

# Clinical decision support with automated text processing for cervical cancer screening

Kavishwar B Wagholikar,<sup>1</sup> Kathy L MacLaughlin,<sup>2</sup> Michael R Henry,<sup>3</sup> Robert A Greenes,<sup>4,5</sup> Ronald A Hankey,<sup>6</sup> Hongfang Liu,<sup>1</sup> Rajeev Chaudhry<sup>7</sup>

► Additional appendices are published online only. To view these files please visit the journal online (<http://dx.doi.org/10.1136/amiainjnl-2012-000820>).

<sup>1</sup>Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota, USA

<sup>2</sup>Division of Family Medicine, Mayo Clinic, Rochester, Minnesota, USA

<sup>3</sup>Division of Anatomic Pathology, Mayo Clinic, Rochester, Minnesota, USA

<sup>4</sup>Department of Biomedical Informatics, Arizona State University, Phoenix, Arizona, USA

<sup>5</sup>Department of Health Science Research, Mayo Clinic, Scottsdale, Arizona, USA

<sup>6</sup>Population Management Systems, Mayo Clinic, Rochester, Minnesota, USA

<sup>7</sup>Division of Primary Care Internal Medicine, Center for Innovation, Mayo Clinic, Rochester, Minnesota, USA

## Correspondence to

Dr Kavishwar B Wagholikar, Division of Biomedical Statistics and Informatics, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA; [waghsk@gmail.com](mailto:waghsk@gmail.com)

Received 9 January 2012

Accepted 31 March 2012

Published Online First

29 April 2012

## ABSTRACT

**Objective** To develop a computerized clinical decision support system (CDSS) for cervical cancer screening that can interpret free-text Papanicolaou (Pap) reports.

**Materials and Methods** The CDSS was constituted by two rulebases: the free-text rulebase for interpreting Pap reports and a guideline rulebase. The free-text rulebase was developed by analyzing a corpus of 49 293 Pap reports. The guideline rulebase was constructed using national cervical cancer screening guidelines. The CDSS accesses the electronic medical record (EMR) system to generate patient-specific recommendations. For evaluation, the screening recommendations made by the CDSS for 74 patients were reviewed by a physician.

**Results and Discussion** Evaluation revealed that the CDSS outputs the optimal screening recommendations for 73 out of 74 test patients and it identified two cases for gynecology referral that were missed by the physician. The CDSS aided the physician to amend recommendations in six cases. The failure case was because human papillomavirus (HPV) testing was sometimes performed separately from the Pap test and these results were reported by a laboratory system that was not queried by the CDSS. Subsequently, the CDSS was upgraded to look up the HPV results missed earlier and it generated the optimal recommendations for all 74 test cases.

**Limitations** Single institution and single expert study.

**Conclusion** An accurate CDSS system could be constructed for cervical cancer screening given the standardized reporting of Pap tests and the availability of explicit guidelines. Overall, the study demonstrates that free text in the EMR can be effectively utilized through natural language processing to develop clinical decision support tools.

Although cervical cancer is highly preventable, it still continues to be a leading cause of death. Guidelines for cervical cancer are complex and not easily recalled by health providers. Consequently, all patients do not receive the optimal screening.<sup>1–4</sup> A potential solution is the use of a clinical decision support system (CDSS) at the point of care, which will recommend the optimal screening/management decision based on the guidelines.<sup>5</sup>

However, an obstacle for developing the CDSS is that application of the national screening and management guidelines involves free-text Papanicolaou (Pap) reports that are not readily amenable for computer interpretation.<sup>6</sup> The objective of this work is to develop a CDSS for cervical cancer screening that can interpret the free-text Pap reports. In this paper, we describe the development and evaluation of the CDSS.

## BACKGROUND

### Cervical cancer screening

Cervical cancer is the fourth leading cause of cancer deaths in women worldwide. Twelve thousand seven hundred and ten new cases were estimated to have occurred in 2011 in the USA alone.<sup>7</sup> Failure to screen with a Pap test is the most common attributable factor for developing cervical cancer.

More than half of the women diagnosed with cervical cancer were found to have inadequate screening.<sup>8–9</sup> Twenty-two per cent of women surveyed in a recent national health review had not had a Pap test in the past 3 years.<sup>10</sup> The challenge now is to ensure that all women get the appropriate screening and management.<sup>8–11</sup> The American Cancer Society, United States Preventive Services Task Force and American College of Obstetricians and Gynecologists have released screening guidelines based on several factors including age, risk of cervical cancer and previous screening test results.<sup>12–15</sup> The American Society for Colposcopy and Cervical Pathology has released guidelines for the follow-up and management of abnormal cervical screening tests.<sup>16</sup>

However, the guidelines developed by the national associations are complex and not easily recalled by health providers. Consequently, all patients do not receive the optimal screening or follow-up of abnormal results. Multiple recent publications have documented poor provider adherence in following guideline-consistent recommendations for cervical screening with cytology and human papillomavirus (HPV) testing.<sup>1–4</sup> As a potential solution, we aim to develop a CDSS that generates a guideline-based recommendation for screening and management for use at the point of care.<sup>5</sup> Dupuis *et al*<sup>17</sup> have shown that an intervention to identify and track patients with abnormal Pap results alone can improve follow-up. Our system will be more comprehensive and will generate recommendations for all female patients.

