The Effect of an Electronic Peer-Review Auditing System on Faculty-Dictated Radiology Report Error Rates

Jonathan H. Chung, MD, Heber MacMahon, MD, Steven M. Montner, MD, Lili Liu, MASc, David M. Paushter, MD, Paul J. Chang, MD, Gregory L. Katzman, MD, MBA

THE PROBLEM: ERRORS IN REPORTS RELATED TO VOICE RECOGNITION

An interpretive narrative report is the main product of a radiologist's work and is the most prevalent communication between the radiologist and the clinical team. In recent years, the importance of turnaround time for radiology reports has been increasingly studied and discussed [1]. Central to such research has been the increased use of voice recognition (VR) software, which is now the predominant method used for the generation of radiology reports [2]. It has been consistently shown that the use of VR decreases turnaround time [3,4]. Given the increased focus on minimizing turnaround time and the ubiquity of VR for report generation, there has been less focus on quality of reports, though many studies have demonstrated a higher rate of errors when using VR [5-7]. Error-ridden radiology reports not only confuse clinicians and create a poor impression among patients who read their reports, but they may also have medicolegal ramifications [8]. *JACR* has recently emphasized clarity in reports by creating a new column, "Speaking of Language," which aims to "improve radiology reporting one meaningless or inappropriate word at a time" [9].

At our medical center, anecdotal evidence suggested that the error rate in radiology reports increased when VR was implemented. To gauge the quality of radiology reports from the standpoint of grammar, clarity, and comprehensibility, we initially implemented a manual system by which reports were proofread by a faculty member who would provide corrective feedback to the individual who generated the report. However, it quickly became obvious that this was too labor intensive and not sustainable. Thus, an IT web-based tool was created to facilitate measurement of the error rate for radiology reports within each section over time, with errors identified systematically by each attending radiologist in the department. In addition, during the period of measurement, a new version of VR was implemented.

The purposes of this study were threefold:

- to gauge the error rate for radiology reports in a tertiary academic medical center;
- to determine whether mandatory, department-wide participation in peer review assessment of radiology reports would in and of itself affect the quality of radiology reports over time; and
- to determine whether introduction of a new VR system in those with prior experience with VR would alter the quality of radiology reports.

We hypothesized that a minority of radiology reports would contain errors; that over time, radiology report error rates would improve because of peer review; and that introduction of a new VR system would lead to a small incremental improvement in radiology report error rates.

WHAT WAS DONE: PEER ASSESSMENT OF ERRORS IN REPORTS

Each month, 10 reports that had been dictated by each radiologist (without a trainee) were randomly selected by the IT tool, anonymized, and submitted into a queue for scoring by other anonymous radiologists in the same subfield of expertise. Reports were scored for the presence of errors on a three-point scale: good, fair, and poor, on the basis of subjective assessment of the number and nature of errors within the report as well as whether the errors were thought to potentially alter the meaning of the report in a way that could be clinically significant.

