Prepopulated Radiology Report Templates: A Prospective Analysis of Error Rate and Turnaround Time

C. M. Hawkins · S. Hall · J. Hardin · S. Salisbury ·
A. J. Towbin

Published online: 25 January 2012
© Society for Imaging Informatics in Medicine 2012

Abstract Current speech recognition software allows exam-specific standard reports to be prepopulated into the dictation field based on the radiology information system procedure code. While it is thought that prepopulating reports can decrease the time required to dictate a study and the overall number of errors in the final report, this hypothesis has not been studied in a clinical setting. A prospective study was performed. During the first week, radiologists dictated all studies using prepopulated standard reports. During the second week, all studies were dictated after prepopulated reports had been disabled. Final radiology reports were evaluated for 11 different types of errors. Each error within a report was classified individually. The median time required to dictate an exam was compared between the 2 weeks. There were 12,387 reports dictated during the study, of which, 1,173 randomly distributed reports were analyzed for errors. There was no difference in the number of errors per report between the 2 weeks; however, radiologists overwhelmingly preferred using a standard report both weeks. Grammatical errors were by far the most common error type, followed by missense errors and errors of omission. There was no significant difference in the median dictation time when comparing studies performed each week. The use of prepopulated

reports does not alone affect the error rate or dictation time of radiology reports. While it is a useful feature for radiologists, it must be coupled with other strategies in order to decrease errors.

Keywords Standardized report · Structured report · Prepopulated reports · Speech recognition · Turnaround time

Introduction

There are many benefits to using speech recognition software to dictate radiology reports. Its use has dramatically decreased the amount of time between performing a radiology study and having a signed, dictated report available to the healthcare providers caring for the patient. Many radiology departments have demonstrated a marked improvement in turnaround time after implementing speech recognition software [1–3].

While there are many benefits to using speech recognition software in the radiology department, one disadvantage is that final radiology reports are often fraught with errors. Prior studies have shown that 4.8–22% of reports generated using speech recognition software contain errors [4–6]. Errors in radiology reports undermine the radiologist in two significant ways. First, the large number of errors can give the person reading the report the impression that the radiologist does not pay attention to detail [7]. Second, even though the vast majority of errors are likely clinically insignificant, errors that alter the meaning of the report do occur and carry potentially significant clinical consequences.

A new feature in commercially available speech recognition software allows standardized reports to be prepopulated into the dictation field based on the radiology information system procedure code. It is thought that prepopulating a

C. M. Hawkins · S. Hall · J. Hardin · A. J. Towbin (✉)
Department of Radiology,
Cincinnati Children's Hospital Medical Center,
3333 Burnet Avenue, ML 5031,
Cincinnati, OH 45229, USA
e-mail: alexander.towbin@cchmc.org

S. Salisbury
Department of Biostatistics and Epidemiology,
Cincinnati Children's Hospital Medical Center,
3333 Burnet Avenue, ML 5031,
Cincinnati, OH 45229, USA

standard report provides a more efficient and less error prone process for radiologists.

The purpose of this study is to evaluate the dictation time and frequency of use of the standard report, as well as the frequency and types of errors in radiology reports by comparing two scenarios: when a standardized report is prepopulated in the dictation field and when no report is prepopulated in the dictation field. We hypothesize that the overall dictation time and error rate of final radiology reports can be decreased by using prepopulated standard reports. In addition, we hypothesize that grammatical errors are the most frequent type of error.

Method and Materials

Study Design

After a waiver was obtained from the institutional review board, a prospective study was performed. All radiology examinations dictated using the departmental speech recognition system (RadWhere, Nuance, Boston, MA, USA) for two consecutive weeks were included in this study.

During the 2 weeks of the study, all radiology reports, digital audio files, and dictation times were logged into a Microsoft Access database (Microsoft, Redmond, WA, USA). During the first week of the study, standard reports were prepopulated into the dictation field without a radiologist needing to manually select the report from a separate list, our institution's current standard practice. During the second week, the prepopulated templates were disabled, forcing users to begin with a blank screen and either dictate freely or manually select a standard report template from a list of standard reports.

Following the study weeks, two radiologists analyzed a random sample of the reports. Reviewer 1 was a second year radiology resident and reviewer 2 was a board-certified pediatric radiologist with 3 years of experience. Each radiologist compared the text report to its corresponding audio file. The two reviewers used the same prespecified definitions of errors with the exception of grammatical errors. Reviewer 1 used a more lax definition of grammatical errors to include all nonstylistic errors. Reviewer 2 used a more strict definition of grammatical errors to include all grammatical errors including purposeful, stylistic errors such as dictating using sentence fragments.

