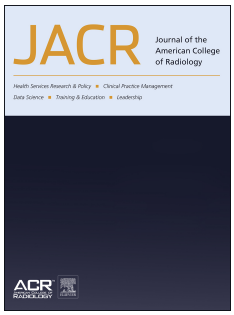


Identifying and Categorizing Diagnostic Errors in Abdominal Imaging using a Diagnostic Imaging-Specific Annotation Process

Ronilda Lacson, MD PhD, Kristine S. Burk, MD, Jeffrey P. Guenette, MD, MPH, Pamela J. DiPiro, MD, Ariadne DeSimone, MD, MPH, Nicole Vetrano, MHA, RT, Ishani Ganguli, MD, MPH, David W. Bates, MD, Ramin Khorasani, MD, MPH

PII: S1546-1440(26)00296-6

DOI: https://doi.org/10.1016/j.jacr.2026.06.006

Reference: JACR 7286

To appear in: *Journal of the American College of Radiology*

Received Date: 17 February 2026

Revised Date: 2 June 2026

Accepted Date: 5 June 2026

Please cite this article as: Lacson R, Burk KS, Guenette JP, DiPiro PJ, DeSimone A, Vetrano N, Ganguli I, Bates DW, Khorasani R, Identifying and Categorizing Diagnostic Errors in Abdominal Imaging using a Diagnostic Imaging-Specific Annotation Process, *Journal of the American College of Radiology* (2026), doi: https://doi.org/10.1016/j.jacr.2026.06.006.

Identifying and Categorizing Diagnostic Errors in Abdominal Imaging using a Diagnostic Imaging-Specific Annotation Process

Ronilda Lacson MD PhD[a,b], Kristine S. Burk MD[a,b], Jeffrey P. Guenette MD, MPH [a,b], Pamela J. DiPiro MD [a,b], Ariadne DeSimone MD, MPH [a,b], Nicole Vetrano MHA, RT [a], Ishani Ganguli MD, MPH [b,c], David W. Bates MD [b,c], Ramin Khorasani MD, MPH [a,b]

[a]Brigham and Women's Hospital, Center for Evidence-Based Imaging, Department of Radiology, Boston, MA, USA
[b]Harvard Medical School, Boston, MA, USA
[c]Brigham and Women's Hospital, Department of Medicine, Boston, MA, USA

Corresponding Author:
Ronilda Lacson MD PhD
Department of Radiology, Brigham and Women's Hospital
One Brigham Circle, 1620 Tremont Street, Boston, MA 02120
Email: rlacson@bwh.harvard.edu
Telephone: 617-525-9712

All 9 co-authors contributed to the design of the study. RL, NV and RK worked on data acquisition and data analysis. All co-authors contributed to data interpretation as well as drafting the article and revising it critically for important intellectual content. All authors have seen and approved the final version of the manuscript being submitted. All authors accept accountability for the overall integrity of the research process and the manuscript.

Source of Support: This work was supported by Agency for Healthcare Research and Quality grant number R18HS029348.

The authors declare that they had full access to all of the data in this study and the authors take complete responsibility for the integrity of the data and the accuracy of the data analysis. The authors declare no conflict of interest.

Acknowledgements: The authors would like to thank Ms. Laura Peterson for reviewing the manuscript. This work was supported by Agency for Healthcare Research and Quality grant number R18HS029348.

Leadership Roles:
Ronilda Lacson MD PhD
Associate Director of Center for Evidence Based Imaging, BWH
Associate Professor of Radiology, HMS

Kristine S. Burk, MD
Medical Director of Safety, MGB Radiology
Assistant Professor of Radiology, HMS

Jeffrey P. Guenette, MD, MPH
Medical Director of Quality for Neuroradiology, MGB Radiology
Associate Director of Center for Evidence Based Imaging, BWH
Associate Professor of Radiology, HMS

Pamela J. Dipiro, MD
Medical Director of Quality, MGB Radiology
Associate Professor of Radiology, HMS

Ariadne DeSimone, MD, MPH
Medical Director of Quality for Thoracic Imaging, MGB Radiology
Assistant Professor of Radiology, HMS

Nicole Vetrano, MHA, RT
Manager, Performance Management
MGB Enterprise Radiology, Quality & Safety

Ishani Ganguli MD, MPH
Associate Professor of Medicine, HMS

David W. Bates, MD
Medical Director of Clinical and Quality Analysis, BWH
Chair, BWH Center for Patient Safety Research and Practice
Professor of Medicine, HMS

Ramin Khorasani, MD, MPH
Radiology Vice-Chair of Quality and Assistant Chief Medical Officer, MGB
Director of Center for Evidence Based Imaging, BWH
Professor of Radiology, HMS

Identifying and Categorizing Diagnostic Errors in Abdominal Imaging using a Diagnostic Imaging-Specific Annotation Process

Abstract

Objective: This initiative aimed to identify and classify contributory factors to diagnostic errors in abdominal imaging using an annotation process to identify, quantify, and categorize process-related errors.

Methods: This was a retrospective, Quality Improvement, IRB-approved study conducted at an academic health system on 200 randomly selected adult patients who completed abdominal diagnostic imaging examinations between 1/1/2022-12/31/2022 in which a radiologist recommended additional imaging (RAI). We used an expert panel to develop a taxonomy based on two previously-described taxonomies for process-related errors. We then applied an iterative process to train annotators to apply the taxonomy and assess medical records for diagnostic errors. We assessed the proportion of patients with diagnostic errors, as well as inter-rater reliability in classifying contributory factors. We also measured percentage agreement in identifying diagnostic errors using an annotation taxonomy vs. a gold standard developed by two abdominal radiologists.

Results: The final sample included 185 patients. A total of 14.6% (27/185; 95% Confidence Interval 10.2%-20.4%) patients undergoing abdominal imaging experienced at least one diagnostic error. Inter-annotator agreement with the gold standard increased from 61.1% (33/54) at baseline for 5 reviewers to 80% (40/50) (chi-square p-value=0.04) after 4 rounds of annotation. “Delay in performing ordered test(s)” is the major contributory factor to diagnostic

errors.

Discussion: Process-related diagnostic errors occurred in almost 1 in 7 patients undergoing abdominal imaging with RAI. After iterative training, an annotation process for identifying patients who have imaging diagnostic errors achieved significant inter-annotator agreement compared to a gold standard of radiology specialist review.

