INVESTIGATION INTO THE OPTIMIZATION OF CT DOSE INDEX THRESHOLDS IN COMMERCIALLY AVAILABLE DOSE MONITORING SOFTWARE

By

Alyssa R. Martinez

A THESIS

Presented to the Department of Medical Physics
And the Oregon Health & Science University
School of Medicine
in partial fulfillment of
The requirements for the degree of

Master of Science

June 2021

Table of Contents

List of Figures ..... *iii*

List of Tables ..... *iv*

ACKNOWLEDGEMENTS ..... *v*

ABSTRACT *1*

1. Introduction *3*

2. Background *5*

2.1 CT Imaging **5**
2.1.1 Image Parameters 5
2.1.2 Scout Images 7

2.2 Dose Metrics **7**
2.2.1 CTDI 7
2.2.2 SSDE 9

2. 3 Risks of Ionizing Radiation **11**

2.4 The Joint Commission **12**

2.5 External Benchmarks **13**
2.5.1 National Dose Reference Levels 13
2.5.2 ACR DIR 14

2.6 Previous Methods of Setting Dose Thresholds **14**

2.7 Radiation Dose Index Monitor Systems **15**
2.7.1 Imalogix™ 15

3. Materials and Methods *17*

3.1 CT Dose Indices Threshold Generation **17**
3.1.1 Statistical Outlier 17
3.1.2 Limiting Parameters 18

3.2 Flagging **19**
3.2.1 Exporting Data from Imalogix 19
3.2.2 Filtering Data 19
3.2.3 Finding True Positives 20

3.3 Comparing True Positive Rates **21**

4. Results *23*

4.1 CT Dose Indices Thresholds **23**

4.2 Protocol Statistics **24**
4.2.1 Abdomen Pelvis with Contrast 24
4.2.2 Abdomen Pelvis without Contrast 25
4.2.3 Chest with Contrast 26
4.2.4 Chest without Contrast 28

4.2.5 Chest Abdomen Pelvis with Contrast 29
4.2.6 Neck with Contrast 30
4.2.7 Head without Contrast 31

5. Discussion 32

5.1 Comparing Methodology 32
5.1.1 Statistical Outlier Versus DRLs 33
5.1.2 Statistical Outlier Versus 98th Percentile 35

5.2 Dose Index Comparison 37

5.3 Future Work 38

5.4 Study Limitations 39

6. Summary and Conclusions 40

7. References 43

List of Figures

Figure 1: Statistics of Statistical Outlier Versus DRLs for SSDE.....34
Figure 2: Statistics of Statistical Outlier Versus DRLs for $CTDI_{vol}$ 35
Figure 3: Statistics of Statistical Outlier Versus 98^{th} Percentile for SSDE 36
Figure 4:Statistics of Statistical Outlier Versus 98^{th} Percentile for SSDE 36
Figure 5:Sensitivity from $CTDI_{vol}$ verses SSDE 37

List of Tables

Table 1: SSDE Dose Thresholds from the Three Methods for the Five Protocols that use SSDE 23
Table 2: $CTDI_{vol}$ Dose Thresholds from the Three Methods for all Seven Protocols 23
Table 3: Statistics for Abdomen Pelvis with Contrast (SSDE) 25
Table 4: Statistics for Abdomen Pelvis with Contrast ($CTDI_{vol}$) 25
Table 5: Statistics for Abdomen Pelvis without Contrast (SSDE) 26
Table 6: Statistics for Abdomen Pelvis without Contrast ($CTDI_{vol}$) 26
Table 7: Statistics for Chest with Contrast (SSDE) 27
Table 8: Statistics for Chest with Contrast ($CTDI_{vol}$) 27
Table 9: Statistics for Chest without Contrast (SSDE) 28
Table 10: Statistics for Chest without Contrast ($CTDI_{vol}$) 29
Table 11: Statistics for Chest Abdomen Pelvis with Contrast (SSDE) 29
Table 12: Statistics for Chest Abdomen Pelvis with Contrast ($CTDI_{vol}$) 30
Table 13: Statistics for Neck with Contrast ($CTDI_{vol}$) 30
Table 14: Statistics for Head without Contrast ($CTDI_{vol}$) 31

ACKNOWLEDGEMENTS

I am deeply grateful for my time at OHSU and throughout my time, I have received lots of support and encouragement.

I would first like to express my gratitude and thanks to my advisor, Dr. Thomas Griglock, for not only his support throughout my thesis but for everything he has done for me during my graduate program. The encouragement and guidance he has given me to continue working hard and persevere through hardships and failures has been undeniably amazing. His insightful feedback and patience has pushed me to become a more intuitive thinker and has brought my work to a higher level than I ever thought it could.

I would also like to express my deepest appreciation to the rest of my advisory committee, Dr. Lindsay DeWeese and Dr. Andrei Pugachev, for their time and effort in working with me on this thesis. Their discussion, ideas, and feedback has been invaluable. I especially appreciate Dr. DeWeese's advice throughout my graduate studies and all of the opportunities she gave me to further my skills in diagnostic imaging. I am extremely grateful to have mentors like them in my life who give me guidance and unwavering support.

I would like to extend my sincere thanks to James Armbruster, the senior director of customer success at Imalogix Inc., for his assistance and aid in understanding how our CT data is interpreted in Imalogix. He was very responsive, and his insight has been invaluable to completing this research.

I could not have completed my graduate program without my fellow students here at OHSU. My special thanks are extended to my cohort who not only has helped my understanding of classes but whose friendships I treasure deeply. They are all wonderful, thoughtful, and brilliant people, and I feel exceptionally lucky to have gotten the chance to work with them.

Finally, I cannot begin to express my thanks to my family and partner. Going through graduate school and writing my thesis has not been an easy task and they have constantly been by my side. They are always there to listen to me, no matter if they understand what I am saying or not. Mom, Matt, and Jenni you have all shaped me to be the person that I am today, and I cannot express my gratitude enough for your unconditional love and support.

ABSTRACT

Introduction:

Monitoring radiation doses from diagnostic Computed Tomography (CT) examinations is essential to maintaining patient safety and ensuring that technologists are using optimal scan parameters. To receive accreditation from The Joint Commission, facilities are required to track incidents where radiation dose indices from CT exams exceed their expected dose index; however, there is no universal definition as to what constitutes an expected dose index. This research focuses on creating a standard of practice for setting CT dose thresholds that maximize the sensitivity and specificity of tracking, ensuring appropriate patient safety without creating an excessive number of false-positive notifications.

Methods:

One year's worth of clinical CT data for seven protocols (9,194 acquisitions containing SSDE data and 13,146 containing $CTDI_{vol}$ data) were tracked via a commercially-available informatics system to determine the acquisitions that needed to be reviewed, known as true positives. Three methods of setting dose thresholds were utilized to compare the number and which acquisitions were being flagged. The first method uses nationally-averaged dose reference levels, which allowed a comparison between site-specific and non-site-specific data, and their impact on the sensitivity of our threshold. The second method set thresholds using our facility's 98^{th} percentile dose indices, which is the current technique of flagging. The third method calculated statistical outliers based on our facility's site-specific protocol data. The accuracy and efficiency of each method were analyzed and compared.

Results:

Seven protocols were compared using $CTDI_{vol}$ as a dose metric while five were compared using SSDE; head and neck protocols were excluded as Imalogix™'s software has not implemented the conversation factors for SSDE for these exams found in TG 293. Following the analysis of this data, there were a total of 138 true positive acquisitions that needed to be reviewed. Dose reference levels flagged all true positive acquisitions, but on average flagged 70% more acquisitions than when using statistical outliers. 98^{th} Percentile flagged 50% fewer acquisitions than statistical outliers; however, this method missed flagging a total of 29 (20%) true positives. Statistical outliers flagged all true positives except for 2 in $CTDI_{vol}$ for Abdomen Pelvis without Contrast, thus flagging 98.6% of the true positives.

Conclusion:

Creating the correct CT dose threshold is imperative to ensuring accurate review and can be a considerable time saver for a facility. Site-specific data accounts for a facility's different technologies and are more efficient than non-site-specific data; however, it is important not to set a threshold too high where true positive exams that need to be reviewed do not get flagged. SSDE accounts for differences in patient size and is more representative of patient absorbed dose than $CTDI_{vol}$, thus is the preferred dose metric when applicable. Setting thresholds based on facilities' statistical outliers is an accurate and efficient method of creating thresholds that ensures patient safety and efficient tracking, while greatly minimizing false positives and saving large amounts of time and effort.

1. Introduction

The significance of Computed Tomography (CT) has increased over the past three decades, due to the increase in technological advances. [19] A CT scan uses a series of x-rays that create a cross-sectional image of a patient's anatomy. CTs account for 63% of patient exposure from radiologic and nuclear medicine procedures in the United States. [9] This diagnostic imaging modality has caused a critical reduction in exploratory medical procedures; however, careful attention must be paid to the safe use and optimization of CT. [2]

CTs generally use low doses of ionizing radiation and are performed only when deemed clinically necessary. Although the risks are small when compared to the benefits, the increased risk from radiation-induced cancers caused responses from journals, scholars, and the public media. The Joint Commission, an organization focused on patient safety and quality improvement, developed diagnostic imaging requirements to ensure that radiation doses from CT scans were tracked and reviewed when above their expected dose ranges. [20] [24] One difficulty brought about by The Joint Commission requirement is that there is no universal definition as to what constitutes an expected dose index. There are available benchmarks that allow sites to evaluate trends and compare practices to national reference values; however, these values are used to compare if dose indices are high and are not target doses. Benchmarks also could have numerous differences from one facility to another due to methodology, patient size, and optimization techniques. For these reasons, facilities need to develop a process to establish expected dose ranges for their specific CT data.

