Laterality Errors in Radiology Reports Generated With and Without Voice Recognition Software: Frequency and Clinical Significance

Marianne T. Luetmer, BA[a], Christopher H. Hunt, MD[b], Robert J. McDonald, MD, PhD[b], Brian J. Bartholmai, MD[b], David F. Kallmes, MD[b]

Purpose: The aim of this study was to determine the incidence, types, and clinical implications of laterality errors and the effect of voice recognition software on the frequency of laterality errors.

Methods: All radiology reports generated between January 2007 and April 2011 were retrospectively evaluated to identify revised reports containing laterality errors. Type of error was catalogued with regard to modality, body part, type of discrepancy (major or minor, with discrepancies considered major if the potential existed to affect patient management), duration of time between report finalization and corrected report, clinical significance, and use of voice recognition. The rate of errors causing major and minor discrepancies between voice recognition–generated reports and nonvoice recognition–generated reports was compared.

Results: Among 2,923,094 reports, 1,607 (0.055%) contained corrected laterality errors, and 56 (0.0019% of the total report volume) were major. A total of 584,878 (20%) were generated using voice recognition. The rate of laterality errors leading to major discrepancies in voice recognition–generated reports was 0.00188%, compared with 0.00192% in nonvoice recognition–generated reports ($P = .9436$). None of the errors led to wrong-sided surgery. However, there were potential adverse effects due to laterality errors in 3 patients with major discrepancies (0.000103% of the total report volume).

Conclusions: Rates of laterality errors were low and, in our population, did not result in wrong-sided surgeries. Rates of laterality errors in reports with major discrepancies were unaffected by voice recognition software, but voice recognition was associated with a significant reduction in the duration of time between report finalization and the issuing of a corrected report.

Key Words: Laterality errors, voice recognition, quality

J Am Coll Radiol 2013;10:538-543.

INTRODUCTION

Discrepancies in the laterality of diagnostic image diagnosis represent a serious source of error that may result in severe clinical sequelae, including wrong-sided surgery, procedure, drain and tube placement, and endovascular treatment [1-6]. Notwithstanding the importance of errors in laterality in radiology reports, the incidence and clinical relevance of such errors remains relatively poorly studied despite significant attention given to the subject in the radiology literature and lay media. Furthermore, to our knowledge, the role of voice recognition software, as opposed to directly transcribed dictation, in the incidence of laterality errors has not previously been evaluated. The purpose of this study was to evaluate the incidence, types, and clinical implications of laterality errors at our institution and to evaluate the effect of voice recognition software on the frequency of laterality errors.

METHODS

Data Acquisition

Study design and execution were subject to institutional review board oversight, and all clinical data were handled

This article is available for CME credit online and CME credit may only be claimed online. Visit www.acr.org, ACR Education, online learning for more information.

[a]Mayo Medical School, Rochester, Minnesota.

[b]Department of Radiology, Mayo Clinic, Rochester, Minnesota.

Corresponding author: Christopher H. Hunt, MD, Mayo Clinic, Department of Radiology, 200 First Street SW, Rochester, MN 55905; e-mail: hunt.christopher@mayo.edu.

0091-2182/13/$36.00 • http://dx.doi.org/10.1016/j.jacr.2013.02.017

in a manner consistent with HIPAA guidelines. We retrospectively reviewed all finalized radiology examination reports, excluding outside films, from our institution between January 1, 2007, and April 28, 2011. Using our proprietary electronic radiology information system (RIS), a search was performed for reports that contained "corrected laterality errors." For a corrected report, our RIS maintains two copies of the report. The first copy is the corrected version without markup. The second copy is the original report with the error struck through and the corrected change in place without strikethrough. Corrected laterality errors, regardless of whether the original report was generated with voice recognition or through direct dictation, can be searched using a word-string search of the marked-up finalized report for "left strikethrough right" and "right strikethrough left" (eg, "right ~~left~~" and "left ~~right~~") as well as the other permutations. For revised reports with corrected laterality errors, we obtained the following data from our RIS: type of laterality error with regard to modality, body part, tissue type, type of examination, duration of time between report finalization and the generation of a corrected report, radiologist, and use of voice recognition software (PowerScribe SDK; Nuance Communications, Inc, Burlington, Massachusetts). Our current level of integration between our voice recognition software and our RIS does not allow a finalized report to be corrected through voice recognition. Before the implementation of voice dictation software, reports were transcribed using either direct dictation or traditional telephone dictation.

