OPEN
ACCESS

Improving the Accuracy of a Clinical Decision Support System for Cervical Cancer Screening and Surveillance

K.E. Ravikumar[1]   Kathy L. MacLaughlin[2]   Marianne R. Scheitel[3]   Maya Kessler[4]
Kavishwar B. Wagholikar[5]   Hongfang Liu[1]   Rajeev Chaudhry[4]

[1] Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, United States
[2] Department of Family Medicine, Mayo Clinic, Rochester, Minnesota, United States
[3] Knowledge and Delivery Center, Mayo Clinic, Rochester, Minnesota, United States
[4] Division of Primary Care Internal Medicine, Department of Medicine, Mayo Clinic, Rochester, Minnesota, United States
[5] Department of Medicine, Harvard Medical School, Boston, United States

Address for correspondence K.E. Ravikumar, PhD, Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, United States
(e-mail: KomandurElayavilli.Ravikumar@mayo.edu).

Abstract

Background   Clinical decision support systems (CDSS) for cervical cancer prevention are generally limited to identifying patients who are overdue for their next routine/next screening, and they do not provide recommendations for follow-up of abnormal results. We previously developed a CDSS to automatically provide follow-up recommendations based on the American Society of Colposcopy and Cervical Pathology (ASCCP) guidelines for women with both previously normal and abnormal test results leveraging information available in the electronic medical record (EMR).

Objective   Enhance the CDSS by improving its accuracy and incorporating changes to reflect the latest revision of the guidelines.

Methods   After making enhancements to the CDSS, we evaluated the performance of the clinical recommendations on 393 patients selected through stratified sampling from a set of 3,704 patients in a nonclinical setting. We performed chart review of individual patient's record to evaluate the performance of the system. An expert clinician assisted by a resident manually reviewed the recommendation made by the system and verified whether the recommendations were as per the ASCCP guidelines.

Results   The recommendation accuracy of the enhanced CDSS improved to 93%, which is a substantial improvement over the 84% reported previously. A detailed analysis of errors is presented in this article. We fixed the errors identified in this evaluation that were amenable to correction to further improve the accuracy of the system. The source code of the updated CDSS is available at https://github.com/ohnlp/MayoNlpPapCdss.

Conclusion   We made substantial enhancements to our earlier prototype CDSS with the updated ASCCP guidelines and performed a thorough evaluation in a nonclinical setting to improve the accuracy of the CDSS. The CDSS will be further refined as it is utilized in the practice.

Background and Significance

In the United States, approximately 12,000 women are diagnosed with cervical cancer each year and 4,000 women die each year from cervical cancer.[1] Ideally, cervical cancer prevention is achieved with regular patient screening and subsequent appropriate management of precancerous findings. Women at greatest risk for cervical cancer are those who have never been screened, have not had regular screening, or have not had guideline-based follow-up of abnormal results.[2–4] Over half of all new cases of cervical cancer occur in these women.[5–7]

One determinant that may contribute to lack of guideline-based follow-up is the complexity of the guidelines, especially for women with a history of abnormal Pap or human papillomavirus (HPV) results on prior screening.[8,9] The American Society of Colposcopy and Cervical Pathology (ASCCP) guidelines involve multiple clinical decision pathways and some recommendations take into consideration the results of abnormal testing from the past 25 years. Primary care clinicians see a high volume of patients daily and have limited time with each patient, which makes compliance with these guidelines challenging.

Clinical decision support systems (CDSS) have the potential to improve patient care, especially for women with prior abnormal results.[8,9] Currently, the majority of CDSS interventions for cervical cancer prevention[10,11] are limited to identifying patients who are overdue for their next routine/normal screening,[9,12] and do not provide surveillance recommendations for follow-up of abnormal results, nor do those systems identify women who need more frequent screening based on other medical conditions.

In earlier works,[13–16] Wagholikar et al described the implementation of a CDSS for Cervical Cancer Screening and Surveillance (CCSS). The system utilized natural language processing (NLP) to extract variables about cervical cancer screening from cytology reports generated after Pap smear screening and pathology reports from colposcopy biopsies, which are predominantly available in text format. Structured data resources, such as coded problem lists (PL) and patient-provided information (PPI), were also used to determine appropriate care recommendations. The system showed potential to reduce the amount of time clinicians needed to determine appropriate follow-up care. However, the system achieved a precision of only 84% for all the patients analyzed. We further observed only 78% precision in patients with an abnormal past history. This showed that there is a substantial room for improvement. Moreover, since the time of the prototype's development, the ASCCP guidelines have been updated requiring changes to the original clinical pathways tested.

Objective

In this study we describe the implementation of the recent updates made in ASCCP guidelines to their prototype and improvements in the precision of care recommendation made by CDSS. We had three main objectives: (1) update the clinical pathways to reflect the latest ASCCP guidelines,[17,18] (2) fix the errors in the previous version of CDSS, (3) improve the accuracy of care recommendation generated by the CDSS in both the abnormal and normal patient population.

Methods

CDSS Workflow Architecture

►Fig. 1 captures the details of the updated CDSS workflow. The CDSS operates in three steps: (1) extraction of primary and derived data elements from the clinical records, (2) computation of the ASCCP guideline-based recommendation, and (3) delivery of the recommendations to the point of care.

