MEASURING PROGRESS IN RESIDENT EDUCATION: PILOT STUDY OF A REPORT COMPARATOR

Jeffrey D. Robinson, MD MBA,
Daniel S. Hippe, MS
UW Medicine: Harborview Medical Center, Seattle, WA

Introduction

- Report generation is the *sine qua non* of diagnostic radiology, and training residents to create accurate reports is one of the key functions of faculty radiologists.
- Due to variability in workflow and personality types, consistent feedback to residents on the quality of their reports and corrections made by faculty before finalization is intermittent at best, with the resident often never knowing what changes the faculty radiologist made to his/her report.

Introduction

- An automated process developed at our institution highlights any changes made in a resident's report by the attending faculty radiologist and places the highlighted reports in a secure online folder for the resident to review.
- The purpose of this project was to determine whether the extent of report corrections could be measured, and whether these measures could be used to assess changes in the number and extent of corrections to residents' reports made by faculty over time.

Methods: Participants

- IRB approved study, informed consent obtained
- R3 and R4 residents block randomized
- Experimental group: access to a secure online folder containing comparator output of reports they dictated in the last week that had been amended by faculty
- Control group: No such access
- All residents received standard instructional feedback in the course of their daily work.
- Attitude survey completed before and after study

Methods: Study period

- Twenty-four week study period
- Divided into 4-week blocks to correlate with resident clinical rotations

Methods: Comparator

- Daily query of institutional database for preliminary and final radiology reports
- Open-source software tool (1) used to compare preliminary and final reports
- Differences between reports quantified using the Levenshtein Distance (LD)

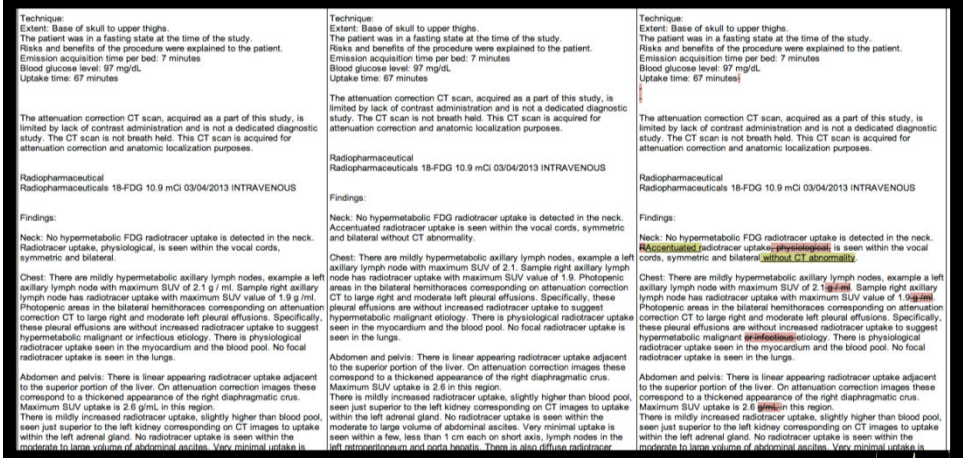
Methods: Levenshtein distance

- Method invented in the 1960's to compare sequences of symbols (2)
- In general, it counts the number of deletions, insertions and translations of symbols in a string
- To correct for different length reports, we normalized the LD:
  - NLD = LD / # characters in longer of the preliminary or final report

Methods: Statistical Analysis

- Categorical variables summarized as count (%)
- Continuous variables summarized as mean +/- sd and median (range)
- Derived additional binary variables: NLD>0, NLD>10%, NLD>20%
- Survey: Seven-point Likert-like scales, compared using Mann-Whitney and Wilcoxon signed-rank tests
- Correction metrics averaged for each 4 week block
- Trends over time were evaluated using linear regression
- Calculations performed for all residents as well as separately for each group.

Results: Sample comparator report



Preliminary report

Technique:
Extent: Base of skull to upper thighs.
The patient was in a fasting state at the time of the study.
Risks and benefits of the procedure were explained to the patient.
Emission acquisition time per bed: 7 minutes
Blood glucose level: 97 mg/dL
Uptake time: 67 minutes

The attenuation correction CT scan, acquired as a part of this study, is limited by lack of contrast administration and is not a dedicated diagnostic study. The CT scan is not breath held. This CT scan is acquired for attenuation correction and anatomic localization purposes.

