Peer Review Impacts Voice Recognition Errors

November 07, 2011
By Deborah Abrams Kaplan [1]

Radiologists at University of Chicago Medical Center set out to decrease their own error rates, and found that with peers scoring and reviewing each others’ reports, and then discussing them at section meetings, error rates dropped.

While voice recognition software can do many things, it can’t correct errors it doesn’t know are there. And according to published research, 30 percent to 42 percent of voice recognition reports contain errors.

So, radiologists at University of Chicago Medical Center set out to decrease their own error rates, and found that with peers scoring and reviewing each others’ reports, and then discussing them at section meetings, error rates dropped.
Rina Patel, MD, a radiology resident at the University of Chicago Medical Center will present the findings from their study later this month at the RSNA meeting in Chicago.
The study was done within the chest section, which has six attending physicians. The reports (20 chest X-rays and five CT scans each) were selected monthly on a random day, and a different radiologist would review a colleague’s reports each month, marking grammatical, typographical and word substitution errors. They did not review scans along with the reports.
The reviewing physician also scored the reports, deducting points for each type of error. The dictating physician received the reports back and the types of errors and strategies to reduce errors were discussed at the section meeting bimonthly. They collected data from September 2010 to April 2011.
“The significant drop (in errors) was after the first intervention,” said Patel, noting that the baseline average score was 86 percent. After the first formal review meeting, there was a significant drop in errors, with a 92 percent score. The scores at the next two bimonthly meetings plateaued at 94 percent.
The scoring was unique to the project, and they categorized errors by significance, with a higher score being better. They did not specifically collect data on errors that would change the meaning of a report.
“Any error, even if it doesn’t change the meaning of the reports, can reflect poorly on the report itself,” she said. “We want high quality reports with no errors.”
The voice dictation software (Nuance RadWhere for Radiology, which includes Dragon NaturallySpeaking) has been used at University of Chicago since September 2007, but this is the first time the radiology group has formally looked at this type of report errors, Patel said.
While the section is no longer collecting data, the peer review continues and Patel said that “having a peer review process is one of the reasons that the error rate has gone down.” The process motivates the radiologists to be more careful in proofreading. “The fact that we still have the peer review process and meetings to discuss the error rates contributes to keeping that error rate low.”
The physicians willingly took part in the study “especially since everyone wants to create high quality reports,” she said. She added that the process was not time-intensive, and that the report discussions took place in regular section meetings. Plus, “since we divided the reviewing work among the attendings, it wasn’t that much work per person.”
The physicians did notice an impact on turnaround time per report. The time to finalize the report increased after that first formal meeting, from 6 hours 15 minutes in November to 7 hours 21 minutes in December. It then returned almost to baseline at the next two meetings, to 6 hours 29 minutes in April. While the time difference might be due to the radiologists more carefully proofreading the reports, it also could be attributed to other factors, like November and December being holiday months and the RNSA meeting, resulting in a staffing decrease. “Even if it was related to peer review, (the time) went back down and the error rates stayed the same at the lower level,” she said.