

Recommendations for Additional Imaging in Radiology Reports: Multifactorial Analysis of 5.9 Million Examinations¹

Christopher L. Siström, MD, MPH
Keith J. Dreyer, DO, PhD
Pragya P. Dang, MD
Jeffrey B. Weilburg, MD
Giles W. Boland, MD
Daniel I. Rosenthal, MD
James H. Thrall, MD

Purpose:

To quantify the rates of recommendation for additional imaging (RAI) in a large number of radiology reports of different modalities and to estimate the effects of 11 clinically relevant factors.

Materials and Methods:

This HIPAA compliant research was approved by the institutional review board under an expedited protocol for analyzing anonymous aggregated radiology data. All diagnostic imaging examinations ($n = 5\,948\,342$) interpreted by radiologists between 1995 and 2008 were studied. A natural language processing technique specifically designed to extract information about any recommendations from radiology report texts was used. The analytic data set included three quantitative variables: the interpreting radiologist's experience, the year of study, and patient age. Categorical variables described patient location (inpatient, outpatient, emergency department), whether a resident dictated the case, patient sex, modality, body area studied, ordering service, radiologist's specialty division, and whether the examination result was positive. A multivariable logistic regression model was used to determine the effect of each of these factors on likelihood of RAI while holding all others equal.

Results:

Recommendations increased during the 13 years of study, with the unadjusted rate rising from roughly 6% to 12%. After accounting for all other factors, the odds of any one examination resulting in an RAI increased by 2.16 times (95% confidence interval: 2.12, 2.21) from 1995 to 2008. As radiologist experience increased, the odds of an RAI decreased by about 15% per decade. Studies that had positive findings were more likely (odds ratio = 5.03; 95% confidence interval: 4.98, 5.07) to have an RAI. The remaining factors also had significant effects on the tendency for an RAI.

Conclusion:

The likelihood of RAI increased by 15% for each decade of radiologist experience and roughly doubled over 13 years of study.

© RSNA, 2009

¹ From the Departments of Radiology (C.L.S., K.J.D., P.P.D., G.W.B., D.I.R., J.H.T.) and Psychiatry (J.B.W.), Massachusetts General Hospital, Boston, Mass. Received February 3, 2009; revision requested March 20; revision received March 26; accepted May 6; final version accepted May 11. Address correspondence to C.L.S., Department of Radiology, University of Florida Health Center, PO Box 100374, Gainesville, FL 32610-0374 (e-mail: sistr@radiology.ufl.edu).

© RSNA, 2009

Radiologists sometimes make recommendations for further imaging tests in their interpretative reports. These include requests for correlation with a different modality to help explain indeterminate findings on the original study or a recommendation that imaging be repeated with the same modality at some time interval for follow-up to evaluate stability, worsening, or resolution of imaging findings. Often the purpose of additional imaging is to reduce uncertainty about equivocal findings at the current examination. Even reports of studies with negative findings may contain recommendations for evaluation with more sensitive modalities or delayed follow-up with the same modality to help detect developing or occult disease. This tendency has been characterized as being

problematic, because it may contribute to increased utilization and cost (1–3).

Baumgarten and Nelson (4) evaluated 545 consecutive abdominal computed tomographic (CT) scans and found radiologist recommendations for additional imaging (RAIs) in 105 (19.3%). Blaivas and Lyon (5) looked at 785 abdominal CT reports, and there were 246 (31%) with at least one RAI. In these two frequently cited studies, RAIs were characterized as radiologist “self-referral” (4,5). There are few other articles detailing the rate and type of RAI in clinical practice. Riddell and Khalili (6) reported on cross-correlation between CT and ultrasonography (US) for patients with acute abdominal pain. During a single year, they found that 8.2% of patients undergoing US underwent subsequent CT scans (6). Of patients undergoing CT first, 4.6% went on to undergo US (6). The largest body of related literature describes incidental findings during screening colonography with CT, and this was recently reviewed by Siddiki et al (7). The rate of highly important incidental findings ranged from 5% to 25%. Though not specifically enumerated, some of these would prompt an RAI (8). At whole-body screening CT, 37% of patients received at least one recommendation for further imaging, which mostly involved the lungs and kidneys (8). Two other major clinical areas in which RAIs have been important for quite some time are chest CT (lung nodule screening and/or follow-up) and breast imaging. We will not attempt to review these complex and controversial subjects here. However, the sample we studied did include imaging of the chest and breast with multiple modalities, and our estimates of

rates and trends in RAI may be informative.

Our group has previously reported on automatic detection and characterization of recommendations contained in radiology interpretations by using a clinical data warehouse supplemented with natural language processing (NLP) of report texts (9–14). These articles described and validated the NLP technique for extracting recommendations (9–11); reported on types, rates, and trends of recommendations (11,12,14); and specifically tested the contribution of examinations induced by recommendations for repeat studies to selected high-cost procedure volumes (13). This study complements and extends these efforts by increasing the sample to include all diagnostic imaging reported by our radiologists since 1995, with multivariable logistic regression to adjust for interaction between factors, and evaluates the effect of radiologist experience and patient age on RAI rates over time.

The purpose of this study was to quantify the rates of RAI in a large number of radiology reports of multiple modalities and to estimate the effects of 11 clinically relevant factors, including change over time, age of patients, and experience of interpreting radiologists.

