Variation in follow-up imaging recommendations in radiology reports: Patient, modality, and radiologist predictors

Laila R. Cochon¹, Neena Kapoor¹, Emmanuel Carrodeguas¹, Ivan K. Ip¹, Ronilda Lacson¹, Giles Boland¹, Ramin Khorasani¹

¹Center for Evidence-Based Imaging, Department of Radiology, Brigham and Women’s Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02115

Abstract

Background: Variation between radiologists when making recommendations for additional imaging and associated factors are unknown. Clear identification of factors that account for variation in follow-up recommendations might prevent unnecessary tests for incidental or ambiguous image findings.

Purpose: Determine incidence and identify factors associated with follow-up recommendations in radiology reports from multiple modalities, patient care settings, and imaging divisions.

Material and Methods: This retrospective study analyzed 318,366 reports obtained from diagnostic imaging exams performed at a large urban quaternary care from 1/1/2016 to 12/31/2016, excluding breast and ultrasound reports. A subset of 1000 reports were randomly selected and manually annotated to train and validate a machine learning algorithm to predict whether a report included a follow-up imaging recommendation (training and validation set consisted of 850 reports and test set of 150 reports). The trained algorithm was used to classify 318,366 reports. Multivariable logistic regression was used to determine the likelihood of follow-up recommendation. Additional analysis by imaging subspecialty division was performed and intra-division, inter-radiologist variability quantified.

Results: The machine learning algorithm classified 38,745 of 318,366 (12.2%) reports as containing follow-up recommendations. The average patient age was 59 years (SD ±17 years); 45.2% (143,767/318,366) of reports were from male patients. Among 65 radiologists, 56.9% (37/65) were male. In multivariable analysis, older patients had higher rates of follow-up recommendations (OR: 1.01 [1.01–1.01] for each additional year), male patients had lower rates (OR: 0.9 [0.9–1.0]), and follow-up recommendations were most common among CT studies (OR: 4.2 [4.0–4.4] compared to X-ray). Radiologist sex (p=0.54), presence of a trainee (p=0.45), and years in practice (p=0.49) were not significant predictors overall. A division-level analysis showed 2.8-fold to 6.7-fold inter-radiologist variation.

Conclusions: Substantial inter-radiologist variation exists in the probability of recommending a follow-up exam in a radiology report, after adjusting for patient, exam and radiologist factors.
Introduction

A critical piece of interpreting a radiologic examination is determining the need for potential follow-up imaging. Reported rates of follow-up recommendations in radiology reports range from 8–37% [1–6] with high rates of inter-observer variability [7]. Even radiology reports with non-critical findings may still contain recommendations for further imaging using different modalities, or recommendations for interval follow-up using the same modality [5]. Furthermore, there are known differences in radiologists’ familiarity and adherence with evidence-based guidelines [8–10]. This may lead to unnecessary additional tests in the follow-up of incidental or ambiguous imaging findings [11,12].

A better understanding of follow-up recommendations in radiology reports is needed, as well as clear identification of factors that might account for variation in rates of follow-up recommendations. Prior literature performed at single institutions and in single subspecialty divisions has suggested that radiologist experience may be associated with fewer follow-up recommendations [5,13]. However, it remains poorly understood how much radiologists vary when making recommendations for additional imaging and what additional radiologist and other factors influence the likelihood of additional imaging. We hypothesized there is a large amount of likely unwarranted variation among radiologists in terms of the probability of making a follow-up recommendation. Therefore, we sought to identify factors associated with follow-up recommendations in radiology reports from multiple modalities, patient care settings, and imaging divisions to better evaluate features associated with a higher amount of follow-up recommendations.

Materials and Methods

The Institutional Review Board at our study institution approved this HIPAA-compliant, observational study; the requirement for written informed consent was waived.

Study Setting and Population

We conducted a 12-month retrospective cohort study (January 1, 2016–December 31, 2016) at a Level-1 urban academic quaternary care hospital and its tertiary care centers, with an outpatient network that spans 183 practices and 1,200 physicians. We used a previously validated machine learning tool to identify reports with follow-up imaging recommendations [14]. The prior work did not analyze possible predictors of follow-up recommendations, which was the purpose of this study. The study population (Figure 1) included all 318,366 radiology reports generated during the study period (the inclusion criteria for the study), excluding those associated with procedures (arthrograms, image guided biopsies, etc.), and reports from the Breast Imaging and Ultrasound divisions (total of 229,714 studies excluded). Procedures were excluded because most will not contain follow-up recommendations and would falsely lower rates of follow-up recommendations. The Ultrasound division was excluded because many of the examinations are for obstetrical patients, which may limit generalizability to other radiology divisions. However, we included ultrasound as a modality because some divisions such as the Emergency Department and Musculoskeletal Imaging include ultrasound as part of their clinical offering. We excluded imaging performed by the Breast Imaging Division because it
represents a distinctive imaging setting with structured assessments and follow-up recommendations (15).