### Clinical decision support

CDSS provide decision suggestions to the health-care provider based on patient data in the electronic medical record (EMR) system.<sup>18</sup> CDSS have been developed for a wide range of decision tasks including diagnostic support,<sup>19–20</sup> preventive care,<sup>21</sup> disease management,<sup>22</sup> and prescription.<sup>23</sup> They have been shown to improve guideline adherence for preventive services,<sup>24</sup> and they are being increasingly adopted in health institutions. However, there are several obstacles to their widespread adoption, which includes lack of accuracy and inability to utilize free-text data.<sup>6–25</sup> The



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://jamia.bmj.com/site/about/unlocked.xhtml>

concerns for accuracy of CDSS have largely been due to the lack of explicit guidelines for decision making, which lead researchers to investigate a variety of paradigms to discover the optimal decision models.<sup>25</sup> However, for the case of cervical cancer screening, the decision guidelines have been developed by several national associations and they can be integrated and implemented in a rule-based CDSS. The other obstacle for the CDSS is that the CDSS will be required to interpret Pap reports that are in free-text form. We have considered natural language processing (NLP) to resolve this problem.

**NLP of clinical text**

Free-text/narrative data constitute a large portion of the patient data in the EMR. These include physicians' notes that describe patient symptoms, physical exam, treatment, and reports of procedures and some of laboratory tests like the Pap test. The narrative data pose a challenge for automated computer processing as they cannot be readily mapped to patient variables. Due to concerns about the validity and accuracy of the extracted information, free-text data have been largely underutilized for computer-based decision support.<sup>26</sup>

To address these concerns, more recently, several NLP tools<sup>27-29</sup> have been developed to facilitate information extraction from the clinical free text.<sup>30</sup> A NLP-based approach has been applied for a variety of problems. These include identifying patient cohorts,<sup>31 32</sup> reporting of notifiable diseases,<sup>35 34</sup> syndrome surveillance,<sup>35</sup> diagnostic classification,<sup>36</sup> identifying co-morbidities,<sup>37</sup> medication event extraction,<sup>38 39</sup> adverse event detection,<sup>40 41</sup> identification of postoperative complications,<sup>42</sup> and disease management. However, application of NLP for clinical decision support has been deferred, possibly because of the requirement for a high accuracy.<sup>6 43</sup>

Many clinical reports like those for radiology are well structured compared with physician notes, and these can be processed more reliably and accurately using NLP approaches.<sup>44</sup> Although the Pap test reports are in narrative form they appear well structured, like the radiology reports that have been the subject of earlier research on NLP. Our work is based on the premise that given the standardization of Pap reporting, it would be possible to construct a free-text processor that is sufficiently accurate for clinical application.

Earlier work by Aronsky *et al*<sup>20</sup> has demonstrated the development of accurate CDSS for diagnosing pneumonia that utilized a free-text processor for pneumonia reports. Demner-Fushman *et al*<sup>43</sup> have discussed the potential of NLP to enhance CDSS. Dupuis *et al*<sup>17</sup> recently constructed a free-text parser to

identify abnormal Pap reports. However, in addition to the identification of abnormal reports our CDSS would be required to interpret several other variables accurately in the report. We hypothesize that given the standardized reporting of Pap tests and the availability of explicit guidelines released by national institutions, the developed CDSS would attain the accuracy required for clinical decision support.

**METHODS**

This study was conducted at Mayo Clinic, which provides care to a million patients annually. It was approved by the institutional review board, and included only those patients who had consented to make their data available for research (IRB number 08-000513).

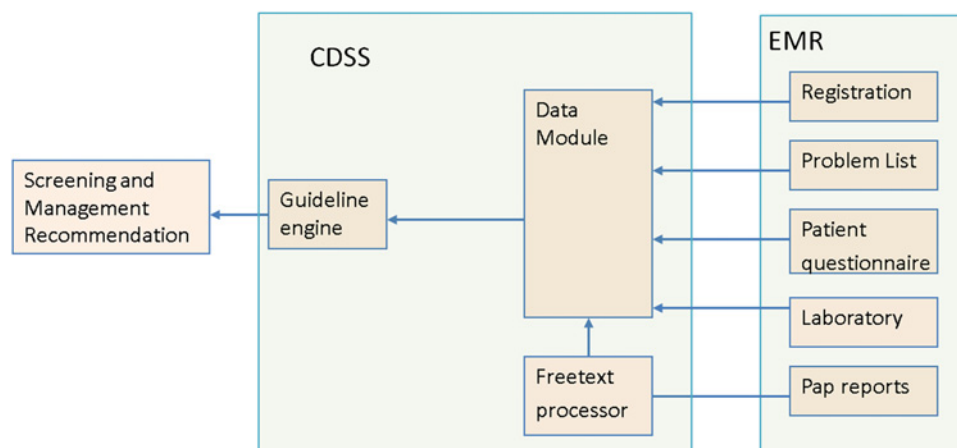
A CDSS was developed on the drools<sup>45</sup> platform. It has three modules (figure 1) and two rulebases: a free-text rulebase for interpreting Pap reports and a guideline rulebase for representing the screening and management guidelines. See the supplementary appendix (available online only) for instructions to download the rulebases. The free-text rulebase was developed using a corpus of 49293 Pap test reports, physician interview and an in-house template used at the Mayo Clinic for Pap report generation.