Additionally, the Mayo Clinical Notes Search Tool was used to perform a patient care *International Classification of Diseases*, ninth rev, procedure and diagnosis code search in the Mayo Clinic Life Sciences System, which is a clinical data repository, to search for wrong-sided surgery (code E8767) in all patients with discrepancies due to laterality errors. Additional searches were performed for "performance of wrong operation (procedure) on correct patient" (code E8765), "other specified misadventures during medical care" (code E8768), and "unspecified misadventure during medical care" (code E8769) [7].

The time of original report generation was classified according to the established institutional work shifts. Reports generated between 7 AM and 5 PM were classified as "day," those between 5 PM and 11 PM as "evening," and those between 11 PM and 7 AM as "overnight." At our institution, all overnight reports by residents are finalized and sent directly to the electronic medical record with staff review performed the following morning starting at 5 AM. Reports generated by residents during the day and evenings are not finalized without staff approval. As a result, laterality errors made by residents during the day and evening, but caught by the staff before being approved and finalized to the electronic medical record, are not included in this study.

Corrected Laterality Error Classifications

At our institution, report discrepancies are graded on the basis of a modified version of the scoring system of Melvin et al [8] for discrepancies (Table 1). Report discrepancies are categorized as either major or minor discrepancies, as determined by the attending radiologist at the time of report revision. Major discrepancies were defined as report errors that have the potential to influence patient management, particularly within the first 24 hours of report generation. As a matter of institutional policy, major discrepancies required direct communication with the physician(s) responsible for direct patient care and documentation of this communication. Minor discrepancies are defined as report errors that are felt to be incidental to treatment and management. The "significant" category advocated by Melvin et al is not used at our institution, as we require grading of an error at the time of the correction, and outcome results are not typically known at the time of the correction. In general, errors that would be classified as significant under this classification scheme would be included in our major discrepancy category. Twenty percent of minor discrepancies were sampled at random to achieve a 5% margin of error using a 95% confidence interval. We also used our institutional electronic medical record to identify the

Table 1. Discrepancy definition with representative examples

Type of Discrepancy	Significance	Examples of Discrepancy
Major	An error that has the potential to be significant, affecting patient management or outcome	• Wrong side of tumor going to surgery • Wrong side of renal mass on CT • Wrong side of suspicious mass on mammogram being sent for diagnostic workup
Minor	An error that is incidental to treatment or management	• Wrong side for a calcified granuloma in the lung • Wrong side for acute fracture in an uncomplicated trauma in an extremity (clinically obvious) • Wrong side for a device (ie, left-sided central venous catheter is said to be on the right side)

Note: The scheme used at our institution is based on a modified version of the scheme suggested by Melvin et al [8].

significance and sequelae of these corrected laterality errors. Clinical records of all patients with major discrepancies were evaluated by a single reviewer (M.T.L.) to determine the clinical consequences of the laterality errors. Similarly, 310 of 1,551 minor discrepancies were sampled at random to achieve a 5% margin of error using a 95% confidence interval.

Examination Classification Scheme

Examination modalities in the study included CT, digital radiography, mammography, MR, nuclear medicine, neuroradiologic procedures, fluoroscopy, ultrasound, nuclear cardiology, and vascular and interventional radiology. Errors were classified by body part into one of the following groups: head and neck, chest, breast, spine, abdomen, pelvis, and extremity. Laterality error rates were determined for each modality, body part, and tissue type (eg, parenchymal or mesenchymal, bone, vascular, or device). Device classifications were used when there was a discrepancy in the location of a peripherally inserted central catheter, shunt, chest tube, and so on.

