ERRORS IN VOICE RECOGNITION GENERATED RADIOLOGY REPORTS: TO MANY MISSED STEAKS?

JONATHAN HICKLE, ALAN BRYDIE

DISCLOSURES

- None

THE PROBLEM

- A survey of referring clinicians showed they appreciate fast turn around times, but note a higher frequency of errors in reports.
- Errors in reports are not tracked at our institution.
- There is no baseline data to evaluate the frequency and severity of errors.
- High quality reports reflect high quality interpretation of imaging.

LITERATURE REVIEW

There are no national or international standards or guidelines.

Basma, 2011:

- 23% of VR reports with errors vs 4% for transcriptionists

Chang, 2011:

- 11% of CR reports contained errors; 2% nonsense errors.
- 36% of non-CR reports contained errors; 5% nonsense errors.

Quint, 2008:

- 22% error rate in cross sectional reports.
- 76% of radiologists believed their error rate was <10%.

METHODS OF COLLECTION

- 10 reports were selected from every radiologist in the CDHA.
- 6 CR and 4 non CR reports were selected in reverse chronological order prior to June 1, 2013.
- Ratio based on proportion of CR vs. non CR reports generated by all radiologists.
- Initial read by stenographer, followed by radiology resident.
- 10 additional reports were collected for radiologists who work with residents.

METHOD OF CATEGORIZATION

- Errors grouped into "**major**" and "**minor**" categories.
- **Major errors**: Any error that affects understanding of the report or could cause harm to a patient.
- All other errors were considered "**minor**".
- Borderline errors were categorized by consensus by ADB and JDH.
- All errors were subcategorized into: *nonsense phrase, wrong units, added word, dropped word, word substitution, punctuation, error in heading, other.*

PRIMARY AUDIT RESULTS

Audit 1	Number	Percentage
Reports with Major Errors	31	3.5
Reports with Minor Errors	176	20
Total Reports	880	100

Target Major error rate (0%) not met
Target Minor error rate (<10%) not met

EXAMINED SOLUTIONS

- Optimize current version:
  - Highlight common substituted and omitted works (no, left, right)
  - Improve grammar checker
  - New microphones
  - Change backround/text colour
  - Retrain Powerscribe
- Quieter reporting environment
- Double read all reports
- Delay signing reports
- Synoptic reporting, macros
- Update to newest version of Powerscribe
- Use another vendor

ACTIONS TAKEN

- Microsoft Word detected 17% of major errors.
- Microphone QA: users asked to adjust volume with every session.
- Minimize background noise in new reporting areas.
- Bottom 50th percentile asked to retrain their voice profiles, other users retrained voluntarily.

RETRAINING PARTICIPATION

- 15 staff retrained their profiles entirely
- 4 staff retrained their adaptation mode
- 5 residents and fellows participated.
- 19/22 of suggested staff and fellows retained as requested.

OVERALL RESULTS

Audit 1	Number	Percentage
Reports with Major Errors	31	3.5
Reports with Minor Errors	176	20
Total Reports	880	100

Audit 2	Number	Percentage
Reports with Major Errors	19	2.2
Reports with Minor Errors	220	25.7
Total Reports	856	100

- Major error rate target not met (0%)
- Minor error rate target not met (<10%)

PERCENTAGE OF REPORTS CONTAINING ERRORS

Percentage error rates		Major	Minor	Combined
Audit 1	*Staff*	4.9%	23.1%	25.1%
	Resident	1.4%	15.7%	17.0%
	Total	3.4%	20.0%	21.7%
Audit 2	*Staff*	2.4%	27.6%	28.2%
	Resident	2.0%	23.0%	23.9%
	Total	2.2%	25.7%	26.4%

- Decrease in Major Errors
- Increase in Minor Errors

ERROR RATES BY MODALITY

Percentage of reports with errors		General	Cross sectional
Audit 1	*Staff*	20.4%	32.8%
	Resident	13.8%	21.9%
	Total	17.7%	28.1%
Audit 2	*Staff*	24.2%	34.7%
	Resident	17.3%	33.8%
	Total	21.4%	34.3%

- Cross sectional reports contain more errors.
  - Length of reports is likely the biggest contributor.
- Resident reports contain fewer errors.
  - Double reader effect.

MAJOR ERRORS

Error Type	Number	Percentage
Nonsense	36	66.7%
Substitution	10	18.5%
Dropped	4	7.4%
Added	2	3.7%
Other	2	3.7%
Total	54	100.0%

Combined results for both staff and resident dictated reports in audits 1 and 2.

- Nonsense errors were the most frequent type of error.
- Substitutions were the most difficult to detect.
- The majority of major errors were not identified when entered into Microsoft Word.

MINOR ERRORS

Error Type	Number	Percentage
Dropped	43	23.4%
Substituted	42	22.8%
Added	38	20.7%
Punctuation	31	16.8%
Heading/Indication	27	14.7%
Other	3	1.6%
Total	184	100.0%

Combined results for both staff and resident dictated reports in audits 1 and 2.

- Dropped, substituted and added word errors occurred with almost equal frequency.
- Frequent minor errors may negatively impact physician perception of the quality of the interpretation.