Radiopharmaceutical
Radiopharmaceuticals 18-FDG 10.9 mCi 03/04/2013 INTRAVENOUS

Findings:

Neck: No hypermetabolic FDG radiotracer uptake is detected in the neck. Radiotracer uptake, physiological, is seen within the vocal cords, symmetric and bilateral.

Chest: There are mildly hypermetabolic axillary lymph nodes, example a left axillary lymph node with maximum SUV of 2.1 g / ml. Sample right axillary lymph node has radiotracer uptake with maximum SUV value of 1.9 g /ml. Photopenic areas in the bilateral hemithoraces corresponding on attenuation correction CT to large right and moderate left pleural effusions. Specifically, these pleural effusions are without increased radiotracer uptake to suggest hypermetabolic malignant or infectious etiology. There is physiological radiotracer uptake seen in the myocardium and the blood pool. No focal radiotracer uptake is seen in the lungs.

Abdomen and pelvis: There is linear appearing radiotracer uptake adjacent to the superior portion of the liver. On attenuation correction images these correspond to a thickened appearance of the right diaphragmatic crus. Maximum SUV uptake is 2.6 g/mL in this region.
There is mildly increased radiotracer uptake, slightly higher than blood pool, seen just superior to the left kidney corresponding on CT images to uptake within the left adrenal gland. No radiotracer uptake is seen within the moderate to large volume of abdominal ascites. Very minimal uptake is

Final Report

Technique:
Extent: Base of skull to upper thighs.
The patient was in a fasting state at the time of the study.
Risks and benefits of the procedure were explained to the patient.
Emission acquisition time per bed: 7 minutes
Blood glucose level: 97 mg/dL
Uptake time: 67 minutes

The attenuation correction CT scan, acquired as a part of this study, is limited by lack of contrast administration and is not a dedicated diagnostic study. The CT scan is not breath held. This CT scan is acquired for attenuation correction and anatomic localization purposes.

Radiopharmaceutical
Radiopharmaceuticals 18-FDG 10.9 mCi 03/04/2013 INTRAVENOUS

Findings:

Neck: No hypermetabolic FDG radiotracer uptake is detected in the neck. Accentuated radiotracer uptake is seen within the vocal cords, symmetric and bilateral without CT abnormality.

Chest: There are mildly hypermetabolic axillary lymph nodes, example a left axillary lymph node with maximum SUV of 2.1. Sample right axillary lymph node has radiotracer uptake with maximum SUV value of 1.9. Photopenic areas in the bilateral hemithoraces corresponding on attenuation correction CT to large right and moderate left pleural effusions. Specifically, these pleural effusions are without increased radiotracer uptake to suggest hypermetabolic malignant etiology. There is physiological radiotracer uptake seen in the myocardium and the blood pool. No focal radiotracer uptake is seen in the lungs.

Abdomen and pelvis: There is linear appearing radiotracer uptake adjacent to the superior portion of the liver. On attenuation correction images these correspond to a thickened appearance of the right diaphragmatic crus. Maximum SUV uptake is 2.6 in this region.
There is mildly increased radiotracer uptake, slightly higher than blood pool, seen just superior to the left kidney corresponding on CT images to uptake within the left adrenal gland. No radiotracer uptake is seen within the moderate to large volume of abdominal ascites. Very minimal uptake is seen within a few, less than 1 cm each on short axis, lymph nodes in the left retroperitoneum and porta hepatis. There is also diffuse radiotracer

Comparator Report

Technique:
Extent: Base of skull to upper thighs.
The patient was in a fasting state at the time of the study.
Risks and benefits of the procedure were explained to the patient.
Emission acquisition time per bed: 7 minutes
Blood glucose level: 97 mg/dL
Uptake time: 67 minutes

The attenuation correction CT scan, acquired as a part of this study, is limited by lack of contrast administration and is not a dedicated diagnostic study. The CT scan is not breath held. This CT scan is acquired for attenuation correction and anatomic localization purposes.

Radiopharmaceutical
Radiopharmaceuticals 18-FDG 10.9 mCi 03/04/2013 INTRAVENOUS

Findings:

Neck: No hypermetabolic FDG radiotracer uptake is detected in the neck. RAccentuated radiotracer uptake ~~physiological~~ is seen within the vocal cords, symmetric and bilateral without CT abnormality.

