# Radiologic Interpretation by Family Physicians in an Office Practice Setting

George R. Bergus, MD; E. A. Franken, Jr, MD; T. Jean Koch, RN; Wilbur L. Smith, MD; Eric R. Evans, MD; and Kevin S. Berbaum, PhD

*Iowa City, Iowa*

*Background.* Radiology is an integral part of the office practice of many family physicians. Nevertheless, data are sparse on the performance of family physicians in this endeavor. This study investigated the performance of family physicians at interpreting radiographs ordered in a free-standing family practice office.

*Methods.* A consecutive series of radiographic studies performed at a family practice office during a 3-year period was surveyed. All radiographic studies included in this analysis (N=1674) were separately interpreted by the family physician ordering the study and an overreading radiologist. If the interpretations agreed, the studies were accepted as having been correctly interpreted. Cases in which the interpretations disagreed were reexamined.

*Results.* Family physicians correctly interpreted 92.4% of the radiographic studies (95% confidence interval, 91.0 to 93.6). Their accuracy with extremity films (96.0%) was significantly higher than their accuracy with chest films (89.3%, $P<.001$). Family physicians were more likely to correctly interpret normal films (95.2%) than abnormal ones (85.9%, $P<.001$). Thirty-five percent of the cases in which there were differences between family physician and radiologist interpretations were correctly interpreted by family physicians.

*Conclusions.* Family physicians showed a high degree of accuracy in radiologic interpretation in an office setting. Chest films were inherently more difficult to interpret than extremity films. Because correct interpretation depends on body part examined and the prevalence of disease, the performance of family physicians will probably vary in different practice settings.

*Key words.* Radiography, primary care, test performance.
(*J Fam Pract 1995; 41:352-356*)

Radiology is an integral part of the office practice of many family physicians. A survey of family physicians in Minnesota found that 87% used radiography in their offices.[1] The same survey noted that most family physicians felt that radiographic interpretation was a fundamental skill of this primary care specialty. Nevertheless, data are sparse on the accuracy of family physicians in this endeavor. Most previous studies investigating the interpretative skills of nonradiologists have focused on primary care physicians in emergency departments.[2-12] These have generally found interpretative accuracy related to prevalence of disease, range of abnormalities, and body part imaged, all of which are factors that differ between an emergency department and a primary care office.

A study performed in an office setting evaluated the performance of family physicians by concurrently collecting data on the interpretation of 100 consecutive studies in multiple offices.[13] The physicians' performance was calculated by using the overreading radiologists' interpretations as the standard. While this standard is a common one, previous research has documented substantial errors in radiographic interpretation made by experienced radiologists. Disagreement rates have been reported to be between 10% and 40%.[14,15]

The present study analyzed the performance of family physicians at interpreting the radiographs obtained in

*Submitted, revised, July 19, 1995.*

*Presented at the 28th Annual Spring Conference of the Society of Teachers of Family Medicine at New Orleans, Louisiana, 1995.*

*From the Departments of Family Practice (G.R.B., T.J.K., E.R.E.) and Radiology (E.A.F., W.L.S., K.S.B.), The University of Iowa, Iowa City. Requests for reprints should be addressed to George R. Bergus, MD, Department of Family Practice, 2133 Steindler Bldg, The University of Iowa, Iowa City, IA 52242-1097.*



ISSN 0094-3509

one free-standing family practice office using a rigorous truth standard. Accuracy of interpretation was analyzed in relation to the body part imaged and whether disease was identified on the radiographic study.

## Methods

The Department of Family Practice at The University of Iowa maintains a free-standing outpatient office where approximately 20,000 patients are seen annually. Only board-certified family physicians and their directly supervised residents practice at this location. This office provides primary care to patients from both the university community and the surrounding rural areas. Within this office there is a basic radiographic unit, consisting of a 125-kVp generator, radiography table, chest film holder, and automatic processor. Radiographs are taken by members of the nursing staff holding a state license to perform chest and extremity studies.

After every radiographic examination was completed, the ordering family physician immediately interpreted the image and recorded his or her radiographic interpretation. Family physicians had access to patients' previous radiographic studies taken at this location. Radiologic studies ordered and interpreted by a resident family physician were reviewed by a faculty family physician to confirm the resident's interpretation before the patient left the office. All films were reread within 1 week by a radiologist. The nurse-radiographers in the family practice department maintained a log of the readings made by the family physician and the radiologist on an ongoing basis.