Extraction of Data Elements from Clinical Reports

The CDSS extracts the data elements based on cytology, HPV, histology, and colposcopy tests. ►Table 1 lists all the data elements extracted by the system. The details of NLP implementation are elaborated elsewhere.[13] The NLP algorithm in the CDSS is a simple rule-based approach where regular expressions are used to extract data elements. Cytology reports indicate epithelial cell abnormalities and the risk factors that predispose a patient toward cervical cancer. Epithelial cell abnormalities fall into a broad category of squamous or glandular cell abnormality. Atypical squamous cells are further 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/high-grade intraepithelial lesion ((LSIL/HSIL) or squamous cell carcinoma (SCC). An atypical glandular cell if present is also reported as glandular epithelial cell abnormality (GECA). Data elements, such as ASC-US, ASC-H, LSIL, HSIL, SCC, and GECA, were extracted from cytology reports using simple regular expressions. From the human papilloma virus (HPV) test reports, the CDSS extracts whether the test outcome is negative or positive based on the descriptions outlined in the report (see example shown in ►Fig. 2). From the biopsy/pathology reports, the CDSS extracts the CIN histology status, namely the CIN2, CIN3 (see example shown in ►Fig. 2), which are high-risk factors that require immediate intervention to prevent the patient to progress to a precancerous or cancerous stage.[19] The CDSS also independently extracts variables about the patient, such as hysterectomy, risk factors, such as cancer, immunodeficiency, and HIV from structured data sources, such as coded problem lists, disposition, and PPI.

Reassembling Patient Variables Longitudinally Over Time

Next, the CDSS further extracts the temporal features about CIN2 and CIN3, and hysterectomy in the past 25 years that play a critical role in making the right care recommendations. In this step, the data elements are arranged longitudinally over a period of time. ►Fig. 2 illustrates the process of temporal assembly of the data elements.

Generation of Care Recommendations

Based on the values of data elements extracted in the first two steps, decision support logic (a set of if-then rules) automates the decision rules in the ASCCP guidelines. ►Fig. 3 illustrates a simple clinical pathway based on the variables extracted by the
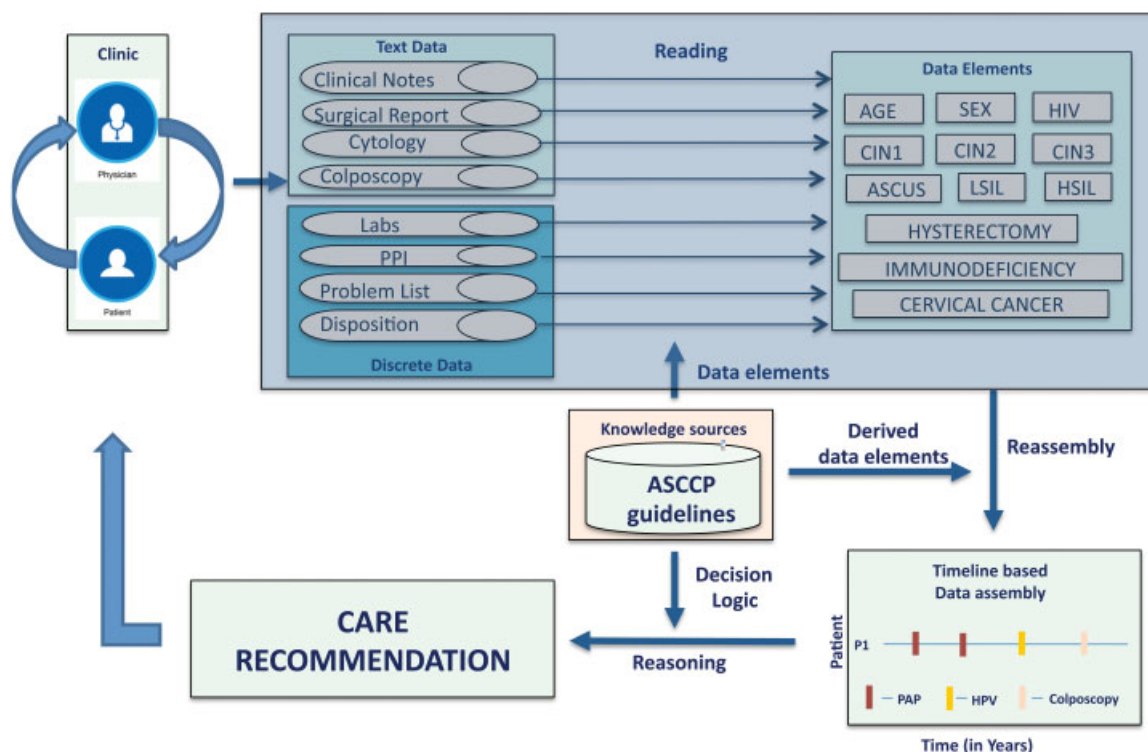
Fig. 1 Cervical cancer screening and surveillance CDSS workflow. The workflow starts with a patient's clinic visit, which results in generation of different data sources. The system automatically reads from multiple sources of information and extracts primary data elements from individual documents. The data elements/variables are then reassembled across different time points. The decision logic rules apply on the temporal data elements to compute the care recommendation. The care recommendations are delivered to the point of care.

