Quantitative Analysis of Uncertainty in Medical Reporting: Creating a Standardized and Objective Methodology

Bruce I. Reiner¹

Published online: 22 December 2017
© Society for Imaging Informatics in Medicine 2017

Abstract

Uncertainty in text-based medical reports has long been recognized as problematic, frequently resulting in misunderstanding and miscommunication. One strategy for addressing the negative clinical ramifications of report uncertainty would be the creation of a standardized methodology for characterizing and quantifying uncertainty language, which could provide both the report author and reader with context related to the perceived level of diagnostic confidence and accuracy. A number of computerized strategies could be employed in the creation of this analysis including string search, natural language processing and understanding, histogram analysis, topic modeling, and machine learning. The derived uncertainty data offers the potential to objectively analyze report uncertainty in real time and correlate with outcomes analysis for the purpose of context and user-specific decision support at the point of care, where intervention would have the greatest clinical impact.

Keywords Report uncertainty · Data mining · Natural language processing · Machine learning

Introduction

It is often stated that the only certainty in medicine is the pervasiveness of uncertainty [1]. While medicine is often viewed as a science driven by objective data, it should also be considered to be part art form, due to its frequent subjective and imprecise nature. This imprecise nature often takes the form of *uncertainty*, which has been described as the “Achilles heel” of the radiology report [2]. A number of studies have shown that the non-standardized language used to express uncertainty in medical reports frequently results in miscommunication and misunderstanding between the report author and reader [3, 4]. Clinicians and other healthcare professionals often interpret and act upon language of uncertainty in different ways, based on their perceived understanding (or misunderstanding) of the intent of the reporting physician [3].

For radiology reporting uncertainty, the resulting gap between radiologist intention and clinician perception may become greater as familiarity and direct interactions between radiologists and clinicians decrease with the adoption of picture archival and

communication systems (PACS), increasing workload demands, and outsourcing of radiologist services through teleradiology [5, 6]. The potential negative outcomes for report uncertainty and misunderstandings include delayed diagnosis or clinical management, increased costs (due to additional and sometimes unnecessary imaging and/or clinical tests), iatrogenic morbidity (associated with interventional procedures), and misdiagnosis [7]. These negative clinical implications associated with report uncertainty can also extend into the patient population, as patient empowerment and information transparency result in greater amounts of medical data being directly shared with patients [8].

One potential solution to address the misunderstanding and negative clinical impact associated with radiology report uncertainty is to standardize the language, context, quantification, and clinical meaning of uncertainty.

Quantitative Analysis of Medical Data

The communication and understanding of medical report data can in theory be improved upon when accompanied by quantitative measures of uncertainty. These uncertainty measures could provide the reader with context relating to the perceived level of diagnostic confidence and accuracy. Similar metrics are routinely provided in the context of laboratory medicine in which the uncertainty measurement provides a quantitative

✉ Bruce I. Reiner
breiner1@comcast.net

¹ Department of Radiology, Veterans Affairs Maryland Healthcare System, 10 North Greene Street, Baltimore, MD 21201, USA

estimate of test result quality and is an integral component of clinical testing quality assurance [9]. In response to the presence of uncertainty and requirement for its measurement, various international standards bodies have jointly developed a *Guide for the Expression of Uncertainty in Measurement* (GUM), for expressing standardized measurement of uncertainty in medical testing laboratories [10].

One can draw a number of parallels between uncertainty measurements in laboratory medicine and medical imaging. In both disciplines, a well-documented variety of external factors that affect data uncertainty have been noted. In medical testing, these factors include patient preparation, biological variation, poor specimen collection or transport, clerical and reporting errors, and various other patient-related variables (e.g., stress, drugs, food intake). In medical imaging, similar external factors affect data uncertainty including poor patient preparation, anatomic variation, suboptimal exam or protocol selection, reporting errors, image quality deficiencies, and patient-related variables (e.g., compliance, body habitus, age, comorbidities). Two important differences between laboratory and medical imaging uncertainty include differences in the way the data is expressed (i.e., numerical versus text) and the increased importance of human factors in medical imaging uncertainty, the most important of which are tied to the radiologist who serves as the primary data source of the report. As a result, quantitative assessment of uncertainty in medical imaging must rely predominantly on text-based analysis and take into account both context and user-specific variables.

