SIIM 2018 ANNUAL MEETING

Demonstrating the potential clinical benefit of a deep learning prioritization algorithm for improving timeliness of positive findings reporting to the ordering physician, in accordance with one of The Joint Committee (TJC)'s identified goals for patient safety.

A clinical study to evaluate the influence of AI-based prioritization on Report Turn Around Time

Gal Yaniv M.D., Ph.D,
Anna Kuperberg M.D
Elad Walach M.Sc.

Table of Contents

1. BRIEF SUMMARY

Aidoc's AI-based decision support solution, uses an algorithm based on deep learning technology to detect and highlight radiological studies with intracranial hyperdense or neck bone hypodense findings.

The purpose

of this study was to evaluate the influence of AI-based prioritization on RTAT (report turnaround time, the most frequently used quality metric in radiology[i]) of prioritized and non-prioritized studies containing positive findings.

Primary endpoint

The RTAT distribution for prioritized workflow queue with various detection accuracy parameters, compared to FIFO (first in, first out) workflow baseline.

Method

In this analysis, we developed a model to describe the workflow queue of a single radiologist. We used this model to study the effect of various accuracy parameters of the prioritization algorithm on the RTAT using a Monte-Carlo simulation .

Hypothesis

Utilization of a prioritization algorithm will reduce RTAT of both prioritized and non-prioritized scans.

Results

The results showed that RTAT of head and neck CT scans containing abnormal hyperdense and neck bone hypodense findings improved significantly when utilizing Aidoc's prioritization algorithm, in comparison to FIFO, with all performance parameters tested. Specifically, the "low" accuracy prioritization algorithm (70% sensitivity, 70% specificity) resulted in over 46% RTAT reduction and the "perfect" accuracy algorithm (100% sensitivity, 100% specificity) resulted in 67.3% reduction in RTAT. When evaluating the algorithm's detection parameters demonstrated in the pivotal clinical trial (~90%, 90% sensitivity, specificity) the results were very similar to those of the "perfect" algorithm, with 60.3% improvement in RTAT. Additionally, the average RTAT of all medical images without findings remained essentially unchanged .

Conclusion

These results demonstrate the potential clinical benefit of Aidoc's prioritization algorithm for improving timeliness of positive findings reporting to the ordering physician, in accordance with one of The Joint Committee (TJC)'s identified goals for patient safety.

2. BACKGROUND

In January 2018, The Joint Commission (TJC) re-iterated its previous National Patient Safety Goal that mandated healthcare providers "report critical results of tests and diagnostic procedures on a timely basis." Improvement in critical results timing reporting is driven by the assumption that timely reporting will lead to timelier clinical interventions with improved treatment outcomes, and prevention of co-morbidities .

A radiology department has a major role in the implementation of this goal, with a large component of the delay in the diagnostic process in the interval between medical-image acquisition and critical finding discovery by a radiologist. Specifically, a study conducted in The Johns Hopkins Medical Institution demonstrated that the mean time from medical image acquisition to critical finding discovery was 62.2 minutes, representing the largest component of the delay in the radiological diagnostic process [ii].

Previous attempts to improve report turnaround time (RTAT) of critical results included prioritization of the worklist (e.g. stat systems)[iii] , improving physician communication, etc. Available prioritization systems are based on the study's metadata, e.g. clinical indication, origin (ED, Inpatient, Outpatient). At the heart of several of these attempts lies the concept of prioritization of diagnostic image interpretation. Radiologists approach the prioritization of their worklist in various ways, frequently without sufficient scientific basis .

Aidoc developed a deep learning algorithm for prioritization based on the study's visual information. The algorithm is based on a convolutional neural network that detects hyperdense intracranial abnormalities and bone hypodense abnormalities in head and neck CTs. Thus, scans suspected to contain an abnormality can be moved-up in the worklist.

