–AAPM–SPR: Practice parameter, European Commission, Radiation Protection No. 185, ICRP Publication 135). While there are other dose metrics used in some settings to measure radiation dose (such as size-specific dose estimate (SSDE) or effective dose), these are not suitable

for a reliable quality measurement because they are not universally or automatically generated by the CT machine, and do not reflect the total dose absorbed by the patient (the most clinically relevant measure). Absorbed doses will vary by sex and weight, but are primarily determined by the doses delivered by the machine. DLP is highly correlated with patient absorbed doses; higher DLPs are associated with higher absorbed dose to the patient's organs and higher patient harm. Conversely, when DLP is lowered, patients are exposed to lower doses of radiation, have correspondingly lower absorbed organ doses, and would be expected to have less detriment from these exposures. Yet while patient absorbed doses are important, they are difficult to quantify. On the other hand, the output of radiation from the machine (DLP) is simple to measure, and most importantly, it can be influenced by the imaging physician or technologist. DLPs generally reflect the average CT dosing at the hospital or outpatient facility level.

To validate this data element, we relied on this published work and tested the availability of clinically plausible and non-missing values for radiation dose in the UCSF International CT Dose Registry.

Empirical validity testing

We have not assessed the relationship between this measure score and other related outcome measures, because there are no validated outcome measures related to radiation dose or medical imaging in children or adults. However, there are organizational factors and processes associated with high quality care at the accountable entity level and we have shown that these are associated with radiation dose levels. Our past work has also shown empirically that providing hospitals with audit feedback on their CT radiation doses, similar to the information provided by this measure score, results in meaningful radiation dose reduction without diminishing the diagnostic usefulness of these tests.

UCSF, as the measure developer, has been working with 26 health care organizations and 161 imaging facilities for 10 years on ways to assess radiation dose and provide feedback to help organizations improve quality and safety of CT imaging. As part of NIH and PCORI funded research, UCSF identified organizational factors and quality improvement strategies associated with radiation doses from abdominal CT. (Solberg 2020) Organizational factors were assessed using Physician Practice Connection (PPC), a questionnaire originally created by the National Committee for Quality Assurance to assess implementation of the Chronic Care Model systems in primary care and from The Change Process Capability Questionnaire (CPCQ), Part 1 on Practice Readiness and Part 2 on Use of Change Strategies. Additional questions about each organization's CT protocols, quality improvement experience and use of the dose measurement software and priority for improving radiation dosage were asked.

To determine organizational leaders' perceptions of barriers to optimizing radiation dose in CT, Whitebird et al. conducted an observational study using semi-structured interviews conducted with 26 organizational leaders from 19 health care systems in the United States, Europe, and Japan. (Whitebird, 2021, Whitebird 2022 in press) Interviews focused on approaches the organizations used to optimize radiation dose and barriers encountered. Data were analyzed using a directed content analysis approach. Detailed qualitative surveys were also conducted with hospital leaders to understand barriers and facilitators to radiation dose optimization.

UCSF also lead a randomized controlled trial involving roughly 1 million CT exams from 100 imaging facilities across 6 countries to determine the impact of multicomponent educational feedback on radiation doses used in CT imaging. (Smith-Bindman 2020) Hospitals and imaging facilities were provided with their median and 75th percentile in dose distribution for head and abdomen CT (highly similar data to what this measure assembles), along with targeted suggestions for improvement, and participation in a quality improvement collaborative.

Citations

1. ACR–AAPM–SPR: Practice parameter for diagnostic reference levels and achievable doses in medical x-ray imaging. Revised 2018.
2. European Commission, Radiation Protection No. 185, European guidelines on diagnostic reference levels for paediatric imaging. 2018.
3. 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.ACR–AAPM–SPR: Practice parameter for diagnostic reference levels and achievable doses in medical x-ray imaging. Revised 2018.
4. ICRP, 2017. Diagnostic reference levels in medical imaging. ICRP Publication 135. Ann. ICRP 46(1).
5. 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.
6. 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.
7. Smith-Bindman R, Yu S, Wang Y, Kohli MD, Chu P, Chung R, Luong J, Bos D, Stewart C, Bista B, Alejandrez Cisneros A, Delman B, Einstein AJ, Flynn M, Romano P, Seibert JA, Westphalen AC, Bindman A. An Image Quality-informed Framework for CT Characterization. Radiology. 2021 Nov 9:210591. doi: 10.1148/radiol.2021210591. Epub ahead of print.

8. Solberg LI, Wang Y, Whitebird R, Lopez-Solano N, Smith-Bindman R. Organizational Factors and Quality Improvement Strategies Associated With Lower Radiation Dose From CT Examinations. J Am Coll Radiol. 2020 Jul;17(7):951-959. doi: 10.1016/j.jacr.2020.01.044. Epub 2020 Mar 17. PMID: 32192955; PMCID: PMC7338232.
9. Whitebird RR, Solberg LI, Bergdall AR, López-Solano N, Smith-Bindman R. Barriers to CT Dose Optimization: The Challenge of Organizational Change. Acad Radiol. 2021 Mar;28(3):387-392. Doi: 10.1016/j.acra.2020.02.016. Epub 2020 Apr 9. PMID: 32278691.
10. Whitebird R, Solberg L, Chu P, Smith-Bindman R. Strategies for Dose Optimization: Views from Health Care Systems. JACR. 2021. In press.

2016 submission:

The three dose parameters (metrics) included in this measure reflects slightly different aspects of dose, and each was included because it provides a unique reflection of dose and can be used to improve quality and safety.

These dose parameters specified in this measure primarily reflect the dose that comes out of the machine and the dose that the patient is exposed to and dictate the absorbed organ doses to the patient. Absorbed doses (these are the doses a patient actually receives) will vary by sex and weight, but are primarily determined by the doses that come out of the machine. These dose metrics are highly correlated with the doses patients receive; higher DLPs, CTDIs, and SSDE are associated with higher absorbed dose to the patient's organs and higher patient detriment (harm). If these doses were lowered patients would be exposed to lower doses of radiation, have correspondingly lower absorbed organ doses and would be expected to have less detriment from these exposures to radiation. While patient absorbed doses are important, they are difficult to quantify.

However, the dose parameters themselves are vitally important as they 1) closely reflect organ doses and 2) are precisely those measurements that the technologist and physician can influence to lower doses. That is why these metrics were chosen for this measure. Estimating absorbed organ doses might be a more precise way to compare doses between two examinations on two patients. However, this is simply not practical. It is much more complicated to estimate these parameters, there are over 30 different organs where these doses can be compared and it does not make sense to measure because the technologist cannot directly influence these measures, and there would be practical way to compare facilities as there ware so many organ doses to compare. Using organ dose might add a very small amount more precision for an estimate of an individual patient, but it's not clear that it's relevant or possible to measure and compare at the facility level. Thus organ dose was not proposed as a practical or useful metric for patient safety assessment.

The output of radiation from the machine is far simpler to measure and in fact is the important variable, as this is what the radiologist and the technologist can influence. The measures are primarily proposed to reflect the average CT dosing at the institutional level and small variations in patient size will average out across institutions.

We have conducted comparison of each of the dose metrics with measures of absorbed dose among a sample of 10,000 CT examinations and the correlations are high (> 90%). Further, the correlation within the metrics is also high. Details of this comparison were provided at the time of consideration of this measure when it was first endorsed. The organ doses were calculated by Dr. Choonsik Lee, PhD an Investigator in the Radiation Epidemiology Branch, in the Division of the Cancer Epidemiology and Genetics at the National Cancer Institute. His research includes the development of dosimetry databases and Monte Carlo dose calculations using human models that permit estimating absorbed radiation dose that takes into account patient weight. His method for estimating organ doses has been validated against direct measurement.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

2022 submission:

Patient/encounter-level (data element) validity

Anatomic area: In developing the framework to assign CT exams to CT categories, we validated the CT categories as being representative of groupings that require different radiation dose and image quality in part by demonstrating significantly greater differences in radiation doses between categories than within categories within the body regions that are subdivided into low, routine, and high dose (N = 4.5 million exams). (Smith-Bindman 2021).

We validated our method of assigning CT category based on CPT® and ICD-10-CM codes against gold standard chart review. The results, weighted by the distribution of CT categories in the UCSF International CT Dose Registry, were: sensitivity = 0.86 and specificity = 0.96 (n=978 CT exams).

Radiation Dose: In the UCSF Registry, in children, DLP was reported and within plausible range for 99.6% of CT examinations.

Empirical validity testing

Process of care surveys were received from 90 imaging centers (90%), and 182,415 abdominal CT scans were collected during the study period. (Solberg 2020) The association between process factors and dose found that after adjusting for patient age, gender, and size, quality improvement strategies were association with mean abdominal CT radiation dose and the odds of a high-dose (> 75th percentile in dose) examination. Completed univariate analyses identified strategies and systems that were significantly associated with lower average doses or lower frequency of high doses for abdominal CT examinations. Imaging centers that (1) measure and track any patient safety measures, (2) assess doses and the impact of any CT changes that are made, (3) set specific goals for improving radiation dose, (4) organize teams to improve doses, (5) pilot test process changes before full implementation, and (6) standardize workflow to encourage dose optimization were associated with at least 30% reduction in high dose examinations in comparison to organizations that did not report these activities. (Table 2b-1) Further, radiation leaders who support dose optimization and organizations with leadership enthusiastic about dose optimization had at least 20% fewer high dose examinations.