**Data Collection**

We queried the imaging data repository populated by the study institution’s electronic health record (Epic Systems Corporation, Verona, WI) for all eligible imaging reports, which we extracted (L.C., physician with 6 years of experience). We used the Epic Data Warehouse (Epic Systems Corporation, Verona, WI) to extract descriptive information for each report including: attending radiologist, modality (X-Ray, Ultrasound, CT, and MRI), imaging division of the interpreting radiologist, whether the report was interpreted with a trainee, patient sex, patient age, and care setting (inpatient, outpatient, or ED) (LC). We used the institution’s physician directory to determine number of years in practice for each radiologist (1–10 or 11+ years).

**Automated Machine Learning Report Classification**

We used a previously developed and validated (14) Support Vector Machine (SVM) algorithm (Precision=0.88, Recall=0.82) to perform Natural Language Processing (NLP) and identify radiology reports with one or more follow-up recommendations. This algorithm was previously optimized for this task, and benchmarked against other machine learning methods and an established information extraction system (iScout) (14). A link to the code used for training and application of this algorithm can be found at [https://github.com/edeguas/radiology-ml-text-classification/blob/master/training_and_predicting_code.py](https://github.com/edeguas/radiology-ml-text-classification/blob/master/training_and_predicting_code.py)

We applied power calculation with Buderer’s formula (95% confidence interval, precision: 0.05, estimated specificity: 0.95) yielding a size of 150 testing samples for evaluation of our algorithm. We selected a corpus of 1,000 reports from the study cohort. Under the guidance of a senior radiologist (R.K., radiologist with 25 years of experience), two raters (E.C., physician with 4 years of experience and RC) reviewed the reports for the presence of any follow-up recommendations. The first rater (E.C.) annotated 780 reports independently, while the second rater annotated 180 reports independently. A total of 40 reports were annotated by both raters to assess inter-rater agreement using the kappa statistic. Any disagreements in the 40 overlap reports were classified as follow-up recommendations. Both raters had access to the full radiology report text and study information. We randomly separated the total annotated reports into a 850 training/validation set and a 150 test set. The SVM model underwent 10-fold cross validation during the parameter optimization process (iteratively training over the 850 set; 765 for training and 85 for validation at each step). The top architecture underwent re-training on the full 850 cohort, with characteristics reported after testing on the 150 unseen test data set.

We defined a follow-up recommendation as any phrase that human reviewers deemed might result in further imaging or procedures (e.g., biopsy, colonoscopy, or cystoscopy). For example, “if clinically indicated, CT can be obtained” and “if clinical suspicion can follow up with MRI” constituted follow-up phrases, while “clinical correlation recommended” did not. We assigned a binary outcome variable to each report, either positive (contained one or
more follow-up recommendations) or negative (contained no follow-up recommendations). We used the validated algorithm on the entire study population of reports.

**Outcome Measures**

The primary outcome was the rate of follow-up recommendations. We also assessed factors associated with follow-up recommendations using patient, exam, and radiologist predictors in the Radiology Department overall, and by imaging subspecialty division. The secondary outcome was the probability of follow-up recommendations for individual radiologists.

**Statistical Analysis**

We determined predictors of follow-up recommendation by modeling factors for all imaging reports in the radiology department and by division: Abdomen, Chest, Cancer Institute, Emergency Radiology, MSK, and Neuroradiology. We analyzed inpatient and outpatient care settings at the division level. We used a nonlinear random-effects model with logit link function to estimate the level of variation in follow-up recommendation across radiologists for both the department and division level models. We used the radiologist as the random effect and generated the logit probability of follow-up recommendation in the models. We used McFadden’s Pseudo R squared to determine the overall variability explained by the model and used delta AIC to quantify the variability explained by the random effect. We evaluated goodness of fit using Hosmer Lemeshow for the department model and each division model. SAS software version 9.3 was used for all statistical analysis (SAS Institute Inc., Cary, NC).

We reported effect size (odds ratios [ORs]) and significance (p-value) of each covariate on the likelihood of follow-up recommendation. For the univariable and multivariable model, $P < 0.05$ was considered significant. However, we entered all variables in the univariable analysis with a $P < .20$ into the multivariable model to obtain adjusted odds ratios (ORs). We reported adjusted probabilities for each radiologist within each division.