Keywords: diagnostic errors, annotation process, diagnostic workflow

Summary Statement: Process-related diagnostic errors occurred in one in seven patients who completed abdominal imaging that had recommendations for additional imaging, with “Delay in performing ordered test(s)” as the major contributory factor to errors.

Introduction

Getting the right diagnosis is often a complex process and may occur in an iterative manner over time, involving multiple specialty consultations and diagnostic examinations, including diagnostic imaging. However, the longer it takes to establish the right diagnosis, the greater the chance for a disease to progress to more advanced stages, which may then lead to worse outcomes.[1] Thus, a National Academy of Medicine (NAM) recommendation for healthcare professionals is reducing diagnostic errors that may harm patients.[2]

NAM defines diagnostic error as the "the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the patient."[3] Schiff et al further define diagnostic errors as "any mistake or failure in the diagnostic process leading to a misdiagnosis, a missed diagnosis, or a delayed diagnosis. This definition could include any failure in timely follow-up and specialty referral or evaluation."[4] In applying these definitions to diagnostic imaging, diagnostic errors can occur due to interpretation errors (misdiagnosis or inaccurate findings),[5,6] but also at other stages in the imaging process, including lack of communication of critical findings,[7,8] or delay or failure to obtain clinically necessary follow-up imaging for incidental or unexplained radiology findings (e.g., a recommendation to re-scan to evaluate a newly identified pulmonary nodule).[9-12]

Despite their importance, diagnostic errors related to diagnostic imaging are difficult to quantify and study, both because there is no gold independent standard, and also because there is no reliable and reproducible method to assess presence and contributory factors to errors (i.e., what went wrong). Most frameworks for assessing diagnostic errors have been developed for the entire medical specialty, specifically in internal medicine.[4,13,14] Radiology-

specific taxonomies have been developed for the subset of diagnostic imaging errors involving understanding interpretation errors related to cognitive and perceptual factors, with little focus on diagnostic errors related to systems and process errors in the diagnostic imaging workflow.[5,15] For abdominal imaging, in particular, there is substantial and significant variation in radiologists' rate of recommending follow-up imaging for findings identified during image interpretation.[16,17] Therefore, the aim of this study was to identify, quantify, and categorize diagnostic errors related to the diagnostic process in patients undergoing abdominal imaging, specifically those that yielded a radiologist recommendation for additional imaging (RAI).

Methods

Study Design and Setting

This Institutional Review Board (IRB)-approved, retrospective, Quality Improvement study was conducted at a large academic medical center with outpatient practices, performing 900,000+ radiology examinations annually. The requirement to obtain consent was waived by the IRB. We adhered to the Revised Standards for Quality Improvement Reporting Excellence (SQUIRE 2.0). The site utilizes two electronic communication systems for certain radiology findings. ANCR (Alert Notification of Critical Results) allows radiologists to complete closed-loop communication with ordering providers about critical results.[8,18] For patients with a RAI, ARRC (Addressing Radiologists' Recommendations Collaboratively) is used to alert the ordering or primary care clinician. If they accept the recommendation, a Collaborative Care Plan is established for further management of the finding.[12,19,20] A Safety Net program is integrated with ARRC,[21] which helped ensure bidirectional communication between radiologists and

referring providers in order to facilitate the timely performance of RAIs that they agreed to be clinically necessary. In addition to coordinating between providers, the Safety Net also engages the patient as part of the escalation process.

Primary Outcome

The proportion of patients with diagnostic errors among all patients reviewed was calculated as the diagnostic imaging error rate. We also measured the number of errors at each step in the diagnostic imaging workflow.

Annotation Framework and Taxonomy Development

An expert panel, comprised of 10 radiologists, internists, and patient safety specialists, was convened to select a definition of diagnostic error and a taxonomy for categorizing these errors. Patient safety specialists hold at least a master's degree and possess a minimum of five years of experience in patient safety. The panel selected the NAM definition of diagnostic error during an in-person panel discussion with 100% consensus, which connotes that a diagnosis is a process that evolves over time and thus, accuracy and timeliness are important to optimize patient management. The panel also elected to utilize the Diagnostic Imaging Workflow described previously.[22,23] Specifically, this workflow lists seven steps in the diagnostic workflow for radiology examinations. We combined Test Ordering and Care Planning due to difficulty in assessing provider intent when ordering an examination (Table 2, column 1).

To guide the development of the taxonomy for diagnostic errors in diagnostic imaging, the panel used the Systems Engineering Initiative for Patient Safety (SEIPS) model,[24-26]

augmented with the Diagnostic Error Evaluation and Research (DEER) taxonomy.[4] SEIPS provides a comprehensive conceptual framework to apply systems engineering in evaluating various causal pathways leading to diagnostic errors. The framework focuses on the "Work System"; a combination of socio-technical factors including Person, Task, Tools and Technology, Organization, and Environment factors, described previously. In our contributory factors for diagnostic errors, we included Person factors such as "*Patient fails to adhere to follow-up plan"* and *"Failure in shared decision-making."* Task factors include *"Test ordered wrong way"* and *"Inability to delete ordered examination."* Tools factors include *"Failure in Critical Result Communication system"* and *"Diagnostic equipment malfunction."* Organization factors include *"Deviation from standard operating procedure."* The DEER taxonomy has been used to categorize diagnostic errors by where in the diagnostic workflow they occurred, including Access/Presentation, History, Physical Examination, Tests, Assessment, Referral/Consultation and Follow-up.

The final diagnostic imaging error taxonomy included 32 categories of contributory factors to error, corresponding to the 7 steps in the diagnostic imaging workflow (Table 2, column 2). An annotation flowchart (Figure 1) and general rules for annotation (Figure 2) were also developed by radiology specialists for guiding trained research personnel who are not health professionals in using the taxonomy during the annotation process. The process also includes an assessment of the degree of harm attributed to each error, based on a score ranging 0-4: where 0=no harm–did not reach patient, 1=no harm–did reach patient, 2=temporary or minor harm, 3=permanent or major harm, and 4=death, published previously.[22]

Annotation Process

The taxonomy was applied to patient medical records via iterative rounds of annotation by research personnel after training, with results compared to a gold standard defined by radiology specialists. Rounds were continued until annotators achieved 80% agreement with the gold standard (Figure 3).