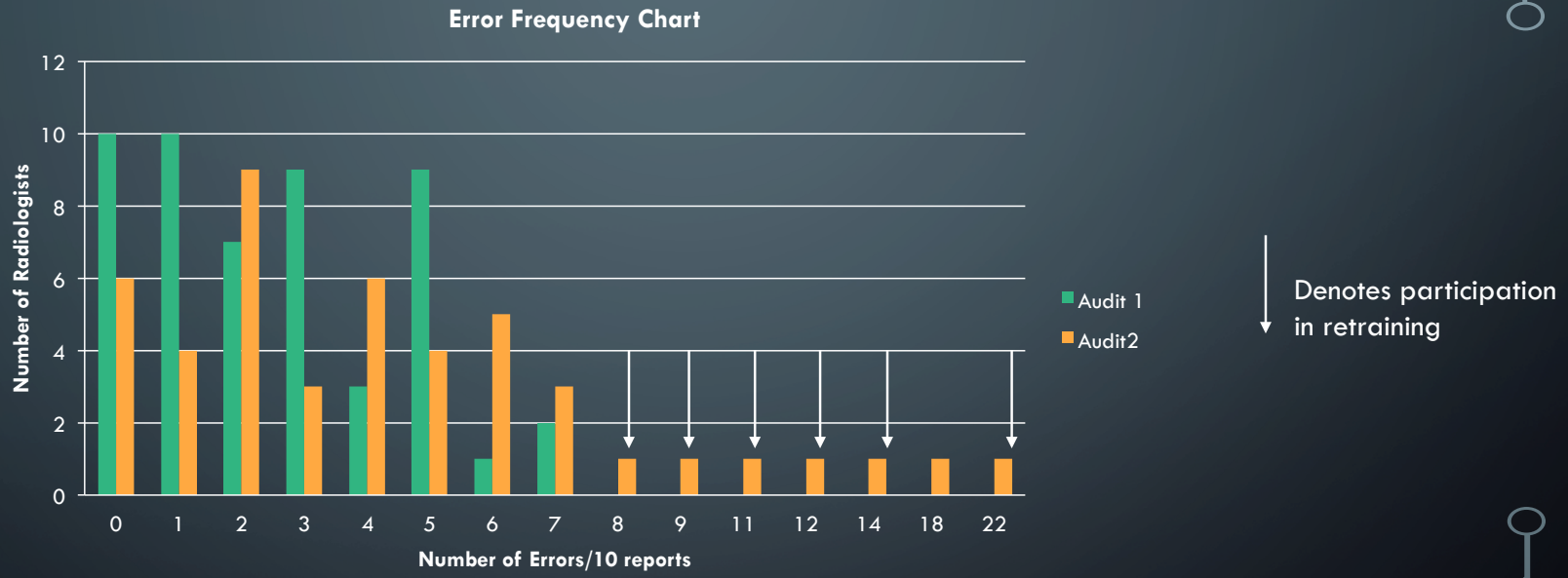
EFFECT OF RETRAINING

Audit 1	Number of Errors	Errors/report
Retraining Group n=19	82	0.43
No Retraining n=30	36	0.12

Audit 2	Number of Errors	Errors/report
Retraining Group n=19	135	0.71
No Retraining n=30	72	0.24

- Number of total errors increased in both groups.
- Larger absolute increase in errors in retraining group.

EFFECT OF RETRAINING



- Smaller standard deviation in the first audit.
- Results are skewed by a number of radiologists with a large number of minor errors.

POTENTIAL FACTORS AFFECTING ERROR RATES

- Staff maintained similar error rates between both audits.
- The quality of initial training and subsequent adaptation may be superior for some users.
- Staff who use templates, canned reports, short dictations are likely to have fewer errors.
- Some individuals are more diligent proofreaders.
- A higher volume of more complex reports may lead to more errors.
- Local environmental factors may contribute to error rates: ie. Background noise and frequent interruptions.
- Staff reporting in offices had fewer errors than those in “common rooms”.

SUMMARY

- Overall **decrease** in major errors between audits
- Major nonsense errors are almost always easily recognized by second readers; they are unlikely to impact patient care if clarified with the reporting radiologist.
- Major errors of substitution are more likely negatively impact patient care.
- Overall **increase** in minor errors between audits
- High participation in retraining
- Retraining was **unsuccessful**.
- **Targets not achieved.**
- A current pilot project allows patients to access their reports; they may doubt the quality of their care if there are frequent errors.

MOVING FORWARD

- The results were reviewed by CDHA departmental leadership.
- A decision was made to upgrade to Powerscribe 360.
- Nuance claims superior voice recognition accuracy over previous versions.
- Other vendors often license the Nuance “Dragon” speech recognition engine; similar error rates can be expected from those vendors.
- Implementation has begun.
- A third audit cycle is suggested to determine if performance meets the desired 0% Major, <10% Minor target levels.

REFERENCES AND ACKNOWLEDGEMENTS

- Basma S, Lord B, Jacks LM, Rizk M, Scaranelo AM. Error rates in breast imaging reports: comparison of automatic speech recognition and dictation transcription. AJR Am J Roentgenol. 2011 Oct;197(4): 923-7. doi: 10.2214/AJR.11.6691. PubMed PMID: 21940580.
- 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. doi: 10.1007/s10278-010-9344-z. PubMed PMID: 20976612; PubMed Central PMCID: PMC3138931.
- Duncan KA, McGurk S, Brauer K, Voice Recognition System Report Accuracy. Audit Template https://www.rcr.ac.uk/audit/voice-recognition-system-report-accuracy
- 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 Dec;5(12):1196-9. doi: 10.1016/j.jacr.2008.07.005. PubMed PMID: 19027683.

- Dr. Alan Brydie
- Marian McKim
- Lorraine Gould
- CDHA IT specialists
- Dalhousie radiologists and residents