Advances in Knowledge

- The likelihood for a report to contain recommendations for further imaging more than doubled from 1995 to 2008 with an odds ratio of 2.16 (95% confidence interval: 2.11, 2.19); this was estimated from logistic regression which holds all other factors equal and was therefore not confounded by radiologist experience, types of study, or other factors.
- Radiologists with increasing levels of experience made fewer recommendations for additional imaging after accounting for all other factors; the size of this effect was estimated to be about a 15% decrease in the odds of recommendation per decade of experience.
- Examinations with positive findings were more likely to contain recommendations for further imaging with an odds ratio of 5.03 (95% confidence interval: 4.98, 5.07) compared with examinations with negative findings; when patients being examined came from the emergency department or were outpatients, the odds of recommendation were roughly 1.5 times higher than for inpatients.

Implication for Patient Care

- This study shows that recommendations for additional imaging are made with different frequency depending on at least 11 relevant factors; any effort to measure and remediate variation among radiologists in recommendation rates must take these factors into account by using case-mix adjustment.

Materials and Methods

This Health Insurance Portability and Accountability Act–compliant research

Published online before print

10.1148/radiol.2532090200

Radiology 2009; 253:453–461

Abbreviations:

NLP = natural language processing

RAI = recommendation for additional imaging

Author contributions:

Guarantors of integrity of entire study, C.L.S., K.J.D.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, C.L.S., K.J.D., P.P.D., G.W.B., D.I.R.; clinical studies, P.P.D., G.W.B., D.I.R., J.H.T.; statistical analysis, C.L.S.; and manuscript editing, C.L.S., K.J.D., P.P.D., J.B.W., G.W.B., J.H.T.

Authors stated no financial relationship to disclose.

See also the editorial by Arensen in this issue.

was approved by the institutional review board under an expedited protocol for analyzing anonymous aggregated radiology data. We queried an existing data warehouse for all diagnostic imaging examinations interpreted by radiologists at our institution from January 1, 1995 through December 31, 2008. The query was designed and validated to return a single row of data for every unique study instance, with no duplication due to billing codes, addendums, or multiple accession numbers. Information concerning each examination included the following: date performed, patient sex, patient age, ordering service, patient location (inpatient, outpatient, emergency department), specialty division of the signing radiologist, medical school graduation date of the signing radiologist, imaging modality, and body region studied. We converted the date variables (study performed and graduation date of radiologist) into years and created a new variable (radiologist experience) by subtraction that quantified the number of postgraduate years at the time the radiologist interpreted the examination.

We obtained information concerning recommendations contained in the interpretative report text by using a previously described and validated NLP system (9–11). This algorithm creates a separate table of recommendation assertions linked to each examination by accession number detailing the class of recommendation (imaging or clinical), the specific type (eg, CT, magnetic resonance [MR] imaging, US, endoscopy, surgery, biopsy), and the suggested time interval for the recommendation (in days, with null for not specified). We created a binary variable for imaging recommendations from each study and set it to “no” when there were no imaging recommendations in the report and to “yes” when there were one or more imaging recommendations (ie, an RAI). We specifically excluded generic statements such as “clinical correlation,” as well as those calling for correlation with surgery, biopsy, or endoscopy and/or colonoscopy. The same NLP system also produces an output for each report that codes any clinically important findings

and returns “negative” when none are detected in the text (9,10,14). This was used to create a “positive findings” variable for subsequent analysis with values of yes and no.

We performed cross-tabulation on all factors (eg, patient sex, patient location, modality, body area) to determine the number and percentage of total examinations represented by each factor level. The binary assertion of whether or not at least one specific imaging recommendation was detected in the report by the NLP algorithm was used as the dependent variable. For each of the factor levels, the number and percentage of examinations with at least one RAI were enumerated. We then used binary logistic regression to analyze the relationship between all factors and the likelihood of RAI. This was done by using standard software (SAS PROC LOGISTIC, version 9.31; SAS, Cary, NC), with the dependent variable being RAI (one or more = 1 and none = 0). Patient age (decade groups), radiologist experience (5-year groups), and year of study (1995–2008) were coded as class rather than numeric variables. For these variables (patient age, year of examination, and radiologist experience), we set the reference level to the lowest value. This allowed us to estimate a separate odds ratio for each successive level (eg, 1996 vs 1995, 1997 vs 1995, 1998 vs 1995, and so on). Otherwise, the reference levels were adjusted to ensure that the odds ratios of all other factor levels would be greater than 1 so that relationships between them would be represented consistently. All of the independent variables are listed as follows, along with reference values: patient age grouped in decades (reference = 0–9 years), patient location (reference = inpatient), patient sex (reference = male), ordering service (reference = anesthesia), modality (reference = angiography), body region (reference = heart), positive findings (reference = no), radiologist division (reference = nuclear medicine), radiologist experience (reference = 5–10 years), year of examination (reference = 1995), and resident dictated (reference = no).

Using a fully specified logistic model allowed us to report the effect size (odds ratios) and significance (confidence intervals) of each independent variable on the likelihood of RAI free of confounding by the other 10 variables. This is especially important for making inferences about the effects of patient age, radiologist experience, and trend over time (year of study).

Results for patient sex, patient location, resident dictated, positive findings, modality, body area, ordering service, and radiologist division were tabulated, while those for year of examination, patient age, and radiologist experience were plotted. In both the tabular and graphic formats, we included the raw percentage of examinations with RAI along with the odds ratio and 95% Wald confidence interval for each factor level. For selected combinations of modality, body area, ordering service, and interpreting radiologist division, we performed stratified analysis to explore specific relationships among subsets of the variable levels. These are described briefly in the Results section where needed.