Physician Specialist Annotation Process

Two abdominal radiology specialists, both specialty-certified with 5 and 30 years of experience, reviewed cases to create a gold standard to be used for the research personnel training. They were provided with annotation materials (i.e., diagnostic imaging workflow, diagnostic error and adverse events taxonomy), a data collection tool, and assigned medical records (see *Data Collection*) for review. The gold standard was based on a final decision by the two radiology specialists after discussion. Discrepancies between the two primary reviewers were resolved by a third independent adjudicator. Using the gold standard cases, the abdominal radiology specialists attended iterative training for annotators until the annotators achieved 80% agreement with the gold standard.

Research Personnel Training

Five annotators (all required to have bachelor's degrees and at least 1 year of clinical research/practice experience) completed five hours of training prior to reviewing patient records for diagnostic errors (Table 1). Each annotator was provided with annotation materials like those provided for the radiology specialists, and at least 10 patient cases at each iteration or round of review. Patients were randomly selected and assignments were restricted so annotators had unique cases at each round. No patients that have been reviewed in prior rounds are re-assigned to any reviewers in future rounds.

After each round, cases were discussed in a one-and-a-half-hour meeting with the study team (including the abdominal radiology specialists and at least one other radiologist and internist) to discuss potential issues faced by annotators and answer relevant questions.

Data Collection and Medical Record Review

For the annotation process, 200 adult patients were randomly selected from those completing diagnostic abdominal imaging examinations with RAI at the study institution between 1/1/2022-12/31/2022. Patient records for the selected examinations, including radiology reports, pathology reports, provider notes, coordinator notes, and automated coordination tool records, were obtained from the institutional electronic health record (EHR: Epic Systems, Madison, WI).

For each patient, we determined whether follow-up was obtained within the timeframe specified in the RAI plus another 50% of that interval. Determination of follow-up was determined from 1/1/2022 to 6/30/2024. If a follow-up timeframe was not specified, we monitored test completion for one and a half years before labeling it a diagnostic error (i.e., diagnostic delay). Annotators did not assess recommended modality or timeframe for validity, but if a provider noted in subsequent medical records that another test should have been performed and this led to a diagnostic error, we considered whether the original modality was inadequate. We noted when ANCR, ARRC or another notification system was accessed by clinicians (e.g., radiologists, primary care clinicians), and whether radiology reports were read by ordering clinicians. Finally, because patients are stakeholders in the decision-making process, if documentation stated that they deviated from a follow-up recommendation or declined follow-up imaging, there was no diagnostic error noted. However, patient “no show” for

scheduled clinically necessary testing was deemed diagnostic error.

Descriptive and Statistical Analyses

To evaluate the annotation process, inter-rater reliability was calculated via percent agreement and Cohen's kappa coefficient[27] adjudicating presence of diagnostic error for each annotator at each round of training. We compared annotator agreement with the gold standard at baseline and on the fourth round of annotation using chi-square analysis. In the last round of training, percent agreement was calculated for categorization of: 1) the harm score, as a binary category where a score of 0 and 1 were categorized as "no harm" and 2-4 as "harm" reaching the patient, 2) the step in the diagnostic imaging workflow, and 3) for categorization into the taxonomy for diagnostic errors, based on the gold standard. Confidence intervals were calculated with Wilson's method.[28] We performed a sensitivity analysis for percent agreement in categorization of the step in the diagnostic imaging workflow and the taxonomy for diagnostic errors, excluding cases when not labeled as errors by annotators.

Results

Study Cohort

There were 7,137 reports with RAI out of 114,630 reports read by abdominal imaging staff (6.2%), from which the 200 cases were randomly selected from. Among the 200 cases, 15 were excluded because they had non-abdominal imaging or had no imaging follow-up recommended but were initially included because of ambiguous report language. Therefore, we analyzed a total of 185 patients.

Diagnostic Error Rate and Harm Scores

A total of 27 study patients (14.6%, 95% Confidence Interval [95%CI] 10.2%-20.4%) had at least one diagnostic error identified based on the gold standard; 10 of these patients had a harm score of 1 (37.0%, 95%CI 21.5%-55.8%) while 17 had a harm score of 0 (63.0%, 95%CI 44.2%-78.5%).

Table 2, column 3 displays the diagnostic errors by imaging workflow step and contributory factors to diagnostic errors. Most of the errors were noted in the *Test Scheduling* step with 14 of the 27 errors identified (51.9%, 95%CI 34.0%-69.3%). Within this step, "Delay in performing ordered test(s)" accounted for the vast majority (13/14; 92.9%, 95%CI 68.5%-98.7%) of diagnostic errors.

Evaluation of the Annotation Process for Presence of Diagnostic Error

Agreement between each annotator and the gold standard in classifying patient cases with diagnostic errors are included in Table 3. An average agreement of 80% with the gold standard was achieved after 4 rounds. One of the annotators did not submit annotations for the first round but was present to review the findings during the case discussion with the study team. Thus, there were only results for 4 annotators in the first round. The mean agreement for all annotators increased at each round. Inter-annotator agreement with the gold standard was 33/54 (61.1%) at baseline for 5 reviewers (the baseline for reviewer 2 was in the 2nd round). This increased to 40/50 (80%) in the fourth round (chi-square p-value=0.04) (Table 3).

Evaluation for Harm Score, Steps in the Diagnostic Imaging Process, and Contributory Factors to Diagnostic Error

Agreement between annotators and the gold standard for the harm score, the step in the diagnostic imaging workflow, and the contributory factors to diagnostic error are provided

in Table 4. Agreement with the gold standard for assessing the harm score was 0.70, for the diagnostic workflow was 0.60, and for the contributory factors to diagnostic error was 0.53. Excluding cases not labeled as diagnostic error by the annotators, agreement with the gold standard for assessing diagnostic workflow was 0.77, and for the contributory factors to diagnostic error was 0.67 (Supplemental Table 1).

Discussion

The observed diagnostic error rate for abdominal imaging with RAI was 14.6%, well within the range of published rates of approximately 5-17%.[29] However, adverse events that were previously reported occurred primarily in the *Imaging Procedure* step of the workflow,[23] as opposed to *Test Scheduling,* which we note in this study. We specifically note that "Delay in performing ordered test(s)" is the major contributory factor to diagnostic errors, which is an actionable issue that can be addressed with diagnostic imaging process enhancement. These enhancements could include continuous adjustments in management and coordination of staff schedules and optimizing imaging time to enhance patient care.[30] Alternatively, if delays are not due to system capacity limits and instead are due to patient preference, more patient-centered initiatives can be introduced. Other adjustments to healthcare settings may be necessary to enhance service accessibility and operational performance.