"Queueing theory", a discipline within the mathematical theory of probability, has been used in various medical fields, including radiology, to investigate the medical workflow[ivv]

Aidoc utilized this theory to model the radiologist workflow in this simulation analysis. Each medical image acquisition was designated as a "queuing event", and each report finalization as "exiting queue". Using queuing theory concepts, the existing workflow management concepts are divided to "First in first out" (first medical image to arrive to the worklist is interpreted first, representative of the approximate standard of care), vs. Priority Queuing (medical images with higher priority are interpreted first), etc .

Utilization of prioritization systems has been shown to significantly improve RTAT of both critical and non-critical findings. These systems change the radiologist workflow management from FIFO to Priority Queuing.

In prioritization systems queuing, the priority level of each case is based on the probability of the medical image to contain a positive finding, and on the degree of urgency of the suspected finding. As noted above, current prioritization systems are based primarily on the medical image's "metadata" criteria - the patient's chief complaint, suspected diagnosis, patient location (ED, Inpatient, Outpatient), etc.

Aidoc developed an artificial intelligence (AI) algorithm that allows radiologists to perform computer-assisted prioritization based on the medical image data itself. The deep learning algorithm is based on a convolutional artificial neural network (CNN). The network was trained to detect hyperdense intracranial and neck bone hypodense abnormalities in CT scans. To build the network, the algorithm was shown a data set of thousands of hyperdense intracranial and neck bone hypodense abnormalities, and learned to detect such findings. Importantly, the algorithm was trained on abnormalities in the anatomical region rather than a specific pathology, allowing it to recognize a variety of significant ("positive") findings. Once the algorithm training was completed, it could be used to interpret relevant scans and indicate whether or not a scan contains a suspected abnormality.

3. STUDY PURPOSE AND RATIONALE

Aidoc's algorithm accurately detects findings in head and neck CT scans prioritizing positive cases for full interpretation by a radiologist. In order to link the measurable performance parameters of the algorithm to the expected clinical outcome of RTAT reduction, a holistic model of the radiologist workflow needs to be taken into account. By simulating that workflow using typical radiological workflow parameters, the company quantitatively analyzed the effect of Aidoc's prioritization on RTAT for the prioritized scans, as well as all scans within the radiologist's queue containing positive findings, and all scans within the radiologist's queue not containing positive findings .

4. PRIMARY ENDPOINTS

The RTAT distribution for prioritized workflow queue with various detection accuracy parameters, compared to FIFO workflow baseline.

5. STUDY DESIGN – MATERIALS AND METHODS

In this analysis, the company developed a mathematical model to describe the workflow queue of a single radiologist, and used a Monte-Carlo simulation to generate realizations of a full work week.

The results of the simulation were analyzed to determine the effect of the queuing method on the distribution of RTAT of medical-images containing positive findings. The trivial queuing method of FIFO (first in first out) was used as a baseline to compare to AI-prioritization based queuing with various sensitivity and specificity parameters for the AI performance.

The mathematical framework is that of "queuing theory". Medical image acquisitions were modeled as independent events with an average time of T_scan between events, where each acquired image is drawn from the known scan type distribution (CT/MRI/US/XR, with or without a positive finding) and enters the queue. The queue is then re-ordered according to the queuing method; in the FIFO workflow queue, it is simply left at the end of the queue until a new image is acquired.

In the AI-prioritization based method, however, image analysis by the algorithm is simulated:

The AI algorithm can either indicate a medical image is either "negative" (no findings were detected by the algorithm) or "positive" (one or more finding was detected by the algorithm). If the AI algorithm indicates the medical image is "negative", it remains at its original position in the queue. If the AI algorithm indicates the medical image is "positive", it is moved up in the queue, behind all other "positive" cases that are already in the queue. Simultaneously, the radiologist reads medical images and finalizes reports with a review time that is assumed to be normally distributed with a known mean value and standard deviation. The accuracy of the radiologist is assumed to be perfect for simulation purposes.