Voice Recognition Software and Transcription

The use of voice recognition software began at our institution on July 1, 2008, and was implemented progressively among the various divisions of the Department of Radiology over the next 10 months. Because of the lack of a uniform start date, the rates of corrected laterality errors were compared in reports generated with and without voice recognition rather than comparing rates before and after the date of voice recognition implementation. The "without voice recognition" category includes the other transcription processes used at our institution (direct dictation and traditional telephone dictation).

Statistical Analysis

All statistical analyses were performed using JMP version 9 (SAS Institute Inc, Cary, North Carolina), StatTools version 5.6 (Palisade Corporation, Ithaca, New York), and Excel 2008 (Microsoft Corporation, Redmond, Washington). Significance was assigned to all statistical tests with P values $<$.05. The influence of voice recognition software use on the rate of laterality errors causing major and minor discrepancies was determined by comparing laterality rates of reports generated with and without voice recognition software using the χ^2 test. Differences in laterality error rates among body part, modality, and finalization time of day were also evaluated using Pearson's χ^2 test followed by a comparison of multiple proportions using the Marascuilo procedure [9]. The delay between preliminary report submission and finalized report generation was analyzed using Wilcoxon's rank-sum test.

Table 2. Rate of laterality errors in revised reports generated with and without voice recognition software

Type of Discrepancy	Laterality Error Rate (Errors/Report Total)	*P*
Major with voice recognition	11/584,878 (0.00188%)	.9436
Major without voice recognition	45/2,338,216 (0.00192%)	
Minor with voice recognition	359/584,878 (0.06138%)	.0020
Minor without voice recognition	1,192/2,338,216 (0.05098%)	

RESULTS

Among 3,085,045 radiology examinations associated with 2,923,094 reports read by a total of 171 different radiologists, 1,607 (0.055%) contained corrected laterality errors, and 56 (0.0019%) were defined as major discrepancies. A total of 584,878 (20%) were generated using voice recognition. The rate of laterality errors leading to major discrepancies in voice recognition–generated reports was 0.00188%, compared with 0.00192% in nonvoice recognition–generated reports (P = .9436). The rate of laterality errors leading to minor discrepancies in voice recognition–generated reports was 0.06138%, compared with 0.05098% in nonvoice recognition–generated reports (P = .0020) (Table 2).

The mean duration of time between report finalization and the issuing of a corrected report in all reports using voice recognition was 134 ± 65.9 hours (median, 5.7 hours; range, 0.02-12,479.4 hours; 95% confidence interval, 4.7-263 hours); the mean duration of time between report finalization and the generation of a corrected report in all nonvoice recognition–generated reports was 279 ± 36.0 hours (median, 9.8 hours; range, 0.02-17,668 hours; 95% confidence interval, 209-350 hours) ($P <$ .0001).

Laterality Errors and Body Part Imaged

Examination for overall effect among body part using Pearson's χ^2 test yielded significant differences ($P <$ .0001). Using comparison of multiple proportions, laterality error rates were highest in reports pertaining to chest imaging ($P <$ .0001), with all other pairwise interactions significant ($P <$ .05) with the exception of the abdomen and pelvis, head and neck, and breast (Table 3). In the chest, 48.9% of errors were discrepancies in parenchymal or mesenchymal tissue, primarily the lungs; 38.3% in device locations; 9.5% in bone, primarily the ribs; and 3.6% vascular.

Laterality Errors and Modalities

Examination for overall effect among report modalities using Pearson χ^2 test yielded significant differences ($P <$ .0001). Using comparison of multiple proportions, fluoroscopy had a significantly lower rate of laterality errors (0.013%) compared with CT, digital radiography, mam-

Table 3. Rate of laterality errors in revised reports and mean duration of time between finalization of original report and time of issuing corrected report with regard to body part

