Clinical Radiology 70 (2015) 1185–1191

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Radiology reports: a quantifiable and objective textual approach

J.A. Scott*, E.L. Palmer

Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

ARTICLE INFORMATION

Article history:
Received 9 March 2015
Received in revised form
20 April 2015
Accepted 5 June 2015

AIM: To examine the feasibility of using automated lexical analysis in conjunction with machine learning to create a means of objectively characterising radiology reports for quality improvement.

MATERIALS AND METHODS: Twelve lexical parameters were quantified from the collected reports of four radiologists. These included the number of different words used, number of sentences, reading grade, readability, usage of the passive voice, and lexical metrics of concreteness, ambivalence, complexity, passivity, embellishment, communication and cognition. Each radiologist was statistically compared to the mean of the group for each parameter to determine outlying report characteristics. The reproducibility of these parameters in a given radiologist's reporting style was tested by using only these 12 parameters as input to a neural network designed to establish the authorship of 60 unknown reports.

RESULTS: Significant differences in report characteristics were observed between radiologists, quantifying and characterising deviations of individuals from the group reporting style. The 12 metrics employed in a neural network correctly identified the author in each of 60 unknown reports tested, indicating a robust parametric signature.

CONCLUSION: Automated and quantifiable methods can be used to analyse reporting style and provide impartial and objective feedback as well as to detect and characterise significant differences from the group. The parameters examined are sufficiently specific to identify the authors of reports and can potentially be useful in quality improvement and residency training.

Introduction

The radiology report has long been a subject of conversation among radiologists. Despite attempts at standardisation,[1–4] clinical radiology reports retain substantial variability, largely reflecting the personal biases of the authors. There are many stories of colleagues anecdotally identifying the author of a report simply by characteristic textual mannerisms. Such differences in style may not be critical to report quality, but their analysis has generally been based upon subjective assessments. A more objective approach to report analysis could complement higher-level subjective analyses by providing impartial and quantifiable characterisation of the radiology report. The present study presents an automated method involving the quantification of the lexical properties of the report.

The primary goal was to improve dialogue among reporting physicians regarding report quality. It is anticipated that an unbiased and quantifiable characterisation of reports might be useful in providing non-confrontational feedback to radiologists regarding their own reporting styles and how these compare with their peers. In the

* Guarantor and correspondent: J.A. Scott, Department of Radiology, White 427, 55 Fruit Street, Massachusetts General Hospital, Boston, MA 02114, USA. Tel.: +1 617 726 8350; fax: +1 617 726 6165.
E-mail address: scott@helix.mgh.harvard.edu (J.A. Scott).

http://dx.doi.org/10.1016/j.crad.2015.06.080
0009-9260/

present study, reports from a group of four experienced radiologists were examined. A randomly selected group of reports was obtained from each radiologist and evaluated using two standard lexical programs to determine the individual characteristics of the reports. The study was limited to bone scintigraphy to ensure that differences between reporting physicians originated from reporting styles rather than varying study complexity. The data were then employed in a neural network to identify the author of unknown reports to validate the specificity of the individual report characterisations more effectively.

Materials and methods

This project was approved by the institutions review board.

The database

Textual analysis was performed using data obtained from Digitext Diction 7.0 (Digitext, Austin, TX, USA) and Microsoft Word 2010 (Microsoft Office Professional 2010, Microsoft, Redmond, WA, USA). The Diction 7.0 program is a commercially available software program that analyses writing for a variety of stylistic elements. The intrinsic tools of Microsoft Word measure simple textual parameters related to reading difficulty, textual length, and passive voice usage. The use of these programs is discussed in more detail below.

Seventy-five consecutive bone scintigraphy reports dating from January 2011 through March 2013 were randomly selected for each of four reporting physicians. Each of these was a staff physician with at least 20 years of clinical experience. All four physicians were board certified in nuclear medicine and three of the four in diagnostic radiology. The database was limited to "staff only" dictations (those without resident involvement) so that the reports reflected only the reporting style of staff physicians. Because of the varying length of reports and the difficulty in meaningfully characterising the occasional short "normal" report, textual selections were constructed from five different reports combined into a single textual body. This process created 15 textual samples from the original 75 reports representing each of the four staff physicians studied, each sample consisting of the text from five independent reports. The report texts consisted of freeform transcriptions and included only the body of the report and the conclusion. All standard entries, such as dosages, clinical information, and standard heading titles, were omitted so that the text would reflect only the author's personal reporting style and not be influenced by standard verbiage common to all reports.

Language processing

These 60 samples were processed using both Digitext Diction and Microsoft Word to evaluate 12 independent variables, described below. Each variable was quantified by the appropriate software, resulting in a numeric value for each report analysed. Microsoft Word was used to identify four variables: the number of sentences in the text, the percent usage of passive voice, the Flesch–Kincaid reading grade level, and the Flesch reading ease.[5] The Flesch–Kinkaid reading grade level and the Flesch reading ease are standard assessments used to quantify the reading difficulty of a passage.

Digitext Diction software was used to analyse text for eight linguistic parameters. Diction 7 is a computer-based textual analysis program that evaluates the "tone" of a verbal message by using dictionaries of words reflecting a particular expressive content. Tone refers to "a tool people use (sometimes unwittingly) to create distinct social impressions via word choice".[6] The assumption made is that "tone is the product of individual word choices that cumulatively build up to produce patterned expectations that tell an audience something important about the author's outlook".[7] The program employs large dictionaries of unique words that characterise certain tonal parameters whose frequency in a text indicates the intensity of a particular tone. The Diction analysis begins by defining several different variables to describe the tone of a text, each evaluating the use of language relating to a specific area.[8] The eight variables used were prospectively chosen because of their anticipated likelihood of relevance to the radiology report. These included: (a) a variable quantifying the degree of "ambivalence" (words expressing hesitation or uncertainty, implying a speaker's inability or unwillingness to commit to the verbalisation being made). Included are hedges, statements of inexactness, confusion, restrained possibility, and mystery; (b) a variable quantifying "communication" (terms referring to social interaction, including both modes and moods of intercourse, as well as terms reflecting various social purposes); (c) a variable quantifying "cognition" (words referring to cerebral processes, both functional and imaginative); (d) a variable quantifying "passivity" (words ranging from neutrality to inactivity, including terms indicating compliance, docility, and cessation). Note that passivity here refers to the use of specific words rather than the grammatical passive voice; (e) a variable quantifying "concreteness" (a large dictionary with little thematic unity excepting words characterised by tangibility and materiality); (f) a variable quantifying "embellishment" (a selective ratio of adjectives to verbs based on the conception that heavy modification slows down a verbal passage by de-emphasising human and material action); and (g) a variable quantifying "complexity" (a measure of the average number of characters-per-word). The eighth parameter that was employed was the number of different words used in the report (although this is not strictly a tonal parameter, it is measured by the Diction software). The four variables obtained from Microsoft Word and the eight variables derived from Digitext Diction comprise the 12 variables examined.

Neural network processing

Having quantified these 12 parameters for each radiologist, it was of interest to determine how specific they were