Results

The query returned 5 948 342 examinations, of which 627 064 (10.54%) had at least one RAI in the dictated report detected by the NLP algorithm. A total of 555 radiologists interpreted these studies, though 229 of them performed less than 1000 each. In aggregate, there were 51 961 (<1.0%) examinations dictated by these low-volume readers. There were 909 111 (15.3%) studies in the sample that were dictated by 193 radiologists who performed between 1000 and 9999 studies. Thus, 133 radiologists interpreted 10 000 or more studies each, and they accounted for the remaining 4 987 270 (83.8%) of all dictations evaluated. There were 810 042 patients; 418 611 (51.68%) were female patients, and 391 368 (48.31%) were male patients (63 patients had *undefined* or *null* recorded in the sex field). At the time of their examinations, the average age of male patients was

Table 1

Imaging Recommendation Details Found by Using the NLP Algorithm

Study Modality	Recommended Modality for Follow-up										Total
	Angiography	CT	Mammography	MR Imaging	Nuclear Medicine	PET	Radiography	US	Current Modality	Other	
Angiography	516 (24.8)	428 (20.6)	12 (0.6)	174 (8.4)	48 (2.3)	13 (0.6)	143 (6.9)	108 (5.2)	395 (19.0)	243 (11.7)	2080 (100)
CT	1237 (0.5)	67 172 (29.1)	2406 (1.0)	52 155 (22.6)	8849 (3.8)	5176 (2.2)	3501 (1.5)	28 034 (12.1)	48 662 (21.0)	13 992 (6.1)	231 184 (100)
Mammography	1 (~0)	140 (0.2)	22 539 (26.1)	1439 (1.7)	9 (~0)	5 (~0)	2361 (2.7)	4230 (4.9)	24 160 (28.0)	31 425 (36.4)	86 309 (100)
MR imaging	757 (1.2)	15 585 (24.7)	1437 (2.3)	21 753 (34.5)	1547 (2.5)	433 (0.7)	1701 (2.7)	4622 (7.3)	9062 (14.4)	6078 (9.7)	62 975 (100)
Nuclear medicine	89 (0.8)	2892 (25.4)	46 (0.4)	1457 (12.8)	522 (4.6)	475 (4.2)	3750 (32.9)	474 (4.2)	590 (5.2)	1110 (9.7)	11 405 (100)
PET	3 (~0)	4595 (35.5)	270 (2.1)	1319 (10.2)	260 (2.0)	1599 (12.4)	85 (0.7)	1158 (9.0)	3070 (23.7)	569 (4.4)	12 928 (100)
Radiography	88 (~0)	74 675 (32.3)	441 (0.2)	30 006 (13.0)	6661 (2.9)	122 (0.1)	51 009 (22.1)	2615 (1.1)	39 115 (16.9)	26 152 (11.3)	230 884 (100)
US	139 (0.2)	10 186 (14.2)	5038 (7.0)	13 611 (18.9)	1073 (1.5)	32 (~0)	1040 (1.4)	18 501 (25.8)	13 970 (19.4)	8240 (11.5)	71 830 (100)
Total	2830 (0.4)	175 673 (24.8)	32 189 (4.5)	121 914 (17.2)	18 969 (2.7)	7855 (1.1)	63 590 (9.0)	59 742 (8.4)	139 024 (19.6)	87 809 (12.4)	709 595 (100)

Note.—Data are numbers of RAIs, with percentages in parentheses. PET = positron emission tomography.

51.7 years ± 22.0 (standard deviation), and the average age of female patients was 54.1 years ± 20.8.

Table 1 details all of the RAIs found by using the NLP algorithm stratified according to study modality and modality of the recommended follow-up imaging. The total ($n = 709\ 595$) exceeds the number of examinations with at least one RAI ($n = 627\ 064$) because some examinations ($n = 73\ 755$) had two or more such recommendations detected. The row percentages (in parentheses) in Table 1 represent the fraction of the recommended modality out of all RAIs (rightmost total column) found in reports from the original modality of the study being interpreted. There are two ways that the algorithm detected recommendation for the same modality as that of the study being reported. This could be either by direct mention of the recommended modality by name or wording taken to mean repeat the modality of the examination currently being reported. For example, in CT reports, RAI with CT occurred 49.1% of the time (explicit mention of CT was 29.1% plus repeat of current modality was 21.0%). By totaling the columns, we get an idea of the overall percentage that various modalities were recommended. For example, additional CT studies were recommended 24.8% of the time, followed by current modality (19.6%) and MR imaging (17.2%). Considering the times that the current modality was itself CT or MR imaging would increase the overall proportion where RAI specified CT or MR imaging was to about half.

The logistic regression procedure reported no convergence errors. The c statistic was 0.76, and the rescaled R^2 was 0.17. The 11 modeled factors (independent variables) were all significant (type 3 χ^2 probability <.0001). The Wald 95% confidence intervals for the 28 odds ratio estimates all excluded 1, which indicated that the differences in recommendation rates that we observed across factors were all simultaneously significant at the 5% level.

We have chosen to list or display the unadjusted rates (as percentages) of examinations with an RAI in the tables and

figures. For each variable, we gave the associated odds ratios and will refer to these when discussing the significance of differences between levels. With the odds ratios, confounding by all other factors has been corrected for by virtue of their inclusion in the multivariable logistic model. The fact that our observed RAI rates were generally consistent with the adjusted odds ratios lends support to the directionality and size of observed effects. Exceptions to this general concordance between unadjusted RAI rates and modeled odds ratios will be addressed individually.