The primary objective of this research is to develop a standardized method for determining appropriate dose thresholds for diagnostic CT examinations. A radiology-based informatics system was used to export Oregon Health and Science University's CT dose data and to investigate specific exams to find cases that needed to be reviewed. Three separate methods that looked at site-specific and non-site-specific data were compared to find the most accurate and efficient method of setting dose thresholds. These methods were also compared with two separate dose metrics to find the more accurate measurement of dose. The purpose was to create a method that maximized sensitivity and specificity, by flagging all exams that needed to be reviewed (true positives) while minimizing flagging those exams that did not require review (false positives).

2. Background

2.1 CT Imaging

Computed Tomography (CT) was first developed by Godfrey Hounsfield in 1967 by using x-ray technology. [1] Diagnostic medicine was transformed, as Hounsfield and his colleague A.M. Cormack made it possible to look at the anatomy of a patient without the need of surgery. Through the use of mathematical constructs of image reconstruction, CT overcomes the limitations of conventional radiography in detail and clarity to produce cross sectional images with superior contrast and anatomic detail. [3]

In CT, the x-ray tube and detector are attached to a gantry that rotates around the patient. The patient is lying on a moving table, known as a couch, and is translated through the beam. The x-ray beam travels through the patient at many angles and positions, and an attenuation profile is then collected by the detector. [2] Once there are enough attenuation measurements, the data can be reconstructed into a volume that represents the patient's anatomy. Contrast between materials is formed by differences in material x-ray attenuation, with low attenuating material appearing dark and high attenuating material appearing bright. After reconstruction, the image can be displayed, stored, and viewed later for analysis.

2.1.1 Image Parameters

In CT scanning the dose levels and image quality are affected by imaging parameters called techniques. Techniques include peak tube potential (kVp), tube current (mA), effective tube current-time product (mAs), and pitch. [2] The peak tube potential determines how penetrating the x-ray beam is thus the quantity and quality of the

photons. The higher the kVp, the higher the average energy of the x-ray beam. [1] Typical tube potentials range from 80 kVp to 140 kVp. 80 kVp and 100 kVp are used when scanning a thin or pediatric patient since there is less material to penetrate. On the other end, 140 kVp is used when imaging larger patients since there is considerably more material to penetrate. An average adult will be scanned with a 120 kVp beam. Similar to tube potential, tube current and tube current-time product determines the quantity of the photons by affecting the beam intensity. Tube current is a measurement of how fast the electrons are flowing through the x-ray tube, while tube current-time product is the result of the tube current and the time per rotation of the CT scan. If current was increased while other parameters were held constant, radiation dose would increase while noise would decrease and vice versa. Radiation dose and noise must be balanced for ensuring patient safety while maintaining diagnostic image quality. [1][2][4]

Pitch describes the movement of the CT table per rotation of the gantry and is defined as:

$$Pitch = \frac{Table\ increment\ per\ rotation\ (mm)}{Beam\ Collimation} \quad (1)$$

The value of pitch represents the degree of overlap of the scan and generally ranges between 0.3 and 2. A pitch of 1 represents a contiguous CT rotation. A pitch greater than 1 represents a gap in the scan and indicates under scanning and results in a faster scan with lower radiation dose, but also lower image quality. A pitch less than 1 represents an overlapping scan and indicates an oversampling, resulting in a longer total scan time with better image quality but higher radiation dose. [2]

2.1.2 Scout Images

A preliminary two-dimensional scan is performed with a stationary gantry and the patient being translated through the bore. This image is called a scout, although it does have several other names such as topogram, scanogram, localizer, and CT radiograph. The scout image is a quick check used to visualize anatomic landmarks of the patient and can be taken anterior-posterior (AP), posterior-anterior (PA), and laterally. CT technologists use the scout to plan the scan parameters, determine the coverage of the scan, and also use it for tube current modulation (TCM). TCM optimizes the current to create consistent image quality over the entire body based on body composition. [2]

2.2 Dose Metrics

Dose indices available in CT are measures of radiation output and not absorbed dose within the patient, as it represents the amount of radiation needed to create an image. There is a difference between CT dose indices and patient dose and should be noted when talking to patients. CT dose is measured on a phantom but can be calculated in multiple ways, as is outlined below.

2.2.1 CTDI

CTDI stands for Computed Tomography Dose Index. This metric was not designed as a direct dose measurement for patient dose assessment, but rather as a means for comparing radiation outputs of different CT scanners. [2] CTDI is a measure from one axial CT scan and illustrates the average absorbed dose along the z-axis. [18]. Measuring this value clinically was difficult due the uncertainty of where and how to quantify it. Discussed below are the enhancements and modifications of CTDI.

2.2.1.1 CTDI100

With $CTDI_{100,}$ the dose is measured using a cylindrical "pencil" ionization chamber 100 mm long and 9 mm in diameter. This is a linear measure of dose along the z-axis using a polymethylmethacrylate (PMMA) cylindrical phantom. There are two standard PMMA phantoms each being 15 cm long. One is 32 cm in diameter to represent an adult body and the other is 16 cm in diameter to represent a pediatric body or adult/pediatric head. Holes are placed in the center and at 1 cm below the surface of the phantom at the 12, 3, 6, and 9 o'clock positions. The pencil ionization chamber is inserted into one of these holes and a CT scan is taken with no table translation. $CTDI_{100}$ measures the amount of radiation exposure to the gas-filled chamber; however, does not consider any variation in the human body and should not be used clinically as a measure of patient dose. [2] [12][18]

2.2.1.2 $CTDI_w$

$CTDI_w$ is a weighted model that was introduced to account for the uneven dose distribution between the center and periphery of the phantom. CTDI varies across the field of view and is higher at the surface than the center and is given by the following equation.

$$CTDI_w = \frac{1}{3} CTDI_{100,center} + \frac{2}{3} CTDI_{100,periphery} \quad (2)$$

Although $CTDI_w$ creates a closer human dose profile when compared to $CTDI_{100}$, there is still no table translation. CT scans are affected by the patient moving through of the gantry and pitch should be taken into consideration. [2] [18]

2.2.1.3 $CTDI_{vol}$

The previous dose metrics have all only considered the dose of a single axial scan; however, most patients are being imaged with a helical CT. A patient translated through the bore as the x-ray beam is continuously rotated. Volume CTDI takes this into consideration and is given by the following equation.

$$CTDI_{vol} = \frac{CTDI_w}{pitch} \tag{3}$$

Pitch considers how far the table translates per full rotation of the gantry and determines if there is an overlap in rotation, thus is a factor when considering radiation dose. As pitch decreasing, causing an oversampling, the $CTDI_{vol}$ will increase. Most scanners have $CTDI_{vol}$ displayed on their consoles and use it as a standard way of measuring dose. CT manufacturers measure $CTDI_{vol}$ in a factory for different tube potentials and pitches, allowing the $CTDI_{vol}$ to be displayed prospectively in a clinical setting. [2] This makes $CTDI_{vol}$ the most accessible dose index displayed on CT scanners and lends itself to directly compare radiation output from multiple protocols.

2.2.2 SSDE

Although $CTDI_{vol}$ is widely accessible, the dose that is measured is for a standard-sized phantom and does not represent radiation dose for different sized patients. [18] There are many different sized patients and if they are not exactly the size of the phantom the dose will be incorrect. When a patient is smaller than the phantom, the $CTDI_{vol}$ shown on the

scanner will be an underestimate of the dose and patients that are larger than the phantom will display an overestimate for the dose output.

To better account for the variety in patient sizes, Size Specific Dose Estimates (SSDEs) were created. SSDE is given by the following equation.

$$SSDE = f \; x \; CTDI_{vol} \tag{4}$$

In equation 4, f represents a correction factor found in The American Association of Physicists in Medicine (AAPM) Report 204 based on the effective diameter of the patient and the $CTDI_{vol}$ is from the scanner. The effective diameter of the patient is taken from the following equation.

$$Effective\ Diameter = \sqrt{AP \; x \; LAT} \tag{5}$$

In equation 5, the AP and LAT represent the length of the patient's body in the AP and lateral direction in centimeters. SSDE is independent of the scanner manufacturer. It is important to check on the correction factor, f, before using it to ensure it is specific to the appropriate 16 or 32 cm phantom. [2]

In more recent years there has been another AAPM report, known as Task Group 220, that uses a water equivalent diameter to measure SSDE instead of geometric size. The absorption of x-rays relies heavily of the x-ray attenuation of different materials in the body and is imperative in determine the radiation dose absorbed by the patient. For example, the thorax and abdomen could have the same external effective diameter; however, since lungs are a low attenuating material compared to abdominal tissues, the

thorax would attenuate less x-rays and thus have a lower dose. The water equivalent diameter method creates a more precise portrayal of patient dose; however, it is not implemented everywhere and will not be used in this study. [26]

2. 3 Risks of Ionizing Radiation

There has been an increased demand for CT scans over the past decades due to the technological advances that increased their capability. In the United States, there are approximately 80 million CT examinations performed yearly. [19] This increased use of CT has decreased the use of invasive exploratory surgeries; however, CT scans do still have risks from ionization radiation. Ionizing radiation is a type of energy that can remove an electron from its atom. When ionizing radiation enters the body it can create hydroxyl radials from x-rays interactions with water molecules. These free radicals can interact with DNA and cause strand breaks. X-rays can also interact with our DNA directly and cause those strand breaks. The majority of radiation-induced damage is rapidly repaired; however, double-strand breaks can lead to mutations, chromosome translocations, and gene fusions. These effects of radiation can be linked to the potential of possible cancer later in life. [14][15]