Table 1 Data elements required for cervical cancer screening and surveillance

Report type	Primary data elements	Derived data elements
Cytology report	LSIL, HSIL, ASCUS	Recent Pap, previous Pap, previous to previous Pap, any previous three cytologies either HSIL, ASCH, or AGC
HPV test	Positive, negative	Recent HPV, prior HPV
Pathology/histology	CIN2, CIN3	History of CIN2/CIN3, history of colposcopy
Surgery	Hysterectomy	History of hysterectomy
Demographics	Age, sex	Age at recent Pap, age at recent HPV
Problem list	Immunodeficiency, HIV, transplant, in utero DES exposure, cervical cancer, AIS	

Abbreviations: AGC, atypical glandular cell; AIS, adenocarcinoma in-situ; ASCUS, atypical squamous cells of undetermined significance; CIN, cervical intraepithelial neoplasia; DES, di-ethylstilbesterol; HPV, human papillomavirus; LSIL, low-grade intraepithelial lesion.

system in the earlier steps. The rules were implemented using Drools rule engine, a widely used open source framework.

CDSS Revisions
To reflect guideline changes in our CDSS and fix errors noted in the previous prototype, we revised our clinical pathways. The figure in the supplementary material (►Supplementary Fig. S1, available in the online version) illustrates the revised pathways. The updated CDSS has 54 clinical pathways with 13 pathways for routine screening (normal) and 41 pathways for high-risk patients (abnormal). Patients are considered to be on "high-risk"/abnormal pathways if they have a history of

abnormal Pap smear, positive HPV, colposcopy reports with CIN II-CINIII in the past 25 years, history of cervical cancer, DES in-utero exposure, or immunosuppression from HIV infection, history of solid organ transplant or on chronic immunosuppressant medications. See ►Supplementary Fig. S1 (available in the online version) for identification of decision end points that are considered normal versus abnormal. The final output of the CDSS for each patient is a recommendation for follow-up Pap smear date when appropriate, colposcopy referral for concerning cytology findings, or flagging the patient as no longer requiring screening due to age or hysterectomy. We elaborate on the changes made to the CDSS below.
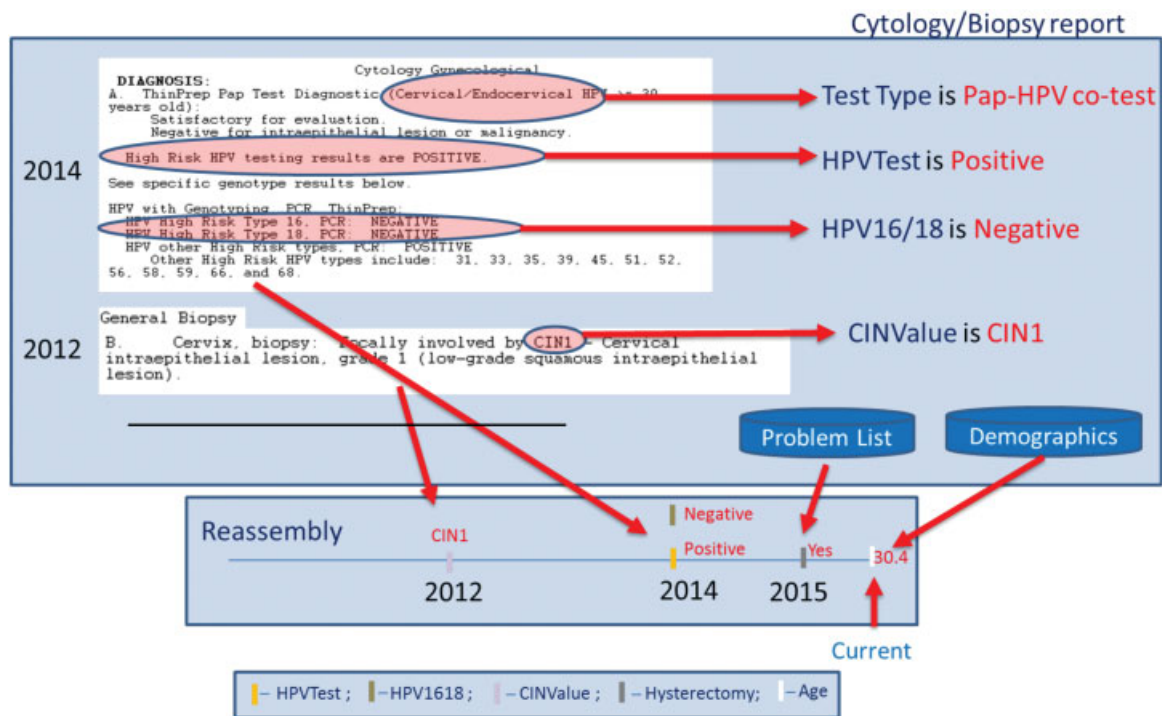
Fig. 2 Temporal assembly of primary and secondary data elements related to cervical cancer screening for a given patient.
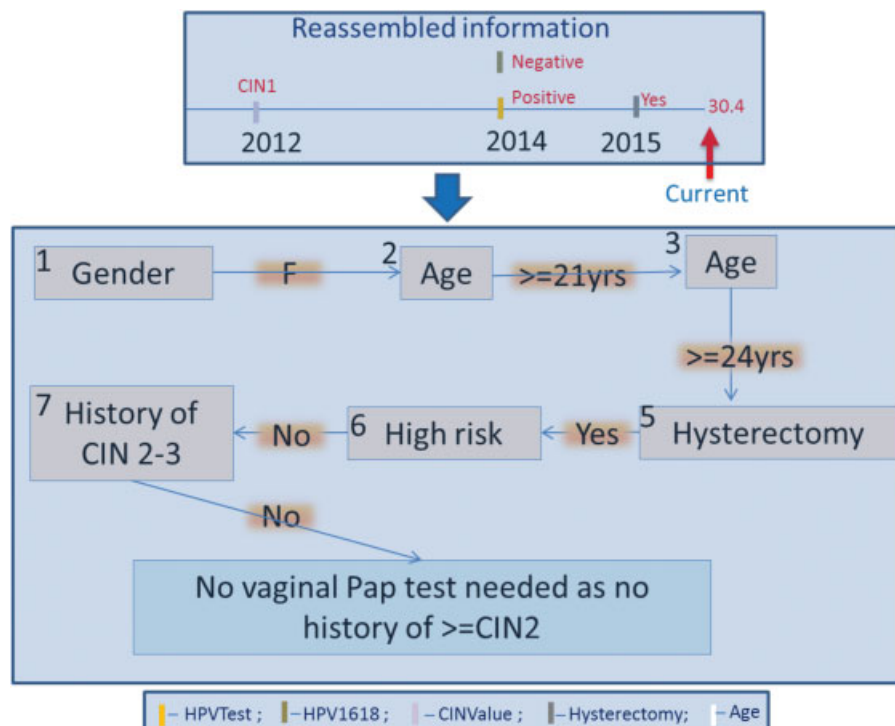