To study the performance of family physicians in this office-based practice, we undertook a cross-section study using all radiographic studies taken in this one office over a 3-year period, between 1990 and 1993. During the study period, 54 different family physicians interpreted radiologic studies at this site. Sixteen were faculty physicians, all of whom were board-certified in family practice. Fourteen of these physicians were residency-trained, and more than one half had entered academic medicine from full-time clinical practice. Thirty-eight of the physicians were residents receiving training in the family medicine program.

Five different faculty radiologists alternated as overreaders. Three of the overreaders were general radiologists, one subspecialized in pediatric radiology, and one subspecialized in chest radiology. The three general radiologists overread approximately 80% of the radiographs included in this study. All radiologists had access to the family physicians' readings.

Radiographic examinations noted to have the same interpretation by the family physician and the radiologist, ie, both readings negative or both positive with the same diagnosis, were assumed to have been correctly read. These determinations were prospectively made by members of the nursing staff who were licensed radiographers. Cases identified as having discrepant readings were reinterpreted for this investigation. At least 6 months passed between the time of the index office visit and the viewing of the discrepant radiologic studies by the investigators.

A radiologist and a family physician, neither of whom had previously viewed the images, reviewed all discrepant studies and attempted to determine the true readings of these films. They viewed each radiographic study together and were blinded from the original readings of the films. The investigators were given information contained in the patient's office chart, including the patient's chief complaint, past medical history, physical findings, laboratory data, results of other imaging studies, and subsequent clinical course. If the study team could not agree on the interpretation of a radiologic study, a separate panel of four radiologists, using only the films and the indication for the study, determined the correct interpretation. The second panel was used to determine the true readings for 7% of the discrepant studies.

After determining the true reading of a radiographic study, the investigators decided which, if any, of the original readings was correct. After reviewing the original readings, the investigators were also allowed to conclude that the original readers were not truly discrepant.

The data were analyzed using chi-square and logistic regression. Reader sensitivity was calculated by determining the percentage of films showing disease that were read correctly, and specificity by determining the percentage of normal films that were read correctly. A normal film was defined as one showing no abnormality or only age-related changes; an abnormal film as one with new, old, or ongoing disease. Because previous studies suggested that accuracy may depend on the part of the body imaged, the analysis included the following two variables: the level of training of the physician and whether the radiograph revealed abnormalities.

## Results

Over the 3-year period, there were 64,690 office visits with 1689 radiographic studies. Thirteen of the studies over the 3-year period were not overread by a radiologist and thus were excluded from this investigation. These films had accompanied patients to urgent consultations with other specialists and were never returned. Two studies that were logged as producing a discrepancy between readings by the family physician and the radiologist could

not be located for review and were excluded. As a result of these exclusions, the database for this study consisted of 1674 cases.

Eight hundred ninety-eight (53.6%) of the radiographic studies included in this investigation were of the chest, all with posterior–anterior and lateral views. Seven hundred seventy-six (46.4%) were extremity examinations. Of these, 411 were of an upper extremity, ie, finger, hand, wrist, forearm, elbow, or upper arm, and 365 were of a lower extremity, ie, toe, foot, ankle, lower leg, or knee. All extremity studies included two or three films with standard views. No other body parts were imaged at this office site.

Of the 1674 radiographic studies, 1179 (70.4%) were normal or showed only age-related changes, and 495 (29.6%) showed acute or chronic disease. A significantly greater proportion of chest studies were determined abnormal than were extremity studies (31.7% vs 27.1%, $P=.04$).

Faculty family physicians had a significantly higher rate of utilization of in-office imaging than did residents. Faculty saw only 26.9% of all patients but ordered 33.9% of the radiographic studies ($P<.001$). Radiographs ordered by faculty and residents had similar distribution by body part ($P>.05$). The percentage of films ordered by faculty and determined to be abnormal was not significantly different from that ordered by residents (31.8% vs 28.4%, $P=.15$).

The family physician and the overreading radiologist made different interpretations in 196 (11.7%) of the radiographic studies. The 8.0% (94/1179) disagreement rate for normal films was significantly lower than the 20.6% (102/495) disagreement rate for abnormal films ($P<.001$). Sixty-nine (35.2%) of the 196 films on which the interpretations differed had been interpreted correctly by the family physician.