To study this system, we generated realizations of a full work week in a medium-sized radiology department, where all the image acquisition (queuing events) and report finalization (exiting queue events) events were drawn from the relevant random distributions .

The parameter values used in the simulation were derived from the relevant literature based on radiological clinical studies and workload statistics:

- The company assumed that interpretation time is normally distributed with a mean time and standard deviation, where the parameter values were based on relevant literature[vi] .
- Prevalence of an abnormal finding in a head CT is based on studies on both trauma[vii] and non-trauma[viii] patients[ix]
- Prevalence of a critical finding in a radiological study was based on a 2014 Lacson et al. study, that demonstrated 13.1% of the study reports evaluated, contained a critical finding[x].
- The company derived distribution of radiological modalities utilization from a mean of two studies looking at Inpatient[xi] , and ED[xii] radiological modalities utilization .

The following parameters derived from the literature cited above were used in the model:

- Distribution of radiological studies per modality:
  - CT = 23%
  - MR = 9%
  - US= 8%
  - XR = 60%

- Proportion of head and neck-CT out of all CT scans = 30%
- Prevalence of a positive significant findings in any study - 13%
- Distribution of radiologist interpretation times –
  - Head and neck CT
    - Positive read time: mean- 10 minutes; standard deviation- 2 minutes
    - Negative time: mean- 5 minutes; standard deviation-1 minutes
  - Non-head and neck CT
    - Positive/Negative read time mean-7 minutes; standard deviation-6.5/3 minutes
  - Other modalities
    - Positive/Negative MR read time mean-20 minutes; standard deviation-10/3 minutes
    - Positive/Negative US read time mean-5 minutes; standard deviation-5/3 minutes
    - Positive/Negative XR read time mean-2 minutes; standard deviation-1/3 minutes
- Rate of incoming medical images - Positive/Negative mean-6 minutes; standard deviation-4/3 minutes

For the AI-prioritization algorithm accuracy performance we used several parameter sets: “perfect” (100% sensitivity, 100% specificity for positive finding detection), “high/Aidoc” (90%, 90%), “medium” (80%, 80%) and “low” (70%, 70%) performance, as demonstrated in Table 1, below.

Table 1: Algorithm Performance Parameters

	Sensitivity [%]	Specificity [%]
Perfect	100	100
High	90	90
Medium	80	80
Low	70	70

The results of a single realization/iteration of the simulation allow the company to infer the RTAT for each of the positive head and neck CT scans generated during the simulated week, as the acquisition time and a report-finalization time for each case are available. The company generated 2000 realizations of the simulations for each performance parameter. Subsequently, the mean and standard deviation of the RTAT over all the cases for each performance parameter were calculated, which were compared between Aidoc’s algorithm and the baseline FIFO method (representing an approximation of the standard of care[xiii]) .

We calculated the average RTAT relative improvement compared to the baseline was calculated as follows:

Average prioritized RTAT relative improvement compared to Baseline=

$$(1 - \frac{\text{PrioritizedRTAT}}{\text{BaselineRTAT}}) * 100$$

6. RESULTS

Using image-data based prioritization, all performance parameters significantly reduced average RTAT for head CTs; and all medical-images with positive findings, compared to the baseline FIFO managed workflow. Full results are detailed below, and in tables 2 and 3.

Perfect Performance (Sensitivity and specificity are 100%) Scenario:

- On average, 1717.2 cases were read, out of which 118.8 were non-contrast CT scans of head or neck, and 15.5 were “positive” non-contrast CT scans of head or neck.
- The average RTAT for positive non-contrast CT scans of head or neck was: 14.83 minutes (0.04 minutes standard deviation)
- The average RTAT for all positive cases was: 39.45 minutes (0.38 minutes standard deviation (
- The average RTAT for all negative cases was: 41.24 (0.4 minutes standard deviation).