Fig. 3 Example of a decision logic rule that computes care recommendation based on data elements and their respective values over a time period by CDSS. CDSS, clinical decision support systems.

(1) Removal of endo-cervical transformation zone (ETZ) criteria: Presence or absence of adequate ETZ impacted follow-up recommendations in the earlier prototype. Previously, a 1-year repeat Pap test was recommended if an inadequate ETZ was noted on the Pap report.[20] The

ASCCP now advises that routine (normal) follow-up is acceptable with absent ETZ in the setting of an otherwise negative test result, as the risk of CIN3+ over time is not higher in this population.[17] The CDSS has been revised to reflect these changes.

(2) Management changes for young women: Significant guideline changes were made in the management recommendations for young women aged 21 to 24 years with minor Pap abnormalities that advise annual follow-up for up to 24 months before proceeding to colposcopy. Management of colposcopic biopsy results preceded by high-grade Pap abnormalities for young women was also altered to reflect a more conservative approach.[17] These major changes are reflected in the updated CDSS.

(3) Updates of management of atypical squamous cells of undetermined significance (ASCUS): The management of ASCUS alone was changed to reflect a 12-month repeat Pap test, rather than 6- and 12-month follow-ups or immediate colposcopy. In the setting of ASCUS with negative HPV, follow-up Pap/HPV cotesting in 3 years, rather than 5 years, was recommended.[17]

(4) Use of Pap/HPV cotesting strategy for follow-up: The algorithms were updated to reflect increased use of Pap/HPV cotesting for follow-up of specific Pap/HPV results and colposcopy biopsy results, even for those under 30 years in some scenarios. Prior guidelines more often advised Pap or HPV follow-up testing only in low-risk scenarios and repeat colposcopy for higher grade abnormal test results.[17]

(5) Early exit for elderly patients with no high-risk factor: To improve computational efficiency, patients with age more than 65 years with no high-risk factors, such as history of cervical cancer or HIV, were removed early in the decision workflow. Previously, these patients were not eliminated until the very end of the pathway.

(6) Changes to fix known errors: We found two types of errors in the previous implementation of CDSS. Some of the variables extracted were not the intended ones that led to the errors in clinical recommendation. For example, age of patient at a specific test is an important criterion for computing the recommendations. We observed an error in the implementation of the previous prototype, where the system instead of considering the patient's age at the last Pap test mistakenly considered the current age of the patient resulting in a wrong recommendation. We also corrected another error related to the NLP algorithm. The algorithm failed to detect the correct HPV test type, which resulted in errors in the recommendations. This was due to changes in the language of our laboratory's HPV reports driven by ASCCP updates recommending reporting of HPV genotypes 16 and 18 in the setting of a negative Pap test.

The output of the CDSS includes three features: (1) the decision endpoint, as labeled in ►Supplementary Fig. S1, available in the online version, (2) a text displaying the clinical recommendation, and (3) the date when the next follow-up was due.

Determination of Due Status of the Patient for the Next Screening

The CDSS also computes the date of follow-up, if it determines that a follow-up is necessary. If the recommended date falls before the date at which the CDSS is run, then the patient is determined to be overdue for their next follow-up. This is an important aspect of CDSS, as it helps the physician in identifying the patients who are overdue for their follow-up.

Implementation

The CDSS framework is implemented in Java. The data retrieval from the respective sources is implemented in two ways (1) as a web service where Java servlets hosted on a web server (2) SQL queries directed against respective data sources for extracting the data elements to compute the care recommendations.

The focus of the current study deals with the evaluation of the precision of CDSS in a nonclinical setting. We have not described the integration of the CDSS recommendation into the clinical workflow.

