

# Workflow-based Data Reconciliation for Clinical Decision Support: Case of Colorectal Cancer Screening and Surveillance

Kavishwar Waghlikar, MBBS PhD<sup>1</sup>, Sunghwan Sohn, PhD<sup>1</sup>, Stephen Wu, PhD<sup>1</sup>,  
Vinod Kaggal<sup>1</sup>, Sheila Buehler, RN CNP<sup>2</sup>, Robert A. Greenes, MD PhD<sup>3,8</sup>,  
Tsung-Teh Wu, MD<sup>4</sup>, David Larson, MD MBA<sup>5</sup>, Hongfang Liu, PhD<sup>1</sup>,  
Rajeev Chaudhry, MBBS MPH<sup>6</sup>, Lisa Boardman, MD<sup>7</sup>

<sup>1</sup>Biomedical Statistics and Informatics, <sup>2</sup>Internal Medicine, <sup>4</sup>Anatomic Pathology,  
<sup>5</sup>Colon and Rectal Surgery, <sup>6</sup>Primary Care Internal Medicine, <sup>7</sup>Gastroenterology and  
Hepatology, Mayo Clinic Rochester, MN  
<sup>3</sup>Biomedical Informatics, Arizona State University, Scottsdale, AZ; <sup>8</sup> Health Science  
Research, Mayo Clinic, Scottsdale, AZ

## Abstract

*A major barrier for computer-based clinical decision support (CDS), is the difficulty in obtaining the patient information required for decision making. The information gap is often due to deficiencies in the clinical documentation. One approach to address this gap is to gather and reconcile data from related documents or data sources. In this paper we consider the case of a CDS system for colorectal cancer screening and surveillance. We describe the use of workflow analysis to design data reconciliation processes. Further, we perform a quantitative analysis of the impact of these processes on system performance using a dataset of 106 patients. Results show that data reconciliation considerably improves the performance of the system. Our study demonstrates that, workflow-based data reconciliation can play a vital role in designing new-generation CDS systems that are based on complex guideline models and use natural language processing (NLP) to obtain patient data.*

## Introduction

A major barrier for computer based clinical decision support (CDS), is the difficulty in obtaining the patient information required for decision making. The lack of adequate information is often due to inherent deficiencies in the clinical documentation. To address this information gap, many systems are designed to seek data entry from the user. However, this approach interrupts the user's workflow and thereby impairs the utility of the system. An alternative solution is to obtain data from related documents or data sources, and then reconcile them. For instance, the history of heart failure in a patient often cannot be reliably determined from the problem list (1). This information can be augmented by the finding of less than 50 percent ejection fraction in the echocardiogram, which is specific to heart failure patients. Such an approach requires a deeper understanding of the documentation quality and domain knowledge which can be discovered with workflow analysis.

In this paper we consider the case of a clinical decision support system (CDSS) for colorectal cancer screening and surveillance (2). We describe the use of workflow analysis to discover needed data sources and reconciliation processes. Further we perform a quantitative analysis of the impact of these processes on system performance using a dataset of 106 patients.

## Background

The ability of care providers to collect and abstract patient information from the clinical record is vital to their clinical decision making. However, the documentation itself is often deficient, which can impact the clinical decisions and consequently the health outcomes. While much of the earlier work on documentation quality has focused on data completeness and accuracy, recently several other measures of documentation quality have been developed (3, 4).

The quality of the patient data/ documentation depends on several factors including: motivation or purpose for the document (5), time available for documentation, use of schemas (6), expertise and qualifications of the documenter and use of documentation tools (7). A deeper understanding of the document characteristics is critical to determining whether the quality of the data contained in the documents is suitable for a particular task. For instance, the diagnostic and billing codes generated by data abstractors for insurance reimbursement lack sufficient accuracy and completeness for clinical decision making, and thus are not suitable for this purpose.

Knowledge of document characteristics can inform the determination of what documents should be gathered and how the data can be reconciled. We define data reconciliation as a process to identify and rectify deficiencies in raw data, to improve the quality of the data. For instance clinicians reconcile inaccuracies in the problem list using the detailed histories recorded in the clinical notes. Any inaccuracies about patient history in a particular clinical note can be resolved by referring to other notes, or by talking to the patient directly. Feblowitz et al have described the need for such a correction process to address documentation errors for data summarization (8). Effective data reconciliation strategies can be designed by using workflow analysis to discover the documentation characteristics.