Table 2b.03 - Results from Solberg et al. showing survey questions that were significantly related to radiation dose in univariate analysis, the mean score for each question, the percent of imaging centers whose respondents strongly agreed or disagreed with the statement, and the percent reduction in average dose and probability of a high dose examination if all facilities strongly agreed with the question.

Survey questions‡	Mean Score	Percent of Respondents who Agree†	Percent of Respondents who Disagree††	% Reduction in Average Dose*	p value	% Reduction in High Dose**	p value
A. Practice Systems	**-**	**-**	**-**	**0.3**	**0.96**	**17**	**0.34**
We track patient safety measure (non-radiology)	55	36%	47%	10	.01	32	<.001
B. Practice Readiness	**-**	**-**	**-**	**8**	**.17**	**39**	**.016**
We assess the impact of CT changes made	73	71%	8%	-	-	31	.01
Rad leaders support dose optimization activities	83	87%	2%	7	<.001	27	<.001
We lowered CT dose in past year	83	85%	3%	7	.01	25	<.001
We review processes & identify areas for improvement	75	71%	8%	3	0.27	18	.03
Leaders of optimization are enthusiastic	69	72%	2%	5	0.13	23	.03
C. Change Strategies	**-**	**-**	**-**	**7**	**.11**	**32**	**.03**
We set specific goals for improving radiation dose and image quality	55	32%	24%	8	.04	37	<.001
We organize a team to improve CT dose	51	33%	31%	8	.01	33	<.001
We tailor decisions to site needs	66	44%	15%	8	.04	31	.02
We pilot test process changes before implementation	58	37%	23%	8	.04	30	.03
We standardized workflows to encourage dose optimization	83	65%	1%	5	.03	31	.01
D. Miscellaneous	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Number of protocols	5	-	-	19	.01	38	001
Number of active QI projects on radiation dose	-	-	-	6	.01	22	.01

Survey questions‡	Mean Score	Percent of Respondents who Agree†	Percent of Respondents who Disagree††	% Reduction in Average Dose*	p value	% Reduction in High Dose**	p value
We obtain clear images of abdomen CT	92	92%	9%	2	<.001	8	<.001

Cells marked by a dash (-) are intentionally blank
QI = quality improvement
*Percent reduction in expected dose length product if all facilities strongly agreed
**Percent reduction in probability of high dose exams (from baseline 25%) if all facilities strongly agreed
‡Abbreviated versions of the questions
†Includes responses of "Yes, and Works Well" or combined responses of "Agree" and "Strongly Agree," depending on question
††Includes responses of "No" or combined responses of "Disagree" and "Strongly Disagree," depending on question

Qualitative analysis identified six primary barriers to dose optimization reported by Hospital and imaging facility leaders (Whitebird 2021): (1) resistance to change, (2) limited time and resources, (3) complex organizational structure, (4) lack of leadership support, (5) variations in CT equipment, and (6) variability in CT protocols. These barriers impeded efforts by health care organizations to optimize radiation doses to patients from CT imaging. Identifying barriers early in any improvement process is an important first step in making meaningful and sustained change. (Whitebird 2021, in press) also describe dose optimization strategies used by health care organizations to optimize radiation dose – and image quality. This identified seven organizational strategies used by these leaders for optimizing CT dose: (1) engaging radiologists and technologists; (2) establishing a CT dose committee; (3) managing organizational change; (4) providing leadership and support; (5) monitoring and benchmarking; (6) modifying CT protocols; and (7) changes in equipment and work rules.
Across all facilities participating in the RCT, the multicomponent intervention resulted in a 23-58% reduction in the proportion of high-dose exams across anatomic areas based on organ dose, with no observed change in image quality. These reductions were both statistically and clinically significant. Audit feedback alone, comparing radiation doses with those of other facilities, also reduced the proportion of high-dose exams and mean doses, but with a smaller magnitude then when coupled with quality improvement.
Head CT examinations in children were assessed before and after the intervention (N=23,860 exams). The median head doses in children were reduced 6.8% (P<.004) after the multicomponent feedback. There were fewer abdomen exams in children (N=9,450), and there were no significant changes in mean dose following the intervention.
2016 submission:
The different metrics are highly correlated (see Keegan JACR 2014, Attachment of UCSF CT DOSE Report). The metrics are highly correlated with absorbed doses.

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]
2022 submission:
Patient/encounter-level (data element) validity
Anatomic area can be assigned with a high degree of accuracy.
Dose length product (DLP), as a measure of radiation dose is standardized, virtually universally available, and well validated and used broadly to reflect the radiation dose delivered to the patient.
Empirical validity testing
Organizations that follow identified practices to improve quality in imaging care, and put systems in place to do so, have lower and more optimized doses for CT. These findings confirm the connections between targeted process improvements (at the facility level) and better performance on our measure of radiation dose.
Providing feedback to hospitals in a format aligned with this measure contributes to meaningful radiation dose reduction.
2016 submission:
The metrics are valid and meaningful and will reflect a facilities average CT doses.
Of note, these are not patient level metrics and, for an individual patient, do not provide information about whether the dose that was used was appropriate.

[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]
2022 submission:
We consider it clinically meaningful to detect entities (hospitals) whose median DLP is more than 0.5 standard deviations greater than the threshold (in our case the 75th percentile of the UCSF International CT Dose Registry within a combination of anatomic area and age stratum, see figures 2b.05 for illustration that such hospitals exist), for classification as hospitals with median over the threshold. For a typical hospital in the UCSF International Dose Registry, skull and abdomen-pelvis exams have mean values roughly 30% higher than the median, while brain exams have mean values roughly equal to the median.
To compute the minimal sample size necessary to be able to detect a high DLP median, we use a modified one-sample t-test. Let $\bar{x}$ be the observed mean of a sample of N exams from an arbitrary hospital, whose exams are identically and independently distributed with mean μ and standard deviation σ. By the central limit theorem we have

$$sqrt(N)(\bar{x}- \mu) \rightarrow N(0,\sigma^2)$$

Meaning, for any fixed scalar k (k=1/1.3 for skull and abdomen-pelvis, k=1 for brain),

$$sqrt(N)(k\bar{x}- k\mu) \rightarrow N(0,k^2\sigma^2)$$

For any arbitrary threshold K, consider a hypothesis test with null and alternate hypotheses

$$H_0: k\mu-K=0$$
$$H_A: k\mu-K>0$$

Suppose we wish to detect $k\mu-K=0.5\sigma$ with 0.8 power and level of significance 0.05. We have, by definition of power

$$0.8=Pr((k\bar{x}-K)/(k\sigma/sqrt(N_M)) > z_{1-0.05} \mid k\mu-K=0.5\sigma)$$
$$\approx 1-Pr(Z < z_{1-0.05}-0.3/(k/sqrt(N_M)))$$

Where Z is a normally distributed random variable, $z_{1-0.05}$ is the 95th percentile of a normally-distributed random variable, N_M is the minimal sample size required to achieve 0.8 power with level of significance 0.05. This value can be computed from the equation above.
We also consider it of clinical interest to assess the proportion of individual exams within a hospital which exceed the 75th percentile benchmarks for their respective anatomic areas and age groups. In this context, it would be of clinical interest to detect entities whose prevalence of “high dose" exams is at least 50%, which is twice the expected prevalence of exams above the 75th percentile benchmark, for classification as hospitals with higher than expected prevalence of “high dose” exams.

To compute the minimal sample size necessary to be able to detect such out-of-range prevalence with 0.8 power, 0.05 level of significance, we use the equation

$$0.8 = Pr[Z < z_{0.025}-H(0.25,0.50) * sqrt(N_{M2})]$$

Where Z is a normally distributed random variable, $z_{0.025}$ is the 2.5th percentile of a normally-distributed random variable, N_{M2} is the minimal required sample size within a hospital to detect a high dose rate of 50%, and

$$H(x,y)=2 * arcsin(sqrt(x)) - 2 * arcsin(sqrt(y))$$

Figure 2b.05a – Distribution of skull examination median DLP in children among hospitals in the UCSF International Dose Registry for hospitals that contributed at least 10 exams during the study period.