Evaluation

We ran the revised CDSS on all female patients 21 years and older visiting primary care clinicians from Employee and Community Health at Mayo Clinic in Rochester, Minnesota between May 1 and 15, 2015, resulting in care recommendations for 3,704 patients. To evaluate the accuracy of the CDSS, we randomly sampled 10% of the patients from each decision end point from this cohort for manual evaluation. Stratified sampling from each clinical recommendation pathway ensured that the evaluation sample had adequate representation of all case scenarios/pathways. We observed an in equal distribution of patients among the different recommendation endpoints. In total, we selected 393 patients for the evaluation. ►Supplementary Table S1, available in the online version lists distribution of patients in the evaluation sample.

We calculated the precision of the recommendations for both routine (normal) and abnormal pathways. The evaluation was performed by an expert clinician assisted by a resident. The expert performed a detailed chart review of all 393 patients including assessment of prior Pap reports, HPV results, and colposcopy biopsy reports. After a review of the clinical record, the evaluators recorded whether the CDSS recommended the appropriate follow-up based on the ASSCP guideline-based computational workflow as displayed by ►Supplementary Fig. S1, available in the online version. The expert clinician led and oversaw the evaluation process to ensure the logic is in compliance with the ASCCP guidelines. The performance of the system was evaluated in terms of precision, which is defined as in the equation given below.

$$\text{Precision (in percentage)} = \frac{\text{Number of patients with correct recommendations}}{\text{Total Number of patients in the sample}} * (1)$$

We performed a detailed error analysis and categorized them on the type of errors, which helped us address them in a systematic manner before integration into the clinical workflow.

Results

CDSS Accuracy

Out of the 393 patients evaluated, the revised system made correct recommendations to 369 patients achieving an accuracy of 93.4%. ►Table 2 stratifies the results for patients with

Table 2 Evaluation of CDSS performance of CDSS among routine/normal and abnormal end points

End point type	Total patients	Total correct	Accuracy
Routine/normal	307	297	96.7%
Abnormal	86	72	83.7%
Total	393	369	93.4%

Abbreviation: CDSS, clinical decision support systems.

normal compared with high-risk (abnormal) end point recommendations. Out of the 393 patients, 307 are considered routine/normal screening and 86 patients deviate from routine screening due to abnormal prior Pap smears (ratio of abnormal to routine/normal is ~1:3). The precision of the CDSS among the patients with endpoints who require only routine/normal screening Pap was higher (96.7%) than among patients who had abnormal results or a high-risk history (83.7%).

Potential Role of CDSS in Determining Women Overdue for Screening

We assessed the possible role that the CDSS could play in identifying women overdue for screening, for both women with normal and abnormal past history. The CDSS generated the suggested next follow-up date for 281 out of the 393 patients. Among the 307 women with routine screening recommendations, the CDSS determined follow-up dates for subsequent screening in 200 patients, while it determined the date for follow-up for 81 out of the 86 patients with abnormal past history. For the 107 (=307–200) women in the routine screening group for whom the CDSS could not determine a follow-up date for, the expert determined that these women no longer required screening, due to the following reasons: (1) a history of a hysterectomy with no history of CIN 2–3 or cervical cancer (70), or (2) age greater than 65 (20). There were 17 women in the routine screening group who had prior screening at an outside facility and their outside Pap record was not accessible in a format that allowed for the CDSS to ascertain the exact date of the next follow-up (17).

Among the 81 women who had abnormal past history, the CDSS identified nearly 67% (54 patients) who were overdue for their next screening (see ►Table 3). Among the women with a normal history, nearly 21% were overdue for their next follow-up. The CDSS could not determine an exact follow-up date for five women who had abnormal past history due to incorrect recommendations by the system.

Error Analysis

We performed a detailed review of any errors that were identified, categorized them based on cause of the error, and tried to think of solutions we encountered while testing our new prototype. We have summarized our error analysis in ►Table 4.

Table 3 Overdue screening and surveillance among routine/normal and abnormal end points

End point type	Total patients (T)	Total patients with definite time for next follow-up (P)	Total patients overdue (O)	Percentage patients overdue (O/P)
Routine/normal	307	200	46	23.0%
Abnormal	86	81	54	66.7%

Table 4 Error analysis and categorization

Error type	Specific error	No. of errors	Ability to address	Solution
Data source errors	Errors in coded problem list	3	No	Feed back to the data sources
Modeling errors	Clinical decision not clearly captured in the decision logic	6	Yes	Altering the implementation based on expert feedback
	Lack of adherence to ASCCP guidelines in past clinical practice	5	No	Such errors will gradually be eliminated once clinical practice strictly adheres to ASCCP guideline
Programming errors	Determination of correct end points	6	Yes (partially solvable)	While programming, errors of simple kind can be permanently eliminated, certain error of correct next follow-up time may not be possible
Evaluation errors	Clinician arriving at a wrong decision	4	Yes	Adoption of such CDSS described in this article has the potential to eliminate such manual errors

Abbreviations: ASCCP, American Society of Colposcopy and Cervical Pathology; CDSS, clinical decision support systems.

The errors can be broadly categorized into four broad categories: data source errors, modeling errors, which includes oversights in our logic as well as gaps due to lack of clinical adherence to guidelines in the past, programming errors, and evaluation errors.