Overall, family physicians correctly interpreted 92.4% (95% confidence interval [CI], 91.0 to 93.6) of the radiographic studies. They were more likely to correctly interpret extremity radiographs than chest radiographs (96.0% vs 89.3% correctly interpreted, $P<.001$). Family physicians were also more likely to correctly interpret normal radiographic studies than abnormal ones ($P<.001$). Their overall sensitivity, the percentage of abnormal studies read correctly, was 85.9% (95% CI, 82.4 to 88.7). The family physicians identified 105 (82.7%) of the 127 cases of pneumonia, 26 (76.5%) of the 34 cases of congestive heart failure, and 178 (97.3%) of the 183 fractures imaged at this office over the 3-year period. Their overall specificity, the percentage of normal studies read correctly, was 95.2% (95% CI, 93.7 to 96.3). Family physicians made 22 false-positive interpretations of pneumonia, 2 of congestive heart failure, and 15 of fracture.

Table 1. Accuracy of Family Physicians at Interpreting Norma and Abnormal Radiographic Studies in an Office Setting

| | Accurate Interpretation by Family Physicians, % | | |
|---|---|---|---|
| Type of Radiograph | Normal Radiographs | Abnormal Radiographs | Total Radiographs* |
| Chest (n=898)† | 93.6 | 80.0 | 89.3 |
| Extremity (n=776)‡ | 96.8 | 93.8 | 96.0 |
| Total (N=1674)† | 95.2 | 85.9 | 92.4 |

*P <.001 comparing accurate interpretation of chest vs extremity films.
†P <.001 comparing accurate interpretation of normal vs abnormal radiographs.
‡P =.06 comparing accurate interpretation of normal vs abnormal radiographs.

The correct interpretation rate for studies showing chronic disease was similar to that for studies showing acute disease (88.3% vs 85.2%, $P>.05$). Further analysis showed that family physicians' interpretive accuracy for normal chest radiographs was markedly higher than for abnormal chest studies, (93.6% vs 80.0%, $P<.001$), but that their accuracy with normal and abnormal extremities studies was more similar (96.8% vs 93.8%, $P>.05$ ). An overview of these data is in Table 1.

Faculty physicians correctly interpreted 90.4% of the studies they ordered, but residents, who interpreted their films with faculty, correctly interpreted 93.5% of their studies ($P=.02$). The resident-faculty pair had both a higher sensitivity, 87.3% vs 83.4%, and a higher specificity, 95.9% vs 93.6%, than did faculty physicians alone, although these differences were not statistically significant.

### *Statistical Analysis*

Factors associated with the family physicians' correct interpretation of radiographs were entered into a multivariate model. Independent variables placed into the model included body part (extremity or chest), presence of an abnormality, and whether the ordering family physician was a faculty member or resident. Logistic regression verified that these factors were significantly associated with a family physician making a correct radiologic interpretation. Radiographic studies that were normal, imaged an extremity, or were ordered by a resident and therefore viewed by a resident-faculty pair were significantly more likely to have been correctly interpreted by the family physician. The odds ratios of these associations are shown in Table 2.

## Discussion

Despite the common use of radiology by primary care physicians, little is known about their accuracy in interpreting radiographs in an office setting. Overall, we found

Table 2. Factors Associated with the Correct Interpretation of Radiographs in an Office Setting

| Factor | Odds Ratio | *P* Value |
|---|---|---|
| Primary reader (faculty alone vs resident-faculty pair) | 1.47 | .04 |
| Body (chest vs extremity) | 2.76 | <.001 |
| True state of the patient (diseased vs normal) | 3.12 | <.001 |

that family physicians correctly interpreted 92.4% of the radiographic studies they ordered in this office. In our study, family physicians showed greater accuracy at interpreting extremity films than chest films and greater accuracy at interpreting normal films than abnormal ones. Both body part and disease state of the body part were independently associated with a correct interpretation.

Halvorsen and co-workers[13] investigated the performance of family physicians at interpreting 508 office radiographs. The rate of agreement between radiologist and primary care physician for normal films found by these investigators (92.1%) was not significantly different from that found in our study (95.2%, $P>.05$). For abnormal films, Halvorsen and colleagues found a 79.4% rate of agreement, compared with our finding of 84.7% ($P>.05$) The agreement rates for chest films and extremity films were also not significantly different.