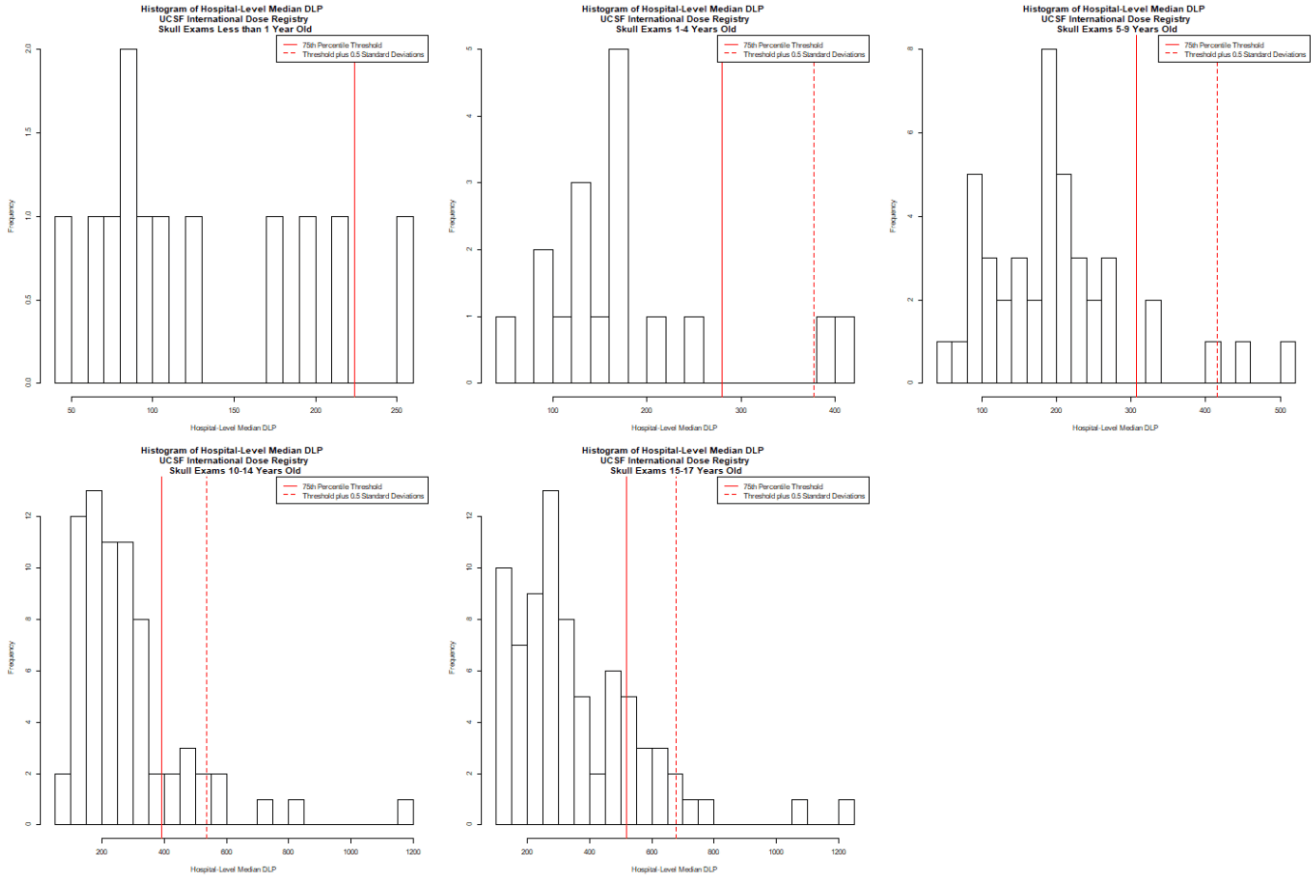


Figure 2b05.b - Distribution of abdomen examination median DLP in children among hospitals in the UCSF International Dose Registry for hospitals that contributed at least 10 exams during the study period.

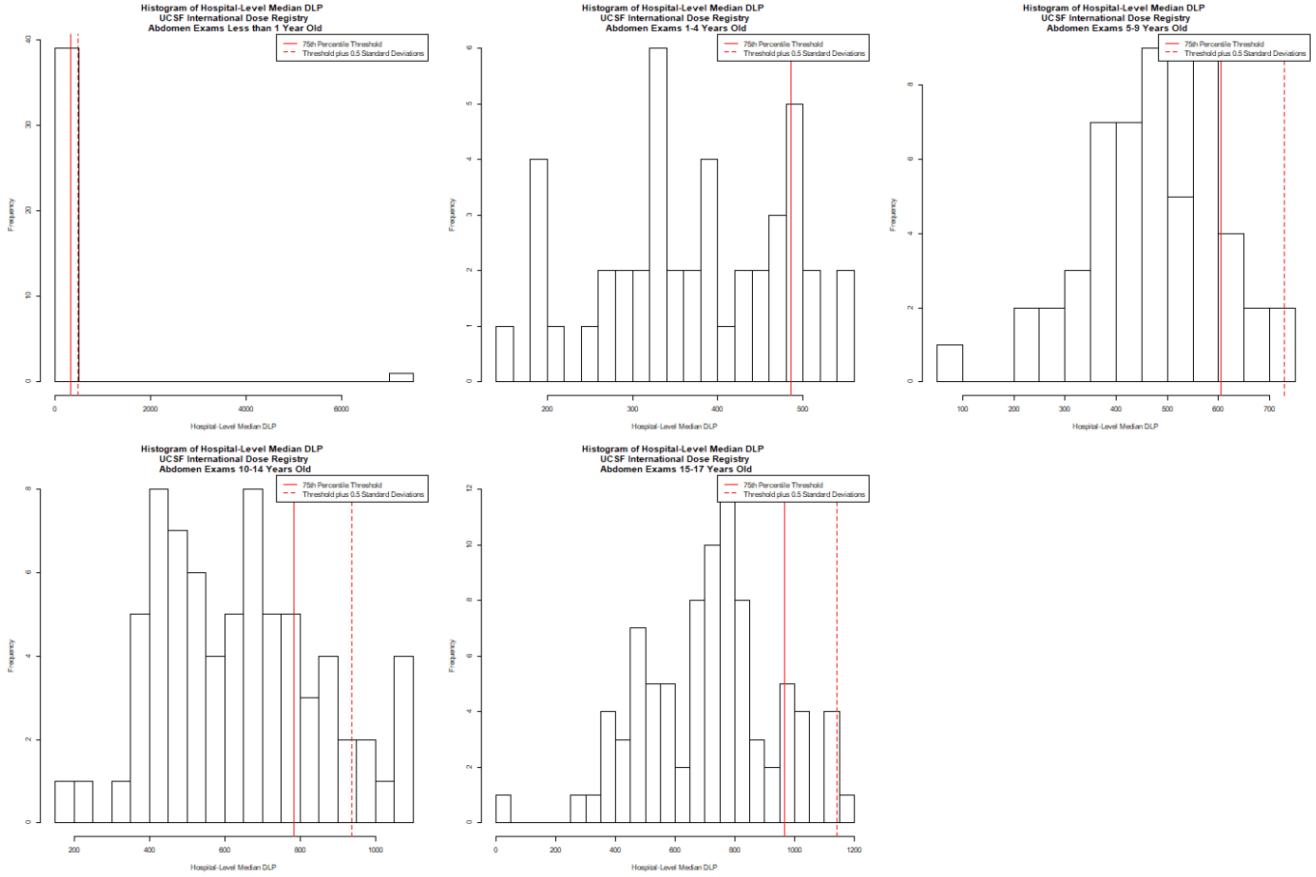
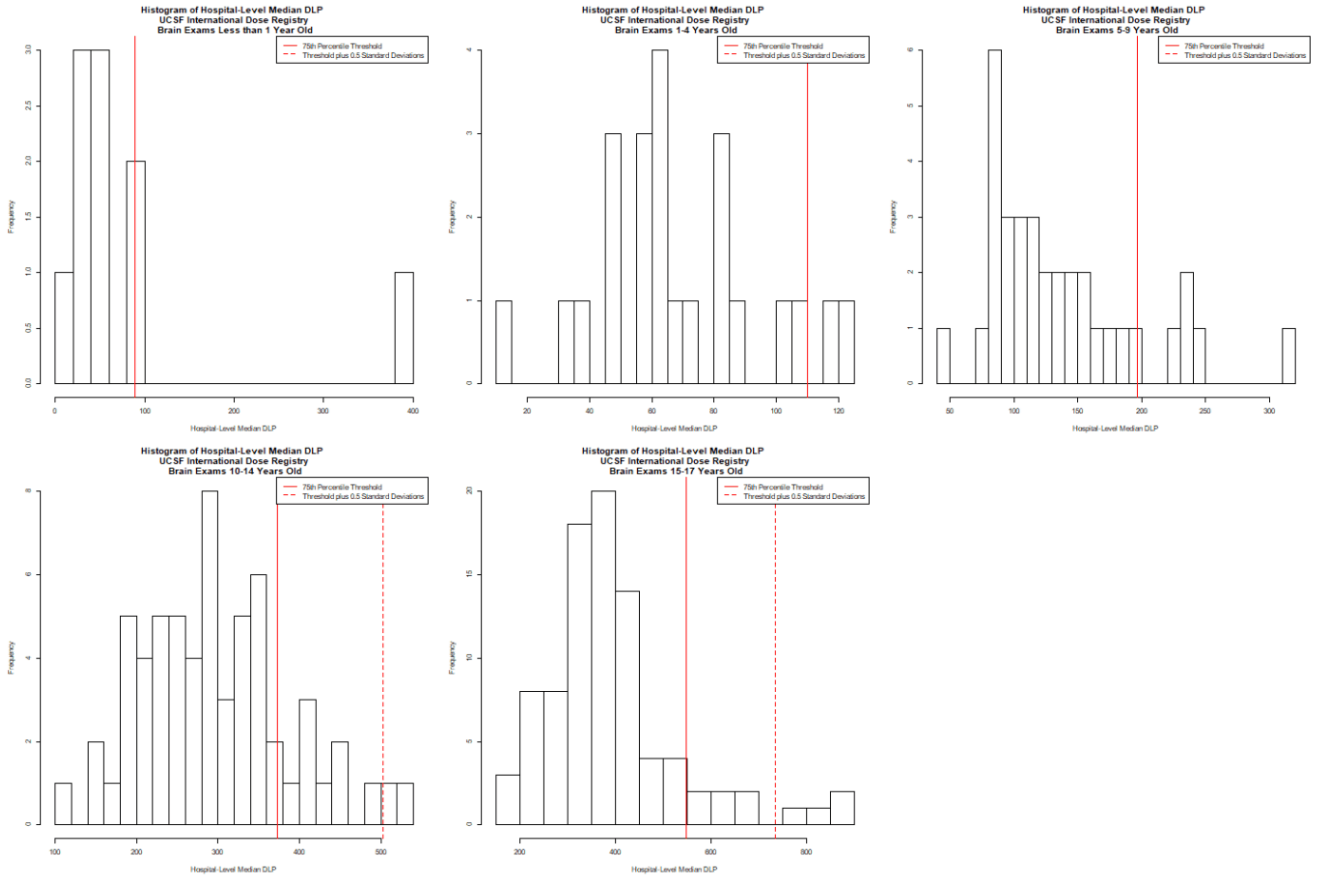


Figure 2b.05.c – Distribution of brain examination median DLP in children among hospitals in the UCSF International Dose Registry for hospitals that contributed at least 10 exams during the study period.



