

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Is It Possible to Automatically Assess Pretreatment Digital Rectal Examination Documentation using Natural Language Processing? A Single Center Retrospective Study
AUTHORS	Bozkurt, Selen; Kan, Kathleen; Ferrari, Michelle; Rubin, Daniel; Blayney, Douglas; Hernandez-Boussard, Tina M.; Brooks James

VERSION 1 - REVIEW

REVIEWER	David Kaufman Arizona State University, USA
REVIEW RETURNED	10-Dec-2018

GENERAL COMMENTS	<p>This is a well-written paper addressing an important problem, namely the automated assessment of digital rectal examinations prior to treatment using NLP techniques. The paper provides a sound rationale for the need for such automated methods given the nature of the problem. The overall approach and study appear to be sound. The analysis of patients' characteristics is appropriate and informative. This study provides a potential contribution to knowledge regarding the efficacy of using NLP to discern quality measures related to proper screening. I have several specific concerns.</p> <p>The rationale for addressing digital rectal examinations is fully adequate. However, the background section should have more information on the use of NLP methods for addressing comparable problems. The review need not be extensive, but it would better situate this work within a growing area of research of considerable importance.</p> <p>The NLP methods are not adequately described in this document. Although the NLP pipeline and related work were previously described in an AMIA paper, more detail should be included here so that the readers could make informed judgments about the approach. It would also be insightful to see how DREs are documented in a clinical note—one example would suffice.</p> <p>The difference between stages in documentation is interesting. It's clear why there would be differences between stages 1 and 4. But the differences between stages 2 and 3 is somewhat anomalous and worthy of commentary.</p>
-------------------------	---

	<p>It would be interesting to know the distinguishing features of a note in which DREs were not documented. That is perhaps a different research question for another study.</p> <p>The authors make a clear and compelling case for the use of NLP to identify (and perhaps improve) adherence to quality guidelines in this and related clinical contexts. They further suggest that the use of such methods could be used to avoid increasing documentation burden to clinicians. If the documentation of DRE is sufficiently important, would that not warrant specific guidelines for documentation or even the use of structured fields? It is reasonable to assume that many medical institutions may not have the necessary expertise or infrastructure to employ these set of methods. The tradeoff of documentation burden versus greater precision for measuring adherence warrants some discussion.</p> <p>On a minor note, I found the phrase “A Key Quality Metric for Prostate Cancer Care” to be a little confusing. It implies the introduction of a new quality metric. “Automatic Assessment of Pretreatment Digital Rectal Examination Documentation” is a comprehensible and sufficiently informative title. Of course, that is up to the discretion of the authors.</p>
--	---

REVIEWER	Raj Ratwani, Phd MedStar Health, Washington, DC United States
REVIEW RETURNED	17-Dec-2018

GENERAL COMMENTS	<p>The authors present an interesting topic developing and testing a method for automatic assessment of a quality metric, provider-documented pretreatment digital rectal examination (DRE), utilizing a natural language processing (NLP) framework. The article is very well written and contains limited typos. In this state, the manuscript requires a minor revision and resubmission.</p> <p>Introduction:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Very well organized and provides a clear rationale for this study <p>Method/Measures:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Figure 1 seems a bit unnecessary and could easily be described as text <input type="checkbox"/> Figure 2 is helpful in explaining the study cohort (i.e. inclusion/exclusion criteria) <p>Results:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Overall, results could be cleaned up and simplified <input type="checkbox"/> The first paragraph in the results section is wordy and confusing (unsure what the N is as the percentages do not add to 100%) <input type="checkbox"/> Not all tables and figures are necessary – consider combining Tables 2 and 3 and deleting Figure 3 <ul style="list-style-type: none"> <input type="checkbox"/> Figure 3 does not clearly visualize any important results <p>Discussion/Conclusion:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Well structured and conveys appropriate results and implications <input type="checkbox"/> Consider alternatives as for the increase in reporting over time – were policies also put in place? Any federal or state requirements?
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: David Kaufman

Institution and Country: Arizona State University, USA

Please state any competing interests or state 'None declared': I have no competing interests to declare.