In addition to discrepancy rates, we are able to provide information on interpretative performance. In the previous office-based study, the physicians' performance was calculated using an overreading radiologist's interpretation as the criterion.[13] This criterion is defensible, but underestimates the accuracy of family physicians. Our investigation indicates that a disagreement between a radiologist and family physician does not necessarily represent error on the part of the family physician. The 92.4% overall accuracy of family physicians is significantly higher than the 88.3% agreement rate between the radiologists and the family physicians ($P<.001$).

Radiologic studies ordered by resident physicians were more likely to be interpreted correctly than were those ordered by faculty physicians. This difference in performance remained statistically significant even after other factors associated with correct interpretation were added to a multivariate model. Although residents received formal radiologic training, we did not detect an improvement in their interpretive performance in our office after this rotation. The most likely explanation for the better performance is that films ordered by residents were actually read by two family physicians—the resident and a faculty member. Dual reading of individual images by radiologists is known to improve interpretative accuracy.[16] Furthermore, the previous study[13] in a family practice office also found that when residents and faculty read radiographs together, their readings were more likely to agree with that of the overreading radiologist than were the readings of either group alone.

Our study has several limitations. With any study of a diagnostic test, it is important to review the criterion.[17,18] As a general principle, the criterion should be based on data beyond the variables being studied.[19] Thus, our access to the medical record with follow-up data provided valuable information for validation. We believe our criterion is closer to the truth of cases than is the report of a single overreading radiologist.

The ideal methodology used to compare the performance of a test with the criterion standard, is receiver operating characteristic (ROC) curve analysis. This approach allows the diagnostic capacity of a test to be separated from the thresholds used to define a positive test.[20,21] We were unable to use this approach because our data contain the radiographic interpretations of the physicians without measure of how confident they were in their interpretations.

Another concern might be the generalizability of our findings. Our study was not performed in a typical community office but, rather, in a family practice training site. The goal of the training office, however, is to model the typical office of a primary care physician with a normal distribution of age groups and a normal cross-section of medical problems. The study site was also unusual in that it was in a small university town, but it draws patients from both the immediate nonuniversity community and nearby rural communities.

The percentage of abnormal studies in our data set was significantly lower than reported by Halvorsen and colleagues[13] (29.6% vs 34.8%, $P=.02$). Because the family physicians correctly interpreted normal films more often than abnormal ones and extremity films more often than chest films, their accuracy in any office setting will depend on the prevalence of disease and case mix in the population being imaged. While our study site might have influenced our findings, we have also described interpretative performance by subcategories. Therefore, our findings should generalize to another office setting with a different disease prevalence.

The use of radiographic imaging in our setting was significantly lower than that of the study conducted in Minnesota family practice training sites[13] (2.61 vs 5.65 studies per 100 patients, respectively $P<.001$). Nearly 20% of the studies performed at the Minnesota sites, however, would have been referred to a hospital-based imaging center in Iowa because of licensing requirements. Another possible reason for low utilization was that none

of the physicians taking part in this study benefited from the revenue derived from radiologic studies.[22]

Accuracy in radiographic interpretation is also known to be a function of image quality.[23] Although there are reports about the poor technical quality of films in the offices of primary care physicians, the images from this study were considered to be, on average, of acceptable quality.[24] Although the radiographic unit and processor were typical of those used in the office of a group practice of primary care physicians, several systems were in place at the study site to maximize the quality of radiographs. The nurse-radiographers were members of the nursing staff at the office and licensed to perform studies of the chest and extremities. Iowa licenses clinical staff to function as limited radiographers, as do 22 other states. The equipment was maintained by biomedical engineers on a scheduled basis, and unreadable film rate was used as an ongoing quality control measure.

