

Is Terminology Used Effectively to Convey Diagnostic Certainty in Radiology Reports?¹

Ramin Khorasani, MD, David W. Bates, MD, MSc, Susan Teeger, MD, Jeffrey M. Rothschild, MD, MPH
Douglas F. Adams, MD, Steven E. Seltzer, MD

Rationale and Objectives. This study was performed to assess the extent of agreement among radiologists and nonradiologists in perception of diagnostic certainty conveyed by words and phrases commonly used in radiology reports.

Materials and Methods. The study was performed in a large academic radiology department. To determine the commonly used terminology for conveying diagnostic certainty in radiology reports, 12 randomly selected radiologists from six different subspecialties were interviewed. The authors identified the 15 most commonly used words and phrases and included these in random order in a questionnaire sent to all staff radiologists ($n = 45$) and to 158 referring physicians. Physicians were asked to rank the 15 phrases in order of the diagnostic certainty conveyed by each, from 1 (most certain) to 15 (least certain), using each number only once. The κ statistic was used to assess agreement in rank order among physicians.

Results. The questionnaire response rate was 76% ($n = 34$) for radiologists and 49% ($n = 78$) for nonradiologists. There was excellent agreement among radiologists ($\kappa = 0.95$) and nonradiologists ($\kappa = 0.93$) in the rank order for the phrase *diagnostic of*. Although there was good agreement ($\kappa = 0.45$) among radiologists for the word *unlikely*, agreement among nonradiologists was poor ($\kappa = 0.27$). There was very poor agreement among all physicians for the rank order of the other 13 phrases.

Conclusion. Among radiologists and nonradiologists, concordance was poor regarding the diagnostic certainty associated with phrases commonly used in radiology reports. Because poor agreement could lead to suboptimal quality of care, the standardization of terminology would benefit all parties.

Key Words. Radiology and radiologists, research; radiology reporting systems.

© AUR, 2003

Effective communication is critical in clinical care. Radiology reports are the main form of communication between radiologists and referring physicians about findings at imaging. These findings are typically reported in free text form. To convey their degree of certainty about a potential finding or diagnosis in their reports, radiologists

use various nonspecific phrases (eg, *possibly*, *probably*, *consistent with*). If these phrases generate confusion about the meaning of a finding or the likelihood of the diagnosis, the imaging results could be misinterpreted, potentially resulting in unnecessary additional testing, suboptimal care delivery, and increased costs.

Findings from previous research studies suggest that language in radiology reports varies widely and may be confusing. One survey found that 40% of referring physicians considered chest radiography reports to be occasionally confusing (1). Sobel et al (2) analyzed the clarity and content of radiologists' reports for chest radiographs and found a high degree of variability: A total of 23 synonyms were used to indicate the presence of abnormal

Acad Radiol 2003; 10:685–688

¹ From the Department of Radiology (R.K., S.T., D.F.A., S.E.S.) and Division of General Internal Medicine and Primary Care (D.W.B., J.M.R.), Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02115; and Partners Healthcare System, Boston, Mass. Received March 7, 2003; revision requested March 11; revision received and accepted March 12. Funded in part by a grant from EMC to the Center for Evidence-based Imaging, Brigham and Women's Hospital. Address correspondence to R.K.

© AUR, 2003

Statistics for Rank Orders Assigned by Physicians to Words and Phrases Commonly Used in Radiology Reports to Convey Diagnostic Certainty

Term	Radiologists			Nonradiologists		
	Mean	Standard Deviation	Range	Mean	Standard Deviation	Range
<i>Diagnostic of</i>	1.1	0.52	1–4	1.1	0.79	1–8
<i>Consistent with</i>	5.2	2.97	1–12	4.8	2.60	2–12
<i>Compatible with</i>	6.3	3.35	3–15	6.5	3.16	1–14
<i>Worrisome for</i>	8.2	2.88	4–14	8.9	3.25	3–15
<i>Suspicious for</i>	6.6	2.69	2–14	7.9	2.78	2–14
<i>May represent</i>	11.4	1.56	7–14	10.8	2.77	3–15
<i>Likely</i>	7.0	2.55	2–12	6.6	2.61	2–15
<i>Most likely</i>	4.3	1.88	2–8	4.1	2.12	2–15
<i>Probably</i>	7.5	2.65	3–13	7.2	2.71	2–15
<i>Possibly</i>	12.0	1.40	9–15	11.5	2.42	4–15
<i>Suggestive of</i>	6.6	2.69	2–14	7.9	2.78	2–14
<i>Highly suggestive of</i>	3.4	1.68	2–7	3.6	2.15	2–15
<i>Question of</i>	13.6	1.16	10–15	12.9	2.40	4–15
<i>Maybe</i>	12.5	1.66	9–15	12.2	2.82	3–15
<i>Unlikely</i>	13.0	3.44	2–15	10.8	4.65	2–15