Please leave your comments for the authors below

This is a well-written paper addressing an important problem, namely the automated assessment of digital rectal examinations prior to treatment using NLP techniques. The paper provides a sound rationale for the need for such automated methods given the nature of the problem. The overall approach and study appear to be sound. The analysis of patients' characteristics is appropriate and informative. This study provides a potential contribution to knowledge regarding the efficacy of using NLP to discern quality measures related to proper screening. I have several specific concerns.

The rationale for addressing digital rectal examinations is fully adequate. However, the background section should have more information on the use of NLP methods for addressing comparable problems. The review need not be extensive, but it would better situate this work within a growing area of research of considerable importance.

Response: Thank you for the suggestion. We expanded the background section to describe in detail the use of our NLP methods.

The NLP methods are not adequately described in this document. Although the NLP pipeline and related work were previously described in an AMIA paper, more detail should be included here so that the readers could make informed judgments about the approach. It would also be insightful to see how DREs are documented in a clinical note—one example would suffice.

Response: Thank you for the suggestion. In the revised manuscript, we present more detail about our NLP methods and now provide examples for DRE documentation to illustrate the method (Figure 2).

The difference between stages in documentation is interesting. It's clear why there would be differences between stages 1 and 4. But the differences between stages 2 and 3 is somewhat anomalous and worthy of commentary.

Response: We agree with the reviewer that it is somewhat surprising that stage 3 patients had the highest rates of DRE. We have added commentary in the discussion: "It is not clear why Stage 3 patients had the highest rates of DRE recorded. By definition, most patients with Stage 3 cancer will have palpable abnormalities on DRE, and the higher rates might reflect a bias in physicians recording positive findings on physical examination or a higher likelihood of performing an exam in a patient referred to their practice with a previously documented abnormality."

It would be interesting to know the distinguishing features of a note in which DREs were not documented. That is perhaps a different research question for another study.

Response: We agree that it would be interesting to understand the features of notes in which DREs were not documented, but we also agree that it would be beyond the scope of the present study. This line of investigation is a logical next step that we will explore since it could allow us to identify the types of notes that "should" have a DRE, and potentially flag them for this documentation so that a DRE can be performed at the next visit. The distinguishing features could include the type/nature of the visit, whether it is a new or return visit, the provider or practice setting and others. DRE

documentation is likely to be most helpful in the initial diagnostic visit since it is used to clinically stage the patient. Once appropriately staged, therapeutic counseling and planning can begin; those visits following assignment of clinical stage will often not have a DRE documented since treatment has been initiated and staging is no longer relevant since DRE is not used to monitor therapy outcome. Patients seen for a second opinion, which comprises a large part of our patient population, often come with an assigned clinical stage, and this might account for a “missed” DRE at the first visit. In some cases, such as cases of low or favorable intermediate risk cancer, patients might elect to delay treatment due to coordination issues, while they seek second opinions, or when they are required to address co-morbidities that require further attention or treatment before their prostate cancer can be addressed. With a long enough delay from initial staging DRE, a repeat exam should be performed since cancer is a dynamic disease. Therefore, the timing of an initial or repeat DRE needs to be close enough to ensure that one has the most up to date clinical staging and therefore the most effective/correct treatment plan.

The authors make a clear and compelling case for the use of NLP to identify (and perhaps improve) adherence to quality guidelines in this and related clinical contexts. They further suggest that the use of such methods could be used to avoid increasing documentation burden to clinicians. If the documentation of DRE is sufficiently important, would that not warrant specific guidelines for documentation or even the use of structured fields? It is reasonable to assume that many medical institutions may not have the necessary expertise or infrastructure to employ these set of methods. The tradeoff of documentation burden versus greater precision for measuring adherence warrants some discussion.

Response: Dr. Kaufman raises an interesting and important point that we have addressed by adding the following to the Discussion section: “Several guidelines recommend performance of a DRE prior to initiation of therapy but are mute on whether and how this quality metric needs to be documented. The rapid rise of the EHR offers new opportunities for measuring compliance with diverse quality metrics that was previously impossible with paper records. We have focused on using the EHR to address the need to measure compliance with several quality metrics in prostate cancer care, including DRE, and have made these algorithms freely available so that other health care systems can modify and use them. In the near future, measurement of compliance with quality metrics, with the goal of improving patient outcomes, will be used to monitor and quantify health care quality. Compliance with quality metrics could be directly tethered to reimbursements, in which case, building structured fields for documentation of quality metrics will increase since these will improve compliance with, and documentation of quality metrics like performance of a DRE prior to initiating therapy.