High Performance Scenario/Representative of the Aidoc Pivotal Study (Sensitivity and specificity are 90%):

- On average, 1717.1 cases were read, out of which 118.6 were non-contrast CT scans of head or neck, and 15.5 were “positive” non-contrast CT scans of head or neck.
- The average RTAT for positive non-contrast CT scans of head or neck was: 18.01 minutes (0.11 minutes standard deviation)
- The average RTAT for all positive cases was: 39.78 minutes (0.38 minutes standard deviation)
- The average RTAT for all negative cases was: 41.16 (0.4 minutes standard deviation).

Medium Performance (Sensitivity and specificity are 80%) Scenario:

- On average, 1717.0 cases were read, out of which 118.7 were non-contrast CT scans of head or neck, and 15.6 were “positive” non-contrast CT scans of head or neck.
- The average RTAT for positive non-contrast CT scans of head or neck was: 20.99 minutes (0.16 minutes standard deviation)
- The average RTAT for all positive cases was: 40.13 minutes (0.39 minutes standard deviation).
- The average RTAT for all negative cases was: 41.08 (0.4 minutes standard deviation).

Low Performance (Sensitivity and specificity are 70%) Scenario:

- On average, 1717.2 cases were read, out of which 118.8 were non-contrast CT scans of head or neck, and 15.5 were "positive" non-contrast CT scans of head or neck.
- The average RTAT for positive non-contrast CT scans of head or neck was: 24.12 minutes (0.2 minutes standard deviation)
- The average RTAT for all positive cases was: 40.48 minutes (0.39 minutes standard deviation)
- The average RTAT for all negative cases was: 41.00 (0.4 minutes standard deviation).

Baseline - No Priority – FIFO:

- On average, 1717.1 cases were read, out of which 118.7 were non-contrast CT scans of head or neck, and 15.6 were "positive" non-contrast CT scans of head or neck.
- The average RTAT for positive non-contrast CT scans of head or neck was: 45.41 minutes (0.46 minutes standard deviation)
- The average RTAT for all positive cases was: 41.05 minutes (0.4 minutes standard deviation)
- The average RTAT for all negative cases was: 40.69 (0.4 minutes standard deviation).

Table 2: Results - Average RTAT values for various Algorithm parameters

Algorithm parameters (Sensitivity[%] specificity[%])	Average RTAT for head and neck CTs with positive findings (standard deviation error) [min]	Average RTAT for all medical images with positive findings (standard deviation error) [min]	Average RTAT for all medical images with negative findings (standard deviation error) [min]
Perfect (100,100)	14.83 (0.04)	39.43 (0.38)	41.24 (0.4)
High/Aidoc (90,90)	18.01 (0.11)	39.78 (0.38)	41.16 (0.4)
Medium (80,80)	20.99 (0.16)	40.13 (0.39)	41.08 (0.4)
Low (70,70)	24.12 (0.2)	40.48 (0.39)	41.00 (0.4)
Baseline (FIFO, no prioritization)	45.41 (0.46)	41.05 (0.4)	40.69 (0.4)