Note.—Numbers indicate ranking from 1 (“most certain”) to 15 (“least certain”).

finding(s), 30 synonyms were used to indicate the possible presence of abnormal finding(s), and three synonyms were used to indicate the absence of abnormalities. However, these authors did not assess agreement among physicians on various phrases used to convey certainty.

To minimize confusion in interpretation, highly standardized approaches have been developed for reporting the results of particular diagnostic tests (eg, cervical smears) (3) and radiology tests (eg, mammography) (4). Such standardization can help quantify the probability that an abnormal diagnostic finding will be associated with a subsequent positive histopathologic finding (5). Consistent use of such reporting systems has been found to result in more effective communication between those reporting the results of tests and those acting on the basis of reported results (6). To our knowledge, however, no uniform terminology standards have been established for conveying diagnostic certainty in reporting of most radiology test results. As a first step toward possible standardization, we sought to assess the degree of agreement among radiologists regarding specific phrases used to convey diagnostic certainty in radiology reports and to evaluate the perception of nonradiologists regarding the same terms.

MATERIALS AND METHODS

The study was conducted in a large academic radiology department at a 720-bed tertiary care hospital per-

forming 400,000 radiology examinations per year. To determine commonly used terminology for conveying diagnostic certainty in radiology reports, we randomly selected 12 radiologists from six different subspecialty sections (neuroradiology, gastrointestinal radiology, genitourinary radiology, musculoskeletal radiology, nuclear medicine, and ultrasonography). Each radiologist was interviewed to determine the terminology he or she used when reporting the full range of diagnostic certainties, and each was asked to state the five most common terms used for this purpose in reports. From these interviews, we tabulated the 15 most common phrases (Table) used to convey diagnostic certainty in radiology reports. We put these phrases in random order in a questionnaire and sent it to all staff radiologists in the radiology department ($n = 45$) and to 158 referring physicians from the physician hospital organization’s membership list, which included primary care providers, internists, and surgeons. All physicians were asked to rank the 15 phrases for level of diagnostic certainty conveyed on a scale from 1 (most certain) to 15 (least certain). The physicians were instructed to use each number only once.

We used the κ statistic to assess agreement in rank orders among radiologists for each word or phrase. A similar analysis was used to analyze the data obtained from nonradiologists. A κ value of >0.7 generally denotes good agreement, and a κ value of <0.4 denotes poor agreement (7).

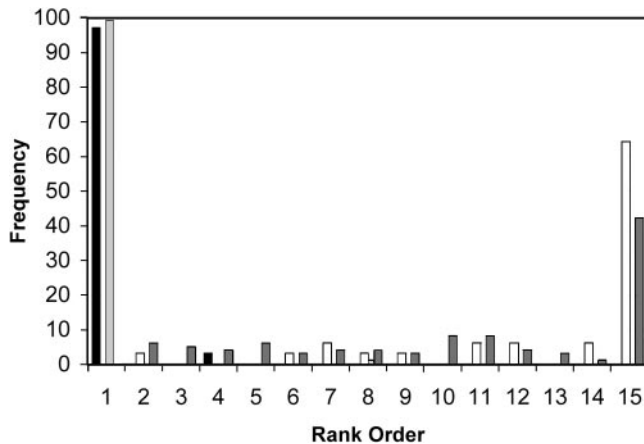


Figure 1. Graph of frequency distribution for rank order (1 = most certain, 15 = least certain) assigned by radiologists and by nonradiologists to the terms *diagnostic of* (black bars = radiologists' ranking, light gray bars = nonradiologists' ranking) and *unlikely* (white bars = radiologists' ranking, dark gray bars = nonradiologists' ranking).