**Results**

**Study Population**

We extracted a total of 318,366 eligible reports from 115,340 patients from the data repository, composed of 52.6% (167,320/318/366) X-Rays, 31.4% (100,050/318/366) CT, 13.5% (42,959/318,366) MRI and 2.5% (8037/318,366) US. These were interpreted by 65 unique attending radiologists, and a trainee was present in 48.0% (152,749/318,366) of reports. Average patient age was 59 years ±17 years (SD). A total of 45.2% (143,767/318,366) reports were from male patients and 54.8% (174,600/318,366) reports were from female patients. Full demographic characteristics are shown in Table 1.

**Automated Machine Learning Report Classification**

The annotated corpus was composed of 432 XR, 249 CT, 223 US and 96 MR reports, with a follow up rate of 12.7% (127 reports). Reports annotated by both raters had a percentage agreement of 97.5, with a Kappa agreement score of 0.78. The optimized SVM model had a
precision of 0.88, a recall of 0.82 and a F1-Score of 0.85 when tested on the test data excluded from training, selection and development of the algorithm.

Rate of Follow-up Recommendations

Of the total 318,366 reports, the algorithm identified 38,745 out of 318,366 (12.2%) as having at least one follow-up recommendation.

Follow-up Characteristics and Univariable Models

In the univariable model (Table 1), all predictors were associated with follow-up exam recommendation (p-values <0.01). For patient factors, older patients had higher rates of follow-up recommendation (OR: 1.01 [1.01–1.01] for each additional year), while male patients had lower rates of follow-up recommendation (OR: 0.9 [0.9–1.0]). Among radiologist and division factors, the presence of a trainee (OR: 1.7 [1.7–1.8]) and having a male radiologist (OR: 1.2 [0.8–1.6]) were both associated with higher odds of follow-up recommendation. Compared to X-Ray, all modalities were associated with higher odds of follow-up: US (OR: 1.6 [1.5–1.7]), CT (OR: 4.6 [4.5–4.7]), and MR (OR: 3.2 [3.1–3.3]). Although several predictors had relatively low effect size (OR close to 1), based on our a priori definition for the univariable model of significance at the p<0.2 level, they were all carried forward to the multivariable model.

Department Multivariable Model

The multilevel logistic regression procedure reported no convergence errors. In multivariable analysis (Table 2), older patients had higher rates of follow-up recommendations (OR: 1.01 [1.01–1.01] for each additional year), and male patients had lower rates of follow-up recommendations (OR: 0.9 [0.9–1.0]). Radiologist sex (p=0.54), presence of a trainee (p=0.45), and years in practice (p=0.49) were not significant predictors in the multivariable model. Among modalities, CT studies (OR: 4.2 [4.0–4.4]) had the highest odds of follow-up recommendation, followed by MRI (OR: 3.2 [3.1–3.4]) and US (OR: 1.2 [1.1–1.3]) in comparison to X-Ray. The overall model explained 7.8% of the variation in follow-up recommendations, while Unique Attending ID (a number given to each individual radiologists) accounted for 4.7% of the variation in our model (goodness of fit p-value <0.03).

Multivariable Analysis by Division

Across all subspecialty divisions, modality remained a significant predictor of follow-up recommendations throughout, but radiologist experience was not significant in any subspecialty division (Table 3). For all other variables, subspecialty divisions differed considerably in which factors were associated with a higher rate of follow-up recommendations. Older patients had higher odds of follow-up recommendations in all divisions (OR: 1.01 [1.00–1.01] for each additional year) except for MSK (OR: 1.00 [0.99–1.00]). Patient sex remained significant in the Abdomen, Chest and MSK divisions, with male patients having lower odds of follow-up recommendations (range of ORs, 0.8 to 0.9).

The attending radiologist was a significant predictor in all divisions except Chest. Probabilities of follow-up recommendations per attending in each division are shown in
Figure 2. Sub-plots for each division show that variation in attending probabilities was significant for all divisions except Chest, \((P\text{-values: 0.01–0.06})\). The difference between lowest and highest inter-radiologist probabilities in subspecialty divisions ranged from 2.8-fold to 6.7-fold. Radiologist sex was significant only for the Cancer Institute, where male radiologists had lower odds of follow-up \((\text{OR: 0.5 [0.4–0.8]})\). The presence of a trainee had higher odds of follow-up in the Chest \((\text{OR: 1.4 [1.3–1.4]})) and Neuroradiology divisions \((\text{OR: 1.3 [1.2–1.3]})\) but had lower odds in the MSK division \((\text{OR: 0.8 [0.7–0.9]})\).