Data Source Errors

Lack of access to accurate information: Delay in access or lack of access to accurate information is one category of error that we encountered during analysis of CDSS recommendation. For example, the colposcopy report and biopsy result may not be accessible to the CDSS for several days due to the provider finalizing the note. While a physician may still have access to nonfinalized data, the CDSS do not have access to such data, which results in temporary errors. Due to these errors, we revised our prototype to update on a daily basis. Another data error was that the database, from which the information was drawn, such as the coded problem list, contains out of date information or an incorrect diagnosis entry. Additionally, the CDSS does not have access to information on women who had their follow-up outside of our health system, which led to erroneous recommendations. Such errors due to nonavailability of patient information are challenging to eliminate.

Modeling Errors

(1) Clinical Scenario not being captured in the decision logic: We found that the CDSS logic/pathways did not capture the case scenarios for four patients. For example, during transition from node 45 (previous cervical cytology) to recommendation 41 "R41" (see ►**Supplementary Fig. S1**, available in the online version), the expert physician felt the need to know an additional criteria of whether the previous HPV was positive or negative.

(2) Lack of adherence to ASCCP guidelines in past clinical practice: The CDSS assumes that clinicians have previously provided care in compliance with the ASCCP recommendations. In situations, where the ASCCP guidelines have not been followed, errors in CDSS recommendations were noted. For example, for a patient with LSIL in the past and age less than 25 years, the system recommended "Pap (only cytology) within a year." However, the primary care physician at the time sent the patient for colposcopy, which is a deviation from the ASCCP guidelines. We strongly believe that adoption of CDSS in clinical practice will streamline clinical practice to adhere to ASCCP guidelines and help overcome such errors in the future. A possible way to resolve this issue is to generate a warning note in the CDSS recommendation that the previous care provided was not in compliance with the guidelines. The major obstacle for this approach is that it would involve modeling past publications of the guideline in the CDSS.

Programming Errors

We also encountered programming errors. For example, the program failed to correctly compute the next test time for certain decision scenarios. For the end point "R38," the recommendation is "Pap-HPV cotest at 1, 2, and 5 years" post-CIN 2–3

treatment. The CDSS makes the right care recommendation, but makes an error while determining the exact date for the 1-, 2-, and 5-year follow-ups. The system by default suggested follow-up dates based on the last Pap date without considering which stage they were in their current follow-up. For this endpoint, the patient should have three different "next test times," namely 1 year after their last Pap date at the initial stage, 1 more year after the last Pap date for stage 2, and 3 years later for the final stage. We corrected this error by leaving the decision making for the next test time recommendation for this endpoint to the physician themselves.

Evaluation Errors

Error in expert's data acquisition: There were four patients when the system's recommendations were marked by the expert physician as incorrect but on further review were resolved to be correct. For instance, the history of hysterectomy was overlooked both by the study team and they arrived at the conclusion that the system's recommendation is wrong. When reassessing the recommendation in light of the date of hysterectomy, they concurred with the system's recommendation. Such errors could have been avoided if more than one physician expert was involved in the review. This draws our attention to the fact that a physician with often very limited time to make a recommendation has the potential chance of overlooking certain data and arrives at a clinical decision that is not aligned to the ASCCP guideline. The ASCCP algorithms for cervical cancer screening and surveillance are too complex for a physician to memorize. CDSS such as the one described in this article is likely to be extremely useful in ensuring that high-risk patients receive appropriate care at the right time.

Computation Time

The system takes an average of 72 seconds to compute a recommendation for one patient. The maximum time the CDSS took for computing a single patient's recommendation is 2 minutes and the minimum time is 9 seconds. The system takes a pause of approximately 100 milliseconds in between processing two successive patients to reduce the load on the secondary sources of enterprise data servers.

Discussion

In the current study, we made enhancements to our earlier prototype CDSS with the updated ASCCP guidelines. We performed a thorough evaluation in a nonclinical setting to improve the accuracy of the CDSS. The NLP enabled decision support system described in this study integrates data from diverse sources (both structured and unstructured data) to arrive at the right clinical decision for cervical screening for patients. The work described in this article is about revising the CDSS and not completely a novel work in terms of CDSS development. However, this effort to revise the implementations of CDSS due to change in the practice of guidelines is a necessary one. It gave us a deep understanding of the systematic process required for a CDSS revision and evaluation required to accomplish before its deployment in clinical practice.

Detailed analysis revealed that the performance of the updated CDSS in both the normal and abnormal population falls in an acceptable range (>95% for normal and > 85% for abnormal), thereby paving a way for its integration into the clinical practice. For abnormal patient population where the CDSS has the potential to make mistakes, the physicians receive a cautionary note as part of the recommendation to verify and determine the appropriate follow-up recommendation for the patient. We believe that this will ensure greater compliance with cervical cancer screening and surveillance to ASCCP guidelines at Mayo Clinic. We acknowledge there were only incremental additions to the number of clinical decision scenarios in comparison to our earlier implementation. Nevertheless, a robust evaluation performed in this study is critical prior to CDSS deployment in a clinical practice. A systematic error analysis helped us characterize the performance of the CDSS, which allowed us to further improve the performance of the CDSS.