In this initiative, we developed and tested an annotation process for reviewing patients who underwent diagnostic imaging to assess for the presence of *process-related diagnostic errors within the imaging workflow*, and a corresponding harm score. We augmented a previously-described annotation taxonomy for contributory factors to diagnostic errors based

on steps in the diagnostic imaging workflow and used this to classify the errors we identified. **We found delay in "Delay in ordering needed test(s)" and "Delay in performing ordered test(s)" as the two most common contributory factors to diagnostic errors.** And we found a one in seven diagnostic error rate. In a previous study analyzing medical litigation cases, 20% alleged a diagnosis-related error.[31] 31% of diagnosis-related ambulatory cases similarly involve a failure or delay in ordering a diagnostic test,[32] and almost half of these diagnostic tests (45%) are from diagnostic imaging.[31] Fortunately, there was no harm to the patient in any of the study cases. This, however, raises the potential that some recommended additional imaging may not be clinically necessary. This should be studied further to understand potential overutilization of high-cost imaging.[33] This may also help address operational efficiency by minimizing recommendations for unnecessary follow-up imaging.

Our results demonstrate the feasibility and importance of reliable, efficient diagnostic error measurement. The annotation training process proceeded smoothly with 5 annotators who achieved 80% inter-rater agreement after 4 rounds of training. Achieving agreement with a gold standard is critical in training safety personnel who can help in documenting and measuring diagnostic errors that are not readily available for review by safety officers. In addition, measuring diagnostic errors efficiently using standard approaches to identify preventable errors are critical to supporting health systems' patient safety initiatives.

Our study builds on prior work by being more inclusive of error types. Specifically, we used a harm score that was based on an existing score for assessing patient safety reports[22] which included errors that did not lead to significant harm. This is different from another harm score adapted for use by Bates et. al.,[34] which included significant, serious, life-threatening and

fatal errors. We were interested in addressing all diagnostic errors even though they may not lead to significant harm at this time because they need to be addressed to avoid potentially significant harm in the future. Studying errors even if they resulted in no actual harm is essential to building system resilience, which contributes to patient safety.[35] It proactively identifies hidden vulnerabilities and prevents future failures. Operationally, addressing these errors within the systems in place includes identification of socio-technical factors that can be addressed as a healthcare system, without added burden to individual providers.

We developed an annotation flowchart for diagnostic imaging, which was highly effective in assisting annotators in completing the iterative annotator training process. It required multiple steps and access to multiple modules within the EHR. Specifically, each annotator relied heavily on the critical result and recommendation communication systems, ANCR and ARRC, as well as provider documentation in the notes. This included documentation of a patient's role in decision-making as well as provider communication regarding reasons for disagreement with a care plan. Institutions without such systems would have to bypass the fourth step in the flowchart (Figure 1) – checking if ANCR utilization is appropriate. This precludes being able to check whether a RAI was acknowledged by the correct provider, was acknowledged by a provider who is not responsible for the RAI, was not sent on time, was not acknowledged on time, was not communicated to patients, or whether there was provider non-agreement with a RAI. This limits the capacity for detection of contributory factors to errors.

Limitations

This study involved a single academic health system. Thus, the incidence rates of reported diagnostic errors may not generalize to other institutions. However, we anticipate that

diagnostic errors at other institutions will occur in various steps in the diagnostic imaging process that involve similar socio-technical factors, and thus be amenable to similar analyses using our taxonomy and annotation training process. In addition, we have access to several automated radiology communication systems. Thus, the annotation flowchart may need to be modified for institutions without access to documentation of provider communications. Observed error rates may not be generalizable to institutions without similar systems. These systems may influence both detection and mitigation of errors. Finally, recommendations with no specified time frame for the follow-up documented in the initial report was assessed to be on time as long as the follow-up imaging was performed within 18 months. This is a very conservative time interval; thus, the true diagnostic error rate may be higher than what we identified in our study. We plan to study the impact of not specifying time frame in future studies. Without time frame included with recommendations, it is difficult to assess the diagnostic error rate precisely because it precludes assessment of timely follow-up completion. Additional studies are also recommended across multi-site institutions with varying Information Technology system capabilities and Electronic Health Records before this process can be standardized.

Take-Home Points

- Process-related diagnostic errors occurred in almost 1 in 7 patients who completed abdominal imaging that had recommendations for additional imaging .
- Research personnel can be trained to detect diagnostic errors such as those related to

delays in follow-up and delays in ordering diagnostic testing using a structured annotation process.