1546-1440/16/$36.00 ■ http://dx.doi.org/10.1016/j.jacr.2016.04.012

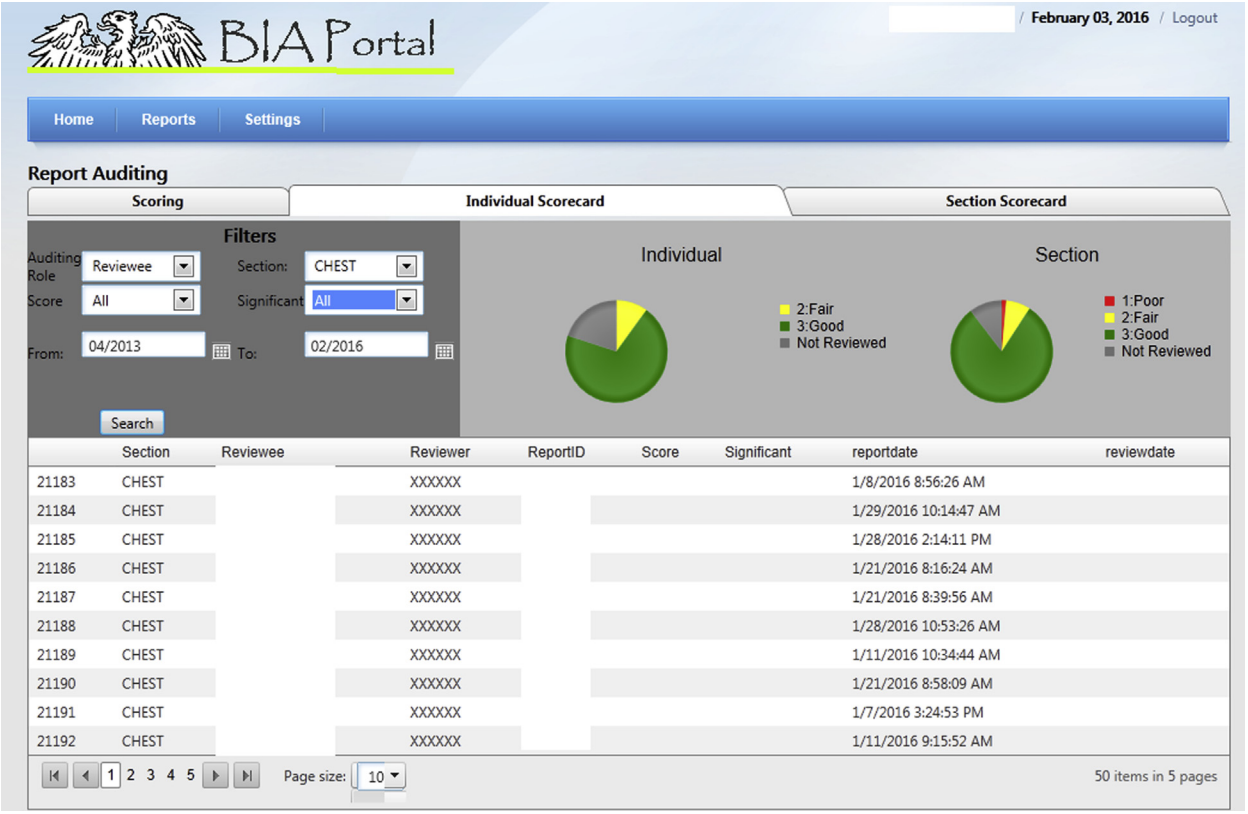

Fig 1. Screen capture of the IT report auditing system summarizing an individual reviewee's evaluations.

A small workgroup determined how many and what types of errors would demote a report from good to fair or poor. A user's manual was created, the faculty members were educated as to appropriate use of the tool, and the project was implemented department-wide. Examples of the types of errors flagged include laterality errors reversing left and right, unit measurement errors (eg, switching millimeters and centimeters between report sections), and VR recognition errors (eg, "the gallbladder wall is tanked with a deer"). Scores were subsequently tabulated for each radiologist over nine months. In addition to being notified of clinically significant errors in reports, all radiologists were notified if any of their reports were rated as poor or fair quality, with the specific errors indicated; reviewees were able to access all of their report evaluations using the online tool (Fig. 1). At the end of this nine-month period, a new VR application (PowerScribe 360; Nuance Healthcare, Burlington, Massachusetts) was instituted. Data were collected for the subsequent seven months after the change in VR. Two-tailed Fisher exact tests and *t* tests were used to compare proportions and means, respectively. *P* values less than .05 were considered to indicate statistical significance.

OUTCOMES

First Nine-month Period After Inception of Peer Review System

In the first time period, an average of 289 ± 18 reports were reviewed per month. After inception of our peer

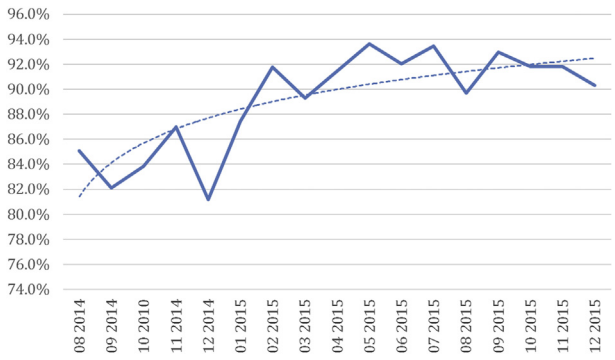

Fig 2. Line graph of percentage of "good" radiology reports over time with logarithmic trend curve.