Among modalities, CT retained the highest odds of follow-up across all divisions \((\text{range of ORs, 1.4 to 14.5, compared to x-ray})\), except for the Neuroradiology Division, in which MRI had the highest odds of follow-up \((\text{OR 1.8 [1.4 – 2.3], compared to x-ray})\), and in the Cancer Institute, in which x-ray had the highest odds of follow-up \((p<0.01)\). Care setting was significant in all divisions except for the Cancer Institute, with reports for inpatients having lower odds of follow-up across all divisions \((\text{range of ORs, 0.4 to 0.8})\) except Chest \((\text{OR: 1.6 [1.5–1.7]})\) and the Cancer Institute \((\text{OR: 1.6 [0.7–3.5]})\). Hosmer-Lemeshow \(p\)-values as follows: Abdomen \(p=0.3\), Chest \(p<0.01\), Cancer Institute \(p=0.19\), ED \(p<0.01\), MSK \(p=0.02\), Neuroradiology \(p<0.02\).

**Discussion**

Leveraging machine learning to identify radiology reports with follow-up recommendations, we found an overall follow-up recommendation rate of 12.2% among over 300,000 radiology reports. Multivariable analysis revealed patient age \((\text{OR: 1.01, } p<0.01)\) and sex \((\text{OR: 0.9, } p<0.01)\), and study modality \((p<0.01)\) predicted the presence of follow-up recommendations. Radiologist sex \((p=0.54)\), trainee involvement \((p=0.45)\), and years in practice \((p=0.49)\) were not significant predictors in multivariable analysis, while unique attending ID was a significant predictor \((p<0.01)\). We found significant inter-radiologist variation in subspecialty divisions, ranging from 2.8 to 6.7-fold. Modality was significant across all division levels, with other factors having varied impact in each division, highlighting considerations that are exclusive to subspecialized imaging. Individual radiologist factors, not including radiologist sex or years of experience, had the largest contribution to the observed variation. Our findings suggest that the same diagnostic finding may yield different follow-up recommendations depending on the radiologist reading the study, which could result in important implications for patient care.

Our overall follow-up recommendations rate is within those previously-reported of 8% to 37\%(1–6). Given the variation at the attending level, our findings suggest that other attending factors are more important than experience or sex and account for a considerable portion of the observed variance. For example, malpractice fears or risk intolerance could lead some radiologists to make increased follow-up recommendations. However, Giess et al. found that concerns about malpractice and individual risk tolerance did not increase screening mammography recall rates.(13) Similarly, Elmore et al. showed that medical malpractice experience or concerns were not associated with mammography recall or false-positive rates(16) Understanding the actual reasons that lead some radiologists to make follow-up recommendations more frequently requires further investigation preferably with large numbers of radiologists from multiple institutions. Our study also contrasts with
previous work showing that radiologists with greater years of experience made fewer follow-up recommendations (5). Several factors may explain this difference, including analysis of different institutions and time periods, and a larger number of reports analyzed in the prior study. However, both the prior study and our study were cross-sectional studies not optimally suited to evaluate the effect of radiologist experience on follow up recommendations. Specifically, a more useful analysis would assess rates of follow up recommendation in a large dataset with time as a variable.

Variable follow-up recommendation rates are an important concern in radiology. Specialty referrals in medicine often reflect established inter-provider relationships where providers are familiar with the practice style, habits and language of their consultants. However, diagnostic radiology referrals are distinct in that the referring provider does not typically select the interpreting radiologist. As clinical responsibilities reduce time for direct in-person consultation, non-standardized follow-up recommendations may contribute to ambiguity and suboptimal clinical decision-making by referring providers, compromise patient care, and raise costs (17,18).

Our study had limitations. First, while our approach to identifying follow-up recommendations has been validated, given our sample size and heterogeneity of radiology reports, results for some radiologists may be skewed due to wide variation in report terminologies. However, considering that our training set was a subset of the same studied population, it is likely this error was small. Also, given our large sample size of reports, several factors were statistically but not clinically significant. Our model accounted for a small amount (7.8%) of the existing variation and we did not control for within patient clustering. This is likely due to the majority of follow-up recommendations in radiology reports stemming from imaging findings and patient conditions as opposed to radiologist- or department-specific factors. Importantly, the purpose of our analysis was to understand the relative size of these different factors in influencing follow-up. We found that a given radiologist can have an important effect on follow-up imaging compared to peers, although imaging findings and clinical history are likely more important. Moreover, because radiologists’ shifts are assigned randomly, each radiologist in a subspecialty division is likely exposed to similar patients, reducing the likelihood of confounding in our analyses. We did not assess follow-up recommendation variation at the disease-level (e.g., renal cysts or incidental pulmonary nodules). The influence of patient, imaging, and radiologist variables on follow-up recommendations may differ by disease. We did not analyze what imaging modalities were recommended in follow-up recommendations; more advanced imaging may have greater cost and quality implications. Finally, our analysis could not identify whether follow-up recommendations were adhered to.