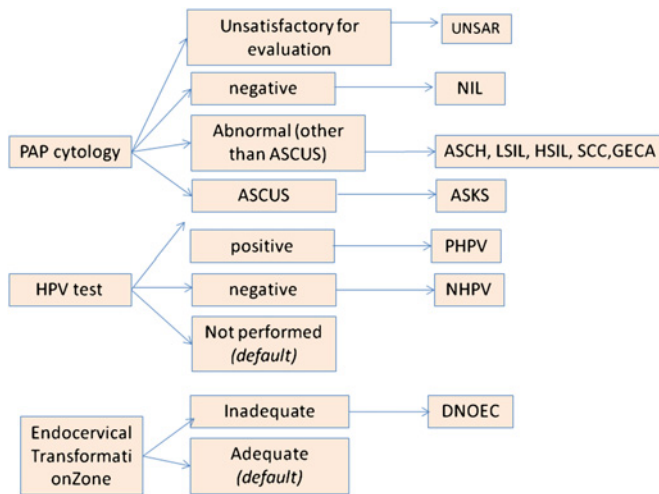
**Free-text rulebase**

A free-text processor conventionally involves two sequentially acting components: lexer and parser. The lexer transforms text to tokens and the parser processes the tokens to generate the output. The free-text rulebase accordingly has two sets of rules: lexer and parser.

The lexer rules were developed from an analysis of 49293 randomly selected Pap test reports from the past 5 years in the Mayo Clinic, Rochester, EMR. The reports were normalized by removing all non-alphanumeric characters and converting to lower case. Patterns of recurring groups of words were identified that occurred in at least 1% of the reports. These were mapped to the template used in the pathology department for report generation. This is an in-house developed system of text codes used by pathologists to annotate Pap test results. The pathology report is generated by translating the codes with text templates, including any ad-hoc comments made by the pathologist. The reports conform to the 2001 Bethesda system of nomenclature that is used for Pap test reporting across the USA.<sup>46</sup> The identified word patterns were essentially mapped to concepts that are conveyed in the Pap report (figure 2). The mappings were represented as a set of if-then rules. For example, the text 'consistent with invasive squamous cell carcinoma' maps to the

**Figure 1** Overview of the clinical decision support system (CDSS). There are three modules: guideline engine, data module and free-text processor. The data module seeks patient information from the Mayo electronic medical record (EMR). It holds the information in a form that is amenable to the guideline engine and depends on the free-text processor to interpret free-text Papanicolaou (Pap) reports. The guideline engine and free-text processor are essentially rule-based.





**Figure 2** Schematic diagram shows some of the concepts in Papanicolaou (Pap) report required for applying the national guidelines for cervical cancer screening. ASC-H, high grade squamous intraepithelial lesion; ASC-US, atypical squamous cells of undetermined significance; (ASKS), ; DNOEC, inadequate transformation zone component; GECA, glandular epithelial cell abnormality; HPV, human papillomavirus; HSIL, intraepithelial lesions categorized as high grade; LSIL, intraepithelial lesions categorized as low grade; NHPV, negative HPV test; PHPV, positive HPV test; SCC, squamous cell carcinoma; UNSAR, unsatisfactory for evaluation; NIL, negative for intraepithelial lesion or malignancy.

concept that the patient has squamous cell carcinoma. This set of rules constituted the lexicon.

Next, the set of parser rules required to interpret the identified concepts was constructed by interviewing the physician. This is a set of if-then rules that encode the logic used by the physician to interpret the concepts expressed in the report. For example, ‘if the report contains the finding of squamous cell carcinoma or a glandular epithelial cell abnormality then the cytology type is abnormal’. The rulebase also contained implicit knowledge used by the physician. For example, ‘if the report does not mention that the endocervical transformation zone is inadequate then it is assumed to be adequate’. The parser also included error checking rules to ensure the logical consistency of the reports. For example, a report cannot state the patient has no epithelial lesion and has squamous cell carcinoma as these two are mutually exclusive events.

To ensure the coverage and accuracy of the free-text rulebase on the corpus of Pap reports used for development, it was applied to the corpus. The reports were grouped into different classes based on the parameter values output by the rulebase (table 1). From each of the classes a fixed number of reports was randomly selected for manual verification and error analysis.

**Guideline rulebase**

The national screening and management guidelines are expressed in the form of natural language and are themselves not computer interpretable. A flowchart was developed to represent the knowledge contained in the guidelines (figure 3). For simplicity, we did not denote offshoots that resulted in no recommendations. For instance, for female patients who have undergone a hysterectomy, the recommendation for Pap test is not generated.

**CDSS evaluation**

For a comprehensive evaluation of the CDSS, we constructed a test set of 74 cases, such that there was at least one instance

**Table 1** Distribution of Pap reports across parameters relevant to the guideline rulebase

Count	%	Cytology	Endocervical zone component	HPV test	
9	0	Abnormal (other than ASC-US)	Inadequate	Negative	
13	0			Positive	
64	0			Not performed	
166	0			Adequate	Negative
168	0				Positive
937	2	Not performed			
1357	3				
12	0	ASC-US	Inadequate	Not performed	
29	0			Positive	
68	0			Negative	
142	0			Adequate	Not performed
453	1	Positive			
966	2	Negative			
1670	4				
57	0	Negative	Inadequate	Positive	
340	1			Adequate	
2325	6			Inadequate	
2953	7			Inadequate	
13212	31			Adequate	
19098	45			Adequate	
37985	91				
383	1	Unsatisfactory for evaluation	-	Negative	
510	1			Not performed	
5	0			Positive	
89	2				
41910	100				

ASC-US, atypical squamous cells of undetermined significance; HPV, human papillomavirus; Pap, Papanicolaou.