Table 2 demonstrates results stratified according to patient sex, patient location, resident dictated, and positive examination findings. Studies performed in female patients had a slightly higher rate of RAI (odds ratio = 1.08 compared with that of male patients). Patients from the emergency department had the highest rate of RAI (odds ratio = 1.60 compared with that of inpatients), closely followed by outpatients (odds ratio = 1.57 compared with that of inpatients). When residents dictated the report, an RAI was more likely (odds ratio = 1.29) than when the attending radiologists dictated the report themselves. The reports contained clinically important findings 66.89% of the time, and these examinations with positive findings were much more likely (odds ratio = 5.03) to also contain an RAI compared with studies with negative findings where no important findings were identified by the NLP algorithm.

Tables 3 and 4 list RAI according to study modality and body area examined, respectively. Diagnostic angiography had the fewest RAIs (3.8%) and was set as the reference category for the odds ratios obtained from logistic regression. The modalities with the most RAIs were PET (with odds ratio = 4.28 compared with that of angiography) followed by CT (odds ratio = 3.96 compared with that of angiography). Studies in the heart had the lowest percentage of RAI (2.4%), and that body area was set as the reference category for odds ratios. Breast imaging had the greatest likelihood of RAI (odds ratio =

Table 2

Imaging Recommendation Rates according to Four Variables

Variable	No. of Examinations	Percentage of Total Examinations	RAI*	Odds Ratio [†]
Patient sex				
Male	2 757 604	46.36	259 065 (9.39)	1.00 (reference)
Female	3 190 640	53.64	367 995 (11.53)	1.08 (1.07, 1.09)
Patient location				
Inpatient	1 543 285	25.94	120 319 (7.80)	1.00 (reference)
Outpatient	3 388 259	56.96	382 588 (11.29)	1.57 (1.55, 1.58)
Emergency department	1 016 798	17.09	124 157 (12.21)	1.60 (1.58, 1.62)
Resident dictated				
No	3 100 218	52.12	275 970 (8.90)	1.00 (reference)
Yes	2 848 124	47.88	351 094 (12.83)	1.29 (1.27, 1.30)
Positive findings				
No	1 960 843	33.11	76 493 (3.90)	1.00 (reference)
Yes	3 961 076	66.89	550 243 (13.89)	5.03 (4.98, 5.07)

* At least one RAI detected. Percentages of examinations with at least one RAI are in parentheses.

[†] Odds ratios, with 95% Wald confidence intervals in parentheses, are from logistic regression holding all other factors equal.

Table 3

Imaging Recommendation Rates according to Study Modality

Modality	No. of Examinations	Percentage of Total Examinations	RAI*	Odds Ratio [†]
Angiography	51 189	0.86	1960 (3.8)	1.00 (reference)
CT	923 956	15.53	197 496 (21.4)	3.96 (3.76, 4.16)
Mammography	449 751	7.56	66 596 (14.8)	1.60 (1.51, 1.69)
MR imaging	407 662	6.85	55 838 (13.7)	2.23 (2.11, 2.35)
Nuclear medicine	195 927	3.29	10 622 (5.4)	3.32 (3.08, 3.58)
PET	43 917	0.74	10 421 (23.7)	4.28 (4.04, 4.53)
Radiography	3 396 441	57.10	219 326 (6.5)	1.25 (1.19, 1.32)
US	479 499	8.06	64 805 (13.5)	2.40 (2.28, 2.52)

* At least one RAI detected. Percentages of examinations with at least one RAI are in parentheses.

[†] Odds ratios, with 95% Wald confidence intervals in parentheses, are from logistic regression holding all other factors equal.

12.38 compared with that of cardiac imaging) followed by imaging of the pelvis (odds ratio = 5.78 compared with that of cardiac imaging).

Examinations of the chest had the third highest level of RAI (10.0%) for any one body region (odds ratio = 4.76 compared with that of cardiac imaging). Because we might have expected higher rates of RAIs in the reports of thoracic studies due to follow-up of lung nodules, we tabulated recommendation rates according to modality with the body area limited to chest. CT comprised 11.6% of chest imaging with an RAI rate of

35.8%. The majority (86.9%) of chest imaging was performed with radiography, and for these, the rate was 6.5%. In absolute terms, 56.3% of all RAIs arising from thoracic studies were found in chest radiograph reports.

The relationship between mammography as an examination modality and breast body region was complex. Mammography appeared to demonstrate a moderate RAI likelihood (odds ratio = 1.60 compared with that of angiography). On the other hand, breast as a body area had the greatest likelihood of RAI (odds ratio = 12.38 compared with

that of cardiac imaging). To examine this relationship further, we cross-tabulated RAI rates according to modality limited to the body area of breast. We found that 87% of all breast examinations were performed with mammography, and these had an RAI rate of 14.8% (as in Table 3). MR imaging and US accounted for virtually all of the remaining breast examinations and demon-

strated much higher RAI rates (38% and 36%, respectively). Many of these were for mammography.

Table 5 gives results stratified according to ordering service. The lowest rate of RAI was found in examinations ordered by anesthesiologists (4.4%), and this was set as the reference for odds ratios. The anesthesia service requested examinations that were pre-

dominantly chest radiography (86.9%), and these demonstrated the same rate (6.5%) as the aggregate. Psychiatry was notable in being responsible for relatively few examinations overall (0.34% of total), yet having the highest rate of RAI (odds ratio = 2.89 compared with that of anesthesia). These examinations were mostly head (23.0%) or chest (44.3%), each with relatively high RAI rates (14.2% and 9.3%, respectively).