In conclusion, we used machine learning to analyze variation in follow-up recommendations in radiology reports, which we found attributable to both radiologist and non-radiologist factors. While radiologist sex, trainee involvement, and experience did not contribute to unwarranted variation in follow-up recommendations, there was substantial variation in follow-up recommendations between radiologists within the same division. Therefore, interventions to reduce unwarranted variation in follow-up recommendations may be most effective if targeted to individual radiologists. Interventions could include feedback reports.
showing follow-up recommendation rates for individual radiologists, educational efforts to improve awareness and acceptance of evidence-based imaging guidelines, and improved decision support tools. Future studies will be needed to assess the impact of multi-faceted interventions on reducing inter-radiologist variation in follow-up recommendations and the impact on quality of care.

Acknowledgements:

Study supported by Association of University Radiologists GE Radiology Research Academic Fellowship. Study supported by Agency for Healthcare Research and Quality (R01HS024722).

Abbreviations:

NLP natural language processing

References

1. Baumgarten DA, Nelson RC. Outcome of examinations self-referred as a result of spiral CT of the abdomen. Acad Radiol. 1997 12;4(12):802–5. [PubMed: 9412692]

2. Blaivas M, Lyon M. Frequency of radiology self-referral in abdominal computed tomographic scans and the implied cost. Am J Emerg Med. 2007 5;25(4):396–9. [PubMed: 17499656]

3. Furtado CD, Aguirre DA, Sirlin CB, Dang D, Stamato SK, Lee P, et al. Whole-body CT screening: spectrum of findings and recommendations in 1192 patients. Radiology. 2005 11;237(2):385–94. [PubMed: 16170016]

4. Shuaib W, Vijayasarthi A, Johnson J-O, Salastekar N, He Q, Maddu KK, et al. Factors affecting patient compliance in the acute setting: an analysis of 20,000 imaging reports. Emerg Radiol. 2014 8;21(4):373–9. [PubMed: 24615661]

5. Sistrom CL, Dreyer KJ, Dang PP, Weilburg JB, Boland GW, Rosenthal DL, et al. Recommendations for Additional Imaging in Radiology Reports: Multifactorial Analysis of 5.9 Million Examinations. Radiology. 2009 11 1;253(2):453–61. [PubMed: 19710005]

6. Levin DC, Edmiston RB, Ricci JA, Beam LM, Rosetti GF, Harford RJ. Self-referral in private offices for imaging studies performed in Pennsylvania Blue Shield subscribers during 1991. Radiology. 1993 11 1;189(2):371–5. [PubMed: 8210362]

7. Michaels AY, Chung CSW, Frost EP, Birdwell RL, Giess CS. Interobserver variability in upgraded and non-upgraded BI-RADS 3 lesions. Clin Radiol. 2017 8;72(8):694.e1–694.e6.

8. Lacson R, Prevedello LM, Andriole KP, Gill R, Lenoci-Edwards J, Roy C, et al. Factors associated with radiologists’ adherence to Fleischner Society guidelines for management of pulmonary nodules. J Am Coll Radiol JACR. 2012 7;9(7):468–73. [PubMed: 22748786]

9. Eisenberg RL, Bankier AA, Boiselle PM. Compliance with Fleischner Society guidelines for management of small lung nodules: a survey of 834 radiologists. Radiology. 2010 4;255(1):218–24. [PubMed: 20308458]

10. Bobbin MD, Ip IK, Sahni VA, Shinagare AB, Khorsanri R. Focal Cystic Pancreatic Lesion Follow-up Recommendations After Publication of ACR White Paper on Managing Incidental Findings. J Am Coll Radiol JACR. 2017 6;14(6):757–64. [PubMed: 28476609]

11. Black WC. Advances in radiology and the real versus apparent effects of early diagnosis. Eur J Radiol. 1998 5;27(2):116–22. [PubMed: 9639136]

12. Welch HG, Schwartz L, Woloshin S. Overdiagnosed: making people sick in the pursuit of health. Boston, Mass: Beacon Press; 2011 228 p.