Workflow is defined in diverse ways in the literature (9). But generally workflow refers to a set of processes that can be represented using a flowchart. Workflow analysis has been used for designing health information technology applications and for measuring the impact of the solutions (10). The wide variety of methods used for workflow analysis includes surveys, focus groups, expert panels, direct observation, interviews and documentation analysis.

We have attempted to use workflow analysis to identify data sources and to design data reconciliation processes in a CDSS for colorectal cancer screening and surveillance (2). The guidelines for colonoscopy are complex, and care providers often fail to make the optimal recommendation (11). Consequently, there is substantial overuse of surveillance colonoscopy among low-risk subjects and underuse among high-risk subjects, and 41% of patients do not receive adequate screening for colorectal cancer (CRC) (12).

Although decision support tools have been developed for a variety of preventive services (13), applications for colonoscopy screening and surveillance have been limited due to complexity of the guidelines (14) and the need for extensive patient information for decision making. The current tools for colonoscopy recommendation are generally limited in scope (15, 16) or require data input by care providers (17), which restricts their utility. The task of constructing a CDSS that automatically collates all the required patient parameters from the electronic health record (EHR) is challenging, since i) several parameters are involved, ii) the documentation is often of inadequate quality or is in freetext form that is not amenable to computer processing, and iii) the system is required to have a high degree of accuracy. To address these challenges we have explored the use of workflow modeling to design and implement a CDSS that can generate guideline-based recommendations by automated analysis of EHR and related data sources. The CDSS uses natural language processing (NLP) to extract some of the required information from textual reports (Figure 1).

## **Methodology**

We performed an observational study of the workflow during colonoscopy and interviewed the involved care providers, in order to identify the documentation that was generated and to understand the content and quality of the data. Further we analyzed the documents in the context of the quality of data for making the colonoscopy decisions.

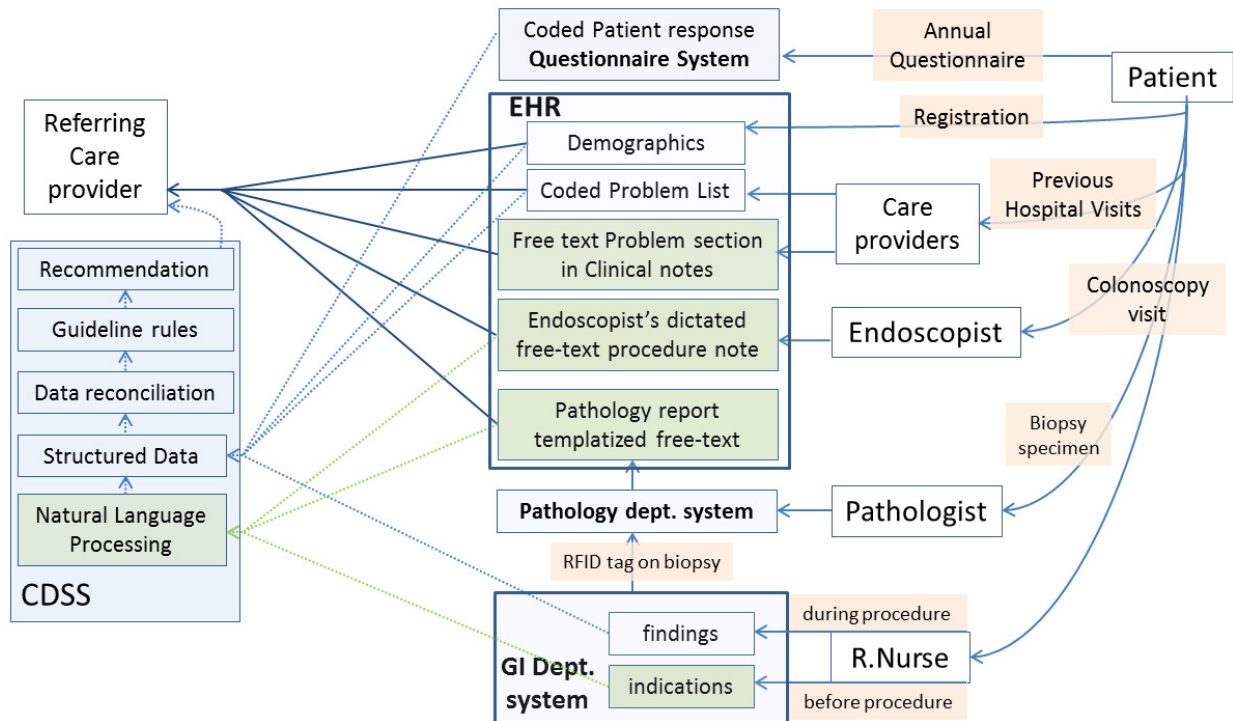
Based on the workflow analysis, we designed several processes for data reconciliation in the CDSS. To determine the utility of these processes, we performed a quantitative analysis by measuring system performance after selectively adding/removing each of reconciliation processes from the CDSS (Table 1). We measured two parameters for each of the processes: i) gain in the accuracy of the patient parameter reconciled by the process, and ii) gain in the accuracy of the recommendations computed by the CDSS. The quantitative analysis was based on a dataset of 106 cases, that was used to iteratively develop and validate the CDSS. The dataset included randomly selected cases representing diverse decision scenarios (2).