On the other hand, many of the >100 quality metrics proposed in prostate cancer are based on expert opinion and other low-level evidence. Given the widespread concerns about “click fatigue” and physician burn-out related to documentation in the EHR, required structured fields need to be deployed wisely and judiciously. As we move to a learning-based health care system with the EHR at its hub, capture of compliance and non-compliance with quality metrics with the automated methods such as ours will provide opportunities to assess the true effects of compliance with disease-related outcomes (such as higher rates of positive surgical margins and recurrence in patients undergoing surgery who did not receive a DRE) as well patient centered outcomes. Ideally, quality metrics that significantly effect clinical endpoints and patient centered outcomes can be tested and incorporated as structured fields, while those that do not can be discarded to decrease documentation burdens on physicians. In addition, understanding which quality metrics lead to improved outcomes and patient experience would allow insurers to tailor reimbursements to incentivize improvements in quality.”

On a minor note, I found the phrase “A Key Quality Metric for Prostate Cancer Care” to be a little confusing. It implies the introduction of a new quality metric. “Automatic Assessment of Pretreatment Digital Rectal Examination Documentation” is a comprehensible and sufficiently informative title. Of course, that is up to the discretion of the authors.

Response: Thank you very much for the suggestion. As mentioned above, based on the editor's suggestion, we changed our title to: “Is It Possible to Assess Pretreatment Digital Rectal Examination Documentation Automatically Using Natural Language Processing? A Single Center Retrospective Study” in order to include the research question, study design and setting.

Reviewer: 2

Reviewer Name: Raj Ratwani, Phd

Institution and Country: MedStar Health, Washington, DC United States

Please state any competing interests or state 'None declared': I have no competing interests to report.

Please leave your comments for the authors below

The authors present an interesting topic developing and testing a method for automatic assessment of a quality metric, provider-documented pretreatment digital rectal examination (DRE), utilizing a natural language processing (NLP) framework. The article is very well written and contains limited typos. In this state, the manuscript requires a minor revision and resubmission.

Introduction:

Very well organized and provides a clear rationale for this study

Response: Thank you for the comment.

Method/Measures:

Figure 1 seems a bit unnecessary and could easily be described as text

Response: We removed Figure 1 from the main text but kept it as a supplementary figure since it concisely illustrates our research question and analytic pipeline.

Figure 2 is helpful in explaining the study cohort (i.e. inclusion/exclusion criteria)

Response: Thank you for the comment. Based on this recommendation, we have added this figure to the main text as Figure 1.

Results:

Overall, results could be cleaned up and simplified

Response: The Results section was reorganized.

The first paragraph in the results section is wordy and confusing (unsure what the N is as the percentages do not add to 100%)

Response: Thank you for pointing out the issue with the percentages as reported in the paper. Since the denominators were different for each percentage, they did not add to 100%. The complete data are now given in Table 1 and the text has been clarified.

Not all tables and figures are necessary – consider combining Tables 2 and 3 and deleting Figure 3. Figure 3 does not clearly visualize any important results

Response: Thank you for the comment. We think Figure 3 nicely displays the change in compliance over time but have moved it to the supplementary materials to simplify the main manuscript. We decided against merging the tables since in so doing, the much smaller font size rendered them unreadable. We decided to keep them separate.

Discussion/Conclusion:

Well-structured and conveys appropriate results and implications

Response: Thank you for the comment.

Consider alternatives as for the increase in reporting over time – were policies also put in place? Any federal or state requirements?

Response: On the federal level, the goals of meaningful use in EHR include improving quality and efficiency in the health care system as well as improving care coordination and overall population health. Appropriate DRE documentation as a means to improving the quality of prostate cancer care falls within these goals, while automating this type of assessment provides opportunities for to both improving efficiency of assessment health care systems locally as part of Quality of Care improvement, as well as and applying this tool to larger populations. In addition, there has been increased emphasis on the secondary use of clinical text, particularly for guideline adherence and quality assessment. However, to our knowledge there were no new policies directly affecting DRE documentation over the study period.