Body Part	Laterality Error Rate (Errors/Report Total)	Odds Ratio*	Mean Duration of Time to Report Correction (h)
Spine	32/2,409,191 (0.00133%)	1.0	9.1 ± 222.9
Abdomen	154/2,681,399 (0.00574%)	4.3	95.2 ± 101.6
Chest	548/2,863,223 (0.01914%)	14.4	113.4 ± 53.9
Pelvis	115/2,418,447 (0.00476%)	3.6	214 ± 117.6
Extremity	327/2,658,842 (0.01230%)	9.3	295 ± 69.8
Head and neck	211/2,395,890 (0.00881%)	6.6	341.1 ± 86.8
Breast	220/2,530,159 (0.00870%)	6.5	568 ± 85

*Relative to the frequency of laterality errors in spine imaging.

mography, MR, and ultrasound ($P < .05$; Table 4). Nuclear medicine had a significantly lower rate of laterality errors (0.034%) compared with ultrasound, CT, and mammography ($P < .05$). Mammography had a significantly higher rate of errors (0.092%) in comparison with digital radiography and vascular and interventional radiology ($P < .05$). CT had a significantly higher error rate (0.071%) than digital radiography (0.046%) ($P < .0001$). Nuclear cardiology, as might be expected, did not have any laterality errors.

Laterality Errors and Time of Day

Laterality error rates among shifts were determined on the basis of a 1-month sample of radiology report volumes in which 83% of reports were finalized during the day, 11% during the evening, and 6% overnight. Laterality error rates were significantly higher in reports finalized during the evening and overnight shifts (0.154% and 0.124%, respectively) compared with the error rate in reports finalized during the day (0.0372%) ($P < .0001$; Table 5). Additionally, the delay between preliminary report generation and the finalization of a corrected report was greatest among reports generated during the day compared with reports generated at night or overnight (Table 5).

Table 4. Rate of laterality errors by modality, standardized by report volumes

Modality	Report Error Rate: Errors/Total Reports	Report Error Rate: %
Nuclear cardiology	0/28,058	0.00
Radio fluoroscopy	3/22,557	0.013
Vascular and interventional radiology	13/38,299	0.034
Nuclear medicine	52/150,863	0.034
Digital radiography	695/1,507,910	0.046
MR	143/235,473	0.061
CT	307/431,532	0.071
Ultrasound	190/294,765	0.064
Mammography	188/204,381	0.092
Neuroradiologic procedures	16/9,256	0.173

Clinical Outcomes

None of the 56 major discrepancies led to wrong-sided surgery or intervention. However, there were adverse effects in 3 (5.4%) of the 56 patients with major discrepancies, or 0.000103% of the entire database. The discrepancy in 1 patient was noted intraoperatively after skin markings were erroneously placed on the incorrect side during the excision of melanoma on the upper back. However, the incision was not made. In 2 separate cases with major discrepancies, laterality errors led to delayed diagnoses of breast cancer and renal cell carcinoma, respectively. In both cases, the disease was allowed to progress for 4 months past the time of initial diagnosis. There were no adverse effects among the 310 randomly audited patients with minor discrepancies. The clinical data repository search failed to identify any additional adverse outcomes or wrong-sided procedures.

DISCUSSION

In the present study, we found that the rate of major laterality errors was similar for reports generated with and without voice recognition software. Although there was a significant difference in rates of minor laterality errors for reports generated with and without voice recognition software, these errors were deemed to be clinically insignificant. This was substantiated as an audit of the minor discrepancies failed to demonstrate any clinical impact with either change in management or outcome. Fortunately, the clinical impact of major laterality errors seems modest, with adverse events limited to an extremely small number of patients in our series (0.000103%). Despite subjective complaints by some radiologists that voice recognition slows down report times, the use of voice recognition software was associated with a significant reduction in the delay both between initial report finalization and the generation of a corrected report.