There are several remaining important questions that should be addressed by future research. The first is about the clinical significance and cost of the interpretive errors made by the family physicians.[25] In the present study, the interpretations were collected in a standardized manner, but outcomes data were not. We know that some cases of pneumonia were missed by family physicians, but we do not have rigorous means of estimating the cost, if any, of these errors. Thus, our analysis is limited to the interpretive performance of physicians. Similarly, although data on the performance of the overreading radiologists are available, we have no way to estimate the cost-effectiveness of having each study formally overread by a radiologist.[26]

## Conclusions

Family physicians have high overall accuracy in interpreting office radiographs of chests and extremities. In our study, the family physicians were correct in over one third of the discrepant cases. This performance is better than that found by studies using the interpretation of a single overreading radiologist as the criterion. Family physicians seem to have more difficulty interpreting chest films than extremity radiographs. We believe that this difficulty may be related to the greater variety of subtle findings that may be captured on chest studies in addition to the higher rate of abnormalities found on chest films. Because correct interpretation depends on the body part being imaged and the prevalence of disease, the performance of family physicians will probably vary in different practice settings. Finally, our data suggest that the interpretive performance of family physicians improves when a colleague concurrently reads the film.

References


1. Halvorsen JG, Kunian A. Radiology in family practice: experience in comunity practice. Fam Med 1988; 20:112–7.
2. DeLacey G, Barker A, Harper J, Wignall B. An assessment of the clinical effects of reporting accident and emergency radiographs. Br J Radiol 1980; 53:304–9.
3. Carew-McColl M. Radiological interpretation in an accident and emergency department. Br J Clin Pract 1983; 37:375–7.
4. Overton DT. A quality assurance assessment of radiograph reading accuracy by emergency medicine faculty. Ann Emerg Med 1987; 16:503
5. Duignan N, Lawson JT. Should all casualty radiographs be reviewed? [letter] BMJ 1985; 291:141.
6. Masel JP, Grant PJ. Accuracy of radiological diagnosis in the casualty department of a children's hospital. Aust Paediatr J 1984; 20: 221–3.
7. Wardrope J, Chennells PM. Should all casualty radiographs be reviewed? BMJ 1985; 290:1638–40.
8. Swain AH. Radiological audit-changes in casualty officer performance during tenure of post. Br J Accid Emerg Med 1986; 1:5–9.
9. Vincent CA, Driscoll PA, Audley RJ, Grant DS. Accuracy of detection of radiographic abnormalities by junior doctors. Arch Emerg Med 1988; 5:101–9.
10. Fleisher G, Ludwig S, McSorley M. Interpretation of pediatric x-ray films by emergency department pediatricians. Ann Emerg Med 1983; 12:153–8.
11. Gleadhill DN, Thomson JY, Simms P. Can more efficient use be made of x-ray examinations in the accident and emergency department? BMJ 1987; 294:943–7.
12. McLain PL, Kirkwood CR. The quality of emergency room radiograph interpretations. J Fam Pract 1985; 20:443–8.
13. Halvorsen JG, Kunian A, Gjerdingen D, Connolly J, Koopmeiners M, Cesnik J, et al. The interpretation of office radiographs by family physicians. J Fam Pract 1989; 28:426–32.
14. Yerushalmy J. The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. Radiol Clin North Am 1969; 7:381–92.
15. Herman PG, Gerson DE, Hessel SJ, et al. Disagreements in chest roentgen interpretation. Chest 1975; 68:278–82.
16. Yerushalmy J, Harkness JT, Cope JH, Kennedy BR. The role of dual reading in mass radiography. Am Rev Tuberc 1950; 61:443–64.
17. Sachett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston, Mass: Little, Brown & Co, 1991:51–68.
18. Begg CB. Biases in the assessment of diagnostic tests. Stat Med 1987; 6:411–23.
19. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978; 299:926–30.
20. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8:283–98.
21. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Invest Radiol 1989; 24:234–45.
22. Levin DC, Edmiston RB, Ricci JA, et al. Self referral in private offices for imaging studies performed in Pennsylvania Blue Shield subscribers during 1991. Radiology 1993; 189:371–5.
23. Kundel HL. How much spatial-resolution is enough—a metaanalysis of observer performance studies comparing plain films and digital hard-copy. In: Jost RG, ed. Medical imaging 1993: design and evaluation. Bellingham, Wash: SPIE–The International Society for Optical Engineering, 1993:86–89.
24. Hopper KD, Rosetti GF, Edmiston RB, et al. Diagnostic radiology peer review: a method inclusive of all interpreters of radiographic examinations regardless of specialty. Radiology 1991; 180:557–61.
25. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991; 11:88–94.
26. Franken EA, Bergus GR, Koch TJ, Berbaum KS, Smith WL. Value-added of radiologist consultation to family practitioners in outpatient setting. Radiology 1995. In press.