Chest: There are mildly hypermetabolic axillary lymph nodes, example a left axillary lymph node with maximum SUV of 2.1 ~~g / ml~~. Sample right axillary lymph node has radiotracer uptake with maximum SUV value of 1.9 ~~g /ml~~. Photopenic areas in the bilateral hemithoraces corresponding on attenuation correction CT to large right and moderate left pleural effusions. Specifically, these pleural effusions are without increased radiotracer uptake to suggest hypermetabolic malignant ~~or infectious~~ etiology. There is physiological radiotracer uptake seen in the myocardium and the blood pool. No focal radiotracer uptake is seen in the lungs.

Abdomen and pelvis: There is linear appearing radiotracer uptake adjacent to the superior portion of the liver. On attenuation correction images these correspond to a thickened appearance of the right diaphragmatic crus. Maximum SUV uptake is 2.6 ~~g/mL~~ in this region.
There is mildly increased radiotracer uptake, slightly higher than blood pool, seen just superior to the left kidney corresponding on CT images to uptake within the left adrenal gland. No radiotracer uptake is seen within the moderate to large volume of abdominal ascites. Very minimal uptake is

Results

Measure	Value
Total reports	39069
Experimental Group	19120 (49%)
Control Group	19949 (51%)
Amended reports	13413 (34%)
NLD > 10%	6426 (16%)
NLD > 20%	3443 (8.8%)

Results: Baseline Survey

	Treatment Groups			Level		
	Experimental (N=11)	Control (N=10)	P-value*	R3 (N=10)	R4 (N=11)	P-value*
Confidence in dictation skills	5.5 ± 1.0	5.4 ± 0.7	0.76	5.5 ± 1.1	5.4 ± 0.7	0.51
Confidence in Radiology knowledge/skills	4.8 ± 0.9	5.1 ± 0.7	0.51	4.8 ± 0.9	5.1 ± 0.7	0.51
Satisfied with feedback on reports	4.3 ± 1.2	4.2 ± 0.8	0.91	4.0 ± 1.2	4.5 ± 0.8	0.33
The feedback helps improve dictation skills	5.5 ± 0.9	4.9 ± 1.3	0.30	5.1 ± 1.2	5.3 ± 1.1	0.77
The feedback helps improve Radiology knowledge/skills	5.7 ± 1.3	5.3 ± 1.3	0.44	5.5 ± 1.2	5.5 ± 1.4	0.88

No significant differences between groups at baseline

* Mann-Whitney Test

Results: Change in Survey Responses

	Combined (N=21)		
	Baseline	Final	P-value*
Confidence in dictation skills	5.4 ± 0.9	5.6 ± 0.7	0.39
Confidence in Radiology knowledge/skills	5.0 ± 0.8	5.0 ± 0.8	0.74
Satisfied with feedback on reports	4.2 ± 1.0	4.3 ± 1.2	0.61
Feedback helps improve dictation skills	5.2 ± 1.1	5.3 ± 1.0	0.60
Feedback helps improve Radiology knowledge/skills	5.5 ± 1.2	5.6 ± 1.2	0.90

No significant changes in survey responses over study period

* Wilcoxon signed-rank

Results: Change in Survey Responses

	Combined (N=21)			Experimental (N=11)		Control (N=10)		
	Baseline	Final	P-value*	Baseline	Final	Baseline	Final	P-value†
Confidence in dictation skills	5.4 ± 0.9	5.6 ± 0.7	0.39	5.5 ± 1.0	5.5 ± 0.7	5.4 ± 0.7	5.7 ± 0.7	0.72
Confidence in Radiology knowledge/skills	5.0 ± 0.8	5.0 ± 0.8	0.74	4.8 ± 0.9	5.1 ± 0.8	5.1 ± 0.7	5.0 ± 0.8	0.43
Satisfied with feedback on reports	4.2 ± 1.0	4.3 ± 1.2	0.61	4.3 ± 1.2	4.5 ± 1.5	4.2 ± 0.8	4.1 ± 0.9	0.33
Feedback helps improve dictation skills	5.2 ± 1.1	5.3 ± 1.0	0.60	5.5 ± 0.9	5.5 ± 1.1	4.9 ± 1.3	5.1 ± 0.9	0.94
Feedback helps improve Radiology knowledge/skills	5.5 ± 1.2	5.6 ± 1.2	0.90	5.7 ± 1.3	5.7 ± 1.4	5.3 ± 1.3	5.5 ± 1.0	0.58