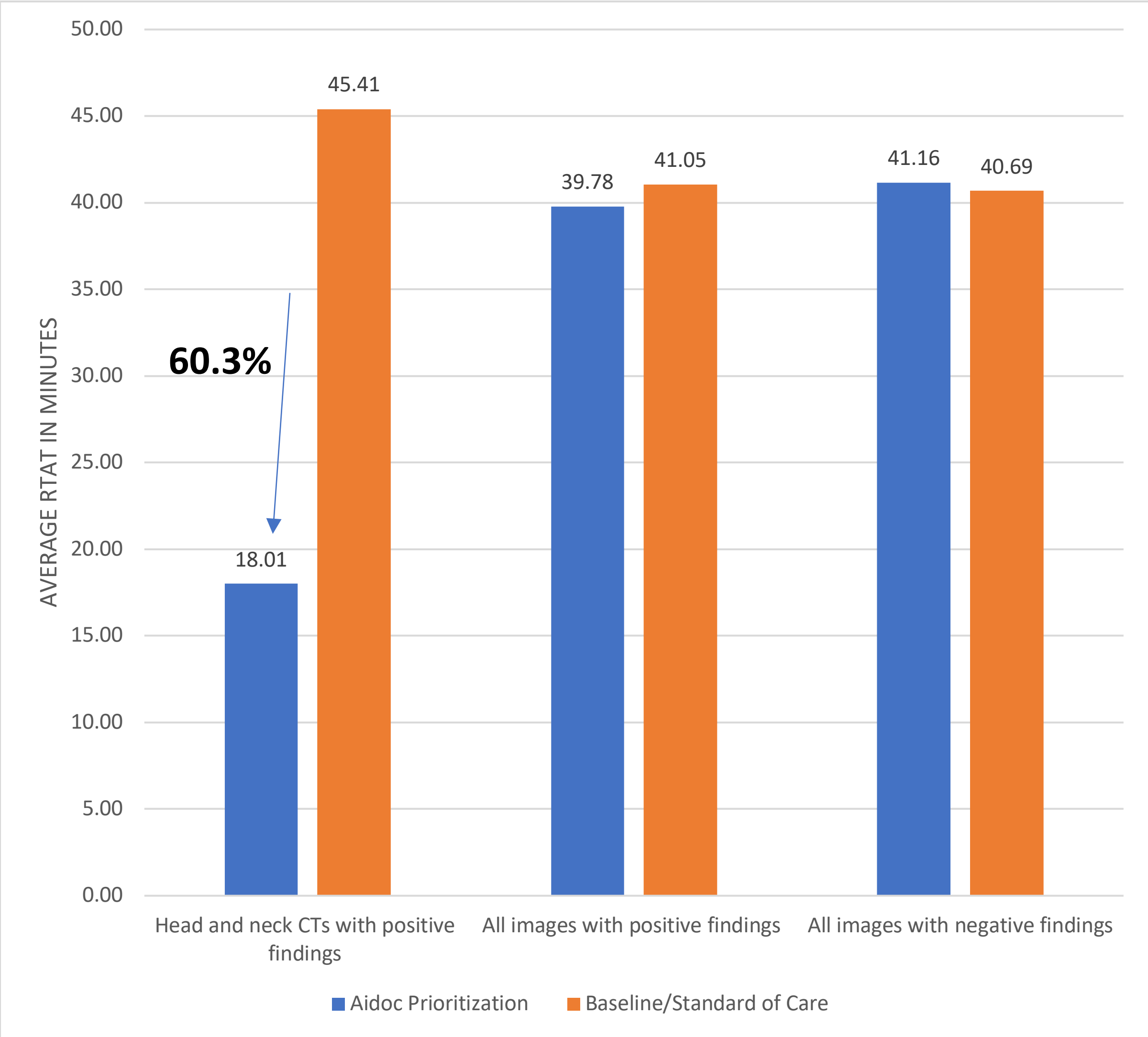
Table 3: Results - Average RTAT relative improvement compared to FIFO

Algorithm parameters (Sensitivity[%] specificity[%])	Average RTAT for head and neck CTs with positive findings **relative improvement** compared to FIFO [%]	Average RTAT for all medical images with positive findings **relative improvement** compared to FIFO [%]	Average RTAT for all medical images with negative findings **relative improvement** compared to FIFO [%]

FIFO	0	0	0
Perfect (100,100)	67.3%	3.9%	-1.4%
High/Aidoc (90,90)	60.3%	3.1%	-1.2%
Medium (80,80)	53.8%	2.2%	-1.0%
Low (70,70)	46.9%	1.4%	-0.8%

Figure 1 below demonstrates the benefits provided by AI-based prioritization. This figure presents the results from Table 1 for the algorithms performance of 90% sensitivity and specificity, as compared to baseline (the approximate standard of care). As shown, the average RTAT drops substantially for head and neck CTs with positive findings when Aidoc's software is used. The RTATs for all positive findings and negative findings are largely unchanged.