Table 6 reports the results stratified according to specialty (division) of the interpreting radiologist. Pediatric and vascular division readers demonstrated the lowest RAI rates (5.7% and 5.8%, respectively). However, during the regression including all other factors, the nuclear medicine division (RAI rate = 6.0%) had the lowest logistic coefficient, and it was set to reference (odds ratio = 1.0). The cardiac division made the most RAIs (19.5%, with odds ratio = 4.19 compared with that of nuclear medicine). The seeming discrepancy with the cardiac body area demonstrating the lowest RAI rate was resolved by noting that most (87.8%) of the cardiac studies were nuclear medicine myocardial perfusion examinations (read by the nuclear medicine division) and had a very low rate (0.65%) rate of RAI.

Figure 1 depicts the trend in RAI rate during the study period. Though there is slight divergence, the curves for RAI rates and odds ratios generally overlap quite convincingly. This allowed us to conclude that the RAI rate did increase over time, holding all other factors equal. Between the start (1995) of the study and its end (2008), the odds of an examination having at least one RAI increased by 2.16 times (95% confidence interval: 2.12, 2.21). This increase was independent of all other factors, including the mixture of modalities and body areas as well as changing distribution of relative experience levels of the radiologists reading the examinations. Furthermore, it seems that the majority of the increase took place in the late 1990s with another small but sustained increase between 2004 and 2008.

One related factor that was positively correlated with the likelihood of RAI was whether there were clinically important abnormalities described

Table 4

Imaging Recommendation Rates according to Body Area Examined

Body Area Examined	No. of Examinations	Percentage of Total Examinations	RAI*	Odds Ratio†
Heart	109 800	1.85	2656 (2.4)	1.00 (reference)
Abdomen	777 375	13.07	87 794 (11.3)	3.21 (3.05, 3.36)
Breast	502 663	8.45	83 333 (16.6)	12.38 (11.74, 13.06)
Chest	2 034 520	34.20	204 241 (10.0)	4.76 (4.54, 5.00)
Extremity	1 133 290	19.05	59 947 (5.3)	2.85 (2.72, 2.99)
Brain	430 467	7.24	64 433 (15.0)	3.15 (3.00, 3.31)
Head and neck (ears, nose, and throat)	115 562	1.94	22 397 (19.4)	4.29 (4.08, 4.52)
Pelvis	299 951	5.04	45 865 (15.3)	5.78 (5.50, 6.07)
Spine	401 158	6.74	41 089 (10.2)	3.65 (3.47, 3.83)
Other‡	143 556	2.41	15 309 (10.7)	5.71 (5.45, 5.98)

* At least one RAI detected. Percentages of examinations with at least one RAI are in parentheses.

† Odds ratios, with 95% Wald confidence intervals in parentheses, are from logistic regression holding all other factors equal.

‡ Other includes dental, bone, whole body, and unspecified.

Table 5

Imaging Recommendations according to Ordering Service

Ordering Service	No. of Examinations	Percentage of Total Examinations	RAI*	Odds Ratio†
Anesthesia	95 321	1.60	4225 (4.4)	1.00 (reference)
Emergency	357 158	6.00	44 759 (12.5)	2.38 (2.30, 2.46)
Medicine	2 732 846	45.94	335 765 (12.3)	2.49 (2.41, 2.57)
Neurology	162 901	2.74	19 421 (11.9)	2.30 (2.22, 2.38)
Neurosurgery	135 375	2.28	11 463 (8.5)	1.68 (1.62, 1.75)
Obstetrics and gynecology	167 614	2.82	23 920 (14.3)	2.36 (2.28, 2.45)
Orthopedics	628 366	10.56	30 911 (4.9)	1.59 (1.53, 1.64)
Pediatrics	196 287	3.30	12 587 (6.4)	2.45 (2.36, 2.55)
Psychiatry	20 224	0.34	2325 (11.5)	2.89 (2.73, 3.05)
Radiation oncology	28 720	0.48	4519 (15.7)	2.35 (2.24, 2.46)
Radiology	33 440	0.56	2047 (6.1)	1.34 (1.26, 1.42)
Surgery	1 137 826	19.13	109 159 (9.6)	2.07 (2.00, 2.14)
Urology	93 246	1.57	8837 (9.5)	2.06 (1.98, 2.14)
Other‡	159 018	2.67	17 126 (10.8)	2.42 (2.34, 2.51)

* At least one RAI detected. Percentages of examinations with at least one RAI are in parentheses.

† Odds ratios, with 95% Wald confidence intervals in parentheses, are from logistic regression holding all other factors equal.

‡ Other includes oral surgery, pathology, physical medicine, and unspecified.

Table 6

Imaging Recommendations according to Division of Interpreting Physician

Radiology Division	No. of Examinations	Percentage of Total Examinations	RAI*	Odds Ratio†
Nuclear medicine	205 648	3.46	12 305 (6.0)	1.00 (reference)
Musculoskeletal	704 767	11.85	44 134 (6.3)	2.35 (2.22, 2.48)
Cardiac	13 606	0.23	2 659 (19.5)	4.19 (3.89, 4.50)
Chest	1 264 620	21.26	137 972 (10.9)	3.26 (3.08, 3.44)
Emergency	658 407	11.07	72 167 (11.0)	3.27 (3.09, 3.46)
Body (gastrointestinal or genitourinary)	765 181	12.86	89 150 (11.7)	2.24 (2.12, 2.37)
Mammography	434 977	7.31	73 508 (16.9)	2.93 (2.77, 3.10)
Neuroradiology	484 618	8.15	78 973 (16.3)	2.88 (2.72, 3.05)
Pediatrics	255 766	4.30	14 487 (5.7)	1.83 (1.73, 1.94)
Vascular	145 510	2.45	8 494 (5.8)	1.62 (1.53, 1.72)
Other‡	1 015 242	17.07	93 215 (9.2)	2.47 (2.34, 2.61)

* At least one RAI detected. Percentages of examinations with at least one RAI are in parentheses.

