

# Patient Safety Tip of the Week

July 24, 2018

## More on Speech Recognition Software Errors

Speech recognition software has the potential to significantly improve efficiencies in multiple healthcare settings. But in our October 4, 2011 Patient Safety Tip of the Week “[Radiology Report Errors and Speech Recognition Software](#)” we highlighted some of the problems associated with use of speech recognition software.

We certainly anticipated that improvements in speech recognition software technology would result in considerably reduced error rates since that time. But our December 2017 What's New in the Patient Safety World column “[Speech Recognition Still Not Up to Snuff](#)” found that both the efficiency and accuracy of speech recognition systems were not quite “up to snuff”.

In that column, we cited a study ([Hodgson 2017](#)) which compared the efficiency and safety of using speech recognition assisted clinical documentation within an electronic health record (EHR) system with use of keyboard and mouse in an emergency department setting. The researchers found that mean task completion times were 18.11% slower overall when using speech recognition compared to keyboard and mouse. For simple tasks speech recognition was 16.95% slower and for complex tasks 18.40% slower. Increased errors were observed with use of speech recognition (138 vs. 32 total errors, 75 vs. 9 errors for simple tasks, and 63 vs. 23 errors for complex tasks). The authors felt that some of the observed increase in errors may be due to suboptimal speech recognition to EHR integration and workflow. They concluded that improving system integration and workflow, as well as speech recognition accuracy and user-focused error correction strategies, may improve SR performance.

Another study of a random sample of 100 notes generated by emergency department physicians using computerized speech recognition (SR) technology also found substantial error rates in notes ([Goss 2016](#)). They found 1.3 errors per note, and 14.8% of errors were judged to be critical. Overall, 71% of notes contained errors, and 15% contained one or more critical errors. Annunciation errors (53.9%) were the most frequent, followed by deletions (18.0%), and added words (11.7%). Nonsense errors, homonyms and spelling errors were present in 10.9%, 4.7%, and 0.8% of notes, respectively.

Now, a new study by Zhou and colleagues ([Zhou 2018](#)) looked at a random sample of operative notes, office notes, and discharge summaries from two hospital systems. They looked at the original speech recognition engine-generated document notes (SR), medical transcriptionist–edited documents, and physician’s signed notes (SN).

The overall error rate in speech recognition generated notes was 7.4% (i.e, 7.4 errors per 100 words). Overall, 96.3% of speech recognition engine-generated document notes (SR), 58.1% of medical transcriptionist–edited documents (MT), and 42.4% of physician’s signed notes (SN) contained errors. Discharge summaries had higher mean error rates than other types and surgeons’ notes had lower mean error rates.

Deletions (34.7%) were the most commonly identified error, followed by insertions (27.0%). Among errors at the SR, MT, and SN stages, 15.8%, 26.9%, and 25.9%, respectively, involved clinical information, and 5.7%, 8.9%, and 6.4%, respectively, were clinically significant. Clinically significant errors often involved medications or diagnoses.

Significantly, the error rate decreased from 7.4% to 0.4% after transcriptionist review and 0.3% in physicians’ signed notes. While stylistic changes or rearrangement of content were fairly commonly made in the editing process, in 27.2% the signing physician added information, and in 17.1% the physician deleted information.

The authors emphasize that this demonstrates the importance of manual review, quality assurance, and auditing.

But we would like to add we are appalled that we still continue to see documents in the EMR with the comment “dictated but not read”! Presumably, that is to make access to notes more timely. But, given the substantial frequency of errors in documents created via speech recognition software (or, for that matter, in documents produced by transcription of regular voice dictation), why would anyone risk the chance that an error could lead to patient harm? Even if those documents are later edited and amended to correct for any mistakes, there is always the possibility that an action may have already been based on the original (unedited) document.

Zhou and colleagues also found evidence suggesting some clinicians may not review their notes thoroughly, if they do so at all. They mention that transcriptionists typically mark portions of the transcription that are unintelligible in the original audio recording with blank spaces (eg, ??\_\_??), which the physician is then expected to fill in. But they found 16 physician-signed notes that retained these marks. In 3 instances, the missing word was discovered to be clinically significant.

In our October 4, 2011 Patient Safety Tip of the Week “[Radiology Report Errors and Speech Recognition Software](#)” we asked how mistakes get overlooked when we review and edit our reports. The number one contributory factor is usually time pressure. In our haste to get the report done and the big queue of other reports to review, we simply don’t review and edit thoroughly. One of the early studies on report errors related to speech recognition systems ([McGurk 2007](#)) noted that such errors were more common in busy areas with high background noise or high workload environments.