CT utilizes relatively low doses of ionizing radiation, with an average exposure from a whole-body CT scan being about 10 mSv. [16] Most of the data concerning radiation-induced cancers come from studies involving survivors of the atomic bombs dropped in Japan in 1945. The BEIR VII report made a lifetime risk model that predicts one person out of one hundred exposed to 100 mSv would develop cancer. Now 100 mSv is much more than a person would receive from any diagnostic imaging exam. Compared to the

potential induction of cancer from radiation exposure it is said that forty-two out of one hundred people would expect to develop cancer due to other causes. Radiation exposure can be dangerous, but CT exams are performed when deemed clinically necessary and when benefits outweigh the hypothetical risks. [15][16]

2.4 The Joint Commission

The Joint Commission is a not-for-profit organization founded in 1951 that is focused on patient safety along with quality improvement in health care. [24] This organization sets standards of practice and provides accreditation for hospitals in the United States, and now certifies over 22,000 healthcare organizations. Accreditation is not mandatory for hospitals; however, it demonstrates that they are committed to a high level of excellence. These standards help increase and strengthen the patient safety efforts in place and may fulfill any regulatory requirements. [24]

The Joint Commission gives out their "Gold Seal of Approval" by inspecting the organization with an on-site survey every three years and requires data be submitted to them every 3 months. During these on-site visits, The Joint Commission evaluates standard compliance by looking at medical records, talking to staff, and looking into patient experiences. [24]

Standards are updated regularly, and The Joint Commission is encouraged to do research when their regulations have any gaps. In order to ensure the safe use of the advanced form of technology from diagnostic imaging, The Joint Commission decided to post new Diagnostic Imaging Requirements on August 10, 2015. [20] One of the new standards,

PI.02.01.01, requires organizations to review and analyze incidents where the radiation dose index from diagnostic CT exams exceeded the expected dose index ranges identified in imaging protocols. These incidents then need to be compared to external benchmarks. Because of this, facilities now need to establish protocol-specific expected dose ranges for their CT exams. [13][20]

2.5 External Benchmarks

The dose index recommendations for each protocol is based on a wide variety of factors; there is no single, all-encompassing source that provides all applicable dose index ranges. [13] External benchmarks enable sites to evaluate trends and compare their practices to national values on a continuous basis. These benchmarks are not facility- or manufacturer- specific and should only be used to see if facility doses are unusually high.

2.5.1 National Dose Reference Levels

The International Commission on Radiological Protection (ICRP) was the first to indicate the need for diagnostic reference levels (DRLs) in 1990. [7] ICRP recommends there be DRLs for all patients and should be used as a simple test for identifying times when the doses are remarkably high. DRLs are usually set at the third quartile values, representing the 25th, 50th, and 75th percentiles of the national standards. [21] These cover a wide spread of patient groups, protocols, scanners, and facilities. This value does not create a standard of "best" practice, nor does it set dose limits, but acts as a means of comparison across many organizations.

2.5.2 ACR DIR

The American College of Radiology (ACR) Dose Index Registry (DIR) is a part of the National Radiology Data Registry (NRDR). The purpose of the NRDR is to provide facilities support in quality improvement efforts and like the DRLs to compare data to national values continually. The DIR is one of eight registries in the NRDR and enables facilities to compare dose indices for CT and fluoroscopy against regional and national values. The CT DIR has received 102 million CT exams to date and has become the gold standard for benchmarking. [6] The DIR is focused on high-quality exam performance along with optimizing patient care. CT dose data is collected and anonymized from many sites across the nation and transmitted to the ACR. A facility can log into the NRDR to look at their performance data which includes graphical feedback, summary statistics, and dose indices. The DIR does not create expected dose ranges and is compared across many different CT technologies and scanners, hence should only be used as a comparison metric of performance. [23]

2.6 Previous Methods of Setting Dose Thresholds

With The Joint Commission's requirement of reviewing CT exams, facilities have needed to decide what their dose threshold should be. Some facilities have chosen to set national DRLs as their dose thresholds. These values represent an average dose based on a large number of CT scans and demonstrate a good picture of what dose looks like. Facilities using this method of setting dose thresholds have stated that DRLs can be outdated and different than their own facility, thus is not well suited for optimization of this tracking dose indices. [10] With the variety in national DRLs, there has been research on creating institutional DRLs based on site-specific data, such as in Liang et al and MacGregor et al.

These values have not been used as dose thresholds but are said to be a benefit in quality improvement and radiation dose optimization efforts in facilities. [28] [29]

There are other studies that have created techniques to develop dose thresholds based on their dose indices data. Crowley et al implemented a radiation dose alert system in CT that create a red alert when an imaging study dose went above that protocols dose threshold. The dose threshold they created was based on two times the median dose length product (DLP) of that protocol. A DLP takes the length of the scan in centimeters and times that by the $CTDI_{vol}$. This research found that procedural documentation errors and patient-related factors, such as size, were associated with false alerts. [27]

2.7 Radiation Dose Index Monitor Systems

Radiation dose index monitoring systems consist of software that collects radiation doses and patient data retrospectively and stores this in a database. These dose monitoring systems collect data either straight from the equipment or through a picture archiving and communication system (PACS). The overall goal of these systems is to be an addition to the quality assurance measures in a facility. This software can help reduce risk, achieve and maintain accreditation compliance, and overall streamline the review process, thus saving time and effort. [5]

2.7.1 Imalogix™

OHSU utilizes Imalogix™ exposure monitoring software to monitor patient dose for CT and Fluoroscopy exams. The goal of Imalogix was to aid in patient care, technologist performance, and overall improvement of quality assurance. Imalogix receives DICOM

headers, radiation dose structure reports (RDSR), and all CT images (scout and reconstructions) through PACS. Each scan that is received is processed to determine scan parameters, patient misalignment, and other acquisition and dose information.

3. Materials and Methods

3.1 CT Dose Indices Threshold Generation

With the Joint Commission requirement of reviewing and analyzing dose indices from diagnostic CT exams, a standardized method of creating thresholds is needed. To develop a more accurate and time-sensitive method, OHSU's site-specific data was used. The threshold developed was based on the dose level a statistician would consider an outlier, known as a statistical outlier.

3.1.1 Statistical Outlier

An outlier is a data point that lies an abnormal distance from other values in a dataset. [17] Statistical outliers are an ideal way of creating a threshold since they should only include diagnostic CT exams that have dose indices that are unusually high compared to the rest of the data. The equation to calculate an upper bound statistical outlier is as follows.

$$Statistical\ Outlier = 75^{th}\ percentile + 1.5 * IQR \quad (6)$$

In equation 6, the 75th percentile is the third, upper, quartile which represents that 75% of the data lies below this margin. The IQR is the interquartile range which depicts the middle 50% of the data. [17] This value was calculated for seven different protocols by exporting dose indices from Imalogix and uploading them into R studio®. R studio is an integrated development environment (IDE) for the programming language R. [25] It allows users to code and develop statistical programs.

3.1.2 Limiting Parameters

Creating statistical outliers as a method of setting thresholds is the first step in the process of determining which exams require review. Once a dose outlier is defined, it can be used as a benchmark for flagging exams with high dose indices. OHSU performs over 35,000 CT exams per year, making it necessary to limit the data examined for the purposes of this study.

The annual period covered for flagging review was the 2019 calendar year, from January 1st to December 31st, 2019. The year 2019 was used due to the pandemic causing inconsistencies in procedure types and number of exams throughout 2020. For simplicity of range of sizes and readily available dose information, only adult exams were looked at.

The seven protocols that were looked at were based on a journal article by Kalpana M. Kanal et al., "U.S. Diagnostic Reference Levels and Achievable Doses for 10 Adult CT Examinations". [8] The purpose of this article was to develop DRLs for the ten most common adult CT exams based on the ACR DIR. When comparing the ten protocols used in this study to the protocols used at OHSU, a total of seven protocols could be compared. OHSU did not have similar protocol names to the other three, so they were omitted. The seven protocols used were Chest Abdomen Pelvis with contrast, Abdomen Pelvis with contrast, Abdomen Pelvis without contrast, Chest with contrast, Chest without contrast, Neck with contrast, and Head without contrast.

All seven protocols were compared using $CTDI_{vol}$ as a dose index, while only five were compared using SSDE. SSDE was not calculated for Neck with contrast and Head

without contrast due to the conversion factors for these bodies areas not being programmed into Imalogix. AAPM report 293 has created SSDE conversion factors for head CTs; however, these conversion factors are still new and has not been widely implemented into systems.

3.2 Flagging

3.2.1 Exporting Data from Imalogix

In Imalogix there is a dashboard known as Analytics, within this section CT examinations can be reviewed and analyzed based on all factors that impact their exam like imaging techniques, dose indices, and patient size. Information can be viewed by acquisition or by study, and since there can be multiple acquisitions with different protocols per study, acquisition level data is what was evaluated for this research. Once all the limiting parameters are filtered, acquisitions can be sorted by both SSDE and $CTDI_{vol}$. The protocol tab within the Analytics dashboard is where all acquisition data is stored, including the date, accession number, protocol, scan number, dose index, effective diameter, average mA, TCM utilization, maximum kVp, minimum pitch, and revolution time. This data can be extracted from Imalogix by downloading it into a CSV file and using Microsoft Excel for analysis.