† Odds ratios, with 95% Wald confidence intervals in parentheses, are from logistic regression holding all other factors equal.

‡ Other includes embolization service and unspecified/general reader division.

Figure 1

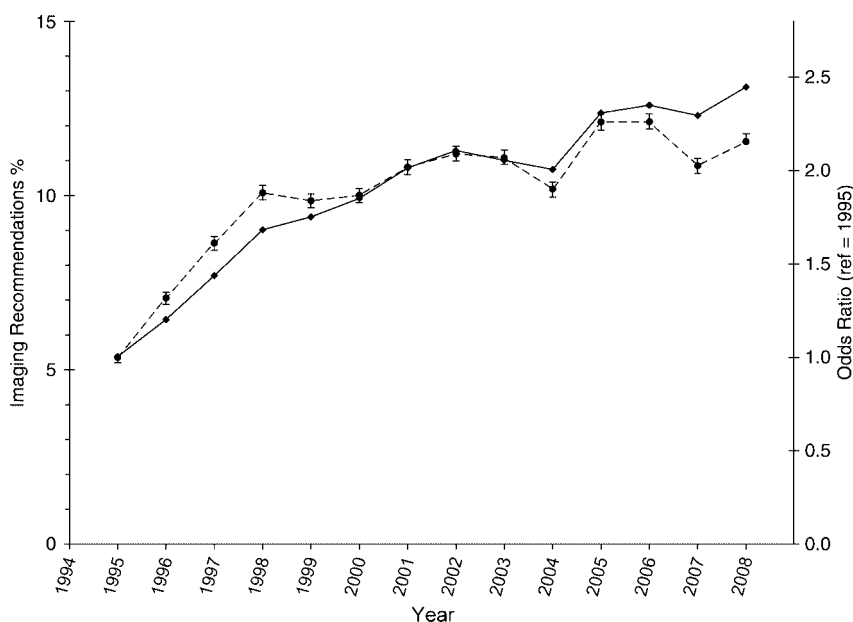


Figure 1: Graph of recommendation rates over time (1995–2008). The diamonds and solid line (left y-axis) represent unadjusted percentage of examinations with an imaging recommendation in each year. The circles and dotted line (right y-axis) depict the odds ratios from logistic regression with 1995 as reference. Error bars on odds ratios are Wald 95% confidence intervals.

(positive findings variable in Table 2). To explore this relationship further, we cross-tabulated the rate of positive examination findings over time and

found it to have actually decreased slightly and gradually (1995 = 68.0%, 2008 = 65.5%). Therefore, in addition to having included that factor in our

regression model, we are confident in asserting that the secular growth in tendency to make an RAI is not confounded by increasing positive examination findings over time.

Figure 2 demonstrates the RAI rate according to patient age, which was grouped into decades. There was considerable divergence between the raw percentage of RAI and the odds ratios. This was due to systematic differences in modality, body area, reader's division, and other factors as patient age increased. However, there was a small and significant increase in the odds (1.52; 95% confidence interval: 1.48, 1.56) of RAI as the patient's age advanced from birth to the 8th decade, which probably reflects a real tendency for patients to acquire reasons for follow-up or correlative imaging as they age.

Figure 3 displays the RAI rate according to radiologist experience. Because we used time between completion of medical school and when the physician signed the report as a proxy for experience, 5 years would be the lowest possible value for someone having a 4-year residency. Therefore, some of the readers in the first interval (5–10 years) may have been post-graduate fellows dictating on their own. This phenomenon may partly account for the large number of radiologists ($n = 422$ of 555) who each read fewer than 10 000 studies and accounted for only about 16% of all examinations included in this study. In Figure 3, both the odds ratios and unadjusted rates show a trend of decrease in making RAIs with advancing reader experience. Any discrepancy between unadjusted percentages and the odds ratios was probably due to differences in the types of examinations that radiologists read throughout their careers. The adjusted odds of RAI were about 15% lower for each additional decade after starting practice.

Discussion

There was a doubling in the proportion of examinations with at least one RAI during the 13 years of our study. This was reflected by unadjusted rates starting at 6% and ending with 12%. Our multivariable analysis confirmed that