Figure1 : Average RTAT for Positive and Negative Studies for Ai -Based Prioritized Queue vs. Baseline (Using Aidoc's 90% Sensitivity and Specificity Performance)



7. DISCUSSION

Aidoc used a deep learning algorithm to build a convolutional neural network (CNN) that was trained to detect suspected hyperdense intracranial and neck bone hypodense abnormalities in CT scans. The purpose of the algorithm was to allow prioritization of medical images suspected to contain abnormalities, thus shortening their RTAT.

This work was designed to estimate the quantitative benefit of such prioritization on RTAT of all medical images with positive findings .

Interestingly, the RTAT of head and neck CT scans containing positive findings improved substantially with all performance parameter sets used for the prioritization algorithm. Moreover, the RTAT of all medical images containing positive findings was shown to improve as well. The improvement in RTAT head and neck CT scans was clinically meaningful from the "low" accuracy algorithm (70% sensitivity, 70% specificity), that resulted in over 46% RTAT reduction, to "perfect" accuracy algorithm (100% sensitivity, 100% specificity) that resulted in 67.3% RTAT reduction. When evaluating Aidoc's parameters (90% sensitivity, 90% specificity) the results were very similar to the "perfect" parameters results, with 60.3% RTAT improvement. Importantly, the change in RTAT of negative studies was minimal as clearly shown in Table 2 and Figure 1, with similar RTATs between the AI algorithm and baseline simulation .

Interestingly, the study suggested that increasing the algorithm accuracy beyond 80-90% results in minimal addition to head and neck CT RTAT improvement, with similar RTAT results between 80%-100% accuracy .

These results demonstrate the significant potential clinical benefit and negligible potential risk of a software such as Aidoc's for prioritization/triage of the clinical radiological workflow.

8. CONCLUSION

This analysis estimated the quantitative benefit of AI based algorithm prioritization on RTAT of studies with positive findings using a Monte Carlo simulation of a "queueing theory" based model of a radiologist worklist .

The results showed that RTAT of head and neck CT scans containing positive findings improved with all performance parameters of the prioritization algorithm. Moreover, the average RTAT of all medical-images containing positive findings was shown to somewhat improve as well, with no significant impact on the RTAT of negative studies .

These results demonstrate the substantial potential clinical benefit of Aidoc's algorithm (detecting hyperdense intracranial and neck bone hypodense abnormalities in CT scans with 90.00% accuracy).

To conclude, the company has provided simulation data on how AI based priority queuing can decrease positive finding RTAT, thus facilitate timely communication of radiological findings, one of TJC's National Patient Safety Goals.

9. REFERENCES

[i] Walker EA, Petscavage-Thomas JM, Fotos JS, Bruno MA. Quality metrics currently used in academic radiology departments: results of the QUALMET survey. Br J Radiol. 2017 Mar;90(1071):20160827. doi: 10.1259/bjr.20160827. Epub 2017 Jan 24. PubMed PMID: 28118038; PubMed Central PMCID: PMC5601514.

[ii] Honig SE, Honig EL, Babiarz LB, Lewin JS, Berlanstein B, Yousem DM. Critical findings: timing of notification in neuroradiology. AJNR Am J Neuroradiol. 2014 Aug;35(8):1485-92. doi: 10.3174/ajnr.A3918. Epub 2014 Apr 10. PubMed PMID: 24722306.

[iii] Gaskin CM, Patrie JT, Hanshew MD, Boatman DM, McWey RP. Impact of a Reading Priority Scoring System on the Prioritization of Examination Interpretations. AJR Am J Roentgenol. 2016 May;206(5):1031-9. doi: 10.2214/AJR.15.14837. Epub 2016 Mar 21. PubMed PMID: 26999578.

[iv] Palvannan RK, Teow KL. Queueing for healthcare. J Med Syst. 2012 Apr;36(2):541-7. doi: 10.1007/s10916-010-9499-7. Epub 2010 May 8. PubMed PMID: 20703697.