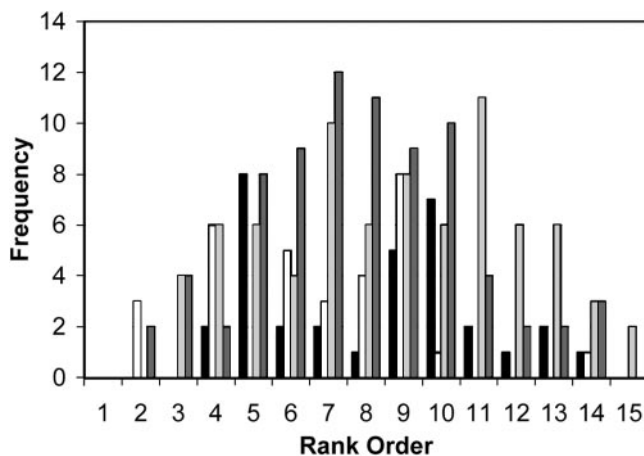


Figure 2. Graph of frequency distribution for rank order (1 = most certain, 15 = least certain) assigned by radiologists and by nonradiologists to the phrases *worrisome for* (black bars = radiologists' ranking, light gray bars = nonradiologists' ranking) and *suspicious for* (white bars = radiologists' ranking, dark gray bars = nonradiologists' ranking). The graph demonstrates the wide variability in the perception of certainty for both physician groups.

RESULTS

The questionnaire response rate for radiologists was 76% ($n = 34$), while the response rate for nonradiologists was 49% ($n = 78$). The 15 commonly used phrases (Table) included both positively framed terms, such as *diagnostic of* and *consistent with*, and negatively framed terms, such as *question of* and *unlikely*. Standard deviations in rank orders were high for most terms, especially

among nonradiologist respondents (standard deviations for 13 of 15 terms were higher among nonradiologists' responses than among responses from radiologists). The wide range in scores assigned to particular terms, even by radiologists, is striking: For example, *unlikely* was ranked from 2 to 15, *highly suggestive of* was ranked from 2 to 14, and *compatible with* was ranked from 3 to 15.

The frequency distribution for the rank orders of sample phrases (Fig 1) shows little consensus for the 13 terms not at the extreme ends of the certainty spectrum. In Figure 2, frequency distributions for the phrases *worrisome for* and *suspicious for* show wide variability for both radiologists and nonradiologists. The κ results (Fig 3) show excellent agreement among both radiologists ($\kappa = 0.95$) and nonradiologists ($\kappa = 0.93$) about the level of certainty conveyed by the phrase *diagnostic of*, but little to no agreement about the other 14 terms.

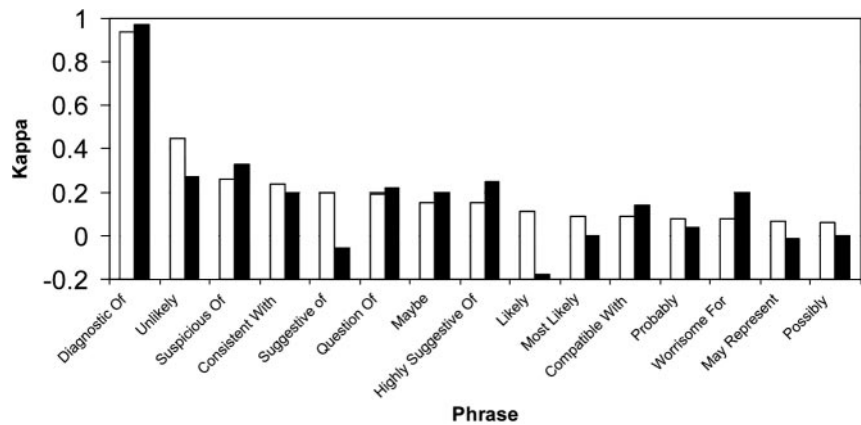
DISCUSSION

Our results demonstrate little or no agreement among radiologists and nonradiologists about the diagnostic certainty associated with phrases commonly used in radiology reports. We found strong agreement only for the phrase *diagnostic of*. We identified phrases with the widest variability in conveying diagnostic certainty, including *consistent with*, *worrisome for*, *suspicious for*, *may represent*, *probably*, and *possibly*.

These results confirm previous research findings that suggested many physicians are confused by terminology in radiology reports (1,2). Sobel et al (2) concluded that a more systematic appraisal of the certainty conveyed in radiology reports would likely improve clinical decision making. Our results also support those of another study showing statistically significant disagreement among physicians in the conclusions reached after reviewing descriptive interpretations of nuclear medicine lung scans, as opposed to standardized reports (8). Similar lack of agreement was found in a survey asking pathologists to associate quantitative values for diagnostic certainty with 21 terms. Agreement was achieved only for terms at the extremes (*always*, *never*) (9). In another study, nonphysician professionals and skilled workers who were asked to quantify the certainty conveyed by descriptive terms that are used often in radiology reports found substantial overlap between *occasionally*, *often*, and *usually* (10).