3.2.2 Filtering Data

Once all dose data is in excel, it is necessary to filter it in order to look only at doses that would have the potential of being flagged. Any acquisition that did not have a dose index included was omitted from the study. The Dose Reference Levels from Kanal's article was used as a guideline as to what acquisitions to look into for flagging. [8] These DRLs are readily available and are the 75^{th} percentile of dose data from around the nation. For

the purposes of determining which exam would likely not be flagged, it may be reasonable to use this value as a benchmark. All of the data in excel was then filtered to only include acquisitions with doses above the DRLs.

3.2.3 Finding True Positives

Once all possible flagged exams are established, the true positives have to be found. A true positive is when an outcome is correctly predicted. In this case, a true positive means that a flagged exam requires review by a physicist or technologist. True positives represent cases when something might have gone awry, or a wrong technique/protocol was used. An example of this could be if a large patient technique was used on a smaller-sized patient. As previously stated, the higher the kVp, the higher the average energy of the x-ray beam, and 140 kVp is used on larger patients because they have more material to penetrate. When using a larger kVp on an average patient this will unnecessarily increase the dose to the patient and these types of exams should be reviewed to see why the technologist might have used large patient protocols.

To determine what acquisitions should be flagged, Imalogix has a search function. Any acquisition that had a dose index higher than the DRLs was individually searched by accession number. When the accession number is found a study level detail view is shown. This provides users with all of the information needed to perform a root cause analysis on this potential outlier. It has an exam dose summary, which shows study level dose information, and provides the highest dose indices from all acquisitions. The patient's effective diameter is determined automatically in the software by measuring the patient's length in the center of the scan boundaries in both the AP and lateral direction.

Effective diameter helps calculate SSDE and correlates to a patient size group in Imalogix (XXS-XXL). Also included are the protocol name, imaging parameters, and dose indices, as well as series details, location information, and patient history. All of this information is useful and helps in determining whether the exam should/should not be flagged, and if it requires follow-up.

The process of finding all true positive exams was done for all seven protocols for SSDE and $CTDI_{vol}$. Once the true positives were found they were marked in the excel file for future analysis.

3.3 Comparing True Positive Rates

The statistical outlier method was then compared to two other methods of setting dose indices thresholds. The first method is the DRLs, as these values are publicly available benchmarks that any site could use. The second method was using the 98^{th} percentile of all of OHSU's CT examination data. In excel, all three dose thresholds were used to see the number of exams that would be flagged. The flagged exams were compared to our true positives to compare the efficacy of each method.

Each thresholding method was compared by sensitivity, specificity, false positive fraction (FPF), and accuracy. Sensitivity is the probability that a test will be positive for those that actually are positive. This can be easily understood by thinking of patients being tested for a disease; sensitivity is the probability that people that have the disease will test positive. [17] In this case, it means how often are acquisitions that should be flagged are

being flagged. Sensitivity is also known as the True Positive Fraction (TPF) and is found by the following equation.

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

In equation 7, the TP stands for True Positive and is the flagged acquisitions that should be flagged. The FN stands for False Negative and represents acquisitions that should have been flagged but did not. False negatives occur when a dose threshold was higher than that of a true positive, thus missed flagging that acquisition. Specificity is similar to sensitivity, except looks at the probability that a test will be negative for those that are actually negative. This refers to how often will a threshold method be able to identify acquisitions that should not be flagged and is calculated by the following equation. [17]

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

In equation 8, the TN stands for True Negative and are the acquisitions that did not get flagged and should not have been flagged. Any exams that were below the DRL were considered TN as this value was used as a benchmark for what exams may not have been flagged. The FP stands for False Positive which are the acquisitions that were flagged but should not have been flagged, i.e., nothing was wrong.

These measurements are important when determining which method is going to be the best for the facility. Along with each of these measurements, false negative acquisitions should be looked into. The goal of the project is to create a method that saves facilities

time and has ideal accuracy. If there are any FN acquisitions, then that method is missing important information that needs to be reviewed.

4. Results

4.1 CT Dose Indices Thresholds

As previously stated, CT dose indices thresholds were created by three separate methods. The first being calculating a statistical outlier based on OHSU's site-specific data, the second being DRLs from a journal publication that is based on the ACR DIR, and the third being OHSU's 98th percentile of their CT data. These dose thresholds were created for both SSDE and $CTDI_{vol}$. A reference table for SSDE dose thresholds for the five protocols that measure SSDE is given in Table 1, and a reference table for $CTDI_{vol}$ dose threshold for all seven protocols evaluated are given in Table 2. All dose thresholds are given in milligray (mGy) as this unit is the measurement for dose and what DICOM headers display.

Table 1: SSDE Dose Thresholds from the Three Methods for the Five Protocols that use SSDE

SSDE Dose Thresholds (mGy)			
	DRLs	**Statistical Outlier**	**98th Percentile**
Abd Pelvis WCon	18	30.47	56.81
Abd Pelvis WOCon	19	22	27.05
Chest WCon	15	18.62	22.68
Chest WOCon	15	18.41	20.7
CAP WCon	18	32.31	54.38

Table 2: $CTDI_{vol}$ Dose Thresholds from the Three Methods for all Seven Protocols

$CTDI_{vol}$ Dose Thresholds (mGy)			
	DRLs	**Statistical Outlier**	**98th Percentile**
Abd Pelvis WCon	15	36.4	82.9
Abd Pelvis WOCon	16	26.95	36.8
Chest WCon	13	22	25.68

Chest WOCon	12	21.53	23.1
CAP WCon	15	33.35	63.71
Neck WCon	19	21.08	27.04
Head WOCon	56	54.4	52.1

4.2 Protocol Statistics

4.2.1 Abdomen Pelvis with Contrast

Data for Abdomen Pelvis with contrast included 1,838 acquisitions for SSDE in 2019 and is shown in Table 3. There was a total of 10 true positive acquisitions showing that 0.54% of all acquisitions needed to be reviewed. When using the DRL of 18 mGy as a threshold a total of 460 of the acquisitions were flagged, of which only 2.17% of the flagged exams were true positives. DRL has a 100% sensitivity however the specificity was 75.37%. The 98th percentile method of 56.81 mGy flagged only 31 acquisitions, with 25% of those being a true positive. However, this method missed 2 true positives causing it to have a sensitivity of 80%. When using statistical outlier 30.474 mGy as a threshold, 76 acquisitions were flagged. Of these flagged exams, 13.16% were truly positive. This method had a 100% sensitivity and 96.39% specificity.

Table 4 is a reference table for the statistics on Abdomen Pelvis with contrast for $CTDI_{vol}$ as a dose index. There were a total of 1,872 acquisitions in 2019 with 11 of them being a true positive. Both DRL and statical outlier methods flagged all 11 true positives creating a 100% sensitivity; however, a DRL of 15 mGy flagged 360 acquisitions compared to the 101 acquisitions for the 36.4 mGy Statistical Outlier. The 98th Percentile of 82.9 mGy had the highest specificity of 98.64% but did miss 7 of the 11 true positives, creating a sensitivity of 36.36%.

Table 3: Statistics for Abdomen Pelvis with Contrast (SSDE)

SSDE for Abdomen Pelvis with Contrast			
Total Exams	**1838**		
Total TP	**10**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	460	76	31
TP	10	10	8
TN	1378	1762	1805
FP	450	66	23
FN	0	0	2
Sensitivity	100%	100%	80%
Specificity	75.37%	96.39%	98.74%

Table 4: Statistics for Abdomen Pelvis with Contrast ($CTDI_{vol}$)

$CTDI_{vol}$ for Abdomen Pelvis with Contrast			
Total Exams	**1872**		
Total TP	**11**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	360	101	34
TP	11	11	4
TN	1242	1771	1831
FP	619	90	30
FN	0	0	7
Sensitivity	100%	100%	36.36%
Specificity	66.72%	95.14%	98.37%

4.2.2 Abdomen Pelvis without Contrast

Data for Abdomen Pelvis exams without contrast had 2,470 acquisitions for SSDE in 2019 and is seen in Table 5. Of those acquisitions, there were 29 true positives. All three methods flagged all 29 true positives creating a 100% sensitivity. DRL had a 19 mGy threshold which flagged a total of 245 acquisitions, with 12% of those being a true positive. The 98th percentile threshold of 27.05 mGy only flagged 58 acquisitions, thus 50% of them were truly positive. Statistical outlier had a threshold of 22 mGy which flagged 136, with 21% of those being a true positive.

For $CTDI_{vol}$, with reference information is in Table 6, had a total of 2,487 acquisitions with 31 true positives. With a low threshold of 16 mGy, DRL flagged 596 acquisitions, and this flagged all true positives creating a sensitivity of 100% with a specificity of 76.99%. 98th percentile of 36.8 mGy flagged 50 acquisitions creating specificity of 98.86%; however, missed 29% of the true positive acquisitions for a sensitivity of 70.96%. Statistical outlier of 29.95 mGy, missed two true positives while flagging a total of 133 acquisitions.