2016 submission:

Comparing institutional performance to benchmarks permits identification of outlying performance. Because the metric is based on summarizing dose for a large number of individuals (> 100 within each stratum) and comparison to benchmarks, the comparisons are stable at identifying outlying performance. In the attached document (UCSF CTDOSE Report), we illustrate the result of comparing institutions (using t-tests and quantile regression) using the NQF measure format. Basically, facilities can be identified and compared with benchmarks, and stable estimates of facilities with outlying performance can be identified. See Miglioretti 2014 JACR, Keegan JACR 2014

While the generation of averages will permit the comparison of facilities to benchmarks, the measure does not specify cutoffs or how the comparisons would be judged. These can be set based on the clinical or quality improvement needs of a facility, organization, etc.

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

2022 submission:

For skull and abdomen and pelvis exams, a minimum of 9 CT exams within each age and anatomic area strata will provide adequate sample size to detect a hospital with median 0.5 standard deviations above the benchmark.

For brain exams, a minimum of 25 exams within each age strata will provide adequate sample size to detect a hospital with median 0.5 standard deviations above the benchmark.

To detect a hospital with a 50% prevalence of high dose exams (i.e. out-of-range rate), the hospital needs a minimum of 23 exams in total.

Many hospitals present in all three data sources do not seem to provide pediatric care, and are therefore excluded from summaries on their sample size adequacy and power. That is, hospitals which do not have more than 10 exams observed in at least one combination of anatomic area and age group are ignored for results shown in this section.
In table 2b.06, important metrics for assessment of power are split in three groups for each of the two mechanisms used for determining "high dose hospitals." These are:
C1 and E1) Hospitals which ostensibly possess the trait which would be detected by statistical significance.
C2 and E2) Hospitals which ostensibly possess the trait which our hypothesis testing assumptions believe provide an 80% chance of statistical significance.
C3 and E3) Hospitals which are statistically significant.
Note that the majority of hospitals under consideration have sufficient sample size to detect statistical significance (B1 and D1 are high). Thus, given the hypothesis testing assumptions made in our sample size calculation, rows C3 and E3 should be roughly equal to at least 80% of rows C2 and E2, though we would hope they approach rows C1 and E1 as closely as possible.
In the Leapfrog Group Implementation Data, this expectation is met.
In the UCSF International Dose Registry, this expectation is substantially exceeded. Nearly all hospitals with median DLP over the threshold in at least one combination of anatomic area and age group (0.39/0.49=80%), and nearly all hospitals with higher than expected prevalence of high dose exams (0.30/0.42=71%) are statistically significant.
Table 2b.06 – Proportion of exams with sufficient sample size, with high median dose, and with high prevalence of high dose exams, for UCSF and the Leapfrog Group

-	-	**UCSF International Dose Registry**	**Leapfrog Group Implementation Data**
A1	Number of Hospitals Total	143	1447
A2	Number of Hospitals with Observed Sample Size >10 **within at least one combination of anatomic area and age group**	106	1340
-	-	-	-
B1	Proportion of Hospitals with sufficient annualized sample size to detect median 0.5 standard deviations above the threshold **within at least one combination of anatomic area and age group**	0.72	0.71*
-	-	-	-
C1	Proportion of Hospitals with observed median above the threshold **within at least one combination of anatomic area and age group**	0.49	0.28*
C2	Proportion of Hospitals with observed median more than 0.5 standard deviations above the threshold **within at least one combination of anatomic area and age group**	0.23	0.10*
C3	Proportion of Hospitals with observed median statistically significantly above the threshold **within at least one combination of anatomic area and age group**	0.39	0.08*
-	-	-	-
D1	Proportion of Hospitals with sufficient annualized sample size to detect 50% prevalence of high dose exams	0.94	0.90
-	-	-	-
E1	Proportion of Hospitals observed with more than expected (25%) prevalence of high dose exams	0.42	Data not supplied
E2	Proportion of Hospitals observed with more than 50% prevalence of high dose exams	0.07	Data not supplied
E3	Proportion of Hospitals observed with statistically significantly more than expected (25%) prevalence of high dose exams	0.30	Data not supplied

Cells marked by a dash (-) are intentionally left blank.
* The Leapfrog Group Implementation Data does not differentiate between skull and brain exams, which are placed into a larger "head" anatomic area. Thus, these numbers pertain only to abdomen/abdomen and pelvis exams.

2016 submission:
Not answered

[Response Ends]

2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]
2022 submission:
The majority of hospitals which provide pediatric care are expected to have sufficient sample size in one year to detect median DLP more than 0.5 standard deviations above the threshold, to classify as median DLP above the threshold, with 80% power.
The majority of hospitals which provide pediatric care are expected to have sufficient sample size in one year to detect a prevalence of high dose exams greater than 50%, to classify as greater than expected (25%) prevalence of high dose exams, with 80% power.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]
2022 submission:
To assess the potential impact of removing exams due to missing radiation dose, we compared the distribution of missing data.
2016 submission:
The measure calls for collecting consecutive scans so that participants cannot choose their best or most optimum dose metrics to quantify. The data will be available, or can be calculated from essentially all (>95%) of CT scans.

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]
2022 submission:
Missing data is extremely rare. Among pediatric examinations in the UCSF International CT Dose Registry, age is missing for 0.05%; anatomic area is missing for 2%; and radiation dose is missing for <1%. (In this registry, CT category is reported in a separate data element. In Leapfrog's implementation of our measure, anatomic area is never missing because it is determined by the CPT code(s) used for billing the scan.)
2016 submission:
Not answered

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.F

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]
2022 submission:
No statistical tests were conducted to assess the impact of missingness because the rates were so small.
2016 submission:
Not answered

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]
No, there is only one set of specifications for this measure
[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]
[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]
[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]
[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]
N/A or no exclusions
[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]
N/A
[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]
N/A
[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]
N/A
[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]
Stratification by risk category (specify number of categories)
[Stratification by risk category (specify number of categories) Please Explain]
3 anatomic areas
5 age groups

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]
N/A
[Response Ends]

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]
[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]
Published literature
Internal data analysis
[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

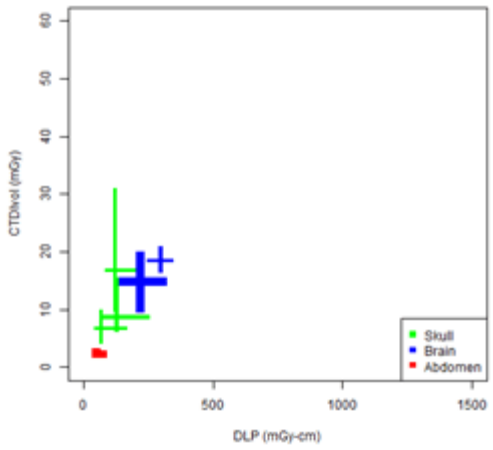
Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]
2022 submission:
The conceptual model underlying the measure and stratification has largely remained unchanged. The radiation doses used for CT in children, as in adults, primarily varies by anatomic area. (Smith-Bindman 2009, Smith-Bindman 2012, Miglioretti 2013, Keegan 2014, Smith-Bindman 2019, Smith-Bindman 2020, Smith-Bindman 2021, Chu 2021, Kanal 2021, Bos 2021, Bos 2022) In adults, there are different indications for CT that require different amounts of radiation and image quality, and 19 distinct CT categories have been created to reflect these different image requirements. (Smith-Bindman 2021) In children there are fewer identified indications for CT scanning (i.e. the list of why children undergo CT is far shorter), and within most anatomic areas there are few differences in image quality requirements that would necessitate different radiation dose levels. This means it is important to stratify by anatomic area, but within anatomic area there is no need to further sub-divide. The exception is head imaging, which includes different clinical indications that require different scanning approaches. Thus, two head CT categories were created for this measure: skull (a "low dose" category for imaging of the bony skeleton) and brain (a "routine dose" category representing all other head indications).
Using data from the UCSF International CT Registry we have summarized doses for each of the categories by age (Bos 2022 in preparation) and shown graphically the different indications within each of the categories.
Table 2b.23. Mean dose length product (and standard deviation) by age and anatomic region using data from the UCSF International CT Registry.

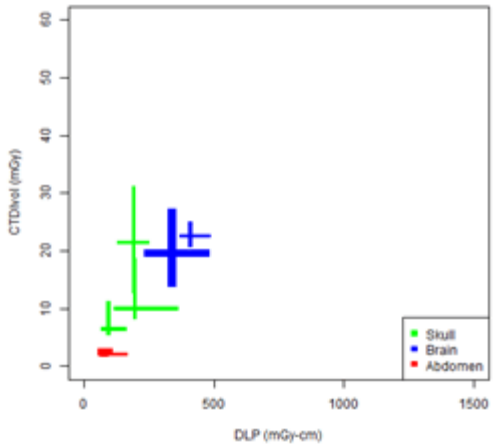
Age	Head, Low	Head, Routine	Abdomen
<1	171 (152)	270 (314)	207 (672)
1-4	232 (184)	388 (222)	92 (75)
5-9	264 (217)	503 (248)	168 (407)
10-14	335 (286)	656 (304)	329 (259)
15-17	401 (323)	787 (352)	471 (370)

Using data from the UCSF International CT Registry we have plotted the doses for the most frequent indications associated with CT imaging of the Skull, Brain and Abdomen and Pelvis, demonstrating substantial differences in doses between the three categories included in the measure, and relatively smaller differences between the specific individual indications within the CT categories (Bos, manuscript in preparation).
Figures 2b.23, 1-5: Graphs show volume CT dose index (CTDIvol) and dose length product (DLP) for the different anatomic regions and age strata. Intersection of arms is median for CTDIvol and DLP.