Innovation Opportunity

Before creating an innovation strategy to quantify uncertainty in medical reporting, the principle goals and objectives should be defined. First and foremost is the creation of a standardized method for uncertainty measurement, which can be applied to all radiology reports regardless of the exam type, clinical circumstances, technology in use, report format, and provider characteristics. If truly standardized, these data can in turn be used to populate referenceable databases which can combine data from multiple institutional providers and in turn be used for a variety of applications including research (e.g., outcome analysis), creation of practice standards and guidelines, development of decision support tools, practice management, and comparative performance assessment. Secondly, the strategy must be adaptable to the diverse community of end-users. If not, adoption will be poor unless mandated by regulatory authorities and/or payers. Thirdly, the technology must demonstrate clear, tangible improvements to everyday users, which ideally should extend beyond the radiologist community to include clinicians, technologists, administrators, and patients. If the ultimate goal is to improve accuracy, understanding, communication, economics, and clinical

outcomes, all involved stakeholders should benefit. Fourthly, the technology should be workflow neutral, meaning that its adoption into everyday practice should not create additional time requirements for the participants. If data input could be automated, this could in theory improve workflow while also removing the potential for manual data entry error.

Taking these objectives into account, one proposed innovation strategy consists of a statistical model which converts text-based uncertainty concepts into standardized numerical measures which quantify the level of uncertainty or diagnostic confidence. In creating such a standardized method of quantifying uncertainty, it is important for both the author (e.g., radiologist) and reader (e.g., clinician) of the report to understand the degree of uncertainty being expressed, as well as the associated clinical implications and recommendations. In theory, this model for quantifying uncertainty in a standardized format could reduce (or potentially eliminate) existing problems relating to miscommunication and misunderstanding of report uncertainty.

Development Strategies

A number of potential strategies can be employed to create a standardized method for quantifying text-based medical uncertainty, involving both human and computerized analysis. Since uncertainty is ubiquitous in medical reporting, the initial method for quantification can utilize input from domain experts in quantifying text-based uncertainty from a large sampling of reports. In this application, the domain experts would be provided with a standardized method for quantifying uncertainty, with the resulting data pooled in an effort to define consensus. In addition to domain experts, everyday end-users can also be used in data collection and analysis with the collection of these data expanding to everyday practice. The ultimate goal is the creation of large sample size statistics which can account for the large diversity of end-users, exam types, clinical contexts, language variation, and institutional practices. The collection of these user-derived uncertainty data can be subsequently extended to everyday practice.

In addition to human-derived uncertainty analysis, a number of computerized technologies are available (and constantly evolving) to automate text-based uncertainty analysis. One of the most simplistic computerized technologies which can be used for uncertainty analysis in text reporting is rules based or string search, in which key words or combinations of words associated with uncertainty are identified. Examples of these would be words such as “possible,” “probable,” “definite,” “uncertain,” “likely,” or “unlikely.” Examples of phrases would include “consistent with,” “compatible with,” “diagnostic of,” and “cannot exclude.” Identifying these and other terms and assigning them uncertainty scores based on radiologist and/or clinician perceptions would provide a relatively

straightforward first pass at addressing the challenge of quantification of uncertainty.

At the next level of technological sophistication for uncertainty, analysis is natural language processing (NLP) and natural language understanding (NLU). These can be used to deconstruct and parse text-based sentences in order to identify and characterize uncertainty on the basis of individual word components while also identifying association relationships between different words and phrases. Whereas, NLP acts by deconstructing and parsing word combinations, NLU acts to extract semantic information by “understanding” the specific context in which words and phrases are used, along with their intended meaning [11, 12]. This use of natural language understanding could be utilized to determine likely meaning of phrases that were not specifically searched for using the strategy of exact matching of words or phrases. The simplest examples of this would involve the application of word negation, such as “not likely,” or “not compatible with.” More complex examples would involve extraction of meaning from phrases expressing uncertainty such as “I have no doubt that,” or “absolutely certain” or “cannot be determined.” These could of course be added to phrases that could be simply searched for; but based on clinical experience, there are so many combinations of these terms that are routinely utilized in radiology reports that a natural language parser is required, especially when negations or word combinations can stretch across a long sentence without being contiguous or even across more than one sentence.

Additional specific computerized technologies can be used for uncertainty analysis including but not limited to) histogram analysis, machine learning, and topic modeling [13–15]. Histogram analysis creates frequency tables, which can quantify and characterize uncertainty terminology as well as correlate the terminology with outcomes analysis. This represents a semi-automated way to discover the majority of words and phrases associated with uncertainty without requiring experts to determine those phrases and terms. Topic modeling is a statistical tool that analyzes words and phrases and their relationships to one another. This represents a more sophisticated tool with which to derive meaning from radiology reports in an unsupervised manner by looking at words and phrases from a large corpus of reports, which are available by mining speech recognition systems or electronic medical records. Without any understanding of the content of radiology reports, topic modeling using techniques such as Latent Dirichlet Allocation can analyze word clusters and analyze the interaction between various word combinations in large numbers of radiology reports [16]. This can be utilized to extract the subset of terms associated with concepts of uncertainty. This approach to extract uncertainty using topic modeling has been applied to industry analyst reports and corporate disclosures [17].