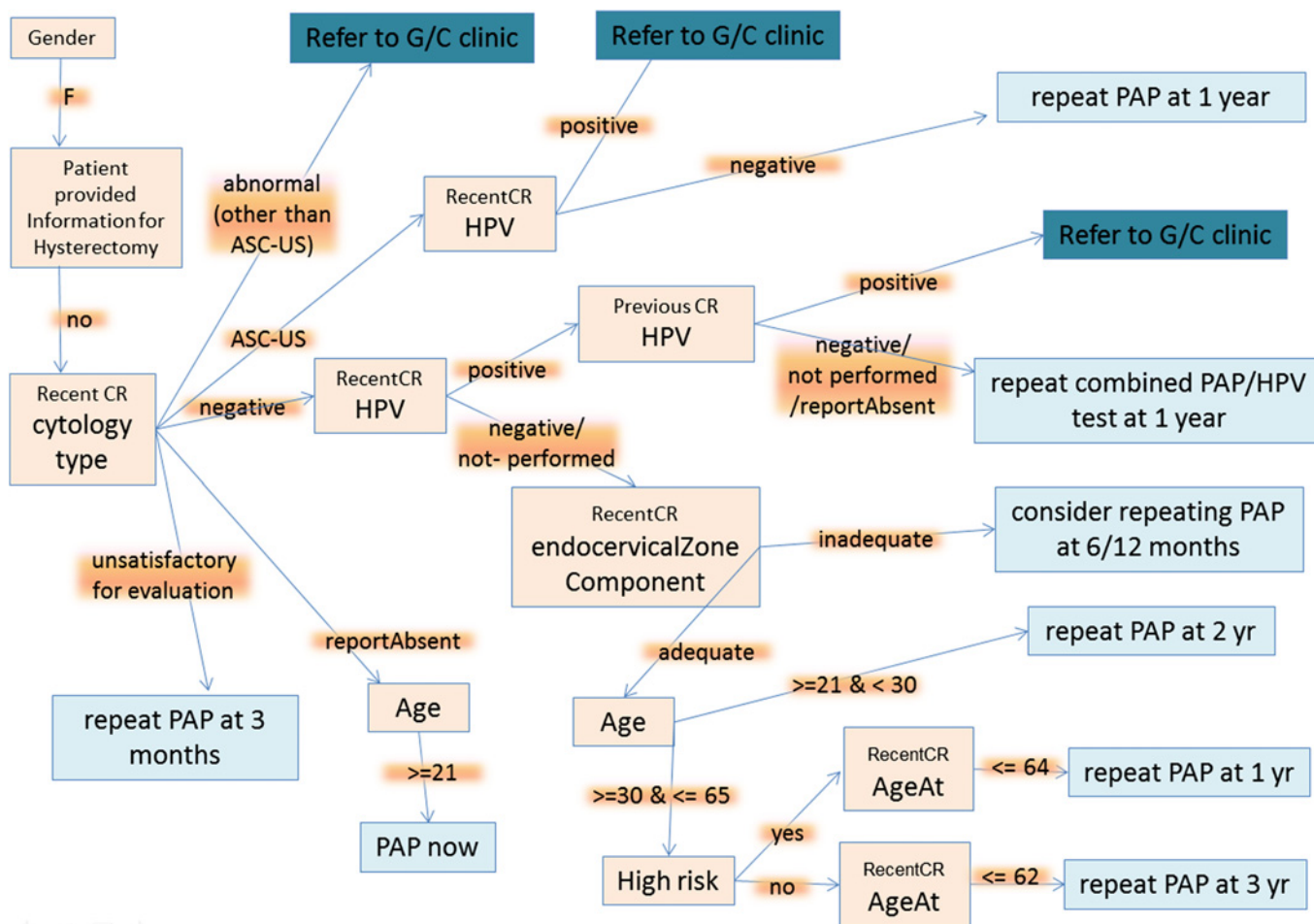
and at most 10 instances for each possible decision path in the guideline flowchart. The test set was constructed by searching patient records in the EMR that matched a particular decision path. For each test case, the task of the CDSS was to generate the optimal recommendation for the indicated decision date. The recommendation was required to be date specific, because the recommendation for a particular patient would change over time reflecting the increase in the patient’s age and changes to the EMR data.

The physician who was interviewed for construction of the rulebase also participated in the evaluation. The recommendation generated by the CDSS was then compared with that made by the physician, who was initially blinded to the recommendation output by the CDSS. When there was a mismatch in the recommendations, the cases were reviewed by the physician and an error analysis was performed.

**RESULTS AND DISCUSSION**

**Free-text rulebase**

The nomenclature of cervical cytology reports has been widely standardized in the USA with the Bethesda system.<sup>46</sup> Slides are classified for specimen adequacy as satisfactory or unsatisfactory for evaluation. When there is no cytological evidence of neoplasia, the slides are designated as ‘negative for intraepithelial lesion or malignancy’. Epithelial cell abnormality is diagnosed when the degree of nuclear atypia is not sufficient to warrant a diagnosis of squamous intraepithelial lesion. Epithelial cell abnormalities fall into a broad category of squamous or glandular cell abnormality. Atypical squamous cells are further



**Figure 3** Flowchart abstraction for the guideline rulebase. ASC-US, atypical squamous cells of undetermined significance; G/C, gynecology clinic; CR, cervical cytology (Pap) report; HPV, human papillomavirus; Pap, Papanicolaou.

categorized as being ‘of undetermined significance’ (ASC-US) or ‘cannot exclude high grade squamous intraepithelial lesion’ (ASC-H). Squamous intraepithelial lesions are categorized as low-grade (LSIL), high-grade (HSIL) or squamous cell carcinoma (SCC). Atypical glandular cells if present are also reported as glandular epithelial cell abnormality (GECA). Absence of endocervical cells is reported as ‘inadequate transformation zone component’ (DNOEC).