The CDSS described in this study is comprehensive, as it can generate recommendations for all patients—screening reminders for average-risk patients and surveillance reminders for patients with past abnormal findings or in a high-risk screening category, which is a critical and significant advancement over the existing systems. Our system performs a complete analysis of the patient data— discrete and free text—and provides explanations for the generated recommendations, thereby providing a greater degree of assistance in clinical decision making. The CDSS is not a replacement for clinicians but can be of great assistance to them in identifying patients at a higher risk for cervical cancer by enabling them to perform appropriate interventions to prevent cancer development.

Our approach represents significant progress from the existing paradigm in decision support of simple average risk screening cohort identification to providing actionable suggestions with comprehensive explanations. In the context of current interventions aiming to improve cervical cancer prevention, our study provides new knowledge about the ability of CDSS to identify high-risk patients and provide specific recommendations to clinicians for guideline-compliant surveillance of abnormal cervical cytology and HPV results.[16]

Given the complex nature of the care recommendation guideline (as illustrated in ►Supplementary Fig. S1, available in the online version) for cervical cancer screening and surveillance, it is very unrealistic for the physician to keep abreast of the recent changes and apply them on their patients all the time. A clinical decision system, such as the one described in this study, may reduce errors in clinician decision making. The greatest value of the CDSS described in this article is in its ability to identify high-risk patients, who are overdue for their follow-up at an appropriate time. Delay in the care for these high-risk patients at an appropriate time may significantly increase the patient's risk toward progression to cervical cancer.

Limitations

The evaluation of the CDSS is not based on a preannotated gold standard dataset. Instead, the expert independently reviewed all the records of the patient and determined whether the recommendation was correct by applying the ASCCP guidelines to each specific patient scenario reviewed.

We could report only the precision of the system and not its recall (a measure of sensitivity or coverage). Evaluation of the performance of the CDSS against a manually created gold standard is ideal. However, creation of gold standard involves a lot of clinician effort and time. Hence, validation based on chart review is the best alternative. Another limitation is the expert was not double blinded and hence may be prone to bias. The evaluation was done only by one clinician and did not involve a wider set of physicians, so we cannot judge inter-rater agreement for this dataset. However, in an earlier formative evaluation of this CDSS, we asked providers to determine a recommendation while blinded to the CDSS recommendation, and we found providers disagreed with the CDSS recommendations on 75 out of 169 (44%) patient scenarios. When there was disagreement with the recommendation made by the CDSS, the cases were decided by an expert. The CDSS recommendation was found to be more often correct than the providers 53/75 (71%) versus 22/75 (29%).[11] The low inter-rater agreement is one of the reasons we limited our evaluation to be compared based on our computational workflow overseen by one expert to verify the workflow is in agreement with the ASCCP guidelines. In a separate study, we are evaluating the impact of sending reminders to patients after multiple physician review.

In this study, we measured the net improvement of all the changes that we made to the system. We did not evaluate the effect of the individual updates that we made to the CDSS workflow and separately evaluated the improvement due to addressing the errors in the previous implementation. The multiple changes that we made to the system make it difficult to attribute the improvement to one particular change. In subsequent studies we will redesign the workflow that enables us to compare the results between two care guidelines.

Another key limitation of the system is that the decision at every point in the clinical care pathway is not stored but recomputed on the fly every time they are processed. This leads to redundancy in computation of data element values. The system on an average takes 72 seconds to compute recommendations for a single patient. It is highly desirable to reduce the redundancy in the computation of values of different data elements for implementation of near real-time care recommendations at the point of care. Ideally, we would like to reduce the per patient computation time to the order of milliseconds. We believe that during the architectural redesign discussed in the following section, we will plan to achieve the computation time per patient in the order of milliseconds.

There are few other important limitations of the current implementation of CDSS. First, a significant amount of manual effort is required to revise the implementation of decision workflow, whenever there is a revision in the care guidelines. Post revision, an extensive evaluation is required before we implement the changes in a clinical setting.

Second, the system works well for the Mayo Clinic data sources. At this point, it requires immense work to make the system interoperable so that CDSS can be seamlessly integrated into other institutions workflow with minimal effort for near real-time clinical use. The proposal by Wagholikar et al[21] regarding adoption of SMART-on-FHIR driven by

REST-API architecture will enable us to overcome some aspects of the two limitations discussed above. We plan to overhaul the software architecture and recast the CDSS into a modular, FHIR-compliant REST-API web architecture so that it is interoperable and can be adopted across institutions.

The CDSS described in the study is a rule-based system and requires extensive manual effort in revising the CDSS implementation. The task of identifying normal versus abnormal patient endpoints is an ideal setting for a machine learning-based classification. However, we did not explore machine learning-based approach to this problem due to the following reasons: (1) the number of patients with normal history is far higher than the number of patients with abnormal history. Hence, there will be an inherent bias in the training sample that may potentially affect the performance of machine learning algorithm. The algorithms are far more complex for the high risk/abnormal endpoints than for the routine/normal endpoint. The complexity of the algorithms creates more opportunities for errors. Hence, the physicians are more comfortable with an open architecture (rule-based approach) than a machine learning approach where the features learned for the specific task are often black box to the physicians. However, we believe that in future work we intend to explore machine learning-based approach to clinical decision support solutions that will generalize well to interinstitution data, thereby facilitating interoperability to a greater extent.