Table 5: Statistics for Abdomen Pelvis without Contrast (SSDE)

SSDE for Abdomen Pelvis without Contrast			
Total Exams	**2470**		
Total TP	**29**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	245	136	58
TP	29	29	29
TN	2225	2334	2412
FP	216	107	29
FN	0	0	0
Sensitivity	100%	100%	100%
Specificity	91.13%	95.61%	98.81%

Table 6: Statistics for Abdomen Pelvis without Contrast ($CTDI_{vol}$)

$CTDI_{vol}$ for Abdomen Pelvis without Contrast			
Total Exams	**2487**		
Total TP	**31**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	596	133	50
TP	31	29	22
TN	1891	2352	2428
FP	565	104	28
FN	0	2	9
Sensitivity	100%	93.54%	70.96%
Specificity	76.99%	95.76%	98.86%

4.2.3 Chest with Contrast

Data for Chest with contrast included 752 acquisitions for SSDE in 2019 and is shown in Table 7. Of those acquisitions, there were 7 true positives. All three methods flagged all 7 acquisitions, creating a sensitivity of 100% across the board. As threshold increases from

each method, with DRL being 15 mGy, Statistical Outlier being 18.62 mGy, and 98th Percentile being 22.68 mGy, the specificity increases.

Table 8 conveys the statistics for $CTDI_{vol}$, where there were 7 true positive acquisitions out of a total of 814. Both DRL and Statical Outlier methods flagged all 7 true positives creating a 100% sensitivity; however, a DRL of 13 mGy flagged 62 acquisitions compared to the 30 acquisitions for the 22 mGy Statistical Outlier. The 98th Percentile of 25.68 mGy had the highest specificity of 98.00% but did miss 3 of the 7 true positives creating a 57.14% sensitivity.

Table 7: Statistics for Chest with Contrast (SSDE)

SSDE for Chest with Contrast			
Total Exams	**752**		
Total TP	**7**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	62	30	19
TP	7	7	7
TN	722	722	733
FP	23	23	12
FN	0	0	0
Sensitivity	100%	100%	100%
Specificity	92.61%	96.91%	98.38%

Table 8: Statistics for Chest with Contrast ($CTDI_{vol}$)

$CTDI_{vol}$ for Chest with Contrast			
Total Exams	**814**		
Total TP	**7**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	223	30	20
TP	7	7	4
TN	591	784	791
FP	216	23	16
FN	0	0	3
Sensitivity	100%	100%	57.14%
Specificity	73.23%	97.14%	98.00%

4.2.4 Chest without Contrast

Chest without Contrast had a total of 3,308 acquisitions for SSDE in 2019 with a total of 11 true positives, shown in table 9. DRL flagged a total of 223 acquisitions with a threshold of 15 mGy, 98th percentile flagged 62 with a threshold of 23.1 mGy, and statistical outlier flagged 103 with a threshold of 18.41 mGy. Both DRL and statistical outlier flagged all 11 acquisitions with a sensitivity of 100%, but statistical outlier had a higher specificity at 97.21%. 98th Percentile had a specificity of 98.36%, however missed 3 true positive acquisitions creating a sensitivity of 72.72%.

With $CTDI_{vol,}$ statistics can be seen in Table 10. There was a total of 3,484 acquisitions with only 5 true positive exams. DRL had a threshold of 12 mGy and flagged over 1000 acquisitions, with 0.46% being truly positive. While the 98th percentile threshold was 23.1 mGy and the statistical outlier was 21.525 mGy, both of those methods flagged less than 100 acquisitions. 98th Percentile had a 7.46% rate of flagging true positives while statistical outliers had a 5.8% true positive fraction. All three methods flagged all 5 acquisitions.

Table 9: Statistics for Chest without Contrast (SSDE)

SSDE for Chest without Contrast			
Total Exams	**3308**		
Total TP	**11**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	223	103	62
TP	11	11	8
TN	3085	3205	3243
FP	212	92	54
FN	0	0	3
Sensitivity	100%	100%	72.72%
Specificity	93.57%	97.21%	98.36%

Table 10: Statistics for Chest without Contrast ($CTDI_{vol}$)

$CTDI_{vol}$ for Chest without Contrast			
Total Exams	**3484**		
Total TP	**5**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	1077	86	67
TP	5	5	5
TN	2407	3398	3417
FP	1072	81	62
FN	0	0	0
Sensitivity	100%	100%	100%
Specificity	69.18%	97.67%	98.21%

4.2.5 Chest Abdomen Pelvis with Contrast

Chest Abdomen Pelvis with Contrast had a total of 826 acquisitions in 2019 for SSDE, with a total of 5 true positives, shown in Table 11. DRL and statistical outliers flagged all 5 true positives creating a sensitivity of 100% while 98th Percentile only flagged 3 of the 5 true positives making sensitivity 60%. Specificity increases as dose threshold increases per method.

When looking at $CTDI_{vol}$ in 2019 there was a total of 833 acquisitions where 12 of those are true positives which can be seen in Table 12. Similarly to SSDE, DRL and statistical outlier flag all true positives compared to the 98th Percentile that missed flagging 3 of the acquisitions. DRL threshold was 15 mGy which flagged 276 acquisitions, Statistical Outlier threshold was 32.31 mGy and flagged 45 acquisitions, while 98th percentile's threshold of 54.38 mGy flagged only 20 acquisitions.

Table 11: Statistics for Chest Abdomen Pelvis with Contrast (SSDE)

SSDE for Chest Abdomen Pelvis with Contrast			
Total Exams	**826**		
Total TP	**5**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	216	31	12

TP	5	5	3
TN	610	795	812
FP	211	26	9
FN	0	0	2
Sensitivity	100%	100%	60%
Specificity	75.62%	98.29%	99.39%

Table 12: Statistics for Chest Abdomen Pelvis with Contrast ($CTDI_{vol}$)

$CTDI_{vol}$ for Chest Abdomen Pelvis with Contrast			
Total Exams	**833**		
Total TP	**12**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	276	45	20
TP	12	12	9
TN	557	788	810
FP	264	33	11
FN	0	0	3
Sensitivity	100%	100%	75%
Specificity	68.11%	96.95%	98.66%

4.2.6 Neck with Contrast

Neck with Contrast is one of the two protocols that do not use SSDE as a dose metric. Looking just at $CTDI_{vol}$ there were 999 acquisitions in 2019, 8 of which were true positives. All three methods flagged all 8 exams, which can be seen in Table 13. DRLs flagged 101 acquisitions with a 7.9% true positive fraction, compared to 98th percentile flagging 28 with a 28.6% true positive fraction, and statistical outlier flagging 64 with 12.5% being truly positive.

Table 13: Statistics for Neck with Contrast ($CTDI_{vol}$)

$CTDI_{vol}$ for Neck with Contrast			
Total Exams	**999**		
Total TP	**8**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	101	64	28
TP	8	8	8
TN	898	935	971
FP	93	56	20
FN	0	0	0
Sensitivity	100%	100%	100%
Specificity	90.62%	94.35%	97.98%

4.2.7 Head without Contrast

Head without Contrast is the second protocol that just uses $CTDI_{vol}$ as a dose metric. In 2019 there was a total of 2657 acquisitions and only 2 of those were true positives. This protocol is the only of the seven whose 98th Percentile threshold is the lowest at 52.1 mGy and the DRL is the highest at 56 mGy. All three methods flagged the 2 true positives and the DRL has 100% sensitivity which can be seen in Table 14.

Table 14: Statistics for Head without Contrast ($CTDI_{vol}$)

$CTDI_{vol}$ for Head without Contrast			
Total Exams	**2657**		
Total TP	**2**		
	DRL	**Statistical Outlier**	**98th Percentile**
Total Flags	2	5	5
TP	2	2	2
TN	2655	2652	2652
FP	0	3	3
FN	0	0	0
Sensitivity	100%	100%	100%
Specificity	100%	99.89%	99.89%

5. Discussion

5.1 Comparing Methodology

When comparing methodology, all factors need to be considered to find the best overall approach of flagging CT dose indices that need to be reviewed. One requirement is that the vast majority, if not all, true positives should be flagged. The purpose of the review process is to flag exams that need to be reviewed, meaning it is imperative to ensure that acquisitions that went wrong are not just slipping by. In this study, there were a total of 138 true positive acquisitions that were determined to require review. When a method flags all possible true positive acquisitions, this represents a sensitivity of 100%. Sensitivity and specificity are factors that are desirable when they are as high as possible.

As previously stated, specificity shows when a method can recognize when acquisitions should not be flagged. This value is not expected to be 100% since the threshold has to be low enough to flag all true positive exams. There are going to be times where an acquisition has a high dose index; however, nothing was wrong with their exam. An example of this would be when a very large patient is being scanned, to distinguish between the different anatomical features more radiation dose has to be used. Large patients could have a high dose that gets flagged but does not need to be reviewed, which will cause specificity not to be 100%. As shown in the results, no matter the method, more than 50% of the flagged exams were false positives and did not need review.

True negatives, as mentioned earlier, represent acquisitions where the dose index was below the protocol DRL. For this reason, the TN value does not depend on the method

being used. True positive acquisition details are essential and can be expressed in sensitivity. Sensitivity also accounts for any false negative acquisitions which are when an exam should be flagged but was not. Every protocol had a 100% sensitivity for DRLs since this was the cutoff of the acquisitions being flagged.

As indicated above, false positives are unavoidable no matter the method that is being used, but the less there is the more time that will be saved. Let's say that it takes around 3 minutes to review a flagged acquisition. Looking at Table 10 for Chest without contrast based on $CTDI_{vol}$, DRLs flagged a total of 1,077 acquisitions which would equate to 3,231 minutes or 53.85 hours of reviewing for the whole year. This is a lot of time for someone to spend just on reviewing diagnostic CT dose indices. Compared to DRLs, statistical outlier only flagged 86 acquisitions equating to 258 minutes or 4.3 hours for a year, corresponding to about 20 minutes per month. 20 minutes is manageable and is extremely reasonable for a person to spend reviewing flagged exams in one month.