A wide variety of other types of Bayesian and machine learning techniques have been described and applied to

uncertainty analysis (e.g., Bayesian analysis, support vector machine), which can combine a number of disparate variables into the analysis. As an example, machine learning can be applied to uncertainty analysis of an abdominal CT report performed for suspected appendicitis just as it has been in the diagnosis of the disease [18]. In addition to analysis of the report findings, additional report data can be included in the analysis (e.g., patient age, identity of the referring clinician), with the goal of determining how uncertainty report language correlates with appendicitis. If abdominal CT reports are analyzed comparing two different radiologists, it is determined that the same language of uncertainty may be associated with entirely different clinical outcomes (i.e., diagnosis of appendicitis). Despite similar degrees of uncertainty language, one radiologist may have a positive predictive value for appendicitis of 95% (i.e., uncertainty language is almost always associated with the diagnosis of appendicitis), whereas the second radiologist report analysis (using the same language expressing uncertainty) has a positive predictive value for appendicitis of only 50%. This illustrates how machine learning in text-based uncertainty analysis can be used to predict clinical outcomes in relationship to a variety of clinical, technical, and demographic variables. Another way of expressing this is that machine learning can determine the relationship between report language and clinical outcomes based on end-user and contextual differences. This ability to combine report language and clinical outcomes using machine learning can result in derived uncertainty measurements, as opposed to rules-based uncertainty measurements which are derived from expert consensus and/or computerized rules-based analysis.

Standardized Uncertainty Metrics

Figure 1 provides an example of how this quantitative analysis of uncertainty language can be created, utilizing both positive and negative numerical values to correlate with “positive” and “negative” report findings. A positive form would refer to the presence of a given finding or disease state, while the negative form would apply to the absence of a given finding or disease state. In accordance with these applications, the standardized numerical quantifier of uncertainty is preceded by a positive (+) or negative (−) sign. While a number of variations can be created to quantify report uncertainty in a standardized fashion, the end result is the same. The syntax and semantics of language is analyzed over large sample volumes (through computerized and/or human methods of analysis) and classified in a standardized fashion in accordance with uncertainty presence and magnitude. The applied “uncertainty measurement” is then validated through end-user feedback and computerized iterative refinement to create a mathematical measure of uncertainty relating to both positive and negative report findings, which can then be recorded in a referenceable

Fig. 1 Representative model for standardized quantification of report uncertainty

database for longitudinal analysis, research, feedback, creation of professional guidelines, technology testing and development, and educational/decision support applications.

The resulting radiologist-derived uncertainty scores and associated text could be presented in summary format and/or highlighted in the report (e.g., color coded) whenever the report is opened and reviewed by a third party. The reader would then have the option to respond to the uncertainty data in a similar standardized manner by rendering their perceived level of agreement, as relating to the uncertainty language in the report. In the event that the report reader was to perceive a difference in uncertainty magnitude, they could input their preferred uncertainty score (Table 1). In cases where multiple end-users have reviewed and replied to the uncertainty scores in question, the original radiologist derived uncertainty score

would serve as the default uncertainty score of record. Since each radiology report would have a large number of uncertainty scores (i.e., in accordance with each individual finding), the analysis could focus on those uncertainty scores which exceed a predefined uncertainty measure threshold and/or are associated with a high-priority clinical finding or disease state.

To illustrate how this might be applied in practice, we can take a radiology report describing a poorly defined density on chest CT. The report language of record states “poorly defined density in the right upper lobe, which could represent either neoplastic mass or an inflammatory process such as infiltrate. Bronchoscopy can be performed for further evaluation if clinically indicated. Before the report is signed and submitted for clinical review, a computerized text analysis is performed which identifies language of uncertainty and presents the highlighted report text (with or without a computer-derived uncertainty grade) for radiologist review and grading using the standardized uncertainty scoring model. In this example, both the radiologist and computer-derived uncertainty scores are recorded as “3 (Intermediate level of uncertainty)”. If the radiologist was to determine that the report uncertainty language did not match their desired intent of uncertainty to be communicated, they could modify the report language in

Table 1 Inter-observer uncertainty score assessment

A: Agree with radiologist recorded uncertainty score
B: Minor disagreement with radiologist recorded uncertainty score (input preferred uncertainty score)
C: Significant disagreement with radiologist uncertainty score (input preferred uncertainty score)

order to arrive at the desired level of uncertainty. This report editing function could be performed independently by the radiologist (i.e., editing the report language without assistance) or through computer-derived assistance. In the latter case, the radiologist could input the desired level of uncertainty to be communicated (e.g., 2-intermediate to high level of certainty). The computer could in turn provide a number of context appropriate text options consistent with the intended level uncertainty. The radiologist could then select the preferred language from the computer-derived text options presented, and the report would be automatically edited to reflect the new text and intended level of uncertainty.