13. Giess CS, Wang A, Ip IK, Lacson R, Pourjabbar S, Khorsanri R. Patient, Radiologist, and Examination Characteristics Affecting Screening Mammography Recall Rates in a Large Academic Practice. J Am Coll Radiol JACR. 2018 7 20;

14. Carrodeguas E, Lacson R, Swanson W, Khorsanri R. Use of machine learning to identify follow-up recommendations in radiology reports. J Am Coll Radiol JACR. 2018 12 29;
15. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA, et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA.: American College of Radiology; 2013.

16. Elmore JG, Taplin SH, Barlow WE, Cutter GR, D'Orsi CJ, Hendrick RE, et al. Does Litigation Influence Medical Practice? The Influence of Community Radiologists’ Medical Malpractice Perceptions and Experience on Screening Mammography. Radiology. 2005 7 1;236(1):37–46. [PubMed: 15987961]

17. Gunn AJ, Sahani DV, Bennett SE, Choy G. Recent Measures to Improve Radiology Reporting: Perspectives From Primary Care Physicians. J Am Coll Radiol. 2013 2 1;10(2):122–7. [PubMed: 23228373]

18. Fischer HW. Better Communication between the Referring Physician and the Radiologist. Radiology. 1983 3 1;146(3):845–845.
| Summary Statement: |
|--------------------|
| There is substantial unexplained variation of almost 7-fold between radiologists in the probability of recommending a follow-up exam in a radiology report, after adjusting for patient, exam and radiologist factors. |
Key Points:

- In multivariable analysis, older patients had higher rates of follow-up recommendations (OR: 1.01 [1.01–1.01] for each additional year), male patients had lower rates (OR: 0.9 [0.9–1.0]), and follow-up recommendations were most common among CT studies (OR: 4.2 [4.0–4.4] compared to X-ray).

- Radiologist sex (p=0.54), presence of a trainee (p=0.45), and years in practice (p=0.49) were not significant predictors of higher follow-up recommendation rates.

- The inter-radiologist variation in the likelihood of making a follow-up recommendation in a radiology report ranged from 2.8-fold to 6.7-fold depending on radiologists’ subspecialty division.
Figure 1: Cohort Flowchart Depicting Inclusion and Exclusion Criteria
Flowchart showing the number of studies performed in 2016, number of studies excluded, and number of studies included in analysis.
Figure 2: Follow-up Recommendation Probabilities per Attending Radiologist (n=65) in each Subspecialty Division.

Adjusted probability of a follow-up recommendation in percent (y-axis) for each radiologist in each division. P values were obtained from the division level model. Radiologists in each division are represented by a Unique Attending ID (x-axis). Figure shows the wide variation within each department, with up to a 6.7-fold difference between the radiologist with the lowest follow-up recommendation probability and the radiologist with the highest probability of making a follow-up recommendation.
### Table 1:

Univariable Analysis of Factors Affecting Follow-Up Recommendations by Department

| Predictor          | Entire Cohort of Radiology Reports | Reports Containing Follow-Up Recommendations | Odds Ratio (95% CI) | P-value |
|--------------------|-----------------------------------|---------------------------------------------|---------------------|---------|
| **Overall**        | 318,366                           | 38,745 (12.2%)                              |                     |         |
| **Patient Factors**|                                   |                                             |                     |         |
| Age, Mean Years (SD) | 59.0 (SD: 17.4)                  | 60.3 (SD: 17.4)                              | 1.01 (1.01–1.01)    | <0.01   |
| Women              | Reference                         |                                             |                     |         |
| Men                | 143,767 (45.2%)                  | 16,573 (42.8%)                              | 0.90 (0.9–1.0)      | <0.01   |
| **Radiologist Factors**|                                |                                             |                     |         |
| Women              | Reference                         |                                             |                     |         |
| Men                | 145,085 (45.6%)                  | 19,246 (49.7%)                              | 1.2 (0.8–1.6)       | <0.01   |
| Trainee, Absent    | Reference                         |                                             |                     |         |
| Trainee, Present   | 152,749 (48.0%)                  | 23,168 (59.8%)                              | 1.7 (1.7–1.8)       | <0.01   |
| Unique Attendings  | 65                                | 65                                          | 1.00                | <0.01   |
| **Years in Practice**|                                |                                             |                     |         |
| 1–10               | Reference                         |                                             |                     |         |
| 11+                | 227,834 (71.6%)                  | 27,071 (69.9%)                              | 0.91 (0.9–0.9)      | <0.01   |
| **Division**       |                                   |                                             |                     |         |
| Neuroradiology     | Reference                         |                                             |                     |         |
| Abdomen            | 23,535 (7.4%)                    | 4,917 (12.7%)                               | 1.3 (1.3–1.4)       | <0.01   |
| Chest              | 84,354 (26.5%)                   | 5,078 (13.1%)                               | 0.3 (0.3–0.3)       | <0.01   |
| Cancer Institute   | 36,561 (11.5%)                   | 6,712 (17.3%)                               | 1.2 (1.1–1.2)       | <0.01   |
| Emergency Radiology| 76,339 (24.0%)                   | 11,084 (28.6%)                              | 0.9 (0.8–0.9)       | <0.01   |
| Musculoskeletal    | 57,793 (18.2%)                   | 4,461 (11.5%)                               | 0.4 (0.4–0.5)       | <0.01   |
| **Modality**       |                                   |                                             |                     |         |
| Predictor | Entire Cohort of Radiology Reports | Reports Containing Follow-Up Recommendations | Odds Ratio (95% CI) | P-value |
|-----------|-----------------------------------|---------------------------------------------|-------------------|---------|
| X-Ray     | Reference                         | 8037 (2.5%)                                 | 1.6 (1.5–1.7)     | <0.01 * |
| Ultrasound| 700 (1.8%)                        | 100,050 (31.4%)                             | 4.6 (4.5–4.7)     | <0.01 * |
| CT        | 21,650 (55.9%)                    | 42,959 (13.5%)                              | 3.2 (3.1–3.3)     | <0.01 * |
| MR        | 6,913 (17.8%)                     |                                             |                   |         |