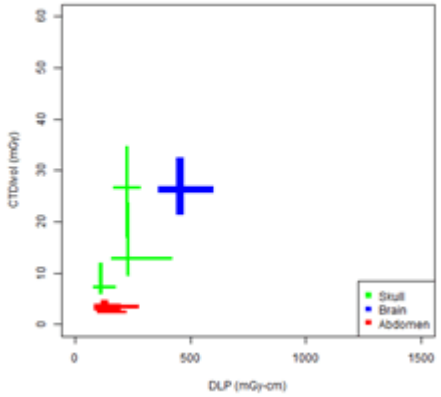
Less than 1 Year Old

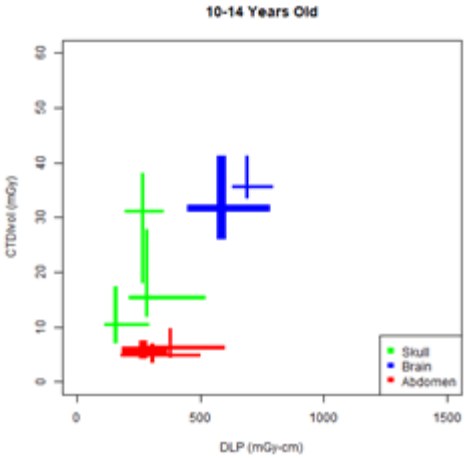


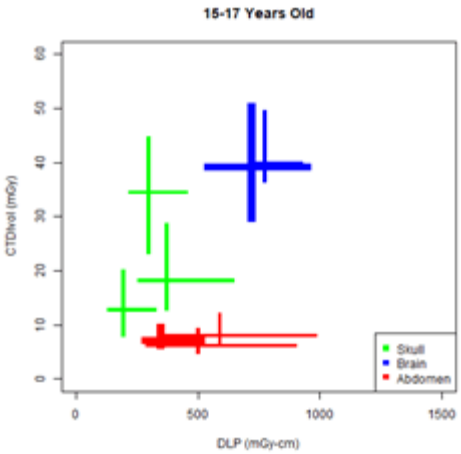
1-4 Years Old



5-9 Years Old







Alternative Text Figures 2b.23, 1-5: These graphs show the relationship between volume CT dose index (CTDIvol) and dose length product (DLP) for the different anatomic regions and age strata, as well as the clinical indications that contribute to each anatomic region. Each indication is shown with a cross, and the clinical indications are shown in different colors based on the body region. The Intersection of arms is the median for CTDIvol and DLP for that indication. These figures demonstrate that there are larger differences between body regions in dose than within body region between the different clinical indications, supporting the use of the three body region categories and age categories used in this measure (eg. while there are many different reasons for abdomen and pelvis CT, the doses for these different reasons are similar to each other.)

Patient size (including height and weight) is strongly associated with radiation dose. Unfortunately there are no automated approaches that can be used to adjusted the dose for patients of different sizes. Age is widely used a proxy for size, which has been supported by the International Commission on Radiation Protection. (ICRP Publications 121 and 135)

Citations

1. Bos D, Yu S, Luong J, Chu P, Wang Y, Einstein AJ, Starkey J, Delman BN, Duong PT, Das M, Schindera S, Goode AR, MacLeod F, Wetter A, Neill R, Lee RK, Roehm J, Seibert JA, Cervantes LF, Kasraie N, Pike P, Pahwa A, Jeukens CRLPN, Smith-Bindman R. Diagnostic reference levels and median doses for common clinical indications of CT: findings from an international registry. Eur Radiol. 2021 Oct 13. doi: 10.1007/s00330-021-08266-1. Epub ahead of print.
2. Bos, D. Pediatric Radiation Dose Benchmarks from the UCSF International CT Dose Registry. 2022, in preparation.
3. Chu PW, Yu S, Wang Y, Seibert JA, Cervantes LF, Kasraie N, Chu CA, Smith-Bindman R. Reference phantom selection in pediatric computed tomography using data from a large, multicenter registry. Pediatr Radiol. 2021 Dec 6. doi: 10.1007/s00247-021-05227-0. Epub ahead of print. PMID: 34866159.
4. International Commission on Radiation Protection. ICRP Publication 121: Radiological protection in paediatric diagnostic and interventional radiology. Ann. ICRP 42(2).
5. International Commission on Radiation Protection. ICRP Publication 135: Diagnostic Reference Levels in Medical Imaging. Ann ICRP 2017;46(1):1–144.
6. Kanal KM, Butler PF, Chatfield MB, Wells J, Samei E, Simanowith M, Golden D, Gress DA, Burleson J, Sensakovic WF, Strauss KJ, Frush D. U.S. Diagnostic Reference Levels and Achievable Doses for 10 Pediatric CT Examinations.

Radiology. 2022 Jan;302(1):164-174. doi: 10.1148/radiol.2021211241. Epub 2021 Oct 26. Erratum in: Radiology. 2022 Jan;302(1):E6.
7. Keegan J, Miglioretti DL, Gould R, Donnelly LF, Wilson ND, Smith-Bindman R. Radiation dose metrics in CT: assessing dose using the National Quality Forum CT patient safety measure. J Am Coll Radiol. 2014 Mar;11(3):309-15. doi: 10.1016/j.jacr.2013.10.009.
8. Miglioretti DL, Johnson E, Williams A, et al. The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk. JAMA Pediatr. 2013;167(8):700-707.
9. Smith-Bindman R, Lipson J, Marcus R, Kim KP, Mahesh M, Gould R, Berrington de González A, Miglioretti DL. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med. 2009 Dec 14;169(22):2078-86. doi: 10.1001/archinternmed.2009.427.
10. Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M, Greenlee RT, Kruger RL, Hornbrook MC, Roblin D, Solberg LI, Vanneman N, Weinmann S, Williams AE. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. JAMA. 2012 Jun 13;307(22):2400-9. doi: 10.1001/jama.2012.5960.
11. 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.
12. 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.
13. Smith-Bindman R, Yu S, Wang Y, Kohli MD, Chu P, Chung R, Luong J, Bos D, Stewart C, Bista B, Alejandrez Cisneros A, Delman B, Einstein AJ, Flynn M, Romano P, Seibert JA, Westphalen AC, Bindman A. An Image Quality-informed Framework for CT Characterization. Radiology. 2021 Nov 9:210591. doi: 10.1148/radiol.2021210591. Epub ahead of print.

2016 submission:

The radiation doses used for CT vary by anatomic area (head, chest, abdomen and pelvis) and age (adult versus various child age groups.) Smith-Bindman (JAMA Int Med 2009; JAMA 2012); Miglioretti JAMA Pediatrics 2013, JACR 2014). Smith-Bindman 2019, Smith-Bindman 2020, Smith-Bindman 2021, Chu 2021, Kanal 20201, Bos 2021, Bos 2022 There are further the categories that have been used in radiation safety programs for data collection In Europe and the UK. While there are other factors that influence radiation doses, these are relatively minor in comparison to these groups, and it is not feasible to collect or report data into smaller stratifications. Nor does the measure lose validity by not stratifying by clinical indication or patient size (see below.)

Why it is not important to stratify for clinical indication or protocol

The way CT scans are conducted should vary by why the patient is being scanned. For example, a search for occult malignancy may require very different parameters than an assessment for bleeding for trauma. Unfortunately there is currently no standardization for either categorizing the indications for imaging, nor for deciding the best way to image given a patient's suspected problem, nor for categorizing the protocols that are used. I am currently leading a project to standardize the protocols we use for imaging across the five University of California Medical Centers and for each indication, the different institutions have adopted dozens of different ways to image patients with similar clinical questions. Smith-Bindman (JAMA Int Med 2009) highlighted the issue of differing ways to image a very standard problem and the resulting radiation dose. For example, while some institutions chose to image patients with suspected stroke using a standard 2 mSv CT, others routinely use a high dose CT where the doses were on average 20, and some facilities used 58 mSv. There are almost no standards for defining how to image different clinical questions and the profound variation reflects physician preferences, and the promotion of certain protocols by the manufacturers, rather than evidence that the higher dose protocols are more accurate or diagnostic or truly needed based on evidence. As an example, in a NEJM article (Smith-Bindman 2010) images were included from a patient who underwent two chest CT examinations for the same clinical indication at the same institution one year apart. The patient had a 1.5 mSv dose study on one occasion and a 15.9 mSv study on the second occasion. Both studies were done for exactly the same reason of the surveillance of a pulmonary nodule, and both were done within a single institution and within the setting of a clinical trial where what was done should be standardized. Thus there is profound variation in how studies are conducted, even in the few situations where the reason for imaging is known and guidelines exist. Further, there are often financial incentives that drive the decision to image using repeated imaging protocols versus single imaging protocols (even though the former could lead to doses that are twice as high as the latter). For example, there were recent reports that some facilities use double imaging protocols (with and without contrast) for conducting Chest CT, thereby double billing and double radiating the patient, in a setting where doing two scans is considered rarely necessary. Thus while some facilities were using double scanning in 1% of patients, others were using this in 80% of patients, and CMS has concluded that this reflected overuse of CT(see http://www.nytimes.com/2011/06/18/health/18radiation.html?_r=1)

Anatomic area, rather than specific indication or protocol, will actually provide the patient with the information they want to know – i.e. if I go to a facility, how high or low will my dose be. It will also allow facilities to identify where they need to explore their doses in greater detail to assess why they are outside the normative range – is it that the are using too high doses within a protocol or using high dose protocols too often. The way the measure is currently written the choice of protocol will be reflected within the facilities metrics, whereas if dose were reviewed only within protocol, the facility that chooses to use high dose studies and repeated studies on most of its patients would appear fine.