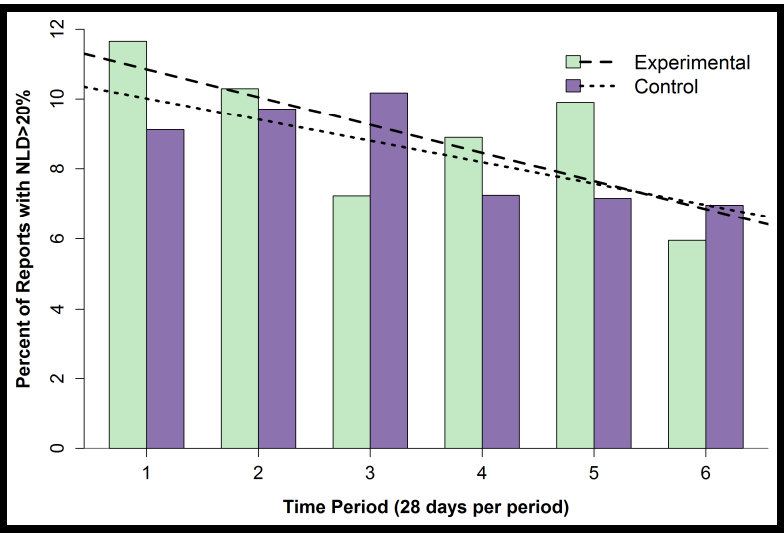
No significant differences in survey responses between groups

* Wilcoxon signed-rank † Mann-Whitney

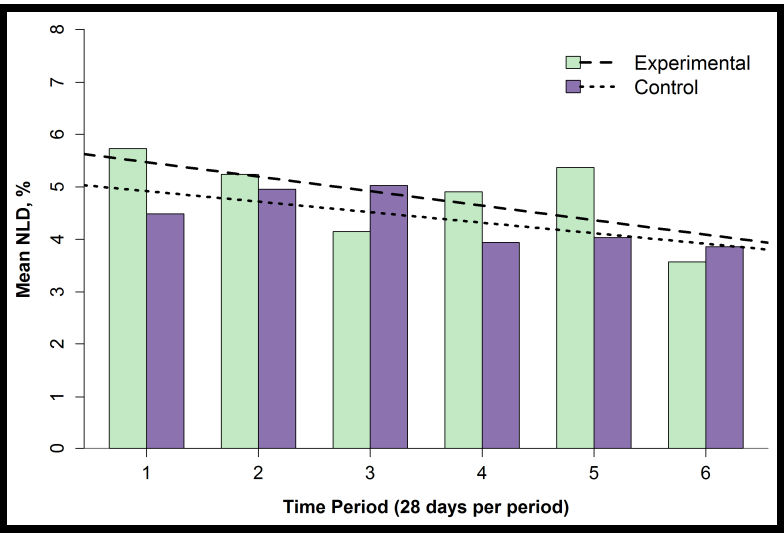
Results: Selected Survey Comments

- I love the idea of the report comparator. In its current form, I do not think it is very useful, because accessing the reports requires many steps that I find a deterrent…
- I think its a great idea and I would use it if it were easier to get to.
- I would have used it more if it had been more accessible.

Results: Trends in Correction Rate



Results: Trends in Mean NLD



Results: Trends in NLD

	First Period		Last Period	
Variable	**Exp.**	**Ctrl.**	**Exp.**	**Ctrl.**
Mean NLD, %	5.7 ±10.8	4.5 ±9.8	3.6 ±8.4	3.9 ±9.1
Percent NLD>0%	39.7	32.3	31.3	31.0
Percent NLD>10%	21.9	16.3	13.3	13.8
Percent NLD>20%	11.6	9.1	6.0	7.0

Results: Trends in NLD

	First Period		Last Period		Combined Group	
Variable	**Exp.**	**Ctrl.**	**Exp.**	**Ctrl.**	**Slope***	**P-value†**
Mean NLD, %	5.7 ±10.8	4.5 ±9.8	3.6 ±8.4	3.9 ±9.1	-0.25	**0.029**
Percent NLD>0%	39.7	32.3	31.3	31.0	-1.00	**0.024**
Percent NLD>10%	21.9	16.3	13.3	13.8	-0.96	**0.005**
Percent NLD>20%	11.6	9.1	6.0	7.0	-0.72	**0.007**