[v] Foster EM, Hosking MR, Ziya S. A spoonful of math helps the medicine go down: an illustration of how healthcare can benefit from mathematical modeling and analysis. BMC Med Res Methodol. 2010 Jun 23;10:60. doi: 10.1186/1471-2288-10-60. PubMed PMID: 20573235; PubMed Central PMCID: PMC2914732.

[vi] Dora JM, Torres FS, Gerchman M, Fogliatto FS. Development of a local relative value unit to measure radiologists' computed tomography reporting workload. J Med Imaging Radiat Oncol. 2016 Dec;60(6):714-719. doi: 10.1111/1754-9485.12492. Epub 2016 Jul 11. PubMed PMID: 27397691; Dora JM, Faccin CS and Fogliatto S. The use of relative value units to monitor radiologists' reporting productivity and workload. 2016; See http://www.dieurope.com/latest_news/the-use-of-relative-value-units-to-monitor-radiologists-reporting-productivity-and-workload.

[vii] Jeret JS, Mandell M, Anziska B, Lipitz M, Vilceus AP, Ware JA, Zesiewicz TA. Clinical predictors of abnormality disclosed by computed tomography after mild head trauma. Neurosurgery. 1993 Jan;32(1):9-15; discussion 15-6. PubMed PMID: 8421561.

[viii] Wang X, You JJ. Head CT for nontrauma patients in the emergency department: clinical predictors of abnormal findings. Radiology. 2013 Mar;266(3):783-90. doi: 10.1148/radiol.12120732. Epub 2012 Nov 30. PubMed PMID: 23204540.

[ix] Lacson R, Prevedello LM, Andriole KP, O'Connor SD, Roy C, Gandhi T, Dalal AK,

Sato L, Khorasani R. Four-year impact of an alert notification system on closed-loop communication of critical test results. AJR Am J Roentgenol. 2014 Nov;203(5):933-8. doi: 10.2214/AJR.14.13064. PubMed PMID: 25341129; PubMed Central PMCID: PMC4426858.

[x] Wang X, You JJ. Head CT for nontrauma patients in the emergency department: clinical predictors of abnormal findings. Radiology. 2013 Mar;266(3):783-90. doi: 10.1148/radiol.12120732. Epub 2012 Nov 30. PubMed PMID: 23204540.

[xi] Shinagare AB, Ip IK, Abbett SK, Hanson R, Seltzer SE, Khorasani R. Inpatient imaging utilization: trends of the past decade. AJR Am J Roentgenol. 2014 Mar;202(3):W277-83. doi: 10.2214/AJR.13.10986. PubMed PMID: 24555625.

[xii] Ali S. Raja, Ivan K. Ip, Aaron D. Sodickson, Ron M. Walls, Stephen E. Seltzer, Joshua M. Kosowsky, Ramin Khorasani. Radiology Utilization in the Emergency Department: Trends of the last Two Decades. AJR Am J Roentgenol. Author manuscript; available in PMC 2016 Jan 26.Published in final edited form as: AJR. Am J Roentgenol. 2014 Aug; 203(2): 355–360. doi: 10.2214/AJR.13.11892 PMCID: PMC4726976; Ondategui-Parra S, Bhagwat JG, Zou KH, Gogate A, Intriere LA, Kelly P, Seltzer SE, Ros PR. Practice management performance indicators in academic radiology departments. Radiology. 2004 Dec;233(3):716-22. PubMed PMID: 15564407.

[xiii] Cunningham, Reid, MacSwain, Clarke Ethics in Radiology: Wait Lists Queue Jumping. Canadian Association of Radiologists Journal Volume 64, Issue 3, August 2013, Pages 170-175 https://doi.org/10.1016/j.carj.2011.12.006

10. DISCLAIMER

This document contains proprietary confidential information. The contents of this document may not be copied or reproduced in any way and may not be disclosed or distributed to any person or entity without the express written permission of Aidoc Medical Ltd.