Misunderstandings generated by inconsistent use of descriptive phrases can cause unnecessary additional testing and adversely affect quality of care. In addition, con-

Figure 3. Graph illustrates the general lack of agreement about the level of diagnostic certainty associated with commonly used phrases in radiology reports, as indicated by the κ statistics for radiologists (white bars) and for nonradiologists (black bars). *Diagnostic of* is the only phrase about which there was a high level of agreement.



fusion can impose extra burdens for translating descriptive phrases into a structured language or numeric representation in computer-based patient records (11).

This study has several limitations. The results come from only one tertiary care center and may not be generalizable to other settings. The specific phrases used and the perception of certainty associated with them may vary by practice setting or by region. Moreover, we did not examine the effect of use of this terminology on additional unnecessary testing or on quality of care.

Several steps could be taken to improve the accuracy of communication among physicians regarding diagnostic certainty. Standardized terminology has been effective, for example, for reporting findings of cervical smears (3) and bone densitometry measurements (6). Within radiology, widespread use of a mammography reporting lexicon is now occurring (12). However, attempts to standardize the reporting of other radiology test results have generally been unsuccessful to date (13). The use of a clearly understood spectrum of modifiers (eg, *most*, *more*, *less*, and *least*) with a word such as *likely*, and the elimination of other words and phrases (particularly those resulting in the widest variation in perception of diagnostic certainty, such as *possibly* and *probably*), is another option. Educating referring physicians and radiologists about an agreed-upon method for conveying diagnostic certainty also could improve the communication of radiology results. However, education is generally a weak strategy for changing physician behavior (14).

Our findings reveal poor concordance among physicians in interpreting the diagnostic certainty of phrases used commonly in radiology reports. This ambiguity may result in additional unnecessary testing or failure to follow up important findings and thus may impair quality of

care. Further studies are needed to determine how best to standardize reporting across the wide range of radiology tests and to assess whether standardized terminology produces more effective communication of diagnostic test results.

REFERENCES

1. Clinger NJ, Hunter TB, Hillman BJ. Radiology reporting: attitudes of referring physicians. *Radiology* 1988; 169:825–826.
2. Sobel JL, Pearson ML, Gross K, et al. Information content and clarity of radiologists' reports for chest radiography. *Acad Radiol* 1996; 3:709–717.
3. Sparkes J, Schoolland M, Barrett P, Kurinczuk JJ, Mitchell KM, Sterrett GF. Trends in the frequency and predictive value of reporting high grade abnormalities in cervical smears. *Cancer* 2000; 90:215–221.
4. D'Orsi CJ. The American College of Radiology mammography lexicon: an initial attempt to standardize terminology. *AJR Am J Roentgenol* 1996; 166:779–780.
5. Liberman L, Abramson AF, Squires FB, Glassman JR, Morris EA, Der-shaw DD. The Breast Imaging Reporting and Data System: positive predictive value of mammographic features and final assessment categories. *AJR Am J Roentgenol* 1998; 171:35–40.
6. El Hajj FG, Stock JL, McClung MR, Saifi G. A national random survey of bone mineral density reporting in the United States. *J Clin Densitom* 2002; 5:3–9.
7. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. 2nd ed. Boston, Mass: Little, Brown, 1991.
8. Bastuji-Garin S, Schaeffer A, Wolkenstein P, et al. Pulmonary embolism: lung scanning interpretation—about words. *Chest* 1998; 114:1551–1555.
9. Kenney RM. Between never and always. *N Engl J Med* 1981; 305:1097–1098.
10. Toogood JH. What do we mean by "usually"? (letter). *Lancet* 1980; 1:1094.
11. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. *J Am Med Inform Assoc* 1997; 4:342–355.
12. Liberman L, Menell JH. Breast Imaging Reporting and Data System (BI-RADS). *Radiol Clin North Am* 2002; 40:409–430.
13. Naik SS, Hanbidge A, Wilson SR. Radiology reports: examining radiologist and clinician preferences regarding style and content. *AJR Am J Roentgenol* 2001; 176:591–598.
14. Soumerai SB, McLaughlin TJ, Avorn J. Improving drug prescribing in primary care: a critical analysis of the experimental literature. *Milbank Q* 1989; 67:268–317.