5.1.1 Statistical Outlier Versus DRLs

Using SSDE as a dose metric, both Statistical Outliers and DRLs had 100% sensitivity for all protocols, as seen in Figure 1. Across the five different protocols, DRLs flagged 70% more acquisitions than the statistical outlier method. This factors greatly in the specificity and FPF by increasing the number of false positives. DRLs had a lower specificity and higher FPF than statistical outliers. As mentioned previously, when the number of flagged exams increases, it results in a large increase in the time required to review these exams. Using $CTDI_{vol}$ as a dose metric has a similar outcome as SSDE except for the protocol Abdomen Pelvis without contrast has two false negatives for

statistical outlier causing the sensitivity to drop from 100%, as seen in Figure 2. This is the only instant where using statistical outlier as a threshold created any false negative acquisitions. Statistical Outliers are based on site-specific data and flagged 98.6% of true positive acquisition. DRLs represent data that is acquired at many sites across the nation with different technologies than OHSUs. Overall when comparing statistical outliers to DRLs, statistical outliers are nearly as accurate in flagging true positives while being immensely more efficient in saving a facility time in the reviewing process.

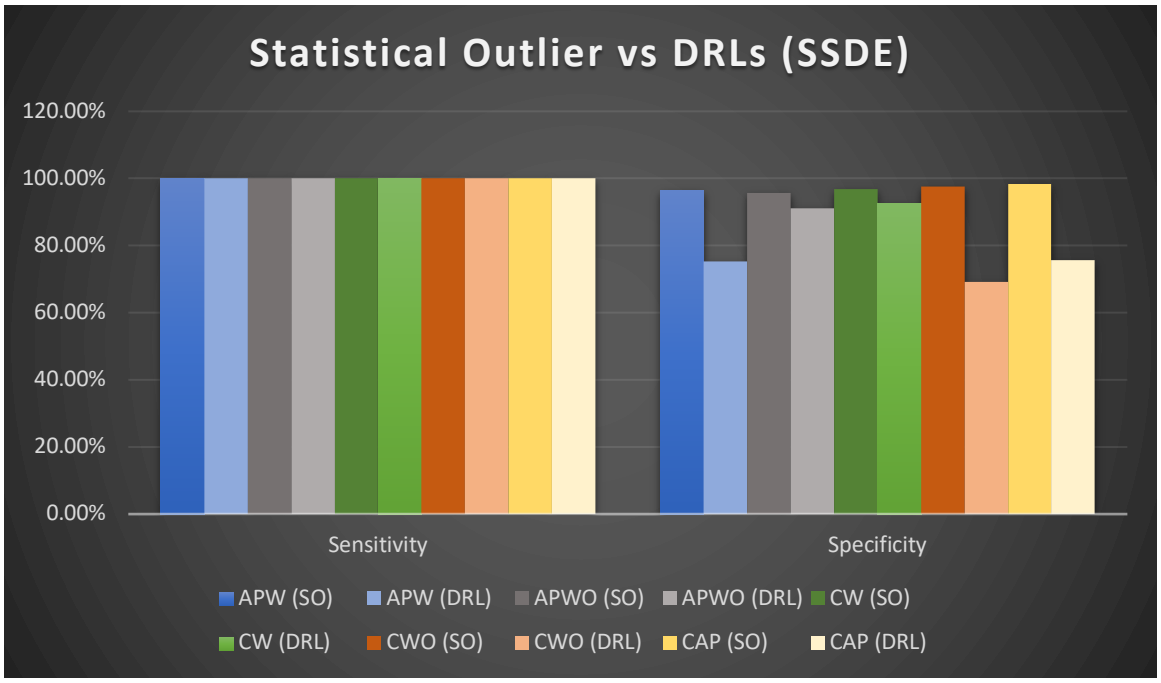

Figure 1: Statistics of Statistical Outlier Versus DRLs for SSDE

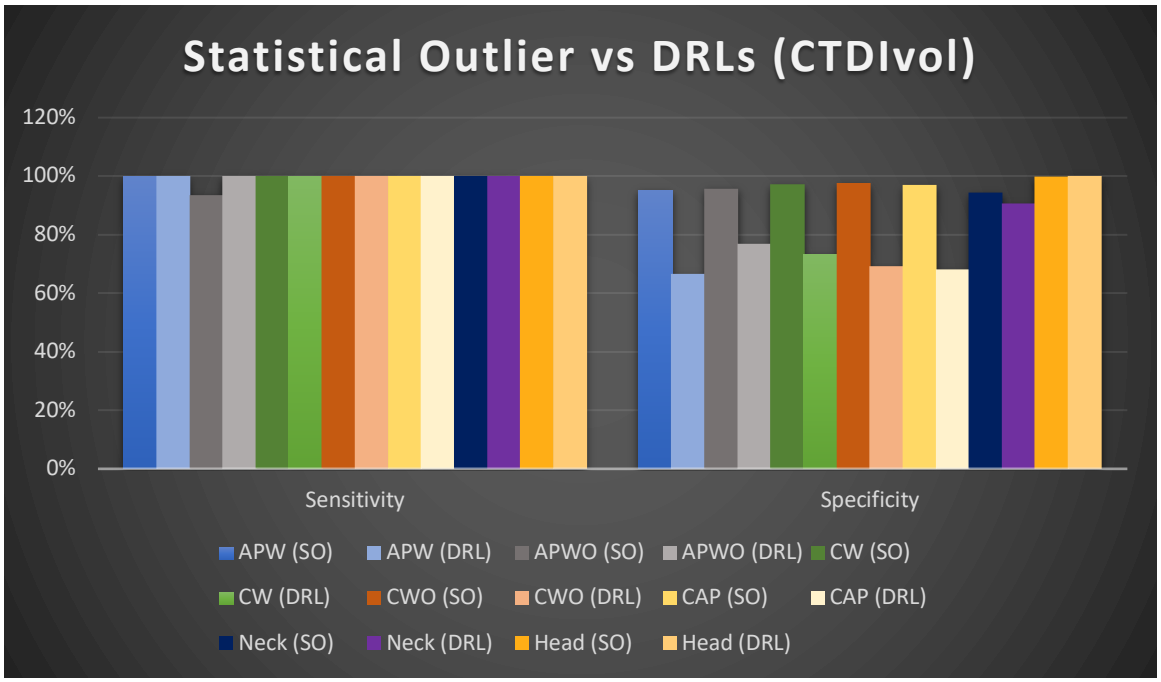

Figure 2: Statistics of Statistical Outlier Versus DRLs for $CTDI_{vol}$

5.1.2 Statistical Outlier Versus 98th Percentile

When comparing statistical outliers to 98th percentile, the statistics are opposite from DRLs since the 98th Percentile has a higher threshold for the majority of protocols. This can be seen in Figures 3 and 4. 98th percentile flags on average 50% fewer acquisitions than statistical outliers. Although flagging fewer acquisitions does help save time spent reviewing flagged exams, there is a higher risk of missing acquisitions that need to be reviewed. Out of the five protocols looked at for SSDE, 98th percentile missed true positives in three, while statistical outliers missed none. For the seven protocols with $CTDI_{vol}$, 98th percentile also missed true positives in three, while statistical outlier missed them in one (Abdomen Pelvis without contrast). 98th percentile missed flagging a total of 20% of all true positive acquisitions while statistical outlier only missed flagging 1.4% of true positives. Overall when looking at statistical outliers versus 98th percentile, statistical

outliers are much more accurate in flagging true positives while staying efficient on saving time.

