MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-922

Syntactic and Semantic Errors in Radiology Reports Associated With Speech Recognition Software

Michael D. Ringler[a], Brian C. Goss[a], Brian J. Bartholmai[a]

[a] *Department of Radiology, Mayo Clinic, Rochester, MN, USA*

Abstract

Speech recognition software (SRS) has many benefits, but also increases the frequency of errors in radiology reports, which could impact patient care. As part of a quality control project, 13 trained medical transcriptionists proofread 213,977 SRS-generated signed reports from 147 different radiologists over a 40 month time interval. Errors were classified as "material" if they were believed to alter interpretation of the report. "Immaterial" errors were subclassified as intrusion/omission or spelling errors. The proportion of errors and error type were compared among individual radiologists, imaging subspecialty, and time periods using χ^2 analysis and multiple logistic regression, as appropriate. 20,759 (9.7%) reports contained errors; 3,992 (1.9%) contained material errors. Among immaterial errors, spelling errors were more common than intrusion/omission errors (P<.001). Error proportion varied significantly among radiologists and between imaging subspecialties (P<.001). Errors were more common in cross-sectional reports (vs. plain radiography) (OR, 3.72), reports reinterpreting results of outside examinations (vs. in-house) (OR, 1.55), and procedural studies (vs. diagnostic) (OR, 1.91) (all P<.001). Dictation microphone upgrade did not affect error rate (P=.06). Error rate decreased over time (P<.001).

Keywords:

PowerScribe, quality control, radiology report, report errors, speech recognition.

Introduction

Speech recognition software (SRS) decreases report turnaround time, but also increases the frequency of semantic and syntactic errors (1). Such errors may negatively impact clinical management, and result in lawsuits (2). We aimed to analyze percentages and types of errors among speech recognition software–generated radiology reports as part of our quality control program.

Materials

We retrieved transcriptionist audits of all speech recognition software–generated radiology reports (PowerScribe; Nuance Communications, Inc.) that were self-edited and signed by 147 different radiologists, between January 3, 2011, and April 16, 2014. 13 trained medical transcriptionists with 1-23 years of experience proofread a random 5% sample of each radiologist's reports each month. Errors were classified as "material" if the transcriptionist could not readily determine the intended meaning in the context of the report. Otherwise, "immaterial" errors were subclassified as intrusion/omission or spelling errors. Reports with multiple errors were only counted once and classified by most egregious error type (material>intrusion/omission>spelling). Punctuation errors were ignored. The proportion of errors and error type were compared among individual radiologists, imaging subspecialty, and time periods using χ^2 analysis and multiple logistic regression, as appropriate.

Results

Of 213,977 reports identified, 20,759 (9.7%) had errors; 3,992 (1.9%) contained material errors. Among 16,767 immaterial errors, spelling errors (10,151, 60.5%) were more common than intrusion/omission errors (*P*<.001). Proportion of errors and fraction of material errors varied significantly among radiologists and imaging subspecialties (*P*<.001). Errors were more common in cross-sectional reports (vs. plain radiography) (OR, 3.72), reports reinterpreting results of outside examinations (vs. in-house) (OR, 1.55), and procedural studies (vs. diagnostic) (OR, 1.91) (all *P*<.001). Dictation microphone upgrade did not affect error rate (*P*=.06). The total error rate also decreased over the 40-month period of this quality control project by 4.3% (*P*<.001).

Conclusion

Speech recognition software–related errors are highly variable among radiologists and imaging subspecialties. One factor may be length of the report. We calculated a relative risk of error in a cross-sectional report of 3.4 (compared to xray),which is consistent with others (3), though we did not specifically track length. Spelling errors are the most common type of immaterial error in self-edited reports, suggesting many radiologists prefer typing over SR. A quality control program with regular feedback may reduce errors over time.

References

[1] Prevedello LM, Ledbetter S, Farkas C, Khorasani R. Implementation of speech recognition in a community-based radiology practice: effect on report turnaround times. J Am Coll Radiol. 2014 Apr;11(4):402-6.

[2] Physician Insurers Association of America. PIAA data sharing reports. Rockville (MD): Physician Insurers Association of America; Jan 1, 1985-Jun 30, 2003.

[3] Chang CA, Strahan R, Jolley D. Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. J Digit Imaging. 2011 Aug;24(4):724-8.

Address for correspondence

Michael D. Ringler, MD, Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905