Reports were first classified as having used the standardized report (structured) or as being freely dictated (prose). For the purpose of this study, a report was classified as structured if it contained any structured element from the standard report.

Each report was then evaluated for the presence of an error or multiple errors. Each error within a report was

individually categorized and assigned one of five main classes: nonsense errors, missense errors, spelling/grammatical errors, translation errors, and errors of omission/commission. Each class of error was further subdivided for a total of 11 specific types of errors. Table 1 provides a definition and example for each type of error. Interpretation errors were not evaluated in this study.

The error rate, frequency of use of the standard report, and report dictation times were compared between the 2 weeks. Several error rates were calculated, including the average number of errors per report, average number of nongrammatical errors per report, the percentage of reports with an error, the percentage of reports with a nongrammatical error, and the average number of errors per dictated word. The formula used to calculate the different error rates is shown in Table 2. In addition to the general error rates, the frequency of each specific type of error was calculated (Table 3).

Dictation time was defined as the length of time between when the dictation was opened in the speech recognition application and when it was signed by the dictating radiologist. It should be noted that the dictation system is integrated with our picture archiving and communication system (PACS) so that when the study is opened in the PACS, it automatically launches the dictation. Thus, in our environment, the dictation time also incorporates the time required to interpret the study. Dictation time was compared between the 2 weeks for all modalities performed Monday through Friday. However, because dictation time is included in the interpretation time, studies that take longer to interpret, such as MRIs and CTs, are less likely to show small differences in dictation time that result from using prepopulated reports. Because of this, dictation time for radiographs was separately analyzed as radiographs are often the shortest reports and it was thought that small differences in dictation times gained by prepopulated reports would be more apparent.

Statistical Analysis

All statistical analyses were performed using SAS® (Version 9.2 Cary, NC, USA). All tests were two sided and were specified a priori. *P* values of less than 0.05 were considered to indicate statistical significance. Reported *p* values are not adjusted for multiple testing.

Nongrammatical errors were compared between the 2 weeks. Initially, the Poisson, negative binomial, zero-inflated Poisson, or zero-inflated negative binomial distributions were all examined using the Schwarz Bayesian Information Criterion (SBC) in the CountReg procedure. These four distributions were evaluated to see which distribution best modeled our zero-inflated data. The negative binomial distribution had the smallest SBC and was considered the best

Table 1 Error types

Error category	Error type	Definition	Intended/spoken phrase	Transcribed/reported phrase
Nonsense	Nonsense	Passages/words/phrases that make no sense or have no sensible meaning	The lungs are clear	The lungs nuclear
Missense	Missense—translational	Translation error that changes the meaning of a phrase/sentence	There is no opacity in the left lung	There is an opacity in the left lung
	Missense—omission	Words not transcribed (omitted) that subsequently change the meaning of a phrase/sentence	There is no pneumonia	There is pneumonia
Spelling/grammar	Missense—human	Human error that changes the meaning of a sentence	Right lower lobe pneumonia	Left lower lobe pneumonia
	Spelling/grammar—typographical error	Grammatical/spelling error resulting in a typographical error that is not a nonsense/missense error	The right lung is clear	The right lung is clear
	Spelling/grammar—homonym Error	Misuse of words or phrases that sound the same, but are semantically distinct	The right lung is clear	The write lung is clear
	Spelling/grammar—grammatical error	Standard grammatical errors including the use of sentence fragments	The lungs are clear	The lungs is clear
	Improper period use error	Duplicate periods or lack of a period at the end of a sentence	The lungs are clear.	The lungs are clear..
Omission/commission	Omission—omission error (other)	Omitted word/phrase that does not result in a missense or nonsense error	The lungs are clear	Lungs are clear
	Commission error	Retained statement from a standardized template that contradicts the dictated findings or impression	There is a hazy opacity obscuring the right hemidiaphragm	The lungs are clear. There is a hazy opacity obscuring the right hemidiaphragm
Translational	Translational (other)	Translation error that does not result in a nonsense/missense error. The resulting sentence still has sensible meaning, as opposed to nonsense errors	The lungs are clear	The lungs are clean