We restricted the scope of the free-text rulebase to only the rules that were required to decide the values of parameters that are necessary for applying the guideline logic described in the previous step. There were four such parameters (figure 2). First, the screening variable indicated that the report was the result of a screening evaluation versus a specific diagnostic test. The referenced guidelines are applicable only for screening reports. Second, the cytology is classified as negative (normal reports), ASC-US, abnormal (other than ASC-US) and unsatisfactory for evaluation. Third, the HPV test could be negative, positive or not performed. Finally, the endocervical transformation zone was adequate or inadequate. The evaluation of the four parameters involved recognition of 11 concepts in the Pap report (figure 2).

To determine coverage and specificity of the identified word patterns on the development corpus, we applied the free-text rulebase to all Pap reports in the corpus. We used 49 293 reports for analysis. Of these, 6988 were either diagnostic or were non-cervical samples. A total of 41 910 (99.1%) reports was classified

by the system into 21 categories corresponding to the combinations of parameter values extracted by the free-text rulebase. The distribution of reports is summarized in table 1. Ninety-one per cent of the reports showed normal (negative for intraepithelial lesion) cytology, 2% of the reports indicated that the specimen was unsatisfactory for evaluation of the cytology, and the remaining 7% reports indicated an abnormal cytology. We manually verified 10 randomly selected reports from each category. A total of 210 (=10×21) reports was selected for the manual verification. The rulebase was found to determine correctly the parameter values for all the examined reports.

The free-text rulebase reported errors for 395 (0.9%) of the reports. Manual examination revealed that these were due to missing data or due to the fusion of multiple reports when data from the EMR were dumped into the research database. Invalid data are not expected in the production EMR, which was verified with spot-checking, and these errors were not considered further. However, in one case, the error was due to mutually exclusive diagnoses in the same report (diagnostic/reporting error). This was considered a rare event and was also not expected to have any significant effect on the CDSS performance.

These results suggest that the free-text rulebase covers all the word patterns in the development corpus and that the patterns are concept specific. These findings are the result of the use of an in-house text template for generating the reports. The identified word patterns closely correspond to the in-house template, barring some spelling variations. As the corpus used for the

**Table 2** Distribution of data points in the guideline rulebase/flowchart across EMR sections

EMR sections	Type of information	No of flowchart nodes
Registration	Sex, age	4
Patient-provided information	History of hysterectomy	1
Patient documents	Pap reports	6
Problem list	Risk of cervical cancer	1
Laboratory reports	HPV test result	3

EMR, electronic medical record; HPV, human papillomavirus; Pap, Papanicolaou.

development represents a very large sample of different Pap reports, the free-text rulebase is expected to perform accurately on nearly all Pap reports in the EMR system. We hope that our study will guide decision-makers to make pathologists' annotations directly available in the EMR. This will facilitate the development of decision support tools by obviating the need for text processing to interpret the free-text reports.

Our choice of a rule-based approach for the text processor was due to: (1) the availability of the in-house template used for generating the reports; (2) the need for representing the physician's logic for report interpretation; (3) the need for providing decision explanations to the physicians that may not have been possible with other approaches; (4) the requirement of high accuracy for the decision support application; and (5) the provision of additional checks (rules) to ensure that the report had logically consistent findings and was not corrupted.

**Guideline rulebase**

Figure 3 shows the flowchart abstraction of the guideline rulebase. The flowchart consisted of 22 nodes (11 leaf nodes) and 20 edges and spanned five different parts of the EMR (table 2). A brief overview is as follows. The flow chart begins with a check in the registration section whether the patient is female. Next, the patient-provided information section is accessed to find

whether the response to the question—'Have your menstrual periods changed in anyway or become abnormal to you?' matches the option 'No, I have had a hysterectomy'. This is to ensure that female patients that have undergone a hysterectomy are not advised to have a routine screening Pap test. The patient-provided information section consists of the patient's response to annually administered questionnaires. Next, for patients with no history of hysterectomy, the list of patient documents is then searched to identify 'Pap reports' and the latest report is analyzed to determine the cytology type. In case the cytology is of type ASC-US or normal, the 'HPV test' and 'endocervical transformation zone component' parameters deduced from the report are used.

If the HPV test is positive, the previous Pap report is analyzed to check for the HPV test. The majority (76%) of patients have a normal cytology, negative or no report of HPV and adequate transformation zone. For these patients, the recommendation depends on the age (accessed from the registration section) and if they are in the 30–65 years age group, whether any of the high-risk conditions (see supplementary appendix, available online only) for cervical cancer appears in their active problem list.

**CDSS evaluation**

The distribution of cases across possible decision paths is shown in table 3. The recommendations made by the CDSS matched those made by the physician for 66 of the 74 cases. The physician was then unblinded to the CDSS recommendations and requested to review the eight 'mismatch' cases. The physician verified that in seven of the eight cases the CDSS recommendation was optimal (table 4) and the CDSS had erred in one case.