Conclusion

In this work, we have implemented the latest updates made to the ASCCP guidelines in 2013 and addressed the errors in the previous implementation. We performed a multistage evaluation and took a systematic approach to identify the causes of errors. We addressed all the errors amenable to correction to improve the clinical performance of CDSS. We believe it is now ready for deployment in clinical practice, and we are in advanced stages of evaluating the impact of the CDSS in a clinical setting.

Clinical Relevance Statement

Our results show that a CDSS can generate recommendations with sufficient accuracy for a complex set of guidelines. The creation of CDSS to identify high-risk individuals in addition to routine screening has tremendous potential. We strongly believe that the comprehensive recommendations generated by the NLP-based CDSS will improve the quality of care for women at risk for cervical cancer.

Multiple Choice Questions

1. Recommendation for managing cervical cancer screening are complex because recommendations are dependent upon:

    a. HPV results
    b. Prior cervical biopsy results
    c. Age of patient

    d. High risk factors such as HIV, immunodeficiency, and cervical cancer
    e. All of the above

    Correct Answer: The correct answer is e, all of the above.

2. Big data empowered natural language processing (NLP)
    a. Can help mine text data in EMR that are needed for clinical decision making
    b. Can help establish real-time clinical decision support
    c. Can help in making the right clinical decision irrespective of the data quality
    d. Option 1 and option 2

    Correct Answer: The correct answer is d, option 1 and option 2.

Authors' Contributions
K.E.R. and K.L.M. designed the experiments. K.E.R. and K.B.W. implemented the new changes. K.L.M. and M.R.S. performed the expert reviews. M.K, H.L., and R.C. participated in the design and analysis. K.L.M., H.L., and R.C. supervised the project. All authors contributed to the manuscript and approved the final version.

Protection of Human and Animal Subjects
This study does not involve any experiments involving human and animal subjects. The institutional review board at the Mayo Clinic, Rochester, MN, approved this study.

Conflict of Interest
None.

References
1 U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2013 Incidence and Mortality Web-based Report. 2016; Available at: http://www.cdc.gov/uscs
2 Duggan MA, Nation J. An audit of the cervical cancer screening histories of 246 women with carcinoma. J Low Genit Tract Dis 2012;16(03):263–270
3 Janerich DT, Hadjimichael O, Schwartz PE, et al. The screening histories of women with invasive cervical cancer, Connecticut. Am J Public Health 1995;85(06):791–794
4 Mema SC, Nation J, Yang H, et al. Screening history in 313 cases of invasive cancer: a retrospective review of cervical cancer screening in Alberta, Canada. J Low Genit Tract Dis 2017;21(01):17–20
5 Siegel R, DeSantis C, Virgo K, et al. Cancer treatment and survivorship statistics, 2012. CA Cancer J Clin 2012;62(04):220–241
6 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(09):675–683
7 Moyer VA; U.S. Preventive Services Task Force. Screening for cervical cancer: U.S. Preventive Services Task Force

recommendation statement. Ann Intern Med 2012;156(12): 880–891

8 Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012;157 (01):29–43

9 Lobach DF. The road to effective clinical decision support: are we there yet? . BMJ 2013;346:f1616

10 Everett T, Bryant A, Griffin MF, Martin-Hirsch PP, Forbes CA, Jepson RG. Interventions targeted at women to encourage the uptake of cervical screening. Cochrane Database Syst Rev 2011; (05):CD002834

11 Kim JJ. Opportunities to improve cervical cancer screening in the United States. Milbank Q 2012;90(01):38–41

12 Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005; 330(7494):765

13 Wagholikar KB, MacLaughlin KL, Henry MR, et al. Clinical decision support with automated text processing for cervical cancer screening. J Am Med Inform Assoc 2012;19(05):833–839

14 Wagholikar KB, MacLaughlin KL, Kastner TM, et al. Formative evaluation of the accuracy of a clinical decision support system for cervical cancer screening. J Am Med Inform Assoc 2013;20(04): 749–757

15 Wagholikar KB, MacLaughlin KL, Casey PM, et al. Automated recommendation for cervical cancer screening and surveillance. Cancer Inform 2014;13(Suppl 3):1–6

16 Wagholikar KB, MacLaughlin KL, Chute CG, Greenes RA, Liu H, Chaudhry R. Granular Quality Reporting for Cervical Cytology Testing. AMIA Jt Summits TranslSci Proc 2015;2015:178–182

17 Massad LS, Einstein MH, Huh WK, et al; 2012 ASCCP Consensus Guidelines Conference. 2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors. ObstetGynecol 2013;121(04):829–846

18 Saslow D, Solomon D, Lawson HW, et al; ACS-ASCCP-ASCP Cervical Cancer Guideline Committee. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. CA Cancer J Clin 2012;62(03):147–172

19 Schiffman M, Wentzensen N. A suggested approach to simplify and improve cervical screening in the United States. J Low Genit Tract Dis 2016;20(01):1–7

20 Davey DD, Cox JT, Austin RM, et al. Cervical cytology specimen adequacy: patient management guidelines and optimizing specimen collection. J Low Genit Tract Dis 2008;12(02):71–81

21 Wagholikar KB, Mandel JC, Klann JG, et al. SMART-on-FHIR implemented over i2b2. J Am Med Inform Assoc 2017;24(02):398–402