Figure 2

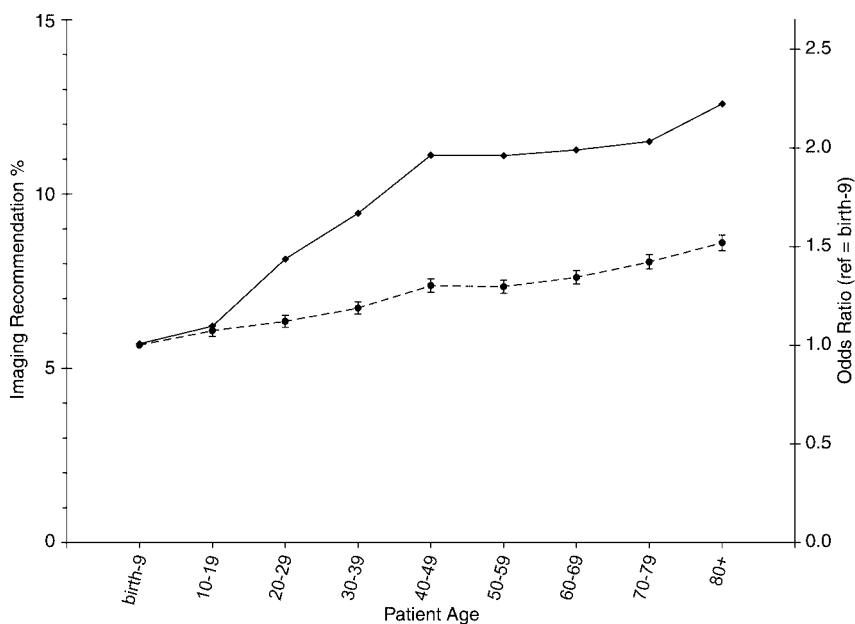


Figure 2: Graph of recommendation rates according to patient age in decades. The diamonds and solid line (left y-axis) represent unadjusted percentage of examinations with an imaging recommendation in each age group. The circles and dotted line (right y-axis) depict the odds ratios from logistic regression with birth to 9 years as reference. Error bars on odds ratios are Wald 95% confidence intervals.

Figure 3

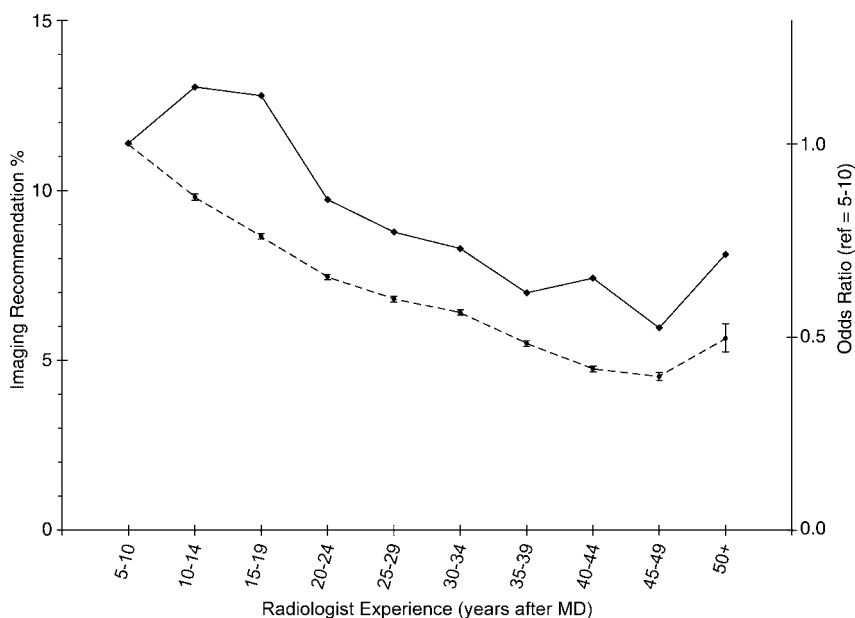


Figure 3: Graph of recommendation rates according to radiologist experience (years after medical school graduation (MD) that the examination results were interpreted). The diamonds and solid line (left y-axis) represent unadjusted percentage of examinations with an imaging recommendation in each successive group. The circles and dotted line (right y-axis) depict the odds ratios from logistic regression with 5–10 years as reference. Error bars on odds ratios are Wald 95% confidence intervals.

there has been a general increase in the tendency for radiologists in our practice to recommend additional or follow-up imaging. The odds ratio of 2.16 (95% confidence interval: 2.11, 2.19) between 1995 and 2008 was estimated from a logistic regression which holds all other factors equal. Changes in the relative mixture of junior versus senior radiologists in the practice over the years were captured in the radiologist's experience variable and did not confound the estimate of changes over time (year variables). Similarly, though the relative proportion of examinations of different types did change over time, inclusion of separate variables for modality and body area in the model should have accounted for them.

One possible causal mechanism for the increase during 13 years, beginning in 1995, was that resolution and information density of images obtained with improving equipment and viewed on more sophisticated soft-copy workstations results in more observations that would prompt an RAI. The nearly linear increase beginning in 1995 corresponded to institutional adoption and expansion of the picture archiving and communication system to near complete penetration by 2000. Other factors that may play a role in the tendency for reports to contain RAIs include malpractice concerns, as well as changes in reporting styles, practices, and mechanics (eg, speech recognition and use of templates). Our study did not include any variables that could assess any of these factors.

Perhaps the most interesting finding was the steady decrease in the tendency to make an RAI as radiologist experience advances. As we have noted, this trend was strongly reflected in the odds ratios from logistic regression and was not confounded by any of the other variables we included in the model. Because the time period covered by the study was 13 years, whereas the range of reader experience was at least 40 years, we have not fully examined a single cohort of radiologists as they mature. We would have preferred to use completion of radiology residency rather than graduation from medical school as the sub-

trahend for our experience variable. However, this information was not routinely stored in our data warehouse. These limitations do not alter the general finding that radiologists tend to recommend less as they gain experience. However, the exact shape of the descent might be somewhat different in a true inception cohort beginning at the 1st year after residency. The slight uptick in RAI by the most senior radiologists is of uncertain importance. Note that the error bar in Figure 3 around the 50+ year experience marker is much wider than the others, indicating a rather small number of such observations (just over 10 000 cases in total). Thus, a few individuals could (and almost certainly did) influence this single data point.