In the first case, the physician missed that the age of the patient was above 65 years and recommended a follow-up Pap test. In the second and third cases, the physician failed to check the previous report and advised a repeat HPV/Pap co-test after

**Table 3** Distribution of cases across the decision table

History of hysterectomy	Last cytology	Last HPV	Last EZC	Age, years	High risk	Previous HPV	Recommendation	No of cases
No	Negative	Negative	Inadequate				Consider repeating Pap at 6/12 months	5
No	Negative	Positive				Report absent	Repeat combined Pap—HPV test at 12 months	4
No	Report absent			≥21			Pap now	3
Yes							No Pap needed	1
No	ASC-US	Negative					Repeat Pap at 12 months	9
No	Negative	Negative	Adequate	≥30 and ≤62	No		Repeat Pap at 3 years	9
No	Negative	Positive				Not performed	Repeat combined Pap—HPV test at 12 months	9
No	Unsatisfactory for evaluation						Repeat Pap at 3 months	3
No	Negative	Negative	Adequate	>64			No Pap needed	4
No	Negative	Negative	Adequate	≥30 and ≤64	Yes		Repeat Pap at 1 year	1
No	Negative	Negative	Adequate	≥21 and <30			Repeat Pap at 2 years	1
No	Negative	Not performed	Adequate	≥21 and <30			Repeat Pap at 2 years	2
No	Negative	Positive				Negative	Repeat combined Pap—HPV test at 12 months	5
No	Negative	Positive				Positive	Refer to G/N	4
No	Abnormal (other than ASC-US)						Refer to G/N	6
No	Report absent				<21		No Pap needed	3
No	ASC-US	Positive					Refer to G/N	5

Each row indicates the set of variables that correspond to a decision path in the flowchart.

Some cells in the table are blank, which indicates that those variables were not required to be evaluated for the particular decision path.

ASC-US, atypical squamous cells of undetermined significance; EZC, endocervical zone component; G/N, gynecology; HPV, human papillomavirus; Pap, Papanicolaou.

**Table 4** Reasons for suboptimal recommendations by the physician

Case	Reason
1	Missed age
2	Missed previous report
3	Missed previous report
4	Missed history of hysterectomy
5	Read report from different date
6	Read report from different date

1 year, while the CDSS correctly noted that both the current and past reports were positive for HPV and referred the patient to the gynecology clinic. In both of these cases the CDSS did well to pick up the need for more evaluation, which is the goal of this computer-aided intervention. The fourth case previously had a hysterectomy, as indicated in a response to the patient questionnaire, which the physician missed. The physician had inappropriately recommended a Pap. In the fifth and sixth cases, the physician had looked up the Pap test result for a different date than was required for the evaluation, and made a suboptimal recommendation. The physician had reflexively looked up the latest report in the EMR, instead of considering the reports dated before the particular decision date as required for the evaluation.

In the seventh case, the physician had recommended follow-up Pap at 3 years, while the CDSS recommended Pap in 1 year, as it had identified that the patient had history of cervical dysplasia and was at high risk of cervical cancer. The physician had closely examined the dates of recent Pap tests and noticed that subsequent to the recording of the risk factor information, the patient was evaluated by a gynecologist and was prescribed screenings at intervals that suggested that the patient was returned to routine screening. In such cases, when the gynecologist returns patients with risk factors to routine screening, it is desirable that the CDSS alerts the care providers to the presence of risk factors and provides an opportunity to reconsider the decision of returning the patient to routine screening. Therefore, for this case the CDSS recommendation was considered to be optimal by the physician.

Finally, in the eighth case the CDSS incorrectly referred the patient to the gynecology clinic when the optimal recommendation that was made by the physician was to repeat Pap/HPV co-test in 1 year. The reason for the failure was that HPV testing is sometimes performed separate from a Pap test and these results are reported in the laboratory system, which was not queried by the CDSS. Although the scenario for the failed case is expected to occur in only a small percentage of patients who have abnormal HPV results, it could lead to overreferral of patients to the gynecology clinic. To address this issue, the CDSS was improved to include the HPV results missed earlier. On including the HPV results from the laboratory, the CDSS generated the optimal recommendation for all test patients.

The physician reviewing the cases for evaluation had led the development of the guideline flowchart, as she was experienced in women's healthcare issues and was very familiar with the guidelines. Despite this, the physician missed the optimal recommendations for six out of 74 cases. Other healthcare providers are expected to be generally less familiar with the required guidelines and would find the CDSS a valuable resource.

For the construction of the evaluation set, we had restricted the number of cases for particular paths in the flowchart. Therefore, the distribution of the test cases differed from the

distribution that would be encountered by the CDSS on deployment. However, our method allowed us to evaluate nearly all possible case scenarios that would be encountered by the system and ensured the validity of possible paths in the flowchart. For instance, as seen from table 1, 76% of Pap reports have a normal cytology, adequate endocervical zone component and are not positive for HPV. The follow-up recommendations for abnormal Pap results are especially critical to ensure that the patients receive appropriate work-up and referral to prevent cancer.<sup>9 47</sup> Therefore, the restriction on the distribution of test cases facilitated a judicious use of the physician's review effort, a focus for the higher impact recommendations and a near complete coverage of possible case scenarios.