Table 2 Error rate formulas and week-to-week comparison

Error rate	Formula	Week 1			Week 2		
		Reader 1	Reader 2	Total	Reader 1	Reader 2	Total
Errors per report	Total errors/total reports	200/219=0.91	550/223=2.47	750/442=1.7	356/350=1.02	906/381=2.38	1,262/731=1.73
Nongrammatical errors per report	Total nongrammatical errors/total reports	122/219=0.56	113/223=0.51	235/442=0.53	227/350=0.65	211/381=0.55	438/731=0.60
Percentage of reports with errors	((Total reports – reports with no errors)/total reports)×100%	46.6%	71.7%	((442–180)/442)×100%=59%	48.6%	70.7%	((731–292)/731)×100%=60%
Percentage of reports with nongrammatical errors	(Total reports – Reports with no nongrammatical errors/Total reports)×100%	31.5%	34.1%	((442–297)/442)×100%=33%	36.9%	35.2%	((731–468)/731)×100%=36%
Errors per dictated word	Total errors/Total dictated words	0.013	0.028	729/34,641=0.02	0.012	0.036	1,229/54,515=0.02

Total errors: week 1=750, week 2=1,262; total nongrammatical errors: week 1=235, week 2=438; reports with no errors: week 1=180, week 2=292; reports with no nongrammatical errors: week 1=297, week 2=468; total reports: week 1=442, week 2=731; total dictated words: week 1=34,641, week 2=54,515

Prepopulated standardized reports were used during week 1. Prepopulated templates were disabled during week 2; however, radiologists could still choose to select a standard template manually

distribution for analysis. Analysis was performed using the GLIMMIX procedure, as this procedure type can be used for a negative binomial distribution.

In order to compare dictation time between the 2 weeks, the nonparametric exact Wilcoxon–Mann–Whitney test was performed using 10,000 Monte Carlo simulations. Dictation time is reported as medians and interquartile range (25th, 75th percentiles). The median was used to better represent the central location of dictation times and is less affected by dictation time outliers. Additionally, the interquartile range (IQR) best represents the variation of the spread of observations.

Results

There were 12,387 reports dictated over the course of the study (4,873 in week 1 and 7,514 in week 2). Of these, 1,208 (9.75%) reports were randomly selected. Thirty-five duplicate reports were removed leaving a total of 1,173 for the study. Reports were randomly assigned to one of two reviewers. Reviewer 1 evaluated 569 reports (219 from week 1 and 350 from week 2) while reviewer 2 evaluated 604 reports (223 from week 1 and 381 from week 2).

Radiologists used a standard report template 96% of the time during week 1 when reports were prepopulated into the dictation field and 86% of the time during week 2 when no report was prepopulated into the dictation field. There were 76.8 dictated words per report in week 1 and 72 dictated words per report in week 2.

Amongst reviewers, there was no significant difference in any of the calculated error rates between the two study weeks (Table 2). The differing error rate between the two reviewers is attributable to difference in definition of grammatical errors. Table 3 demonstrates the distribution of all error subtypes as well as the number of reports required to see one error. Overall, there were 1.70 errors per report during week 1, when the standard templates were prepopulated, and 1.73 errors per report during week 2 when prepopulated templates were disabled.

When looking at the frequency of the different types of errors, grammatical errors were by far the most common, accounting for nearly 60% of the total errors each week and approximately 1 error per report. The vast majority of these errors were due to the style of dictation (i.e., using sentence fragments instead of complete sentences). When grammatical errors were excluded, there was no significant difference in the error rate between the 2 weeks ($p=0.7106$). There were 0.53 nongrammatical errors per prepopulated report and 0.6 nongrammatical errors per report when prepopulated templates were disabled.

Missense errors were the most common nongrammatical type of error, accounting for 12% of all errors (10% during

Table 3 Distribution of Error Types in all Edited Reports

Error type	Total number of errors			Percentage of total errors (%)			Errors per report			Reports per 1 error		
	Reader 1	Reader 2	Total	Reader 1	Reader 2	Total	Reader 1	Reader 2	Total	Reader 1	Reader 2	Total
Nonsense error	43	76	119	7.7	5.2	5.9	0.08	0.13	0.1	12.5	7.7	10
Missense—translational Error	88	111	199	15.8	7.6	9.9	0.15	0.18	0.17	6.7	5.6	6
Missense—omission error	23	2	25	4.1	0.1	1.2	0.04	0.003	0.02	25	333	50
Missense—human error	11	6	17	2	0.4	0.8	0.02	0.01	0.014	50	100	67
Spelling/grammar—typographical error	8	15	23	1.4	1	1.1	0.01	0.02	0.02	100	50	50
Spelling/grammar—homonym error	0	14	14	0	1	0.7	0	0.02	0.01	0	50	100
Spelling/grammar—grammatical error	207	1027	1234	37.2	70.5	61	0.36	1.7	1.05	2.8	0.6	1
Improper period use error	64	55	119	11.5	3.8	5.9	0.11	0.09	0.1	9.1	11.1	10
Omission—omission error (other)	49	92	141	8.8	6.3	7	0.09	0.15	0.12	11.1	6.7	8
Omission—commission error	13	26	39	2.3	1.8	1.9	0.02	0.04	0.03	50	25	33
Translational (other)	50	32	82	9	2.2	4.1	0.09	0.05	0.07	11.1	20	14