We demonstrated significant differences in rates of laterality errors by body part, with the chest suffering the highest rate. The majority of these errors were discrepancies between lungs, followed by discrepancies in the location of peripherally inserted central catheters, shunts, and chest tubes. This study confirms the need for careful

Table 5. Laterality error rate and mean duration of time between finalization of original report and time of issuing corrected report with regard to time of day of finalization of original report

Time of Day	Laterality Error Rate	*P* Day	*P* Evening	*P* Overnight	Time to Corrected Report (h), Median (IQR)	*P* Day	*P* Evening	*P* Overnight
Day	0.0372% (906/2,433,235)		<.0001	<.0001	21.5 (1.5-119.1)		<.0001	<.0001
Evening	0.154% (475/307,808)	<.0001		.0212	8.3 (1.0-14.3)	<.0001		.0634
Overnight	0.124% (226/182,051)	<.0001	.0212		5.3 (2.6-8.4)	<.0001	.0634	

Note: IQR = interquartile range.

scrutiny of laterality when describing the location of indwelling devices.

The laterality error rate in reports finalized during the evening and overnight shift was 4 times higher than that of reports finalized during the day; however, the duration of time between report finalization and the generation of a corrected report was >2 times as great in the day reports in comparison with those originally finalized during the evening and overnight. The reason for this is likely multifactorial. First, there is likely a significant component of fatigue in the rate of laterality errors. In addition, with the increasing nocturnal work volume being seen over the past decade, careful screening for laterality errors, especially during the overnight shifts, becomes even more important. Second, there is likely a significant underestimation of the rate of laterality errors that occur in the day and are not corrected. Reports generated and finalized during the evening and overnight shifts are made by residents. The next morning, a staff radiologist reviews these reports, thereby providing a second level of quality control that is not present in the day, when the staff radiologist is reading alone. As a result, the day laterality error rate is almost certainly spuriously low, as the laterality error probably comes to the point of detection only if a clinician notices it and notifies the staff radiologist, who then corrects the error. The importance of careful assignment of laterality regardless of the time of the day is especially important for staff radiologists.

Previous studies have evaluated the use of voice recognition software in radiology [7,10-18], but the majority focused on cost-effectiveness and reporting time. Of those analyzing accuracy, several found increases in the frequency of errors without the use of voice recognition [10,11,14]. However, these studies evaluated for errors in syntax, grammar, spelling, and so on, and did not specifically address laterality errors, which are more likely to be clinically relevant. Furthermore, they were limited by small sample sizes and evaluated fewer radiologic modalities.

Quint et al [13] documented the frequency of laterality errors in radiology reports generated using voice recognition software. Those authors found a 1.5% (4 out of 265) incidence of laterality errors. However, they did not compare this rate with the rate of laterality errors in reports using other transcription methods, and they only evaluated a limited number of CT images, limiting the ability to generalize their study. Our overall laterality error rate of 0.055% was considerably higher than the 0.00008% incidence of corrected laterality errors reported by Sangwaiya et al [1]. This may be because they defined a laterality error to be a discrepancy between the impression and body of the radiology report and only included reports that mentioned both the right and left sides, which we feel greatly underestimates error rates.

Laterality error rates were significantly higher in reports finalized during the evening and overnight shifts, which may be due to several factors. At our institution, the majority of cases read during the evening shift are read by second-year and third-year residents, and the majority of cases read during the overnight shift are read by fourth-year residents. All trainee reports, approximately 17% of total report volume, are audited by staff radiologists within 24 hours, which leads to a higher rate of error detection because each imaging study as well as each report is essentially reviewed twice. Second, trainees most likely make a higher percentage of laterality errors. This finding is supported by our finding that there was a significant reduction in laterality errors between reports read by second-year and third-year residents during the evening and those read by fourth-year residents overnight. Finally, the increased overnight laterality error rate could be a result of increased fatigue inherent to most overnight shifts.

Our retrospective study had a number of important limitations. Because of the large number of reports and examinations (our institution performs >1 million radiologic examinations per year), examinations interpreted by staff radiologists during the day were not rechecked by other staff members. Thus, the true laterality error rate may be underrepresented. Second, because of the large report volume, we reviewed the clinical records of 310 patients rather than all patients with minor discrepancies to determine the clinical significance of errors. Although we have a structured grading system based on accepted guidelines, classification of a discrepancy as major or minor depended on the opinion of the attending radiologist at the time of the correction and therefore was inherently subjective. However, the primary focus of this study was not the clinical significance of errors but rather our best efforts to determine, in a large series of radiology reports, the approximate incidence of laterality errors and voice recognition's effect on the frequency and type of laterality errors.