Results suggest that the guideline rulebase contains the logic required to generate the optimal recommendation for all patients. This was because the guidelines were comprehensive and detailed, and that allowed us to make an explicit flowchart representation required to construct the guideline rulebase.

A limitation of the proposed approach is that the developed CDSS depends on the Pap report and may not be readily portable to other institutions that have different word patterns in the Pap reports. Also, the CDSS depends on the availability of other data elements in the EMR like a well-defined problem list and patient-provided information, which may not be present at other institutions. Nevertheless, we expect that the proposed approach may be applied to construct a system tailored to the individual hospital.

Another limitation of our study is that only one physician with an expertise in the cervical cancer screening/management guidelines participated in this study. To validate the system further, it is necessary to include others physicians with an expertise in this domain, to review the guideline flowchart and to evaluate the CDSS.

Nonetheless, the results indicate that the developed free-text processor for the Pap smear report was accurate, as a result of the standardization of reporting the Pap test. Evaluation revealed that the CDSS made the optimal screening recommendations for 73 of 74 patients and it identified two cases for gynecology referral that were missed by the physician. The CDSS aided the physician to modify recommendations in six cases. The failure case was because HPV testing was sometimes performed separate from the Pap test and these results were reported in the laboratory system that was not queried by the CDSS. Subsequently, the CDSS system was corrected to include the HPV results missed earlier, and it generated the optimal recommendation for all patients. Given the high accuracy of the system, the authors consider it a suitable candidate for deployment in clinical practice.

## CONCLUSION

This study outlines the development and validation of a CDSS that performs automated text processing for cervical cancer screening. The results of the evaluation indicated that the developed CDSS performed accurately, given the standardized reporting of Pap tests and the availability of explicit guidelines. Overall, our study demonstrates that free text in the EMR can be effectively utilized with NLP methods to develop useful clinical decision support tools.

## Future directions

After validation with other physicians, the developed CDSS would be deployed at the outpatient departments at the Mayo Clinic, Rochester, and user feedback about CDSS performance would be collected. The CDSS would be integrated with an

interface in the EMR, which lists all preventive care reminders for the healthcare providers. An impact analysis would be carried out by comparing the screening and referral rates before and after deployment of the system. In the near future, our approach would be replicated for other free text-based decision problems such as colon cancer screening, sleep disorder management and asthma management.

**Acknowledgments** The authors are grateful to Chad Richter for extracting the corpus of Pap reports used in this study.

**Contributors** KBW developed the system, led the study design and analysis and drafted the manuscript. KM provided the expertise for the guideline flowchart and carried out the case reviews for the evaluations. RC and HL conceived the study. MH and RH contributed to the design of the system. KM, HL, RAG and RC participated in the study design, analysis and manuscript drafting. All authors read and approved the final manuscript.

**Competing interests** None.

**Patient consent** Obtained.

**Ethics approval** Ethics approval was granted by the Mayo Clinic institutional review board.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