Total errors=2,012 (reader 1=556, reader 2=1,456). Total reports=1,173 (reader 1=569, reader 2=604)

Sample calculation (nonsense error) reader 1: (43 nonsense errors/556 total errors)×100%=7.7% of all errors; 43 nonsense errors/569 total reports=0.08 errors per report; 1/0.08 errors per report=1 error per 12.5 reports

week 1 and 13.2% during week 2). Overall, there were 0.21 missense errors per report. Missense errors were categorized into three subtypes: translational, omission, and human errors. The most common subtype of missense error was a translational missense error, which accounted for 82.6% of missense errors and 9.9% of all errors (0.17 errors per report).

Errors of omission were the third most common error and accounted for 10.1% of all errors. Errors of omission that changed the meaning of a phrase or sentence (missense omission errors) accounted for 1.2% of all errors (0.02 errors per report). Table 3 describes the frequency of the remaining types of errors.

There was no difference between the 2 weeks in dictation time for all modalities combined ($p=0.9156$). The median (IQR) dictation time in minutes was 4 (2, 10) for both weeks (Table 4). Plain radiographs were the most common type of study obtained in our department and the one most likely to show small differences in dictation time. However, there was not a significant difference between the weeks ($p=0.9617$). The median (IQR) dictation time for radiographs in minutes was 3 (1, 5) for week 1 and 3 (2, 5) for week 2 (Table 4).

Table 4 Turnaround time in minutes

	Week 1 median (IQR) ^a	Week 2 median (IQR) ^a
Turnaround time (all modalities)	4 (2, 10)	4 (2, 10)
Turnaround time (plain radiographs)	3 (1, 5)	3 (2, 5)

^a Interquartile range: 25th and 75th percentiles

Discussion

The most striking result of this study was the strong preference radiologists demonstrated for using the departmental standard report templates. This preference was further supported by the similar number of dictated words per report between the two study weeks. When reports were not prepopulated, radiologists still manually selected a standard template 86% of the time when presented with a blank screen, even though they had the option to dictate in prose. This result is surprising as it runs counter to one of the major arguments against standardized reporting—the lack of autonomy of the radiologist to dictate freely [7]. While it is possible that this finding is unique to our institution due to our preexisting practice of prepopulating a standard report in the dictation field, we believe this still demonstrates the appeal that a standardized report template offers to the radiologist.

Because of the strong preference for using standard reports, having a prepopulated structured report in the dictation field did not significantly affect the report dictation time. The time spent manually selecting or using a voice command to select the standard report was not significant when compared to the time required to interpret and dictate a report. In order to better compare small differences in time, the dictation time for plain radiographs was evaluated. Radiography reports were highlighted because these reports tend to be shorter than other types of studies. Even when comparing these reports, there was no difference in dictation time between the 2 weeks.

Errors were common in this study. There were approximately 1.7 errors per report in this study and errors were found in approximately 60% of all reports evaluated. The

error rate observed in this study is much higher than the 4.8–22% error rate published in other studies [4–6].

The higher error rate in this study may be attributed to several factors. First, errors were broadly defined and categorized into one of 11 different error types. The number of different types of errors that were evaluated in this study was greater than in prior studies [4–6]. We believe that by looking for all types of errors, we are able to achieve a more exact error rate.

The second potential reason we believe our error rate was higher than other studies is the high incidence of grammatical errors. Even if the data for reviewer 1 (the reviewer who used a more lax definition of errors) is used, there were a large number of nonstylistic grammatical errors in this study. Examples of these nonstylistic grammatical errors include incorrect verb tenses, missing or inappropriate commas, double periods, extra spaces, and incorrect uppercase or lowercase letters. Grammatical errors occurred much more frequently than other errors, accounting for 61% of all errors. While portions of our definition of what constitutes a grammatical error are controversial, we believe that it best fits the underlying purpose of a radiology report: to communicate results of a radiology examination accurately and clearly.