The main limitation of this study was that, though quite precise, our estimates of RAI rates, trends, and variable relationships may not generalize to other institutions. As described in the introduction, there are few other published studies from outside our department to compare with. These investigators relied on manual review of report texts, had small sample sizes, looked at RAI arising from abdominal CT interpretations, and found rates of 19.3% and 31% (4,5). In our entire sample, there were 307 300 abdominal CT scans, of which 49 145 (16%) had at least one RAI. One reason for our lower rate could be that the automated NLP method for detecting RAI may be somewhat insensitive compared with manual abstraction, though our method studies have consistently yielded sensitivities well above 90% (9–14). However, if our NLP method had a systematic and consistent false-negative rate for RAI, it would not affect odds ratio estimates, and our inferences based on them would hold true.

Finally, we have not made any attempt to assess differences between individual radiologists in their relative tendency to make an RAI. Our multivariable regression explained 17% of the variance in the outcome (presence of at least one RAI in the report). Even

though we did include radiologist experience as a predictor, a substantial fraction of the remaining unexplained variance probably arises from inter-radiologist differences in tendency to recommend. Quality improvement and physician profiling initiatives will serve as motivation to quantify, report, and remediate performance measures, and these may include recommendation rates in the case of radiologists. We are currently developing procedures for using our multivariable logistic model as a risk adjustment tool. This will enable creation of observed-to-expected ratios for recommendations that will adjust for each radiologist's case mix. In addition to producing fair comparisons of recommendation rates between radiologists, these techniques will allow us to make meaningful statements about the patterns of variation among radiologists after accounting for experience, specialty, and case mix.

Very large sample sizes such as in this study will often yield hypothesis tests with small *P* values. Furthermore, confidence intervals on parameter (effect size) estimates will be quite narrow. Therefore, the fact that results are statistically significant does not imply that they are meaningful for clinical application or policy considerations. This has been called the “*P* value fallacy” (15,16). More relevant are the quantitative estimates of the effect sizes. At the same time, nearly complete samples of the population of interest do not eliminate confounding and interaction between variables. This is why we have produced and reported odds ratios from multivariable logistic regression in addition to unadjusted percentages of recommendations. Though the unadjusted numbers and percentages of recommendations are intuitively meaningful, we urge readers to focus on the odds ratios when making inferences about the effect of factors such as modality, radiologist experience, patient age, and so forth on the relative tendency to make an RAI. Such issues will likely become more commonplace as computerized patient records and administrative

data are aggregated and analyzed with increasing frequency.

References

1. 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;189(2):371–375.
2. Varipapa RJ. Physicians and outpatient diagnostic imaging: overexposed? *JAMA* 1993;269(13):1633–1634.
3. Levin DC, Rao VM. Turf wars in radiology: other causes of overutilization and what can be done about it. *J Am Coll Radiol* 2004;1(5):317–321.
4. Baumgarten DA, Nelson RC. Outcome of examinations self-referred as a result of spiral CT of the abdomen. *Acad Radiol* 1997;4(12):802–805.
5. Blaivas M, Lyon M. Frequency of radiology self-referral in abdominal computed tomographic scans and the implied cost. *Am J Emerg Med* 2007;25(4):396–399.
6. Riddell AM, Khalili K. Assessment of acute abdominal pain: utility of a second cross-sectional imaging examination. *Radiology* 2006;238(2):570–577.
7. Siddiki H, Fletcher JG, McFarland B, et al. Incidental findings in CT colonography: literature review and survey of current research practice. *J Law Med Ethics* 2008;36(2):320–331.
8. Furtado CD, Aguirre DA, Sirlin CB, et al. Whole-body CT screening: spectrum of findings and recommendations in 1192 patients. *Radiology* 2005;237(2):385–394.
9. Thomas BJ, Ouellette H, Halpern EF, Rosenthal DI. Automated computer-assisted categorization of radiology reports. *AJR Am J Roentgenol* 2005;184(2):687–690.
10. Dreyer KJ, Kalra MK, Maher MM, et al. Application of recently developed computer algorithm for automatic classification of unstructured radiology reports: validation study. *Radiology* 2005;234(2):323–329.
11. Dang PA, Kalra MK, Blake MA, Schultz TJ, Halpern EF, Dreyer KJ. Extraction of recommendation features in radiology with natural language processing: exploratory study. *AJR Am J Roentgenol* 2008;191(2):313–320.
12. Dang PA, Kalra MK, Blake MA, et al. Natural language processing using online analytic processing for assessing recommendations in radiology reports. *J Am Coll Radiol* 2008;5(3):197–204.
13. Lee SI, Saokar A, Dreyer KJ, Weilburg JB, Thrall JH, Hahn PF. Does radiologist recommendation for follow-up with the same imaging modality contribute substantially to high-cost imaging volume? *Radiology* 2007;242(3):857–864.
14. Dang PA, Kalra MK, Blake MA, et al. Use of Radcube for extraction of finding trends in a large radiology practice. *J Digit Imaging* doi: 10.1007/s10278-008-9128-x. Published online June 10, 2008.
15. Dixon P. The p-value fallacy and how to avoid it. *Can J Exp Psychol* 2003;57(3):189–202.
16. Goodman SN. Toward evidence-based medical statistics. I. The P value fallacy. *Ann Intern Med* 1999;130(12):995–1004.