Why it is not important to adjust for patient size

Weight will contribute to the variation in dose used for CT, and if individual patients were compared, it would be extremely important to assess weight when deciding about optimum ways to set up CT scans. Differences in weight may account for a 1-3 fold difference in the radiation used. Dr. Huda has published several relevant recent papers showing that doses vary up to 2 fold based on patient weight . "Radiation related cancer risks in a clinical patient population undergoing cardiac CT" AJR 2011 and "Estimating cancer risks to adults undergoing body CT examinations" Radiation Protection Dosimetry 2011. However, its important to point out that it is in no way established exactly how to increase doses for larger patients – i.e. there is no clear standard. A recent and interesting article found that machines that automatically adjust for patient weight seem to be giving too much radiation so that the organ doses increase even more so than does the weight (Israel, G. M., Cicchiello, L., Brink, J. and Huda, W. Patient size and radiation exposure in thoracic, pelvic, and abdominal CT examinations performed with automatic exposure control. Am. J. Roentgenol. 195, 1342–1346 (2010).

We have assessed the association between weight and the doses used, and presented at the initial submission of this metric, with an explanation of why it is not important to adjust for weight. When we compared the radiation dose used among patients in the top quartile of weight, to the radiation dose used in the bottom quartile of weight, the average doses increased by a factor of less than 2. For example among adult patients age 25 and older in the lowest quartile of weight (i.e. those under 152 lbs) the mean DLP among patients who underwent an abdominal and pelvic CT was of 781 . Among patients in the largest quartile of weight (ie those between 220 and 425 lbs, reflecting a mean weight twice as high), the mean dose was 1282 DLP or around 60% higher. However, within each of the weight groups, there was much more dramatic variation within group, then between groups. For example, among the smallest patients (those <25%) the range in dose between the 1st and 99th distribution was 54 – 1890 (40 fold variation between the highest and lowest group), and in variation in the highest quartile of weight was 352 – 2885 (8 fold variation). Thus the variation in dose based on weight was small in comparison to weight based on other factors (such as physician and facility preferences). We have a paper in press in Radiology that assesses, among a large sample of 800,000 CT scans factors that influence dose, site variation contributes far more variation to the model than even patient size.

These weight differences are not relevant at the facility level, as while patient size may influence dose by 2 fold (between the smallest and largest patients) other factors, can influence the dose by up to 100 fold (based on our data), and these factors, rather than individual patient weight, will drive the facility level dose indices measures. Even if a facility had ALL patients of a size <25%, versus all patients over the 75% the influence would be very modest.

However, while I do not believe including weight would influence a facility's measures, there have been several recent publications which provide simple ways to account for size when reporting radiation dose, and including one of these metrics in the measure may allow greater adoption of the measure by various stakeholders. These measures essentially have determined for a fixed amount of machine dose, how the absorbed dose to the patient varies by their size; larger patient will tend to have a lower adjusted dose (because the same dose is spread out in their larger body) whereas a smaller patient will have a correspondingly larger dose (because the same dose is distributed in a small volume of tissue.) Using these adjustment factors, it is possible to get a more precise estimate of the dose absorbed by the patient based on the machine output and a conversion factor based on the patient's size. The SSDE measurement (AAPM Report 204, Size Specific Dose Estimates in Pediatric and Adult Body CT Examinations) is now included in this measure and is a measure that accounts for patient size.

In our work across the University of California Medical Centers (as part of the UCDOSE, PI Smith-Bindman), among > 100,000 CT scans, there was no difference in facility level conclusions about performance when any of the metrics were used (i.e. SSDE, CTDIvol, DLP and ED) all characterized facilities the same.

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

N/A

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]
2022 submission:
A comprehensive review of the published literature was performed to identify of patient-level or exam-level risk factors. Because decisions are made at the level of patient groups, rather than individual patients, the logic model does not include varying technical parameters for individual patients. To the extent they have been studied, social factors including race/ethnicity and socioeconomic status are not predictive of radiation dose for CT exams. (Strauchler 2012, Freeman 2012, Hou 2014, Messenger 2015) Messenger et al. (2016) used a cohort of 3442 CTs for calcium scoring to assess the relationship between effective dose (dose length product multiplied by a fixed conversion factor) and a variety of patient characteristics including age, sex, ethnic group, and body mass index. Each continuous independent variable was converted into categories, and the means of each category was reported. They reported no substantial differences between effective dose and any categorical/categorized patient characteristic, except age among those >75 years old. There is a potential concern that the age of CT machines may be associated with increased radiation dose, as newer machines sometimes offer dose reduction software. Theoretically, this could lead to higher doses and poorer performance on the measure in safety-net settings that may have older machines. However, there is no evidence to support a strong association between CT machine factors, including the age of the machine, and increased radiation dose. (Catalano 2007) In a study of over 2 million CT exams from 151 institutions, including 290 machines from the four largest machine manufacturers and 49 machine models, Smith-Bindman et al. evaluated the contribution of machine characteristics to radiation dose variation. (Smith-Bindman 2019). No patient or machine characteristics explained the variability of effective dose to any notable extent. The authors concluded that differences in observed dose were almost entirely associated with how institutions used the machines, reflecting different choices of technical scanning parameters and not the machines themselves.
Another study showed, among institutions performing low-dose CT exams for lung cancer screening, a significant proportion of institutions and patients had doses that exceeded guideline-recommended dose levels. However, the type of institution, including whether the hospital was a public hospital, was not associated with the radiation dose used. (Demb 2019.) Lastly, several analyses are underway using data from the UCSF International CT Dose Registry demonstrating that optimized doses have been observed across all machine makes and models in the Registry, regardless of machine characteristics.
Thus, our decision to not include social risk factors was based on review of the literature and finding no empirical evidence supporting the influence of social risk factors (including provider-level proxies for social risk factors, such as machine characteristics) on radiation dose. Providers who see a disproportionate number of patients from disadvantaged backgrounds, or in safety-net settings which may have older CT machines, are not expected to fail the measure more frequently because of these factors.
Citations

1. 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.
2. Demb J, Chu P, Yu S, Whitebird R, Solberg L, Miglioretti DL, Smith-Bindman R. Analysis of Computed Tomography Radiation Doses Used for Lung Cancer Screening Scans. JAMA internal medicine 2019;179(12):1650-1657. doi: 10.1001/jamainternmed.2019.3893
3. Freeman K, Strauchler D, Miller TS. Impact of socioeconomic status on ionizing radiation exposure from medical imaging in children. J Am Coll Radiol. 2012 Nov;9(11):799-807. doi: 10.1016/j.jacr.2012.06.005. PMID: 23122347.
4. Hou, J.K., Malaty, H.M. & Thirumurthi, S. Radiation Exposure from Diagnostic Imaging Studies Among Patients with Inflammatory Bowel Disease in a Safety-Net Health-Care System. Dig Dis Sci 59, 546–553 (2014). https://doi.org/10.1007/s10620-013-2852-1
5. Messenger B, Li D, Nasir K, Carr JJ, Blankstein R, Budoff MJ. Coronary calcium scans and radiation exposure in the multi-ethnic study of atherosclerosis. Int J Cardiovasc Imaging. 2016 Mar;32(3):525-9. doi: 10.1007/s10554-015-0799-3. Epub 2015 Oct 29. PMID: 26515964.
6. Strauchler D, Freeman K, Miller TS. The impact of socioeconomic status and comorbid medical conditions on ionizing radiation exposure from diagnostic medical imaging in adults. J Am Coll Radiol. 2012 Jan;9(1):58-63. doi: 10.1016/j.jacr.2011.07.009. PMID: 22221637.

7. 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.

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]
Our stratification approach was developed a priori, based on the literature review and clinical expert input described in 2b.23 above. It was validated through descriptive and visual analyses of the distribution of doses within and across strata, as described above.
[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]
N/A
[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]
N/A
[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]
N/A
[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]
2022 submission:
Stratification is necessary because the benchmark dose values vary dramatically across strata as shown above. With stratification, we ensure that the outcome rate is equal for all 5x3=15 strata, despite markedly different dose distributions across these strata. Stratification thus removes the bias that would otherwise adversely affect hospitals that do a higher proportion of brain scans, or hospitals that treat a higher proportion of adolescents (and are therefore required to use higher overall median doses).
As described in 2b.25, there are no guidelines or evidence that lead us to expect different doses by factors other than the anatomic area imaged and patient size (and for reasons described previously, we use age as a surrogate for patient size). Thus, we don't need to adjust by any other factors. Machine make and model are not important determinants of dose and thus do not need to be accounted for.