Significant downward trends detected when groups were combined
Implies decreasing correction rate, presumably learning

*Based on linear regression; represents change in dependent variable per 28 day period;
†Test of non-zero slope (assumed periods were independent);

Results: Trends in NLD

	First Period		Last Period		Combined Group		Experimental Group		Control Group		Diff. btw. Groups
Variable	**Exp.**	**Ctrl.**	**Exp.**	**Ctrl.**	**Slope***	**P-value†**	**Slope***	**P-value†**	**Slope***	**P-value†**	**P-value‡**
Mean NLD, %	5.7 ±10.8	4.5 ±9.8	3.6 ±8.4	3.9 ±9.1	-0.25	**0.029**	-0.28	0.18	-0.20	0.11	0.71
Percent NLD>0%	39.7	32.3	31.3	31.0	-1.00	**0.024**	-1.13	0.14	-0.83	0.17	0.71
Percent NLD>10%	21.9	16.3	13.3	13.8	-0.96	**0.005**	-1.14	0.13	-0.74	**0.031**	0.55
Percent NLD>20%	11.6	9.1	6.0	7.0	-0.72	**0.007**	-0.80	0.11	-0.61	0.059	0.69

Did not detect significant difference in improvement in correction rate,
But experimental group had somewhat steeper slope (faster correction reduction)

*Based on linear regression; represents change in dependent variable per 28 day period;
†Test of non-zero slope (assumed periods were independent);
‡Test of difference between slopes (assumed periods and groups were independent within periods, which is likely conservative in this case).

Discussion

- This comparator is closely modeled after that of Sharpe, et al (3)
- Their comparator was eventually incorporated into a results reporting dashboard (5)
- Our effort studies residents longitudinally to measure change in performance over time.

Discussion

- Our data showed statistically significant improvement in overall resident performance over the study interval. This finding conforms with the expectation of the educational process in general.
- The experimental group showed a trend towards a greater degree of improvement than the control group.

Discussion: Limitations

- While the individual comparator reports are quite useful for didactic purposes, the NLD results should only be viewed using time-averaged data, as report-to-report or day-to-day variation in NLD is quite high.
- Variability between faculty members' approaches to editing resident reports can skew NLD data.

Discussion: Limitations

- We were able to access data from two of our four teaching hospitals, so residents entered and left the study as they rotated into and out of the two hospitals in the study.
- Residents' use of the comparator reports was hampered by difficulty accessing the report folders, potentially reducing the impact of the reports on resident learning.

Discussion: Next Steps

- Evolution of our enterprise RIS and PACS infrastructure will ease access to report data, simplifying data capture, network security and maintenance of the comparator application.
- Redesign of the resident user interface is expected to ease access to the comparator reports, increasing their utilization.
- No plans to incorporate NLD data into resident evaluations at this time.

Conclusion

- The report comparator can document changes in resident dictation performance over time.
- A six month pilot confirmed the expected overall improvement in resident performance, with trending data supporting greater improvement by residents who had access to the comparator reports.
- Resident use of the comparator reports was hampered by a difficult user interface.

Acknowldegement

- Dr. Angelisa Paladin and Dr. Norman Beauchamp for generous financial support in the development of the comparator application.

References

- 1. Google-diff-match-patch. [cited 2013 January 1]. Available from: http://code.google.com/p/google-diff-match-patch.
- 2. Levenshtein A. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. Soviet Physics Doklady. 1966;10(8):707-10.
- 3. Sharpe RE, Jr., Surrey D, Gorniak RJ, Nazarian L, Rao VM, Flanders AE. Radiology report comparator: a novel method to augment resident education. Journal of Digital Imaging. 2012;25(3):330-6.
- 4. Surrey D, Sharpe RE, Jr., Gorniak RJ, Nazarian LN, Rao VM, Flanders AE. QRSE: a Novel Metric for the Evaluation of Trainee Radiologist Reporting Skills. Journal of Digital Imaging. 2013;26(4):678-82.
- 5. Gorniak RJ, Flanders AE, Sharpe RE, Jr. Trainee Report Dashboard: Tool for Enhancing Feedback to Radiology Trainees about Their Reports. Radiographics. 2013;33(7):2105 - 13.