* Statistically significant; CI = Confidence Interval; SD = Standard Deviation
Table 2:
Multivariable Analysis of Factors Affecting Follow-Up Recommendations by Department

| Predictor                  | Odds Ratio (95% CI) | P-value |
|----------------------------|---------------------|---------|
| **Patient Demographic Factors** |                     |         |
| Age, Years                 | 1.01 (1.01–1.01)    | <0.01   |
| Women                      | Reference           |         |
| Men                        | 0.9 (0.9–1.0)       | <0.01   |
| **Radiologist Demographic Factors** |                 |         |
| Women                      | Reference           |         |
| Men                        | 0.9 (0.7–1.2)       | 0.54    |
| Trainee, Absent            | Reference           |         |
| Trainee, Present           | 1.0 (1.0–1.0)       | 0.45    |
| Unique Attending ID        | <0.01               |         |
| **Years in Practice**      |                     |         |
| 1–10                       | Reference           |         |
| 11+                        | 0.9 (0.6–1.3)       |         |
| **Division**               |                     | <0.05   |
| Neuroradiology             | Reference           |         |
| Abdomen                    | 1.7 (1.0–2.8)       |         |
| Chest                      | 0.8 (0.4–1.4)       |         |
| Cancer Institute           | 1.6 (0.9–2.8)       |         |
| Emergency Radiology        | 1.5 (0.9–2.5)       |         |
| Musculoskeletal            | 1.2 (0.7–2.1)       |         |
| **Modality**               |                     | <0.01   |
| X-Ray                      | Reference           |         |
| Ultrasound                 | 1.2 (1.1–1.3)       |         |
| CT                         | 4.2 (4.0–4.4)       |         |
| MR                         | 3.2 (3.1–3.4)       |         |
The overall model explained 7.8% of the variation in follow-up recommendations, while Unique Attending ID (a number given to each individual radiologists) accounted for 4.7% of the variation in our model (goodness of fit p-value <0.03).
Table 3: Multivariable Analysis by Radiology Division