CONCLUSIONS

Rates of laterality errors were low and, in our population, did not result in wrong-sided surgeries, but they have the potential to cause significant patient harm. Rates of laterality errors in reports with major discrepancies were unaffected by voice recognition software; however, voice recognition was associated with a significant reduction in the duration of time between report finalization and the issuing of a corrected report.

TAKE-HOME POINTS

- Laterality error rates are low but have the potential to cause significant patient harm.
- Voice recognition does not change the rate of laterality errors.
- Voice recognition is associated with a shorter duration between making a laterality error and issuing a revised report.

REFERENCES

1. Sangwaiya MJ, Saini S, Blake MA, Dreyer KJ, Kalra MK. Errare humanum est: frequency of laterality errors in radiology reports. AJR Am J Roentgenol 2009;192:W239-44.
2. Kwaan MR, Studdert DM, Zinner MJ, Gawande AA. Incidence, patterns, and prevention of wrong-site surgery. Arch Surg 2006;141:353-7.
3. Seiden SC, Barach P. Wrong-side/wrong-site, wrong-procedure, and wrong-patient adverse events: are they preventable? Arch Surg 2006;141: 931-9.
4. Jhawar BS, Mitsis D, Duggal N. Wrong-sided and wrong-level neurosurgery: a national survey. J Neurosurg Spine 2007;7:467-72.
5. Christensen JD, Hutchins GC, McDonald CJ. Computer automated detection of head orientation for prevention of wrong-side treatment errors. AMIA Annu Symp Proc 2006:136-40.
6. Clarke JR, Johnston J, Blanco M, Martindell DP. Wrong-site surgery: can we prevent it? Adv Surg 2008;42:13-31.
7. Centers for Medicare and Medicaid Services. Diagnosis and procedure codes: abbreviated and full code titles. Baltimore, Maryland: Centers for Medicare and Medicaid Services; 2011.
8. Melvin C, Bodley R, Booth A, Meagher T, Record C, Savage P. Managing errors in radiology: a working model. Clin Radiol 2004;59:841-5.
9. Zwick R, Marascuilo L. Selection of pairwise multiple comparison procedures for parametric and nonparametric analysis of variance models. Psychol Bull 1984;95:148-55.
10. Rana DS, Hurst G, Shepstone L, Pilling J, Cockburn J, Crawford M. Voice recognition for radiology reporting: is it good enough? Clin Radiol 2005;60:1205-12.
11. McGurk S, Brauer K, Macfarlane TV, Duncan KA. The effect of voice recognition software on comparative error rates in radiology reports. Br J Radiol 2008;81:767-70.
12. Lemme PJ, Morin RL. The implementation of speech recognition in an electronic radiology practice. J Digit Imaging 2000;13:153-4.
13. Quint LE, Quint DJ, Myles JD. Frequency and spectrum of errors in final radiology reports generated with automatic speech recognition technology. J Am Coll Radiol 2008;5:1196-9.
14. Pezzullo JA, Tung GA, Rogg JM, Davis LM, Brody JM, Mayo-Smith WW. Voice recognition dictation: radiologist as transcriptionist. J Digit Imaging 2008;21:384-9.
15. Bhan SN, Coblentz CL, Norman GR, Ali SH. Effect of voice recognition on radiologist reporting time. Can Assoc Radiol J 2008;59:203-9.
16. Gutierrez AJ, Mullins ME, Novelline RA. Impact of PACS and voice-recognition reporting on the education of radiology residents. J Digit Imaging 2005;18:100-8.
17. Langer S. Radiology speech recognition: workflow, integration, and productivity issues. Curr Probl Diagn Radiol 2002;31:95-104.
18. Chang CA, Strahan R, Jolley D. Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. J Digit Imaging 2011;24:724-8.

CME: This article is only available for CME credit online and CME credit may only be claimed online. Visit www.acr.org, ACR Education, online learning for more information.