## **Results**

Figure 1 is an overview of the information flow delineated by the workflow studies. There are three information systems that are involved: the EHR and the internal databases in the departments of pathology and gastroenterology. A brief summary of the chronologic flow of information is as follows.

When the patient is scheduled for a colonoscopy, the indications assessed by the referring care provider are recorded in the gastrointestinal (GI) system. The colonoscopy procedure lasts for about 30 minutes, and is performed by an endoscopist, who is either a gastroenterologist (GE) or a colorectal cancer surgeon. The endoscopist is assisted by a registered nurse (RN) and licensed practical nurse (GI assistant). The RN operates a computer terminal throughout the procedure, entering information into the GI database, including the patient's vital signs, medications, equipment usage, adequacy of the cleanliness of the colon, specimen collection and procedure findings. The procedure findings are usually explicitly communicated to the RN by the endoscopist, but at times the RN independently observes and records the findings. This information is primarily collected for administrative, medico-legal purposes and research.

Immediately after the procedure is completed, the endoscopist dictates a voice recording of the description of the procedure (procedure/endoscopy note) which is manually transcribed into a preliminary textual report. A GI pathologist examines the specimens and records all histological findings using a system of in-house codes and ad hoc comments, which are ultimately converted into a textual pathology report via a template. The endoscopist reviews and edits the transcribed dictation of the procedure note, but rarely adds a recommendation for follow up to the colonoscopy procedure note. This is because the endoscopist lacks a review of the patient's family history of CRC, prior polyp history and the pathologist's report about any tissue removed at the procedure. Determination of the appropriate surveillance and future colonoscopy evaluations is then made by the referring/ordering care provider, who is expected to review the endoscopy note, the pathology report and pertinent clinical notes before making the decision.



**Figure 1.** Overview of clinical information flow and architecture of the CDSS. Freetext data sources are shaded in green.

We noted that there were redundant work processes for recording colonoscopy and other patient data. For instance the RNs recorded the preparation and findings of polyp number, site and size, as structured elements in the GI database. The same information was recorded in the endoscopist's freetext procedure note. These findings motivated us to use the overlapping data sources for data reconciliation, as follows – i) To improve data completeness we gathered complimentary data elements from overlapping data sources or assumed default values when no data was found. ii) To improve accuracy we resolved inconsistent data elements during the collation process, by prioritizing the data source that is more reliable. Table 1 lists the patient parameters, data sources and accuracy gains due to the multiple-source data gathering and reconciliation processes.

The quantitative analysis revealed that with the data reconciliation processes the CDSS failed to compute any recommendation for only 6 of 106 patients, due to errors in interfacing with the EHR webservice. We excluded these cases from further analysis. For the remaining 100 cases, CDSS processes computed the optimal recommendation correctly for 95 patients (accuracy=95%).