| Predictor                  | Abdomen Odds Ratio (95% CI) | Abdomen P-value | Chest Odds Ratio (95% CI) | Chest P-value | Cancer Institute Odds Ratio (95% CI) | Cancer Institute P-value |
|----------------------------|-----------------------------|-----------------|---------------------------|---------------|-------------------------------------|--------------------------|
| Patient Age, Years         | 1.01 (1.00–1.01)            | <0.01 *         | 1.01 (1.00–1.01)          | <0.01 *       | 1.01 (1.00–1.01)                    | <0.01 *                  |
| Patient Sex, Women         | Reference                   |                 | Reference                 |               | Reference                            |                          |
| Patient Sex, Men           | 0.9 (0.9–1.0)               | 0.02 *          | 0.8 (0.8–0.9)             | <0.01 *       | 1.0 (0.9–1.0)                       | 0.09                     |
| Radiologist Sex, Women     | Reference                   |                 | Reference                 |               | Reference                            |                          |
| Radiologist Sex, Men       | 1.3 (0.8–2.3)               | 0.30            | 1.1 (0.4–2.7)             | 0.87          | 0.5 (0.4–0.8)                       | <0.01 *                  |
| Trainee, Absent            | Reference                   |                 | Reference                 |               | Reference                            |                          |
| Trainee, Present           | 0.9 (0.9–1.0)               | 0.06            | 1.4 (1.3–1.4)             | <0.01 *       | 1.1 (1.0–1.2)                       | 0.13                     |
| Unique Attending ID        | 0.01 *                      |                 | 0.06                      |               | 0.02 *                              |                          |
| Years in Practice          |                             |                 |                           |               |                                     |                          |
| 1–10                       | Reference                   | <0.5            | Reference                 | 0.99          | Reference                            | 0.21                     |
| 11+                        | 0.8 (0.5–1.4)               |                 | 1.0 (0.3–3.9)             |               | 1.6 (0.8–3.5)                       |                          |
| Modality                   |                             |                 |                           |               |                                     |                          |
| X-Ray                      | Reference                   |                 | Reference                 |               | Reference                            |                          |
| MR                         | 4.2 (3.7–4.8)               |                 | 5.3 (3.2–8.5)             |               | 0.7 (0.6–0.8)                       |                          |
| CT                         | 4.9 (4.4–5.6)               | <0.01 *         | 14.5 (13.4–15.7)          | <0.01 *       | 0.8 (0.7–0.9)                       | <0.01 *                  |
| Ultrasound                 | 3.3 (2.7–4.0)               |                 | 0.9 (0.7–1.0)             |               |                                     |                          |
| Care Setting                |                             |                 |                           |               |                                     |                          |
| Outpatient                 | Reference                   | <0.01 *         | Reference                 |               | Reference                            |                          |
| Inpatient                  | 0.8 (0.74–0.9)              |                 | 1.61 (1.5–1.74)           | <0.01 *       | 1.61 (0.73–3.52)                     | 0.24                     |
| Predictor                  |                             |                 |                           |               |                                     |                          |
| Emergency Radiology        |                             |                 |                           |               |                                     |                          |
| Musculoskeletal            |                             |                 |                           |               |                                     |                          |
| Neuroradiology             |                             |                 |                           |               |                                     |                          |
| Predictor                        | Abdomen | Chest | Cancer Institute |
|---------------------------------|---------|-------|------------------|
|                                 | Odds Ratio (95% CI) | P-value | Odds Ratio (95% CI) | P-value | Odds Ratio (95% CI) | P-value |
| Patient Age, Years              | 1.01 (1.00–1.01) | <0.01* | 1.00 (0.99–1.00) | <0.01* | 1.01 (1.00–1.01) | <0.01* |
| Patient Sex, Women Sex          | Reference | Reference | Reference |       | Reference | Reference |
| Patient Sex, Men Sex            | 1.0 (1.0–1.0) | 0.81  | 0.9 (0.9–1.0) | <0.01* | 1.0 (0.9–1.1) | 0.63   |
| Radiologist Sex, Women Sex      | Reference | Reference | Reference |       | Reference | Reference |
| Radiologist Sex, Men Sex        | 0.59 (0.3–1.2) | 0.15  | 1.2 (0.5–3.2) | 0.71  | 1.9 (0.4–8.6) | 0.43   |
| Trainee, Absent                 | Reference | Reference | Reference |       | Reference | Reference |
| Trainee, Present                | 1.03 (1.0–1.1) | 0.31  | 0.8 (0.7–0.9) | <0.01* | 1.3 (1.2–1.3) | <0.01* |
| Unique Attending ID             | 0.02* | 0.03* | <0.01* |
| Years in Practice               |         |       |                   |
| 1–10                            | Reference | 0.14  | Reference | 0.93  | Reference | 0.66   |
| 11+                             | 0.6 (0.3–1.2) |       | 1.0 (0.4–2.6) |       | 1.3 (0.4–4.1) |       |
| Modality                        |         |       |                   |
| X-Ray                           | Reference | Reference | Reference |       | Reference | Reference |
| MR                              | 1.9 (1.6–2.3) |       | 2.1 (1.9–2.3) |       | 1.8 (1.4–2.3) |       |
| CT                              | 3.7 (3.5–3.8) | <0.01* | 2.5 (2.2–2.8) | <0.01* | 1.4 (1.1–1.9) | <0.01* |
| Ultrasound                      | 0.5 (0.4–0.5) |       | 1.0 (0.6–1.7) |       |       |       |
| Care Setting                    |         |       |                   |
| Outpatient                      | Reference | Reference | Reference |       | Reference | Reference |
| Inpatient                       | 0.4 (0.4–0.5) | <0.01* | 0.7 (0.7–0.7) | <0.01* |       |       |

*Statistically Significant; CI=Confidence Interval