1. Yabroff KR, Saraiya M, Meissner HI, et al. Specialty differences in primary care physician reports of papanicolaou test screening practices: a national survey, 2006 to 2007. *Ann Intern Med* 2009;**151**:602–11.
2. Saraiya M, Berkowitz Z, Yabroff KR, et al. Cervical cancer screening with both human papillomavirus and papanicolaou testing vs papanicolaou testing alone: what screening intervals are physicians recommending? *Arch Intern Med* 2010;**170**:977–85.
3. Lee JW, Berkowitz Z, Saraiya M. Low-risk human papillomavirus testing and other non recommended human papillomavirus testing practices among U.S. health care providers. *Obstet Gynecol* 2011;**118**:4–13.
4. Roland KB, Soman A, Benard VB, et al. Human papillomavirus and papanicolaou tests screening interval recommendations in the United States. *Am J Obstet Gynecol* 2011;**205**:447.e1–8.
5. Hoeksema LJ, Bazyz-Asaad A, Lomotan EA, et al. Accuracy of a computerized clinical decision-support system for asthma assessment and management. *J Am Med Inform Assoc* 2011;**18**:243–50.
6. Sittig DF, Wright A, Osheroff JA, et al. Grand challenges in clinical decision support. *J Biomed Inform* 2008;**41**:387–92.
7. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin* 2011;**61**:69–90.
8. Janerich DT, Hadjimichael O, Schwartz PE, et al. The screening histories of women with invasive cervical cancer, Connecticut. *Am J Public Health* 1995;**85**:791–4.
9. Leyden WA, Manos MM, Geiger AM, et al. Cervical cancer in women with comprehensive health care access: attributable factors in the screening process. *J Natl Cancer Inst* 2005;**97**:675–83.
10. Nelson W, Moser RP, Gaffey A, et al. Adherence to cervical cancer screening guidelines for U.S. women aged 25–64: data from the 2005 Health Information National Trends Survey (HINTS). *J Womens Health (Larchmt)* 2009;**18**:1759–68.
11. Wheeler CM. Natural history of human papillomavirus infections, cytologic and histologic abnormalities, and cancer. *Obstet Gynecol Clin North Am* 2008;**35**:519–36; vii.
12. Saslow D, Runowicz CD, Solomon D, et al. American Cancer Society guideline for the early detection of cervical neoplasia and cancer. *CA Cancer J Clin* 2002;**52**:342–62.
13. U.S. Preventive Services Task Force. Screening for cervical cancer: recommendations and rationale. *Am J Nurs* 2003;**103**:101–2; 105–6; 108–9.
14. ACOG Committee on Practice Bulletins — Gynecology. ACOG practice bulletin no. 109: cervical cytology screening. *Obstetrics Gynecol* 2009;**114**:1409–20.
15. U.S. Preventive Services Task Force. Screening for Cervical Cancer. 2003. <http://www.uspreventiveservicestaskforce.org/uspstf/uspsscerv.htm> (accessed 15 Sep 2011).
16. Wright TC Jr, Massad LS, Dunton CJ, et al. 2006 Consensus guidelines for the management of women with abnormal cervical cancer screening tests. *Am J Obstet Gynecol* 2007;**197**:346–55.
17. Dupuis EA, White HF, Newman D, et al. Tracking abnormal cervical cancer screening: evaluation of an EMR-based intervention. *J Gen Intern Med* 2010;**25**:575–80.
18. Greenes RA, ed. *Clinical Decision Support: The Road Ahead*, 1st edn. New York: Academic Press, 2006.
19. Elkin PL, Liebow M, Bauer BA, et al. The introduction of a diagnostic decision support system (DXplain) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging Diagnostic Related Groups (DRGs). *Int J Med Inform* 2010;**79**:772–7.
20. Aronsky D, Fisman M, Chapman WW, et al. Combining decision support methodologies to diagnose pneumonia. *Proc AMIA Symp* 2001:12–16.
21. Chaudhry R, Tulledge-Scheitel SM, Parks DA, et al. Use of a web-based clinical decision support system to improve abdominal aortic aneurysm screening in a primary care practice. *J Eval Clin Pract*. Published Online First: 15 March 2011. doi:10.1111/j.1365-2753.2011.01661.x
22. Campion TR Jr, Waitman LR, May AK, et al. Social, organizational, and contextual characteristics of clinical decision support systems for intensive insulin therapy: a literature review and case study. *Int J Med Inform* 2010;**79**:31–43.
23. Griffey RT, Lo HG, Burdick E, et al. Guided medication dosing for elderly emergency patients using real-time, computerized decision support. *J Am Med Inform Assoc* 2012;**19**:86–93.
24. Lau F, Kuziemsky C, Price M, et al. A review on systematic reviews of health information system studies. *J Am Med Inform Assoc* 2010;**17**:637–45.
25. Waghlikar KB, Sundararajan V, Deshpande AW. Modeling paradigms for medical diagnostic decision support: a survey and future directions. *J Med Syst*. Published Online First: 1 October 2011. doi:10.1007/s10916-011-9780-4
26. Meystre SM, Savova GK, Kipper-Schuler KC, et al. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearbook Med Inform* 2008;**47**:128–44.
27. Friedman C. A broad-coverage natural language processing system. *Proc AMIA Symp* 2000:270–4.
28. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001:17–21.
29. 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* 2010;**17**:507–13.
30. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc* 2011;**18**:544–51.
31. Denny JC, Choma NN, Peterson JF, et al. Natural language processing improves identification of colorectal cancer testing in the electronic medical record. *Med Decis Making* 2012;**32**:188–97.
32. Himes BE, Kohane IS, Ramoni MF, et al. Characterization of patients who suffer asthma exacerbations using data extracted from electronic medical records. *AMIA Annu Symp* 2008:308–12.
33. Effler P, Ching-Lee M, Bogard A, et al. Statewide system of electronic notifiable disease reporting from clinical laboratories: comparing automated reporting with conventional methods. *JAMA* 1999;**282**:1845–50.
34. Lazarus R, Klompas M, Campion FX, et al. Electronic support for public health: validated case finding and reporting for notifiable diseases using electronic medical data. *J Am Med Inform Assoc* 2009;**16**:18–24.
35. Gerbier S, Yarovaya O, Gicquel Q, et al. Evaluation of natural language processing from emergency department computerized medical records for intra-hospital syndromic surveillance. *BMC Med Inform Decis Mak* 2011;**11**:50.
36. Liao KP, Cai T, Gainer V, et al. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010;**62**:1120–7.
37. Mishra NK, Cummo DM, Arzen JJ, et al. A rule-based approach for identifying obesity and its comorbidities in medical discharge summaries. *J Am Med Inform Assoc* 2009;**16**:576–9.
38. Boytcheva S, Tcharaktchiev D, Angelova G. Contextualization in automatic extraction of drugs from hospital patient records. *Stud Health Technol Inform* 2011;**169**:527–31.
39. Li Z, Liu F, Antieau L, et al. Lancet: a high precision medication event extraction system for clinical text. *J Am Med Inform Assoc* 2010;**17**:563–7.
40. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc* 2005;**12**:448–57.
41. Bates DW, Evans RS, Murff H, et al. Detecting adverse events using information technology. *J Am Med Inform Assoc* 2003;**10**:115–28.
42. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011;**306**:848–55.
43. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform* 2009;**42**:760–72.
44. Jha AK. The promise of electronic records: around the corner or down the road? *JAMA* 2011;**306**:880–1.
45. JBoss Community. *Drools*. 2012. <http://www.jboss.org/drools/> (accessed 15 Sep 2011).
46. Solomon D, Davey D, Kurman R, et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* 2002;**287**:2114–19.
47. Zapka J, Taplin SH, Price RA, et al. Factors in quality care — the case of follow-up to abnormal cancer screening tests—problems in the steps and interfaces of care. *J Natl Cancer Inst Monogr* 2010;**2010**:58–71.