- Enabling early and systematic detection of process-related diagnostic errors in diagnostic imaging may help mitigate these errors and facilitate evaluation of potential impact of process enhancements to engage clinicians and patients in proactively addressing diagnostic errors.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 01 2022;72(1):7-33. doi:10.3322/caac.21708
2. IOM. Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care. Balogh EP, Miller BT, Ball JR, editors. Washington (DC): National Academies Press (US); 2015 Dec 29. PMID: 26803862. 2015.
3. IOM. *Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine; Balogh EP, Miller BT, Ball JR, editors. Improving Diagnosis in Health Care. Washington (DC): National Academies Press (US); 2015 Dec 29. Available from: https://www-ncbi-nlm-nih-gov.ezp-prod1.hul.harvard.edu/books/NBK338596/ doi: 10.17226/21794*. December 2015 ed. National Academy of Sciences; 2015.
4. Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med*. Nov 9 2009;169(20):1881-7. doi:10.1001/archinternmed.2009.333
5. Lee CS, Nagy PG, Weaver SJ, Newman-Toker DE. Cognitive and system factors contributing to diagnostic errors in radiology. *AJR Am J Roentgenol*. Sep 2013;201(3):611-7. doi:10.2214/AJR.12.10375
6. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive Error in Radiology. *AJR Am J Roentgenol*. Apr 2017;208(4):739-749. doi:10.2214/AJR.16.16963
7. Waite S, Scott JM, Drexler I, et al. Communication errors in radiology - Pitfalls and how to avoid them. *Clin Imaging*. Sep - Oct 2018;51:266-272. doi:10.1016/j.clinimag.2018.05.025
8. Lacson R, Prevedello LM, Andriole KP, et al. Four-year impact of an alert notification system on closed-loop communication of critical test results. *AJR Am J Roentgenol*. Nov 2014;203(5):933-8. doi:10.2214/AJR.14.13064
9. Nemer JS, Kazam JK, Wong TT. Trends of Follow-Up Recommendations Made on Musculoskeletal MRI Reports. *AJR Am J Roentgenol*. 03 2020;214(3):630-635. doi:10.2214/AJR.19.21770
10. You JJ, Laupacis A, Newman A, Bell CM. Non-adherence to recommendations for further testing after outpatient CT and MRI. *Am J Med*. Jun 2010;123(6):557.e1-8. doi:10.1016/j.amjmed.2009.11.018
11. Mabotuwana T, Hall CS, Hombal V, et al. Automated Tracking of Follow-Up Imaging Recommendations. *AJR Am J Roentgenol*. Mar 12 2019:1-8. doi:10.2214/AJR.18.20586
12. Kapoor N, Lynch EA, Lacson R, et al. Predictors of Completion of Clinically Necessary Radiologist-Recommended Follow-Up Imaging: Assessment Using an Automated Closed-Loop Communication and Tracking Tool. *AJR Am J Roentgenol*. Mar 2023;220(3):429-440. doi:10.2214/AJR.22.28378
13. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf*. Oct 2013;22 Suppl 2:ii21-ii27. doi:10.1136/bmjqs-2012-001615
14. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med*. Jul 11 2005;165(13):1493-9. doi:10.1001/archinte.165.13.1493

15. Kim YW, Mansfield LT. Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors. *AJR Am J Roentgenol*. Mar 2014;202(3):465-70. doi:10.2214/AJR.13.11493
16. Cochon LR, Kapoor N, Carrodeguas E, et al. Variation in Follow-up Imaging Recommendations in Radiology Reports: Patient, Modality, and Radiologist Predictors. *Radiology*. May 2019:182826. doi:10.1148/radiol.2019182826
17. Kapoor N, Lacson R, Eskian M, et al. Variation in Radiologists' Follow-Up Imaging Recommendations for Small Cystic Pancreatic Lesions. *J Am Coll Radiol*. Oct 2021;18(10):1405-1414. doi:10.1016/j.jacr.2021.06.007
18. Lacson R, O'Connor SD, Andriole KP, Prevedello LM, Khorasani R. Automated critical test result notification system: architecture, design, and assessment of provider satisfaction. *AJR Am J Roentgenol*. Nov 2014;203(5):W491-6. doi:10.2214/AJR.14.13063
19. Hammer MM, Kapoor N, Desai SP, et al. Adoption of a Closed-Loop Communication Tool to Establish and Execute a Collaborative Follow-Up Plan for Incidental Pulmonary Nodules. *AJR Am J Roentgenol*. Feb 2019:1-5. doi:10.2214/AJR.18.20692
20. Desai S, Kapoor N, Hammer MM, et al. RADAR: A Closed-Loop Quality Improvement Initiative Leveraging A Safety Net Model for Incidental Pulmonary Nodule Management. *Jt Comm J Qual Patient Saf*. May 2021;47(5):275-281. doi:10.1016/j.jcjq.2020.12.006
21. Abbasi N, Kapoor N, Lacson R, et al. Cumulative Effect of Targeted Interventions on Radiologist Recommendations for Additional Imaging. *Radiology*. Jun 2025;315(3):e243750. doi:10.1148/radiol.243750
22. Cochon L, Lacson R, Wang A, et al. Assessing information sources to elucidate diagnostic process errors in radiologic imaging - a human factors framework. *J Am Med Inform Assoc*. 11 2018;25(11):1507-1515. doi:10.1093/jamia/ocy103
23. Lacson R, Cochon L, Ip I, et al. Classifying Safety Events Related to Diagnostic Imaging From a Safety Reporting System Using a Human Factors Framework. *J Am Coll Radiol*. Mar 2019;16(3):282-288. doi:10.1016/j.jacr.2018.10.015
24. Carayon P, Karsh BT, Gurses AP, et al. Macroergonomics in Healthcare Quality and Patient Safety. *Rev Hum Factors Ergon*. Sep 1 2013;8(1):4-54. doi:10.1177/1557234X13492976
25. Carayon P, Wetterneck TB, Rivera-Rodriguez AJ, et al. Human factors systems approach to healthcare quality and patient safety. *Appl Ergon*. Jan 2014;45(1):14-25. doi:10.1016/j.apergo.2013.04.023
26. Holden RJ, Carayon P, Gurses AP, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*. 2013;56(11):1669-86. doi:10.1080/00140139.2013.838643
27. Shabankhani B. Assessing the inter-rater reliability for nominal, categorical and ordinal data in medical sciences. . *Acta Pharma Pract*. 2020;11(S4):144-148.
28. Brown L, Cai T, DasGupta A. Interval Estimation for a Binomial Proportion. *Statistical Science*. 2001;16:101-133.
29. Hall KK, Shoemaker-Hunt S, Hoffman L, et al. Making Healthcare Safer III: A Critical Analysis of Existing and Emerging Patient Safety Practices. 2020.
30. Crowley C, Guitron S, Son J, Pianykh OS. Modeling workflows: Identifying the most predictive features in healthcare operational processes. *PLoS One*. 2020;15(6):e0233810. doi:10.1371/journal.pone.0233810

31. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med*. Oct 3 2006;145(7):488-96.
32. (CRICO) CRIC. 2014 CBS Report: Malpractice Risks in the Diagnostic Process. 2014. Accessed October 1, 2015. http://www.rmfstrategies.com/Clinician-Resources/Article/2014/CBS-Intro?sc_camp=0EB2AF4DEE834EB995C210A024A5DB29
33. Carter EA, Morin PE, Lind KD. Costs and Trends in Utilization of Low-value Services Among Older Adults With Commercial Insurance or Medicare Advantage. *Med Care*. Nov 2017;55(11):931-939. doi:10.1097/MLR.0000000000000809
34. Bates DW, Levine DM, Salmasian H, et al. The Safety of Inpatient Health Care. *N Engl J Med*. Jan 12 2023;388(2):142-153. doi:10.1056/NEJMsa2206117
35. Fairbanks RJ, Wears RL, Woods DD, Hollnagel E, Plsek P, Cook RI. Resilience and resilience engineering in health care. *Jt Comm J Qual Patient Saf*. Aug 2014;40(8):376-83. doi:10.1016/s1553-7250(14)40049-7
36. IOM. Institute of Medicine. Improving Diagnosis in Health Care. National Academy of Sciences; 2015. http://iom.nationalacademies.org/~/media/Files/Report%20Files/2015/Improving-Diagnosis/DiagnosticError_ReportBrief.pdf
37. Singh H, Sittig DF. Advancing the science of measurement of diagnostic errors in healthcare: the Safer Dx framework. *BMJ Qual Saf*. Feb 2015;24(2):103-10. doi:10.1136/bmjqs-2014-003675
38. Folli HL, Poole RL, Benitz WE, Russo JC. Medication error prevention by clinical pharmacists in two children's hospitals. *Pediatrics*. May 1987;79(5):718-22.

Figure Legend:

Figure 1: Annotation Flowchart
Figure 2: General Rules for Annotation
Figure 3: Annotator Training Process

Tables

Table 1: Annotation Training Process for Categorizing Diagnostic Errors in Diagnostic Imaging

Training Process (Steps)	Study Materials	Duration
Describe Data Definitions and Frameworks	Definitions of diagnostic errors[4,14,36,37]	1.5 hours
	Definitions of adverse events and harm score[22,23,34,38]	
	Taxonomies for coding imaging diagnostic errors[4,22]	
	Diagnostic imaging workflow[22]	
Describe Clinical Studies and Resilient Systems to Reduce Diagnostic Errors	Diagnostic Excellence Center on Diagnostic Error (DECODE)	1.5 hours
	Alert Notification of Critical Results (ANCR)[18]	
	Addressing Radiologists' Recommendations Collaboratively (ARRC)[12]	
Manual Review Process	Electronic Health Record Review (Clinical notes, provider acknowledgement, scheduler notes)	2.0 hours

Table 2: Diagnostic Imaging Workflow and Taxonomy for Contributory Factors to Diagnostic Errors

Diagnostic Imaging Workflow Step[22,23]	Contributory Factors to Diagnostic Error[4,22]	Number of Errors (N=27)
Test Ordering and Care Planning	Patient fails to adhere to follow-up plan	1
	Failure/delay to act on follow-up	
	Communication gap	
	Incorrect clinical information	
	Delay in communication to patient	
	Failure/delay in ordering needed test(s) and Inadequate handoff	7
	Error in test sequencing	
	Ordering of wrong test(s)	
	Test ordered wrong way	
	Incorrect clinical information	
	Duplicate orders	
	Inability to delete ordered examination	
Test Scheduling	Failure in performing ordered test(s)	1
	Delay in performing ordered test(s)	13
Test Protocoling	Technical errors/poor processing of test	
	Wrong protocol	1
Imaging Procedure	Procedure not performed	
	Image mislabeled (e.g., wrong patient/test)	
	Delayed procedure	
	Patient unable to cooperate	
	Wrong procedure performed	
	Incorrectly performed procedure	
	Diagnostic equipment failure/malfunction	
	Deviation from standard operating procedure	
Image Interpretation	Erroneous radiology reading of test	
	Erroneous follow-up recommendation	
	Incorrect clinical information	
Reporting	Failed/delayed reporting of result	
	Lack of report availability	
Result Communication	Delay in communication to Responsible Physician	3
	Delayed acknowledgement of result	
	Delay in communication to patient	
	Failure in Critical Result Communication system	1

Table 3: Agreement between Annotators and Gold Standard for Presence of Diagnostic Error

Annotators	Presence of Error	
	Percent Agreement	Kappa
Round 1* (n=11 cases)		
1	0.45	0.00
3	0.45	0.00
4	0.82	0.62
5	0.91	0.81
Round 2 (n=10 cases)		
1	0.40	0.09
2	0.40	0.09
3	0.60	0.20
4	0.60	0.31
5	0.50	0.07
Round 3 (n=10 cases)		
1	0.60	0.29
2	0.80	0.54
3	0.30	0.00
4	0.80	0.54
5	0.90	0.74
Round 4 (n=10 cases)		
1	0.90	0.74
2	0.80	0.54
3	0.70	0.40
4	0.70	0.21
5	0.90	0.74