Without data reconciliation, prep (preparation) was not correctly extracted from the endoscopy note in (16 of 100) 16% cases, which decreased the recommendation accuracy by 19%. With addition of reconciliation processes to use prep data from GI findings database when the prep was not mentioned in the endoscopy note, 5% more cases were resolved and the recommendation accuracy gained by 3%. The improvement was due to improved recognition of inadequate preparation, which was either not documented in the note or missed by the NLP algorithm. The accuracy of the system after integrating the data sources (i.e. endoscopy and GI findings) as described above was greater than restricting the system to use either of the sources.

We added a reconciliation process for assuming “adequate prep” when the prep was not recorded by either the endoscopist or the RN. This increased the accuracy for prep and the recommendation by 12% and 10% respectively. With both the multi-source reconciliation processes for prep, there was an overall accuracy gain of 16% and 13% for detecting prep and the recommendation, respectively. However there were 2 cases where the CDS system failed to compute the optimal recommendation due to incorrect assessment of the prep. In both cases the endoscopist had used idiosyncratic or complex expressions e.g., “there was some liquid stool that was suctioned” to indicate adequate prep, and the RN had indicated inadequate prep or not recorded information.

**Table 1.** Quantitative analysis of the data reconciliation process showing accuracy without the reconciliation process for the parameter and recommendation.

Patient Parameter	Primary data source in EHR	Reconciliation processes Description	% accuracy (% gain)	
			parameter	recommendation
Age	Demographics	-	-	-
Inadequate prep	Endoscopy note	Use data from GI findings, if no data about prep is found in endoscopy note	93 (5)	92 (3)
		Assume good prep. if no data is found in endoscopy note or GI findings	86 (12)	85 (10)
CRC risk	Problem list	Combine with data from GI indication	90 (10)	88 (7)
		Combine with data from Patient Questionnaire database	99 (1)	95 (0)
CRC syndrome	Problem list	-	-	-
IBD	Problem list	Combine with GI indications	99 (1)	94 (1)
Polyp cytology	Pathology report	-	-	-
Polyp size	Endoscopy note Pathology report*	Combine with data from GI findings and consider the largest polyp size reported	69 (31)	93 (2)
Polyp number	Endoscopy note*, Pathology report*	Exclusively use GI findings	-	-
Polyp removal	Endoscopy note	-	-	-

\* these data sources are excluded in the CDSS design based on the workflow analysis.

Similarly, using the record of GI indications helped considerably in determination of patient risk for colorectal cancer (CRC), but was less useful for supplementing data about history of inflammatory bowel disease (IBD). Patient questionnaire information was not useful.

With the workflow analysis, we noted that the RN is likely to record the number of polyps excised during the colonoscopy more accurately than the endoscopist, since the RN is required to label biopsy specimens as they are collected in real time. As the endoscopist dictates the findings after the procedure, he or she is likely to forget the exact number of polyps. Hence we designed the CDSS to use the structured recording made by the RN instead of developing an NLP algorithm to extract this information from the endoscopy note. As expected, there were several inconsistencies in the polyp size as documented by the endoscopist and the RN. However, the data reconciliation only improved the recommendation accuracy by 2%, since small differences in polyp size did not affect decision making. Overall as shown in Table 1, the results indicate that incorporation of the data reconciliation processes significantly improved the accuracy of the recommendations computed by the CDSS.

## Discussion

The workflow analysis informed the design of data reconciliation processes, which significantly improved the performance of the CDSS. Our study demonstrates that data reconciliation guided by workflow analysis may be helpful to overcome the lack of good quality documentation that is an obstacle for CDSS development.

Conventionally, software engineering practices involve workflow studies, when the software is primarily for workflow support. However CDSSs are generally triggered into action, after the care providers create documentation in the EHR, and the CDSS design is not concerned with the user’s workflow during the documentation phase. Supporting the user workflow for document creation is the task of the EHR system. Consequently, the focus of workflow studies for CDSS design has been on effective delivery of decision recommendations with minimum interruption of the users’ workflow and to aid communication of the CDSS development team with the care providers (18). However, our study demonstrates that when there is lack of adequate data quality, workflow studies of the documentation phase can help discover data reconciliation strategies that can resolve the data deficiencies.