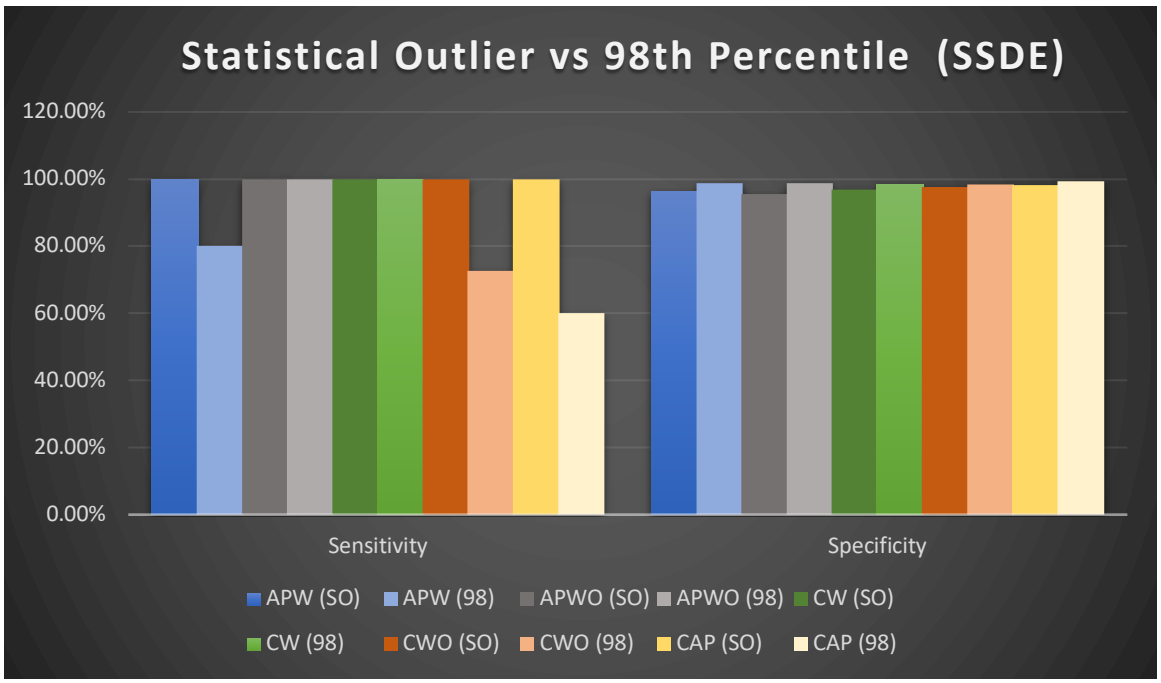

Figure 3: Statistics of Statistical Outlier Versus 98th Percentile for SSDE

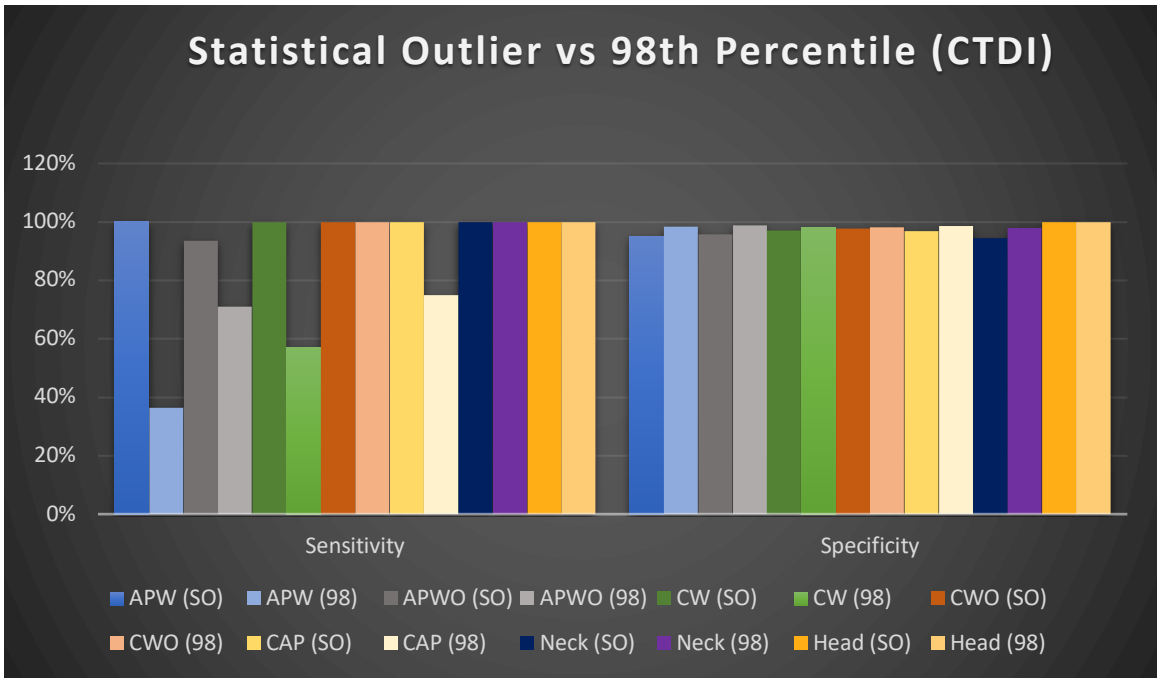

Figure 4:Statistics of Statistical Outlier Versus 98th Percentile for SSDE

5.2 Dose Index Comparison

SSDE is the most accurate dose metric since it considers patient size and its' effect on the dose required to obtain images of sufficient quality. When patients are large their $CTDI_{vol}$ is an overestimate of the radiation dose which will make flagging inaccurate. As previously stated, many of the acquisitions being flagged are going to be due to larger patients requiring a larger amount of dose. Since all of these methods are based on actual patient data if the dose is being overestimated then these thresholds are too, which could cause more missed true positives. This can be seen in Figure 5, the sensitivity fluctuates, and using $CTDI_{vol}$ can cause sites to miss flagging their necessary exams.

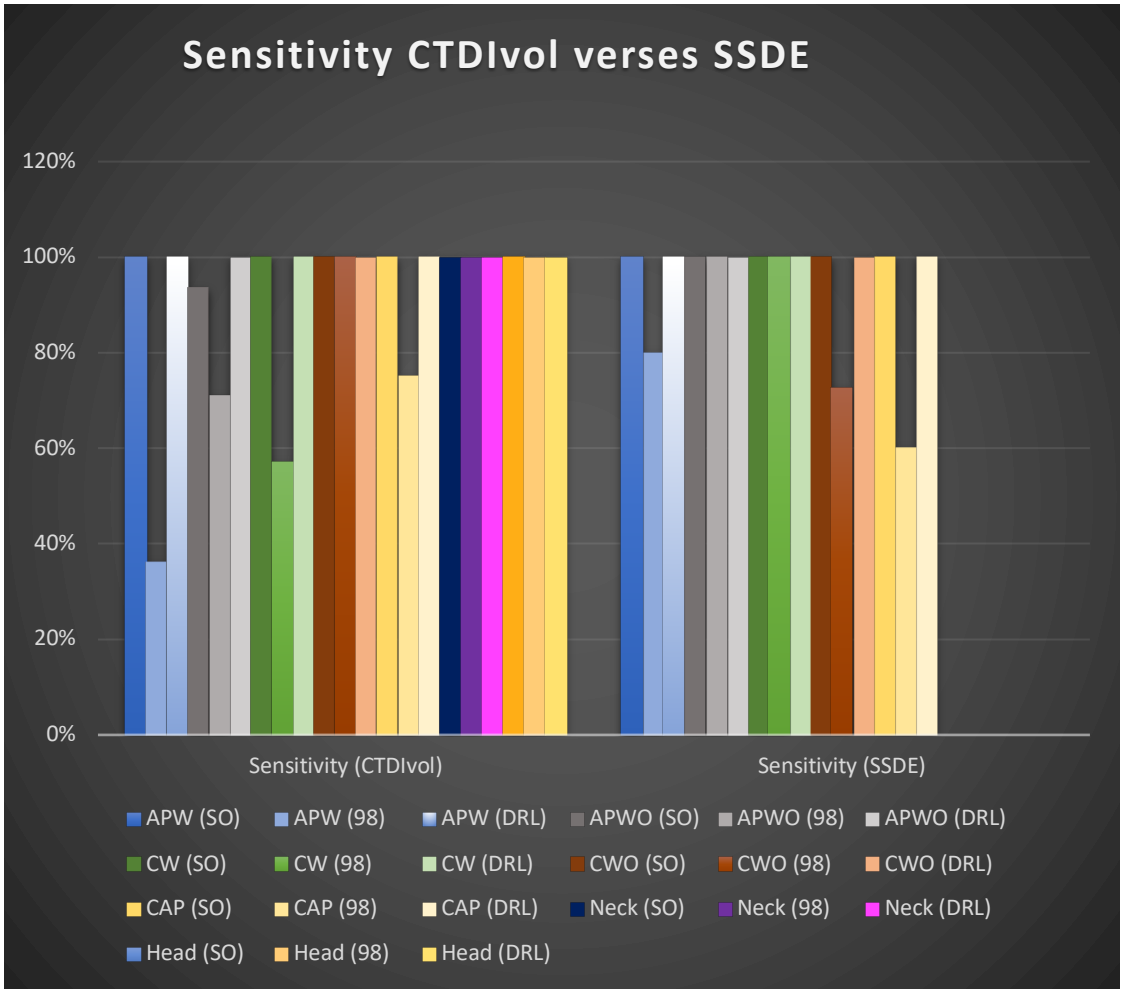

Figure 5:Sensitivity from $CTDI_{vol}$ verses SSDE

5.3 Future Work

One of the most important aspects that need to be addressed moving forward is the test of the statistical outlier method in all other protocols. In 2019, OHSU had over 25 protocols that imaged over 100 exams and these protocols would be the next step to test if this method is accurate. Statistical outliers provide a simple effective solution to tracking high dose indices from diagnostic exams and are promising for future more specific studies, as this research could be continued and expanded on.

Over time this method could become more exact by creating a different threshold for different groups of people. As mentioned above, size plays a huge factor in the amount of radiation dose output so why not create thresholds based on size. Many of the false positive exams found in all methods were due to large patients, separating thresholds based on size could decrease the rate that facilities are flagging unnecessary exams. Diminishing this rate would then considerably decrease the time it takes to review. Since pediatric cases vary a good deal in size, this continuance of the research would allow this method to expand to pediatric exams.

There are machine characteristics that affect the amount of dose patients are receiving which are CT manufacturer and model. Not all manufacturers are the same and within a manufacturer, there are different models with different levels of technological advancements. The newer models tend to have more dose reduction software and strategies, as the technical capabilities of machines have changed over time. [11] Some dose reduction technologies are automatic tube current modulation, iterative reconstruction algorithm, additional beam filters, and dynamic collimation. [1][10][22]

With different dose reduction strategies, it would be interesting to see if at a facility that multiple manufacturers and models had, if dose threshold would differ significantly between them.

5.4 Study Limitations

When carrying out this study, there were over 13,000 acquisitions that could be reviewed across all seven protocols. As previous stated, the reviewing process can be a time-consuming process, and due to this and for the purposes of this research limiting the number of acquisitions reviewed was needed. This research used DRLs as a benchmark in determining which exams would likely not be flagged. By limiting the number of acquisitions reviewed, there is a possibility of missing false negative acquisitions. There could be acquisitions that had a lower dose than the DRLs that were not looked at; however, there could have still had used the wrong technique and need to be reviewed. False negative could also be missed due to not separating the data by size before analysis. For example, if a small patient had a high dose compared to other patients of their size, such as 10 mGy compared to others only receiving 5 mGy, their exam would not be caught in flagging because this research is excluding data below a certain dose for all patients. Statistical outliers had a high sensitivity overall and with this potential of missing false negative acquisitions, this sensitivity value could have been an overestimate to the actual value.