The resulting uncertainty score data (including the specific uncertainty language and corresponding scores of individual end-users) could in turn be recorded into a report uncertainty database for the purpose of longitudinal analysis. By incorporating the identities of each participating end-user, one could in theory create both context and user-specific uncertainty analyses, which can be used for a number of subsequent applications (e.g., computer-derived uncertainty scores).

Complementary data which can be used to support the standardized quantitative uncertainty metrics include probability statistics which can be provided in the form of a numerical range and provide the reader with context relating to the degree of uncertainty and its outcome probability. These probability statistical ranges can be created commensurate with the degree of uncertainty and can be presented in generalized, context-specific, and user-specific forms. As an example, suppose a radiologist were to report the following on a hip radiograph in the setting of trauma. “No obvious hip fracture identified, CT correlation recommended if an occult fracture remains of clinical concern”. In addition to the standardized uncertainty score of -4 , an associated probability range can be presented to the report reader related to the probability of fracture detected on CT in the presence of a “negative” finding on a radiograph. These probability statistics can be reported in a range in a variety of formats including generalized (e.g., the overall probability range for all radiographic exams), context specific (e.g., the probability range for only hip radiograph, with and without patient age considerations), or user specific (e.g., probability statistics for the individual reader). The ultimate goal is to standardize uncertainty in quantitative terms, provide clinical context through probability statistics, and improve communication and understanding through collaborative data collection and analysis.

References

1. Luther VP, Crandall SJ: Ambiguity and uncertainty: neglected elements of medical education curricula? *Acad Med* 86:799–800, 2011
2. Reiner B: Uncovering and improving upon the inherent deficiencies of radiology reporting through data mining. *J Digit Imaging* 23: 109–118, 2010
3. Lindley SW, Gillies EM, Hassell LA: Communicating diagnostic uncertainty in surgical reports: disparities between sender and receiver. *Pathol Res Pract* 210:628–633, 2014
4. Reiner B, Siegel E, Protopapas Z et al.: Impact of filmless radiology on the frequency of clinician consultations with radiologists. *AJR* 173:1169–1172, 1999
5. Reiner BI: Strategies for radiology reporting and communication. Part I: challenges and heightened expectations. *J Digit Imaging* 26: 610–613, 2013
6. Khorasani R, Bates DW, Teeger S et al.: Is terminology used effectively to convey diagnostic certainty in radiology reports? *Acad Radiol* 10:685–688, 2003
7. Reiner B: A Crisis in confidence: A combined challenge and opportunity for medical imaging providers. *J Am Coll Radiol* 2: 107–108, 2014
8. Hanauer DA, Liu Y, Mei Q et al.: Hedging their bets: the use of uncertainty terms in clinical documents and its potential implications when sharing the documents with patients. *AMIA Annu Symp Proc* 2012:321–330, 2012
9. White GH, Farrance I: Uncertainty of measurement in quantitative medical testing: A laboratory guide. *Clin Biochem Rev* 25:S1–S24, 2004
10. Guide to the Expression of Uncertainty in Measurement. ISO: Geneva, 1995. ISBN 92-67-10188-9
11. Pons E, Braun LMM, Hunink MGM et al.: Natural language processing in radiology: a systematic review. *Radiology* 279:329–343, 2016
12. Cai T, Giannopoulos AA, Yu S et al.: Natural language processing techniques in radiology research and clinical applications. *RadioGraphics* 36:176–191, 2016
13. Savova GK, Masanz JJ, Ogren PV et al.: Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation, and applications. *J Am Med Inform Assoc* 17:507–513, 2010
14. Meystre SM, Sakova GK, Kipper-Schuler KC et al.: Extracting information from textual documents in the electronic health record: a review of recent research. *IMIA Yearbook of Medical Informatics* 47:128–144, 2008
15. Chau M, Chen H: A machine learning approach to web page filtering using context and structure analysis. *Decision Support Systems* 44:482–494, 2008
16. Blei D, Ng A, Jordan M: Latent Dirichlet allocation. *Journal of Machine Learning Research* 3:993–1022, 2003
17. Huang A, Leavy R, Zang A, et al: Analyst information discovery and interpretation roles: A topic modeling approach. *Manag Sci*, 2017 doi:[10.1287/mnsc.2017.2751](https://doi.org/10.1287/mnsc.2017.2751)
18. Hsieh CH, Lu RH, Lee NH et al.: Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. *Surgery* 149:87–93, 2011