NLP has been largely under-utilized for CDS, as it often lacks sufficient accuracy (14). However, our study has demonstrated that it is feasible to use high precision rule-based NLP algorithms in CDSS. The lack of sensitivity of the NLP algorithm can be overcome with multi-source data reconciliation. For instance, the NLP on the endoscopy note for extracting polyp size was rendered effective when it was supplemented with the structured GI findings. The findings of the workflow analysis also obviated the laborious effort to develop complex NLP for determining some of the required information. For instance, the RNs' structured recording of GI findings obviated the need to develop an NLP algorithm for detecting number of polyps from the endoscopy note.

A limitation of our study was that it was small in scale and was localized to a single institution. However the described methods are generalizable and can be useful for other researchers.

## Conclusion

The workflow analysis was useful to identify data reconciliation strategies to address documentation gaps, and the reconciliation processes considerably improved the performance of the CDSS. Our study demonstrates that, workflow-based data reconciliation can play an important role in designing new-generation clinical decision support systems, which are based on complex guideline models and utilize text processing.

## References

1. Wright A, Maloney FL, Feblowitz JC. Clinician attitudes toward and use of electronic problem lists: a thematic analysis. *BMC Med Inform Decis Mak.* 2011;11:36.
2. Waghlikar KB, Sohn S, Wu S, et al. Clinical Decision Support for Colonoscopy Surveillance Using Natural Language Processing. *IEEE Healthcare Informatics, Imaging, and Systems Biology Conference.* University of California, San Diego, CA; 2012.
3. Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing Electronic Note Quality Using the Physician Documentation Quality Instrument (PDQI-9). *Appl Clin Inform.* 2012;3(2):164-74.
4. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc.* 2012 Jun 25.
5. Reiser SJ. The clinical record in medicine. Part 2: Reforming content and purpose. *Ann Intern Med.* 1991 Jun 1;114(11):980-5.
6. Rosenbloom ST, Denny J, Xu H, Lorenzi N, Stead W, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc.* 2011 Mar-Apr;18(2):181-6.
7. Tang PC, LaRosa MP, Gorden SM. Use of computer-based records, completeness of documentation, and appropriateness of documented clinical decisions. *J Am Med Inform Assoc.* 1999;6(3):245-51.
8. Feblowitz JC, Wright A, Singh H, Samal L, Sittig DF. Summarization of clinical information: a conceptual model. *J Biomed Inform.* 2011 Aug;44(4):688-99.
9. Unertl KM, Novak LL, Johnson KB, Lorenzi NM. Traversing the many paths of workflow research: developing a conceptual framework of workflow terminology through a systematic literature review. *J Am Med Inform Assoc.* 2010 May-Jun;17(3):265-73.
10. Unertl KM, Weinger MB, Johnson KB, Lorenzi NM. Describing and modeling workflow and information flow in chronic disease care. *J Am Med Inform Assoc.* 2009 Nov-Dec;16(6):826-36.
11. McFarland EG, Levin B, Lieberman DA, et al. Revised colorectal screening guidelines: joint effort of the American Cancer Society, U.S. Multisociety Task Force on Colorectal Cancer, and American College of Radiology. *Radiology.* 2008 Sep;248(3):717-20.
12. Cancer screening - United States, 2010. *MMWR Morb Mortal Wkly Rep.* 2012 Jan 27;61(3):41-5.
13. Lau F, Kuziemy C, Price M, Gardner J. A review on systematic reviews of health information system studies. *J Am Med Inform Assoc.* 2010 Nov-Dec;17(6):637-45.
14. Sittig DF, Wright A, Osheroff JA, et al. Grand challenges in clinical decision support. *J Biomed Inform.* 2008 Apr;41(2):387-92.
15. Imperiale TF, Sherer EA, Balph JA, Cardwell JD, Qi R. Provider acceptance, safety, and effectiveness of a computer-based decision tool for colonoscopy preparation. *Int J Med Inform.* 2011 Oct;80(10):726-33.
16. 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 Jan;32(1):188-97.
17. Terraz O, Wietlisbach V, Jeannot JG, et al. The EPAGE internet guideline as a decision support tool for determining the appropriateness of colonoscopy. *Digestion.* 2005;71(2):72-7.
18. Jalote-Parmar A, Badke-Schaub P, Ali W, Samset E. Cognitive processes as integrative component for developing expert decision-making systems: a workflow centered framework. *J Biomed Inform.* 2010 Feb;43(1):60-74.