The measure score seeks to identify very poor performance, varying widely from the population average (i.e. median radiation doses above the 75th percentile by strata, or at least 50% of pediatric CT exams overall dosed above the 75th percentile). The most likely reason for this to occur is that facilities are not child-sizing their doses, and performance on this measure will provide clear feedback that they must.

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]
2022 submission:
The differences in hospital scores are a valid representation of their quality, because they factor out the impact of variation in the anatomic area of the scan and the age of the patient (as a proxy for size).

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]
N/A
[Response Ends]

Criteria 3: Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]
Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)
Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.
[Response Begins]
ALL data elements are in defined fields in a combination of electronic sources
[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]
N/A
[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]
We have plans to develop this measure into an eCQM by the next round of maintenance review, to include and address measurement of image quality in CT exams.
We acknowledge a common concern in the radiology community related to radiation dose reduction is a subsequent deterioration of image quality, as too little radiation dose can produce inadequate image quality. The actual risk for this is low, as research suggests doses may be lowered between 50-90% without impacting image diagnostic utility (den Harder 2018, Rob 2017, Konda 2016, Huppertz 2015). Nevertheless, a quality measure assessing CT radiation dose should also include a balancing measure of image quality to preserve the diagnostic usefulness of these exams.
Currently, no existing measure in the NQF inventory of endorsed measures or in any known quality program addresses image quality because (1) there is a lack of consensus on what constitutes image quality in the radiology community, and (2) measuring image quality would require assessment of CT images, which has been technologically infeasible in the absence of electronic tools. We have developed a related, adult CT quality measure (submitted for NQF endorsement in the Fall 2021 cycle) that incorporates both radiation dose and image quality into an electronic clinical quality measure (eCQM). We intend to develop these methods specifically for the pediatric population, and when this measure is due for the next round of maintenance review, we aim to submit it as an eCQM that assesses both radiation dose and image quality.
References

1. Den Harder AM, Willemink MJ, van Doormaal PJ, et al. Radiation dose reduction for CT assessment of urolithiasis using iterative reconstruction: A prospective intra-individual study. Eur Radiol. 2018;28(1):143-150.
2. Huppertz A, Lembcke A, Sariali el H, et al. Low Dose Computed Tomography for 3D Planning of Total Hip Arthroplasty: Evaluation of Radiation Exposure and Image Quality. J Comput Assist Tomogr. 2015;39(5):649-656.
3. Konda SR, Goch AM, Leucht P, et al. The use of ultra-low-dose CT scans for the evaluation of limb fractures: is the reduced effective dose using CT in orthopaedic injury (REDUCTION) protocol effective? Bone Joint J. 2016;98-B

4. Rob S, Bryant T, Wilson I, Somani BK. Ultra-low-dose, low-dose, and standard-dose CT of the kidney, ureters, and bladder: is there a difference? Results from a systematic review of the literature. Clin Radiol. 2017;72(1):11-15.

[Response Ends]

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]
The data were widely available to Leapfrog for data collection. Because the data can be obtained from a special report generated by ACR and through dose monitoring software, there have not been any significant data reporting issues.
[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),

Attach the fee schedule here, if applicable.

[Response Begins]
N/A
[Response Ends]

Criteria 4: Use and Usability

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

4a.01. Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]
Public Reporting
[Public Reporting Please Explain]
This measure is currently used by the Leapfrog Group The results are publicly reported at https://ratings.leapfroggroup.org

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
[Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Please Explain]
This measure is currently used by the Leapfrog Group and is reported publicly.

Quality Improvement (Internal to the specific organization)
[Quality Improvement (Internal to the specific organization) Please Explain]
N/A

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Public reporting
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
Quality Improvement (internal to the specific organization)
[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

The assessment of radiation dose for diagnostic imaging is a relatively new concept. Several large and small organizations, hospitals, and hospital associations (particularly those that focus on children) are beginning to assess radiation. The Joint Commission, starting this year (2015), will start asking the facilities they oversee to begin assessing radiation doses as well. But otherwise, the concept is new.
[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]
There are innumerable organizations that would potentially be interested in collecting and using this data for accreditation and benchmarking, if the measure were adopted and endorsed. These organizations range from the Joint Commission, to innumerable hospital organizations, to insurers and to CMS (if the measure were applied to adults). Further, UCSF has a large project entitled Partnership for Dose that Dr. Rebecca Smith-Bindman leads (the title author of this measure), and she would be willing to commit to allowing any organizations who are interested to submit their data to this project, for use in performing benchmarking and certification. The Partnership for Dose currently have 150 hospitals/outpatient facilities that participate in the project, and we have the team and expertise to be able to do this. If the measure is endorsed, Dr. Smith-Bindman will work closely with all of these mentioned organizations to try to move ahead to submit an accountability application.
[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]
Leapfrog Group provides results to individual hospitals.
For Leapfrog, all participating hospitals can access their scored results on our public reporting website at https://ratings.leapfrogroup.org and can understand how they were scored using our publicly reported scoring algorithms document at https://www.leapfroggroup.org/survey-materials/scoring-and-results
Survey opens April 1. First reporting deadline is June 30. Hospitals see results July 25 and are able to update/correct responses up until November 30. If they submit an update, results are updated on public reporting website within first 5 days of the next month.

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]
Leapfrog Group provides results to individual hospitals.
For Leapfrog, all participating hospitals can access their scored results on our public reporting website at https://ratings.leapfrogroup.org and can understand how they were scored using our publicly reported scoring algorithms document at https://www.leapfroggroup.org/survey-materials/scoring-and-results
Survey opens April 1. First reporting deadline is June 30. Hospitals see results July 25 and are able to update/correct responses up until November 30. If they submit an update, results are updated on public reporting website within first 5 days of the next month.

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]
Leapfrog has received feedback on the measure via our their Help Desk. Biggest reported issues – using age instead of size and rewarding low doses in scoring w/o indication of image quality. This year, Leapfrog added a series of fact finding questions to the Survey to help account for image quality – particularly for those reporting very low doses – new question available at https://www.leapfroggroup.org/survey-materials/survey-and-cpoe-materials in Section 9B of the Survey.
[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]
The biggest reported issues were 1) the use of age instead of size as a way to stratify results and 2) rewarding low doses in scoring w/o indication of image quality. This year, Leapfrog added a series of fact finding questions to the Survey to help account for image quality – particularly for those reporting very low doses – new question available at https://www.leapfroggroup.org/survey-materials/survey-and-cpoe-materials in Section 9B of the Survey. Age is an excellent surrogate for size for the two categories that rely on head size which varies only modestly within age groups. Age is an approximate surrogate for size in the abdomen and because abdomen size is not routinely collected would create a large burden on sites to have to collect.
[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]
N/A
[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]
1.Changed the specifications to standardize use of specific phantoms for head and body exams (this had been a source of confusion and variability for reporting entities). This was based on comments submitted to Leapfrog and on UCSF research.
2.We have subdivided the previously specified "head" anatomic area into two new anatomic areas: "skull" representing low radiation dose indications that are focused on imaging of the facial and skull bones, sinus, temporal bone, and for assessment of patency of a ventricular shunt; and "brain" representing all other head indications which have the brain as the primary focus rather than the bony skeleton. This decision was primarily based on analysis of data from the UCSF International CT Dose Registry and recent publications from the American College of Radiology Dose Registry demonstrating meaningful differences in radiation doses used for skull versus brain CT. However, it was also frequently noted by reporting entities that a single "head" category did not adequately reflect necessary variation by clinical indication.

[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]
Leapfrog does not yet have data on changes over time. They have said that they will have improvement data this year when they compare 2021 to 2022 survey results
[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]
2022
Given the relationship of radiation dose and image noise, there is concern that dose reduction will result in deteriorated image quality. Theoretically, this reduces the diagnostic utility of CT images and could harm patients by requiring repeated scanning (thus doubling the dose). While this measure does not directly assess image quality, *we have not been informed by reporting entities of any negative impact on image quality.*
As noted in 3.04, we have developed a related, adult CT quality measure (submitted for NQF endorsement in the Fall 2021 cycle) that incorporates both radiation dose and image quality into an electronic clinical quality measure (eCQM). In field-testing for that measure, failures due to inadequate image quality (measured in image noise) were exceedingly rare, with less than 1% of exams, on average, across all reporting entities. This findings aligns with prior research, which suggests doses may be lowered between 50-90% without impacting image diagnostic utility (den Harder 2018, Rob 2017, Konda 2016, Huppertz 2015).
Nevertheless, the measure steward is attentive to this concern and is working to develop similar image quality specifications specifically for the pediatric population, and when this measure is due for the next round of maintenance review, we aim to submit it as an eCQM that assesses both radiation dose and image quality.
References

1. Den Harder AM, Willemink MJ, van Doormaal PJ, et al. Radiation dose reduction for CT assessment of urolithiasis using iterative reconstruction: A prospective intra-individual study. Eur Radiol. 2018;28(1):143-150.
2. Huppertz A, Lembcke A, Sariali el H, et al. Low Dose Computed Tomography for 3D Planning of Total Hip Arthroplasty: Evaluation of Radiation Exposure and Image Quality. J Comput Assist Tomogr. 2015;39(5):649-656.
3. Konda SR, Goch AM, Leucht P, et al. The use of ultra-low-dose CT scans for the evaluation of limb fractures: is the reduced effective dose using CT in orthopaedic injury (REDUCTION) protocol effective? Bone Joint J. 2016;98-B
4. Rob S, Bryant T, Wilson I, Somani BK. Ultra-low-dose, low-dose, and standard-dose CT of the kidney, ureters, and bladder: is there a difference? Results from a systematic review of the literature. Clin Radiol. 2017;72(1):11-15.