Even if grammatical errors are excluded from our error rate, the percent of reports with a nongrammatical error was 33% during week 1 and 36% during week 2. This leads to the third potential reason why the error rate was much higher in this study as compared to other studies: to our knowledge, this is the first study to compare the audio files to the transcribed reports directly. By listening to the audio files, we were able to classify many errors that traditionally have been difficult to identify, including many translational errors such as errors of omission (errors that occurred when a spoken word was not transcribed by the speech recognition system). Errors of omission accounted for 10.1% of all errors detected. These errors have the potential to be significant when the omitted word changes the intended meaning of the phrase/passage. Significant errors of this type (mis-sense omission errors) accounted for 1.2% of all errors.

The final potential reason our error rate was higher than other studies lies in our study design. In some studies, emphasis was placed only on “significant” error types, which were defined as errors that could potentially lead to the conveyance of incorrect or confusing information. In our study, all error types, regardless of their effect on message conveyance, were evaluated. If only missense errors were evaluated, our error rate would fall within the range of previously published reports [4, 6]. We believe that because the radiology report is the single most important basis on which radiologists are judged by referring clinicians, every effort should be made to evaluate and eliminate *all* error types [7]. Error-free reports help the radiologist to convey an

aura of professionalism, thoroughness, and attention to detail in each and every report.

This study highlights the need to develop strategies to prevent errors within radiology reports. There are several potential methods to prevent errors, some of which are currently available and some of which can be developed. The first potential method is to separate the dictating and correcting processes. Radiologists often find it difficult to correct their own reports as their mind reads what they think was said rather than what was transcribed. One strategy to decouple the dictating and editing process is to employ transcriptionists to act as correctionists. In this model, the transcriptionist listens to the dictated report and uses the audio file to correct the transcribed report. While a correctionist model may help to decrease the error rate of reports, it does so at considerable expense both financially (hiring correctionists) and in terms of system efficiency (lengthening the turnaround time) [1, 5, 8]. These drawbacks prevent most departments from using a correctionist model.

A second potential method of error prevention is to dictate as few words as possible. This can be achieved by creating and using structured reports. A well-designed structured report allows the radiologist to dictate a study using a small number of words or phrases. Structured reports can be designed so that the dictated phrase incorporates a previously crafted phrase, sentence, paragraph, or even an entire section. By creating the dictation ahead of time, it is often easier to edit the report for errors, including grammatical errors. Since the completion of this study, our department has worked to create department-wide structured reports. These reports are now the prepopulated standard report that is present in the dictation field when a study is opened. We have not yet evaluated if the use of more structured reports has decreased the error rate of reports in the department.

The third potential method for error prevention is for the dictation software to incorporate more error cues into the editing process. While an error cue for misspelled words is already incorporated into most speech recognition software (e.g., word underlined in red), error cues for grammar and translation errors are not. Most users are already familiar with error cues for spelling and grammatical mistakes in word processing applications. Grammatical errors would likely decrease if speech recognition software incorporated the same grammar rules and error cue (e.g., word underlined in green) as the word processing software. In addition to alerting the user to grammatical errors, speech recognition applications could use a unique error cue if the software determined that the transcribed word was unlikely to match the dictated word. At least one speech recognition software company has employed this type of error cue in the past.

There are several limitations to this study. Manual error logging opened the possibility of both undercounting and overcounting errors. An error could be undercounted if it

was not recognized by the reviewer (false negative) and it could overcounted if the reviewer incorrectly graded an error (false positive). Despite this limitation, it should be noted that both graders had nearly identical error rates for all nongrammatical error types. The two reviewers did differ in the number of grammatical errors due to one reviewer using the more strict definition of a grammatical error (to include purposefully dictated sentence fragments as an error). When using a lax definition, approximately 0.4 grammatical errors were detected per report with grammatical errors occurring in approximately 11% of all reports. Using a more strict definition resulted in a detection rate of 1.7 grammatical errors per report with approximately 30% of reports having grammatical errors. While an automated error detection system would be more likely to provide an accurate error rate, a system of this type would have difficulty classifying errors of omission [9].