6. Summary and Conclusions

The main goal of this research was to develop a standardized method of setting dose thresholds for diagnostic CT examinations that complied with the Joint Commission's requirements of reviewing and analyzing high dose incidents. The aim was to create a method that maximized flagging the number of exams that went awry and needed to be reviewed while minimizing the number of high dose exams that are flagged but are appropriate in technique. This goal was accomplished by using OHSU site-specific data to create a statistical outlier as a threshold. Statistical outliers were compared to two other methods of creating thresholds by looking at statistics of the flagging of seven protocols. These protocols were also separated by dose metric, $CTDI_{vol,}$ and SSDE.

OHSU's CT dose data from 2019 was exported from Imalogix into R studio where those statistical outliers were calculated. The second method was based on a journal article that developed DRLs for the ten most common adult CT exams based on the ACR DIR. The third was OHSU's current threshold method of using the 98^{th} percentile of all of OHSU's CT data.

Based on 2019 data, true positives were found by searching each accession number into Imalogix and doing a root cause analysis on the outlier. In excel, all three dose thresholds were used to see the number of exams that would be flagged. The flagged exams were compared to our true positives to make sure the thresholds were flagging all necessary exams. Five of the seven protocols were compared with SSDE and all seven of them were compared with $CTDI_{vol}$.

Statistical tests were used to give a measurement of a method's accuracy and flagging assessment. Sensitivity gave us the percentage of true positives that a method was flagging and if any true positives were missed, ideally this value was 100%. The lower the specificity, the more false positives flags were found in a method. These two tests were key in assessing each method.

DRLs had 100% sensitivity across the board in both dose metrics; however, this method flagged many more exams than either of the other techniques. There is no clear definition ad to what constitutes an expected dose index, so as a public available dose benchmark, DRLs will flag the necessary dose incidents. The downside of using DRLs is the time it will take someone to review these flagged exams, and for this reason, a site should use its own data to create more accurate thresholds.

98th Percentile is based on OHSU site-specific data and the statistics do show evidence of flagging fewer exams overall to save time by having the highest specificity over almost all protocols. This is due to the 98th percentile having the highest dose value except for the protocol Head without Contrast. As the threshold rises, there will be fewer exams to go through, but it raises the risk of missing true positive exams. This method missed 20% of the true positives looked at, showing that it is not an accurate technique.

From this study, it was found that statistical outlier thresholds have the highest accuracy at flagging true positive exams while minimizing the number of false positive alerts. This method had 100% sensitivity in eleven out of the twelve samples, only missing 1.6% of true positives, and had a specificity above 95% in all twelve.

When comparing SSDE to $CTDI_{vol}$ as a dose metric, SSDE creates a more accurate picture of dose for all body sizes. $CTDI_{vol}$ either under or overestimates depending on patient size and can cause facilities to miss exams that should be flagged by skewing the dose threshold.

In conclusion, having a correct CT dose threshold set is a big-time saver for a facility and this can be seen when using site-specific data compare to non-site-specific data. When using site-specific data, it is necessary to be careful not to set the threshold too high where exams that need to be reviewed are missed. SSDE is a more accurate dose metric compared to $CTDI_{vol}$ and should be the preferred method of flagging exams.

There are many different ways of setting CT dose thresholds, and this research presents an accurate and efficient way of maximizing a facility's effort in tracking CT dose incidents. The use of statistical outliers based on SSDE dose data is a promising method for creating thresholds that comply with the Joint Commissions' quality assurance requirement and ensures productivity.

7. References

[1] Goldman LW. Principles of CT and CT Technology. *Journal of Nuclear Medicine Technology*. 2007;35(3):115-128. doi:10.2967/jnmt.107.042978

[2] Bushberg JT, ed. *The Essential Physics of Medical Imaging*. 3rd ed. Wolters Kluwer Health/Lippincott Williams & Wilkins; 2012.

[3] Seeram E. *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*.; 2016.

[4] Lira D, Padole A, Kalra MK, Singh S. Tube Potential and CT Radiation Dose Optimization. *American Journal of Roentgenology*. 2015;204(1):W4-W10. doi:10.2214/AJR.14.13281

[5] Gress DA, Dickinson RL, Erwin WD, et al. AAPM medical physics practice guideline 6.a.: Performance characteristics of radiation dose index monitoring systems. *J Appl Clin Med Phys*. 2017;18(4):12-22. doi:10.1002/acm2.12089

[6] Jones AK, Wunderle KA, Gress DA, et al. A diagnostic medical physicist's guide to the American College of Radiology Fluoroscopy Dose Index Registry. *J Appl Clin Med Phys*. 2021;22(4):8-14. doi:10.1002/acm2.13227

[7] McCollough CH. Diagnostic Reference Levels. *American College of Radiology*. Published online 2010. https://www.imagewisely.org/Imaging-Modalities/Computed-Tomography/Diagnostic-Reference-Levels

[8] 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. doi:10.1148/radiol.2017161911

[9] Mettler FA, Mahesh M, Bhargavan-Chatfield M, et al. Patient Exposure from Radiologic and Nuclear Medicine Procedures in the United States: Procedure Volume and Effective Dose for the Period 2006–2016. *Radiology*. 2020;295(2):418-427. doi:10.1148/radiol.2020192256

[10] Parakh A, Kortesniemi M, Schindera ST. CT Radiation Dose Management: A Comprehensive Optimization Process for Improving Patient Safety. *Radiology*. 2016;280(3):663-673. doi:10.1148/radiol.2016151173

[11] Smith-Bindman R, Wang Y, Chu P, et al. International variation in radiation dose for computed tomography examinations: prospective cohort study. *BMJ*. Published online January 2, 2019:k4931. doi:10.1136/bmj.k4931

[12] Bauhs JA, Vrieze TF, Primak AN, Bruesewitz R, McCollough CH. CT Dosimetry: Comparison of Measurement Techniques and Devices. *RSNA*. Published online 2008.

[13] Bohl MA, Goswami R, Strassner B, Stanger P. Meeting The Joint Commission's Dose Incident Identification and External Benchmarking Requirements Using the ACR's Dose Index Registry. *Journal of the American College of Radiology*. 2016;13(8):936-942. doi:10.1016/j.jacr.2016.04.026

[14] Brenner DJ, Hall EJ. Computed Tomography — An Increasing Source of Radiation Exposure. *N Engl J Med*. 2007;357(22):2277-2284. doi:10.1056/NEJMra072149

[15] Hall EJ, Giaccia AJ. *Radiobiology for the Radiologist*. Eighth edition. Wolters Kluwer; 2019.

[16] National Research Council (U.S.), ed. *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2*. National Academies Press; 2006.

[17] Rosner B. *Fundamentals of Biostatistics*. 8th edition. Cengage Learning; 2016.

[18] Report No. 096 of the American Association of Physicist's in Medicine (AAPM). "The Measurement, Reporting, and Management of Radiation Dose in CT" (2008).
[19] Report No. 204 of the American Association of Physicist's in Medicine (AAPM). "Size-Specific Dose Estimates (SSDE) in Pediatric and Adult Body CT Examinations." (2011) https://www.aapm.org/pubs/reports/RPT_204.pdf
[20] Diagnostic Imaging Requirements. Published online 2015. https://www.jointcommission.org/-/media/deprecated-unorganized/imported-assets/tjc/system-folders/topics-library/ahc_diagimagingrpt_mk_20150806pdf.pdf?db=web&hash=43B614B4AA29ADDF44F8B8C1C2DE59B6
[21] About Diagnostic Reference Levels (DRLs). *IAEA*. Published online 2018. https://www.iaea.org/resources/rpop/health-professionals/radiology/diagnostic-reference-levels/about-diagnostic-reference-levels#1
[22] "Comparing CT Manufacturers and Models: Choosing the Best CT for Your N." *PrizMed Imaging*, prizmedimaging.com/pages/comparing-ct-manufacturers-and-models-choosing-the-best-ct-for-your-needs.
[23] "Dose Index Registry: American College of Radiology." *American College of Radiology | American College of Radiology*, www.acr.org/Practice-Management-Quality-Informatics/Registries/Dose-Index-Registry.
[24] "FAQ." *The Joint Commission*, www.jointcommission.org/.
[25] *RStudio*, www.rstudio.com/products/rstudio/.
[26] Report No. 220 of the American Association of Physicist's in Medicine (AAPM). "Use of Water Equivalent Diameter for Calculating Patient Size and Size-Specific Dose Estimates (SSDE) in CT" (2014) https://www.aapm.org/pubs/reports/rpt_220.pdf
[27] Crowley C, Ekpo EU, Carey BW, et al. Radiation dose tracking in computed tomography: Red alerts and feedback. Implementing a radiation dose alert system in CT. *Radiography*. 2021;27(1):67-74. doi:10.1016/j.radi.2020.06.004
[28] MacGregor K. identifying institutional Diagnostic reference levels for CT with radiation Dose index Monitoring software1. *RSNA*. 2015;276(2).
[29] Liang CR, Chen PXH, Kapur J, Ong MKL, Quek ST, Kapur SC. Establishment of institutional diagnostic reference level for computed tomography with automated dose-tracking software. *J Med Radiat Sci*. 2017;64(2):82-89. doi:10.1002/jmrs.210