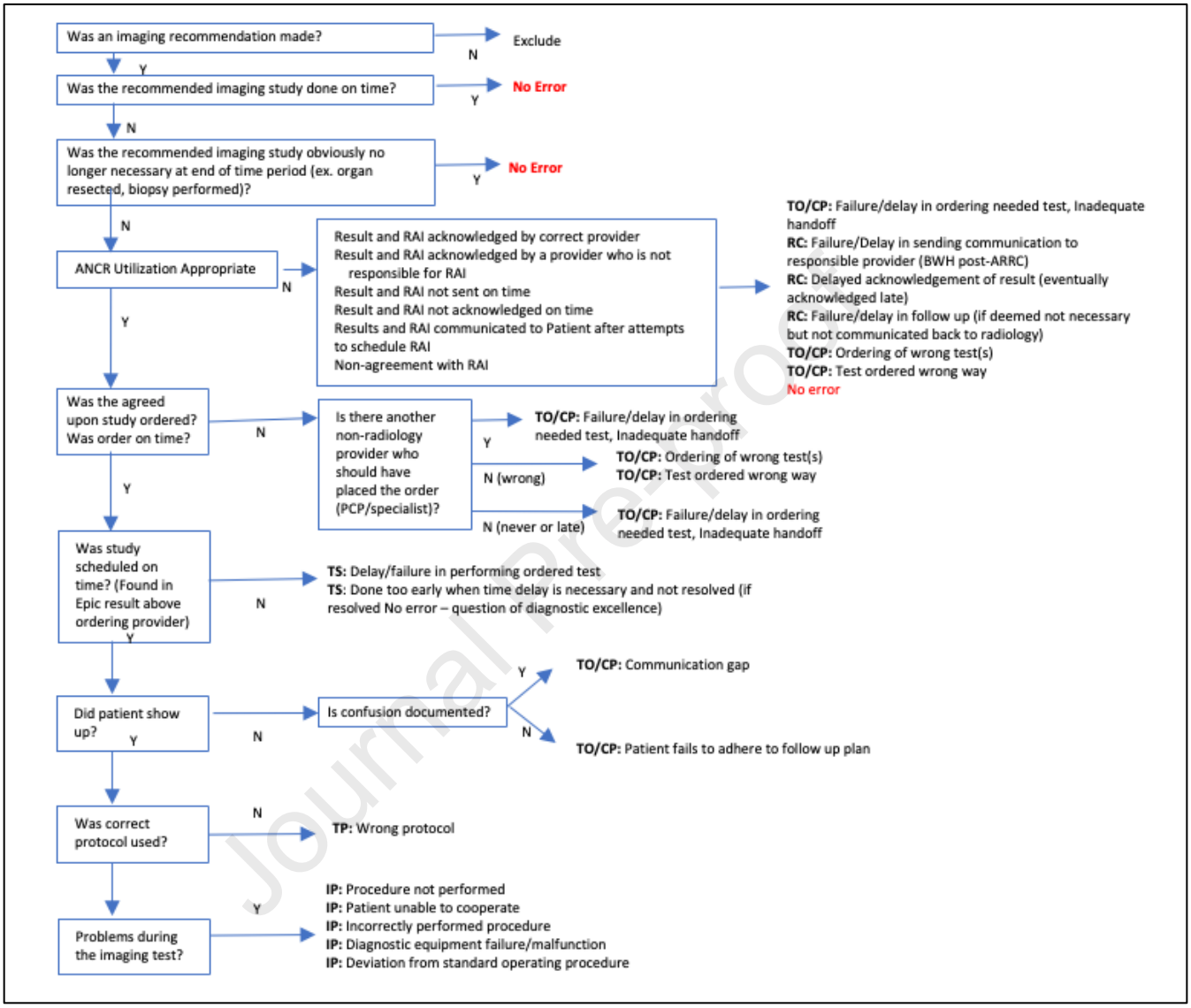
*Reviewer 2 lacks round 1 data and did not submit annotations for the first round

Table 4: Agreement between Annotator and Gold Standard for Harm Score, Diagnostic Workflow, and Contributory Factors to Diagnostic Error at Round 4 of Annotation

Annotators	Harm Score	Diagnostic Workflow	Contributory Factors to Diagnostic Error
	Percent Agreement	Percent Agreement	Percent Agreement
1	0.62	0.63	0.63
2	0.88	0.50	0.38
3	0.62	0.75	0.63
4	0.62	0.38	0.25
5	0.75	0.75	0.75
Mean	**0.70**	**0.60**	**0.53**

Figures

Figure 1: Annotation Flowchart



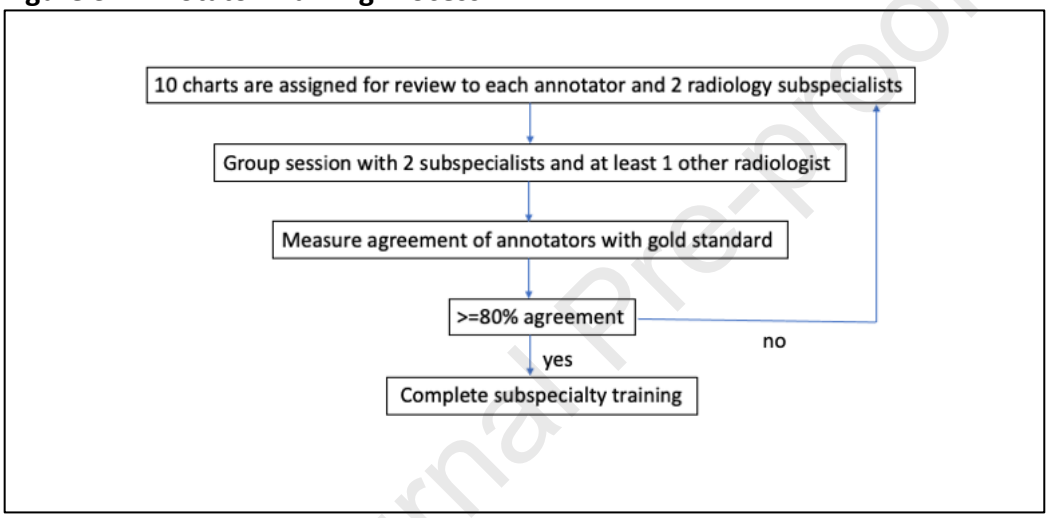
TO/CP: Test Ordering/Care Planning; TS: Test Scheduling; TP: Test Protocoling; IP: Imaging Procedure; RC: Result Communication

Figure 2: General Rules for Annotation

General Rules:

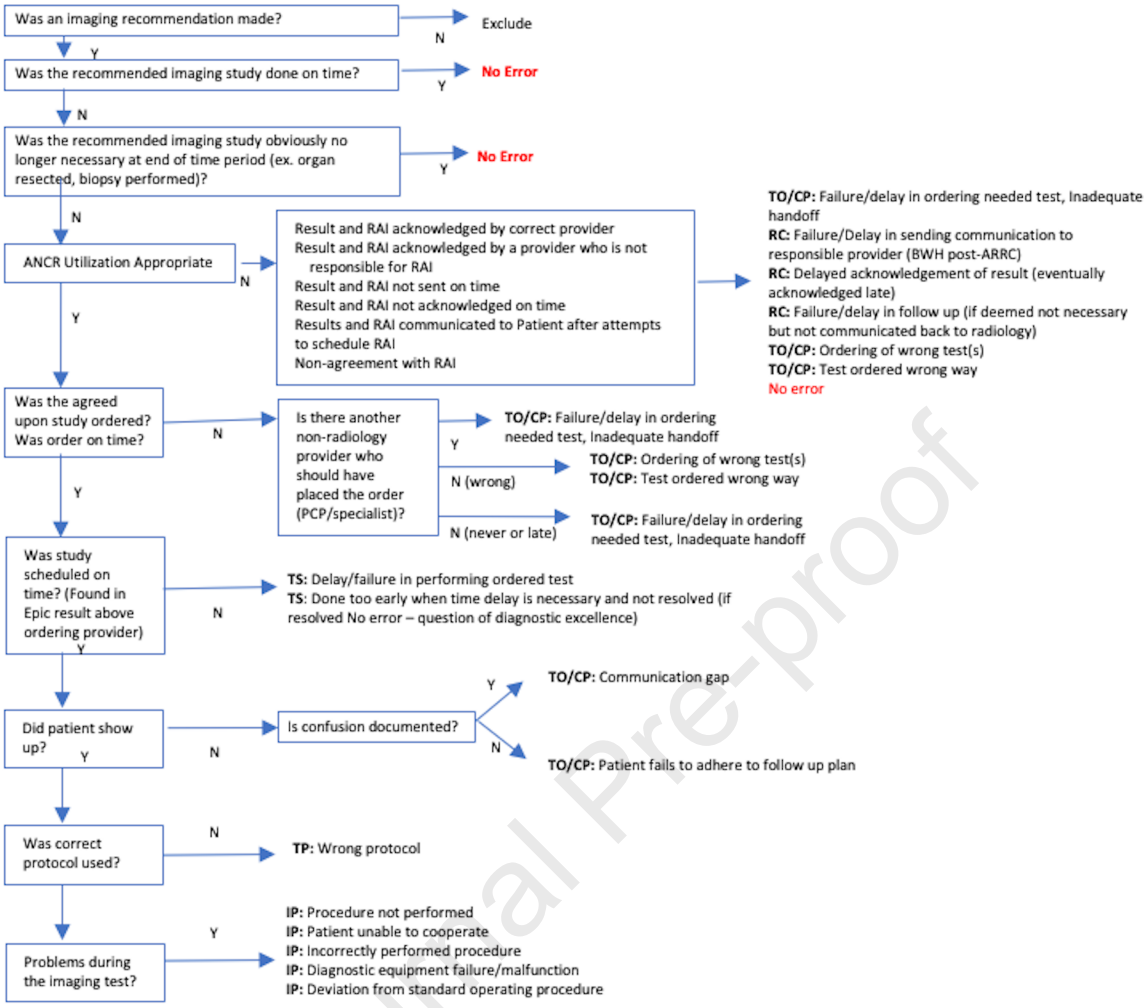
- Allow for 1.5 times the maximum recommended follow up time frame before considering delay in follow-up.
- Each patient is a stakeholder in the decision-making process; if documentation states that they deviated from a follow up recommendation after discussion with their physician, this is not an error.
- Recommendations for imaging ordered by an outside provider is included in the study scope.

Figure 3: Annotator Training Process



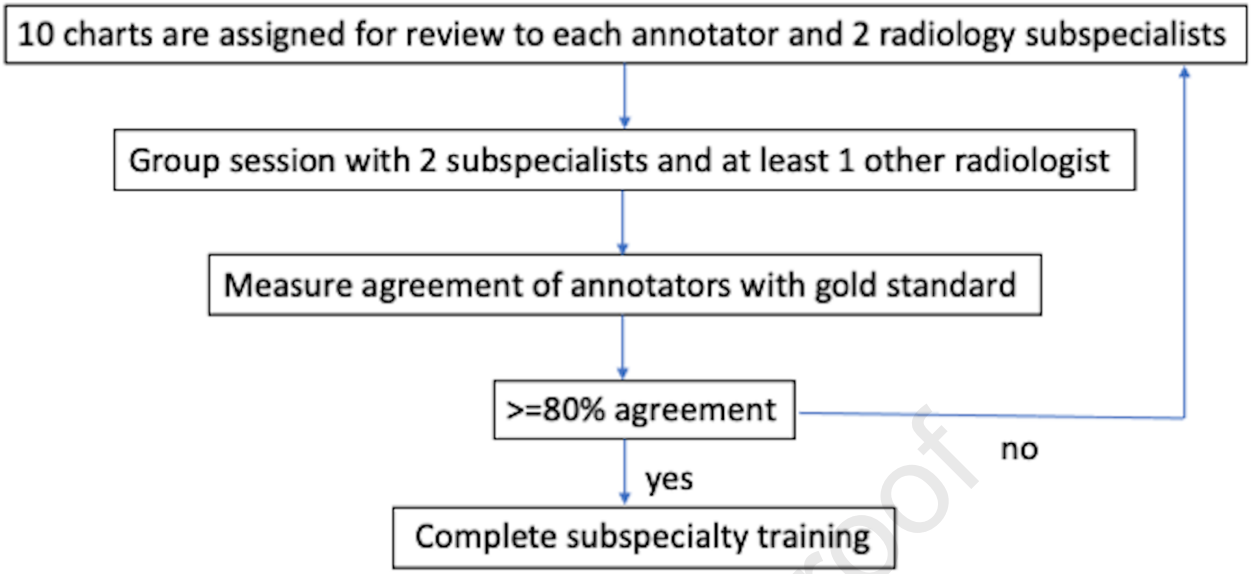
Supplemental Table 1: Agreement between Annotator and Gold Standard for Diagnostic Workflow and Contributory Factors to Diagnostic Error at Round 4 of Annotation (Excluding Cases Not Labeled as Errors)

Annotators	**Diagnostic Workflow**	**Contributory Factors to Diagnostic Error**
	Percent Agreement	**Percent Agreement**
1	0.71	0.71
2	0.67	0.50
3	0.86	0.71
4	0.60	0.40
5	1.00	1.00
Mean	**0.77**	**0.67**



General Rules:

- Allow for 1.5 times the maximum recommended follow up time frame before considering delay in follow-up.
- Each patient is a stakeholder in the decision-making process; if documentation states that they deviated from a follow up recommendation after discussion with their physician, this is not an error.
- Recommendations for imaging ordered by an outside provider is included in the study scope.



Summary Statement

Process-related diagnostic errors occurred in one in seven patients who completed abdominal imaging that had recommendations for additional imaging, with “Delay in performing ordered test(s)” as the major contributory factor to errors.

Take-Home Points

- Process-related diagnostic errors occurred in almost 1 in 7 patients who completed abdominal imaging that had recommendations for additional imaging.
- Research personnel can be trained to detect diagnostic errors such as those related to delays in follow-up and delays in ordering diagnostic testing using a structured annotation process.
- Enabling early and systematic detection of process-related diagnostic errors in diagnostic imaging may help mitigate these errors and facilitate evaluation of potential impact of process enhancements to engage clinicians and patients in proactively addressing diagnostic errors.