But a second phenomenon happens as well. Our mind plays tricks on us and we often “see” what we think we should see. We show many examples during some of our

presentations of orders or chart notes that have obvious omissions where the audience unconsciously “fills in the gaps” and thinks they saw something that wasn’t there (“of course they meant milligrams”). It is easy for us to do the same thing when we are reading our own reports. In addition, the “recency” phenomenon probably comes into play, where the physician perceives he/she sees what he/she just dictated. The Quint paper noted below ([Quint 2008](#)) suggests that mistakes like this may actually be more frequent the sooner you are reviewing your report. They even suggest that reviewing your report 6-24 hours after dictation rather than immediately may reduce the error rate.

Dictating in an environment with minimal background noise can help reduce errors. And McGurk et al note that use of “macros” for common standard phrases also reduces the error rates.

We’re willing to bet that most of you have no idea what your error rates are, regardless of whether you are using automated speech recognition software or traditional dictation transcription services.

Obviously, you need to include an audit of report errors as part of your QI process, not only for radiology but for any service that does reports of any kind, whether done by speech recognition software or more traditional transcription. While random selection of reports to review is a logical approach, there are other approaches that may make more sense. Part of the peer review process in radiology is to have radiologists review the images that a colleague had reported and see if the findings concur. One could certainly add checking for report errors as part of that process.

One older study ([Quint 2008](#)) found errors in 22% of radiology reports where radiologists estimated the error rates would be well less than 10% for the radiology department as a whole and even less frequent for themselves. In the Quint paper, the reports were analyzed as they came up as part of their weekly multidisciplinary cancer conference. Reviewing them in a fashion like this makes the review more convenient but also adds context to the review. One gets to see how the errors could potentially impact patient care adversely. We like that approach where such multidisciplinary conferences take place. It also tends to raise the awareness of the existence and scope of report errors among not only the people generating the reports, but also those reading the reports.

**Integrating evaluation of your reports into your QI program, thus, is critical.** So make sure you are determining your error rates in all your dictated reports (whether traditional or speech recognition format) and feeding back those error rates to the providers doing the reports. Such feedback to the providers doing the reports was important in reducing the error rates in the study by McGurk et al.

**Some of our past columns relating to speech recognition software:**

- October 4, 2011 “[Radiology Report Errors and Speech Recognition Software](#)”
- December 2017 “[Speech Recognition Still Not Up to Snuff](#)”

## References:

Hodgson T, Magrabi F, Coiera E. Efficiency and safety of speech recognition for documentation in the electronic health record. *Journal of the American Medical Informatics Association* 2017; 24(6): 1127-1133  
<https://academic.oup.com/jamia/article/24/6/1127/4049461/Efficiency-and-safety-of-speech-recognition-for>

Goss FR, Zhou L, Weiner SG. Incidence of speech recognition errors in the emergency department. *International Journal of Medical Informatics* 2016; 93: 70-73  
[http://www.ijmijournal.com/article/S1386-5056\(16\)30090-9/abstract](http://www.ijmijournal.com/article/S1386-5056(16)30090-9/abstract)

Zhou L, Blackley SV, Kowalski L, et al. Analysis of Errors in Dictated Clinical Documents Assisted by Speech Recognition Software and Professional Transcriptionists. *JAMA Network Open* 2018; 1(3): e180530 July 6, 2018  
<https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2687052>

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: 767-770  
<http://bjr.birjournals.org/cgi/reprint/81/970/767?maxtoshow=&hits=10&RESULTFORMAT=1&author1=mcgurk&andorexacttitle=and&andorexacttitleabs=and&andorexactfulltext=and&searchid=1&FIRSTINDEX=0&sortspec=relevance&resourcetype=HWCIT>

Quint LE, Quint DJ, Myles JD. Frequency and Spectrum of Errors in Final Radiology Reports Generated With Automatic Speech Recognition Technology. *Journal of the American College of Radiology* 2008; 5(12): 1196-1199  
<http://www.jacr.org/article/S1546-1440%2808%2900361-X/abstract>



The  
Truax  
Group  
Healthcare Consulting  
[www.patientsafetysolutions.com](http://www.patientsafetysolutions.com)

<http://www.patientsafetysolutions.com/>

[Home](#)

[Tip of the Week Archive](#)

[What's New in the Patient Safety World Archive](#)