2016:
There are two potential limitations of the proposed measures that need to be described. CT radiation dose will vary by patient size, and the specific protocols used, and yet we are not suggesting the dose indices be collected in separate strata for size (other than for children) nor for different protocols. These two issues will be addressed separately.
PATIENT SIZE
One factor that influences the radiation dose in CT is patient size. In general higher doses are used in large patients in order to maintain the same image quality as can be achieved with lower doses in smaller patients. It simply takes higher doses of radiation to penetrate (get through) larger sized patients. Thus the recorded radiation doses in part will reflect the size of the patients seen.
If a facility sees a very high proportion of obese patients, their doses will be higher than a facility that sees very thin patients. This issue will be important when facilities compare their dose indices to normative data (to the diagnostic reference level data), as they should compare their actual data to data of facilities that assess similar patients. This is the reason that facilities should note the state where their facility is located if they submit their data to a national organization. Diagnostic reference levels should be generated at a local enough level (state, or region of the country) so they are most useful and relevant with respect to the size of patients scanned. Thus diagnostic reference data should reflect geographic differences and be appropriate to the typical patients seen in a given area, as called for in the FDA white paper on radiation safety. Thus if patients tend to be larger in the Northwestern states, the diagnostic reference levels may be higher in that region. As long as a given facility is compared to the correct area, this would have no impact unless a facility differs profoundly from the other facilities in its geographic region. Of note, the differences in patient size will only have a relative small impact on dose (around a two fold difference between the smallest and largest adult patients,) whereas variation in dose by 20-50 fold have been seen unrelated to patient size (Smith-Bindman, JAMA 2012; Smith-Bindman, JAMA IM 2009; Miglioretti, JAMA Pediatrics 2013; and Miglioretti, JACR 2014). Thus, while the current metrics does not perfectly account for size, size is a small contributor to dose, in comparison to much larger, unexplained and unjustified variation. (Smith-Bindman, Radiology, 2015).

Thus the validity of the proposed NQF measure dose not require consideration individual level adjustment of patient size. Facilities (even without consideration of external data) can compare their own data from one year to their data from prior years, and unless there is a profound shift in the weight of their patients, this will have no impact on their data. Facilities should still perform in-depth analysis of patient's who receive high radiation doses (perhaps above the 75% distribution at their own institution) to determine if those doses were appropriate and justified, or if they could have been reduced.

Further, none of the quality control programs in existence and described above (UK, European or American College of Radiology Programs) assess patient weight in conjunction with CT dose measures. It is simply not feasible, and would make it far more difficult for facilities to assemble dose data, as this information is not recorded as part of the radiology medical record, and is typically not available anywhere for most patients seen in outpatient settings. Difference in patient size is only one factor contributing to dose, and likely accounts for only a small amount of the large variation in dose within and between facilities.

The issue of the validity of this measure without consideration of patient size was vetted with a large number of physicists. There was widespread agreement that this measure as specified was highly valuable. Three letters of support originally submitted with this measure (from the ACR, NCRP and FDA) supporting the measure as specified were included with the initial submission of this NQF measure when it was first approved.

CT PROTOCOLS

The way CT studies are conducted (the "protocols" using the language of CT) leads to the radiation doses patients will receive. These are the specific instructions the radiologist or other physician and technologists program into the CT machine at the time of scanning. The instructions include how large an area to scan, how many times to scan each area and the settings of kVp and mAs to use. If a larger anatomic area is imaged, the dose the patient receives will be higher. If a multiphase study is done (meaning a single anatomic area is imaged many times) the dose will be higher than if a single-phase study is done. If a facility chooses to use multiphase protocols frequently, or to scan large anatomic areas frequently, their doses will be higher than facilities that try to minimize the area imaged or number of scans taken. The type of scans done in Los Angeles California and Huntsville Alabama that led to the extreme radiation dose exposures for CT, were perfusion scans, a type of scan where a small area of the brain is imaged dozens, and sometimes hundreds of times.

The two ways to collect and compare CT dose index information would be first to compare doses WITHIN the specific study type - thus compare doses for routine single phase studies and compare doses for multiphase studies, or second, to compare typical doses for all patients who undergo a CT within a single anatomic area (ignoring considering of the specific protocol used).

The latter method is far more practical. It's a large amount of work to determine the specific protocol, why a study was done, whether it was routine or not, how many phases were used, and it is simply not practical to have a data abstractor or technologist necessarily know how distinguish the study type. However, the latter method is far more valid, reproducible and a reliable measure of quality. This is particularly true as there are no evidenced based guidelines about when particular protocols should be used. In particular the multiphase, higher dose protocols are not clearly indicated in particular clinical situation, studies have not shown they lead to improved diagnoses or quantified the potential harm in their use, and differences reflect practice variation more than any objective criteria of the need for these multiphase, studies. That's not to say that these higher dose protocols don't have any value – but only that decisions about when to use different protocols are more based on physician preferences that patient outcomes, and choosing to frequently use these higher dose protocols should be reflected in the radiation dose quality metrics generated at a facility. Comparing doses within protocol would profoundly mask true differences to patients. In the example provided above relating to renal protocol CT is an example. While most institutions indeed had on their books a low dose protocol, these protcols were infrequently used. Comparing dose within protocol would have masked the actual doses patients receive.

To highlight this issue, a concrete and very realistic example has been provided below of two facilities and their choice regarding imaging patients with head CT. Keep in mind that the question a patient, a referring clinician, a radiologist, a hospital administrator or payer might wonder is what is the dose Ms. Smith will likely receive if she goes to a particular facility for a head CT.

Two facilities (A and B) will have different doses for different exam types and will have a different distribution of how often the different exam types are used.

For the sake of this example, we will estimate that a basic head CT has a dose of 2-3 mSv and a multiphase head CT has a dose of 20-30 mSv (Smith-Bindman, Arch Intern Med, 2009) did you want something else shown from this section?

Routine head CT 2-3 mSv

Multiphase head CT 15-20 mSv

For the sake of this example, we will estimate that facility "A "uses the routine head CT for most of their patient's (95%) and that at facility "A" the dose for the basic head CT is 2.5 mSv, and is 20 mSv for a multiphase head CT .

At facility "B" they use the routine head CT less often (50%) and use the multiphase CT more often (also 50%.). Their dose for the basic head CT 2 mSv (lower at this facility as they use the much higher dose, multiphase study more often, so can get away with lowering the dose on the routine study). They also have a lower dose for the multiphase study, at 15 mSv. For the sake of this example, we estimate each facility will conduct 100 head CTs over the course of a week.
If the two facilities were compared within protocol study type, facility "B" would appear to be doing a better job at dose reduction, as they have a lower dose for a routine head CT (2 mSv versus 2.5 mSv) and have a lower dose for a multiphase head CT (15 mSv versus 20 mSv). And yet this facility is using the higher dose multiphase protocol far more often which results in higher doses on average to patients.
Thus if we would compare the average dose per head CT at facility A (which is the clinical quality question a patient and payer would care about), it would be far lower at facility "A." Facility "A" has an average dose of 3.4 mSv (95% low dose studies * 2.5 mSv + 5% high dose studies*20 mSv)/100] whereas facility "B" has an average dose that is substantially higher at 8.5 mSv (50% low dose studies * 2 mSv + 50% high dose studies*15 mSv)/100].
Thus the dose patients receive will be driven by the choice of protocol more than the dose within protocol and doing comparisons only within protocol with mask real and important differences. Thus comparing overall exposure within anatomic area is not only more feasible, it is more appropriate if the goal is to identify facilities where the typical doses are simply too high. The facility with atypical doses could explore why their doses are high.
Cited in this section:
Smith-Bindman 2015, Predictors of Computed Tomography Radiation Dose and Their Impact on Patient Care. In Press, Radiology
Smith-Bindman R, Lipson J, Marcus R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med 2009;169:2078-86.
“ Radiation doses from commonly performed diagnostic CT examinations are higher and more variable than generally quoted, highlighting the need for greater standardization across institutions.” – Conclusion statement from Abstract

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]
N/A
[Response Ends]

Criteria 5: Related and Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)
[Response Begins]
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 (for CT Abdomen-pelvis with contrast/single phase scan, CT Chest without contrast/single
[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)
[Response Begins]
[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]
One existing process measure in the CMS Merit-based Incentive Payment System (MIPS) program is related (not competing) in that it addresses patient safety related to radiation exposure in CT imaging:

1. Optimizing Patient Exposure to Ionizing Radiation: Count of Potential High Dose Radiation Imaging Studies: Computed Tomography (CT) and Cardiac Nuclear Medicine Studies (CMIT # 2286, steward: American College of Radiology)

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]
Yes
[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]
The ACR measure compares performance to their own size-specific DRLs (diagnostic reference levels), which is the 75th percentile. (As opposed to ADs, which is the 50th percentile). The benchmarks are given in the Kanal 2017 paper – which is specific to ADULTS. These benchmarks were developed using exams in adults.
The smallest head diameter given is 12-14 cm.
Smallest chest is 21-25 cm.

Smallest abdomen is 21-25 cm.
I do not know how they would apply these benchmarks to a baby or young child.
https://pubs.rsna.org/doi/10.1148/radiol.2017161911?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed
Another difference is that the ACR has one category of "head and brain" and we separated this into (1) skull and (2) brain as the doses vary between these categories. I think this makes it a related, and not competing measure (different denominators).

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]
N/A
[Response Ends]