The second limitation of this study is that the interreader and intrareader consistency was not performed. Intrareader consistency could have been evaluated by having each reviewer grade a sample of the reports twice. Interreader agreement could have been performed by evaluating a random sample of overlapping reports. Because, the goal of this project was to classify errors that occur in radiology reports rather than assess radiologists' ability to identify errors in reports inter- and intrareader consistency was not evaluated. It should be noted that while interreader consistency was not evaluated, the error rates for all error types (with the exception of grammatical errors) were similar between the two reviewers.

Perhaps the largest limitation of this study is that it did not take into account the editing of reports after dictation. It is possible that editing would lead to a higher number of false-positive error classifications. An example of a false-positive error would be if a dictated word was transcribed but then purposely deleted in the editing process. Because the reviewers were able to listen to the dictation and read the reports, they felt they could often tell when postdictation editing had taken place. It should be noted that because reports were edited after being transcribed by the speech recognition system, that the error rate does not represent the error rate of the speech recognition system. It is possible that the editing process both introduced and corrected many errors.

The final limitation of this study is that it was not possible to confidently detect small differences in dictation time. The main reason that differences could not be detected is due to the integration and automation of the dictation process. When a study is opened in the departmental PACS, it automatically launches the dictation window; thus, the dictation time reported in this study also includes the time required to interpret the image. While it is clear that having a prepopulated report would save at least 3–5 s of time

compared to the manual process, this difference in time is too small to be detected when dictation time also incorporates interpretation time. If only the true dictation time was obtained, it is more likely that small differences in the dictation time would be apparent.

There are several strengths of this study. We believe that the main strength is the study design. To our knowledge, this study is the first to compare audio files with actual dictated clinical reports. Using the audio files allowed us to classify all types of potential errors. In addition, we believe that this study represents the largest prospective study evaluating errors in radiology reports. The second major strength is our strict definition of errors, particularly grammatical errors. While one could argue the merits of classifying sentence fragments as an error, the high incidence of grammatical errors observed even when a less strict definition was used highlights the frequency of this type of error and serves as a notice for radiologists and speech recognition software vendors to focus on this type of error in the future.

Conclusions

The use of prepopulated reports alone did not affect the error rate or dictation time of radiology reports. Radiologists overwhelmingly chose to use standard reports regardless of whether or not they were prepopulated. This preference runs counter to many of the proposed disadvantages of standard reports and suggests that once radiologists are familiar with the standard report that they prefer using it, even at the cost of manually selecting or using a voice command to select the report.

Errors were a frequent occurrence in the radiology reports in this study. Grammatical errors are by far the most common type of error, occurring nearly twice as frequently as all other types of errors combined. This is significant as grammatical errors are not typically caused by a speech recognition error. Missense errors were the second most common type of error. This type of error is important to minimize as it has the potential to be clinically significant by altering the meaning of the dictated phrase.

This study highlights the frequency of errors in radiology reports and suggests that strategies for decreasing errors should be considered. Further studies are required to determine if employing these strategies affects the error rate in finalized radiology reports.

References

1. Krishnaraj A, Lee JKT, Laws SA, Crawford TJ: Voice recognition software: effect on radiology report turnaround time at an academic medical center. *AJR Am J Roentgenol* 195:194–197, 2010

2. Bhan SN, Coblenz CL, Norman GR, Ali SH: Effect of voice recognition on radiologist reporting time. *Can Assoc Radiol J* 59:203–209, 2008
3. Koivikko MP, Kauppinen T, Ahovuo J: Improvement of report workflow and productivity using speech recognition – A follow-up study. *J Digit Imaging* 21:378–382, 2008
4. Quint LE, Quint DJ, Myles JD: Frequency and spectrum of errors in final radiology report generated with automatic speech recognition technology. *J Am Coll Radiol* 5:1196–1199, 2008
5. McGurk S, Brauer K, Macfarlane TV, Duncan KA: The effect of voice recognition software on comparative error rates in radiology reports. *Br J Radiol* 81:767–770, 2008
6. Kanal KM, Hangiandreou NJ, Sykes AMG, Eklund HE, Araoz PA, Leon JA, Erickson BJ: Evaluation of the accuracy of continuous speech recognition software system in radiology. *J Digit Imaging* 13:211–212, 2000
7. Reiner BI, Knight N, Siegel EL: Radiology reporting, past, present, and future: the radiologist's perspective. *J Am Coll Radiol* 4:313–319, 2007
8. Reinus WR: Economics of radiology report editing using voice recognition technology. *J Am Coll Radiol* 4:890–894, 2007
9. Voll K, Atkins S, Forster B: Improving the utility of speech recognition through error detection. *J Digit Imaging* 21:371–377, 2008