

# Consolidated EHR Workflow for Endoscopy Quality Reporting

Shorabuddin SYED<sup>a,1</sup>, Benjamin THARIAN<sup>b</sup>, Hafsa Bareen SYEDA<sup>a</sup>, Meredith ZOZUS<sup>c</sup>, Melody L. GREER<sup>a</sup>, Sudeepa BHATTACHARYYA<sup>d</sup>, Mahanazuddin SYED<sup>a</sup>, and Fred PRIOR<sup>a,c</sup>

<sup>a</sup>*Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA*

<sup>b</sup>*Division of Gastroenterology and Hepatology, University of Arkansas for Medical Sciences, Little Rock, AR, USA*

<sup>c</sup>*Department of Population Health Sciences, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA*

<sup>d</sup>*Department of Biological Sciences and Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR, USA*

<sup>e</sup>*Department of Radiology, University of Arkansas for Medical Sciences, Little Rock, AR, USA*

**Abstract.** Although colonoscopy is the most frequently performed endoscopic procedure, the lack of standardized reporting is impeding clinical and translational research. Inadequacies in data extraction from the raw, unstructured text in electronic health records (EHR) pose an additional challenge to procedure quality metric reporting, as vital details related to the procedure are stored in disparate documents. Currently, there is no EHR workflow that links these documents to the specific colonoscopy procedure, making the process of data extraction error prone. We hypothesize that extracting comprehensive colonoscopy quality metrics from consolidated procedure documents using computational linguistic techniques, and integrating it with discrete EHR data can improve quality of screening and cancer detection rate. As a first step, we developed an algorithm that links colonoscopy, pathology and imaging documents by analyzing the chronology of various orders placed relative to the colonoscopy procedure. The algorithm was installed and validated at the University of Arkansas for Medical Sciences (UAMS). The proposed algorithm in conjunction with Natural Language Processing (NLP) techniques can overcome current limitations of manual data abstraction.

**Keywords.** Colonoscopy, quality improvement, natural language processing, data integration, electronic health records

## 1. Introduction

About 3.4 million luminal Gastrointestinal cancers (esophageal, stomach, colorectal) are detected globally every year [1, 2]. These cancers represent a substantial health challenge for society, with a mortality rate of about 63%, with colorectal cancer being the third most common cause of cancer related mortality among both women and men [2].

---

<sup>1</sup> Corresponding Author, Shorabuddin Syed, University of Arkansas for Medical Sciences, Little Rock, AR, USA; E-mail: ssyed@uams.edu.

Screening colonoscopy plays a critical role in diagnosis of colorectal cancers. Quality improvement and screening efficiency are dependent upon the study of quality metrics. The American College of Gastroenterology has published a list of 15 quality indicators to improve colonoscopy safety and performance [3, 4]. The established quality metrics such as adenoma detection rates, bowel preparation, cecal intubation rate and scope withdrawal times are documented in endoscopy and pathology reports. Procedure indicators, medical history such as comorbidities, active medication, and socio-economic status require review of clinical history and radiology reports. The inability to extract information from unstructured text in electronic health records (EHR) is a barrier to quality improvement and secondary research related to colonoscopy. EHR contains a wealth of insight into patients [5] and the integration of disparate documents that contain pertinent data.

Recently computational linguistic techniques such as Natural language processing (NLP) and Machine Learning (ML) have been used as an alternative to manual data abstraction from unstructured free text [6]. Extracting clinical predictors from colonoscopy procedures poses an additional challenge, as vital details related to the procedure are often stored in disparate documents. It is crucial to integrate various document types generated from each colonoscopy visit before any data extraction techniques can be applied. Previous studies have attempted to relate pathology documents to the colonoscopy procedure [7, 8]. However, the linkage algorithms had the following limitations: 1) for multiple pathology orders placed on the procedure day, the algorithms were not able to identify the pathology report specific to the procedure; 2) for upper endoscopy and colonoscopy performed on the same day, the algorithms were not able to differentiate between the procedures. Moreover, imaging reports were not integrated to gather procedure indications and other related data, opening doors for potential source of study bias. To address this limitation, we built an automated and generalizable EHR workflow that links colonoscopy, pathology and imaging documents by analyzing the chronology of various orders placed relative to the colonoscopy procedure.

## 2. Methods

To link pathology and imaging orders associated with the procedure, an algorithm was built using a set of most commonly used order attributes. The algorithm is based on two steps: 1) link pathology orders 2) link imaging orders. Pathology orders are generally placed during or immediately after the colonoscopy, if any biopsies or samples are taken. Thus, zero, one or several pathology reports can be associated with the procedure. Whereas, imaging orders can constitute procedure indication or follow-up.

### 2.1. Link Pathology Orders

As shown in Figure 1, to identify and link pathology orders generated from the colonoscopy procedure an algorithm with the following steps was built.

**Step 1:** Identify all completed colonoscopy procedures and collect data elements related to patient's medical record number, gastroenterologist performing the procedure, date and time of procedure.

**Step 2:** Extract pathology orders placed for a patient within 48 hours of the colonoscopy procedure. Although 95% of the orders are placed during the procedure, the 48 hours

window was chosen to account for delayed orders and time lag due to overnight emergency procedures.

**Step 3:** A procedure can have zero, one or multiple pathology orders. If no pathology orders are found then confirm with the procedure billing codes.

**Step 4:** For procedures that have one or more associated pathology orders, validate if the orders are related to the colonoscopy procedure by verifying “specimen type”, “pathology order location”, and “authorizing provider”. Eliminate orders that are not related to the colonoscopy procedure. Archive colonoscopy orders that are not associated with pathology orders and contradicts with billing codes for manual review.

**Step 5:** For procedures that have only one pathology order, link them.

**Step 6:** For procedures with multiple pathology orders, verify that the orders was placed on the same day and link the orders based on the specimen type, and classify orders as “primary” or “secondary” based on the pathology ordering time.

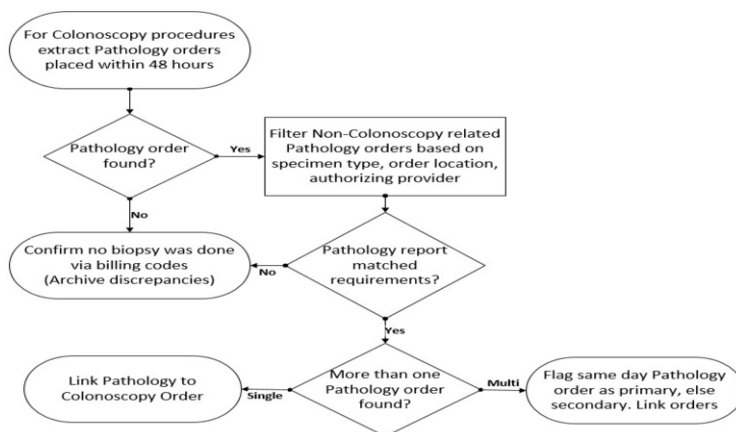


Figure 1. Workflow to link pathology orders to colonoscopy orders.

## 2.2. Link Imaging Orders

As imaging orders (i.e., Abdominal-pelvis CT scan, Abdominal USG) can constitute procedure indication or follow-up, it is vital to identify imaging orders placed before colonoscopy and the orders placed as part of follow up from the procedure. To link imaging reports a two-step approach was followed.

First, to link radiology reports done prior to colonoscopy procedure, we built a corpus of colonoscopy procedure indications, i.e. the reason for performing the procedure. The corpus was built by a panel of gastroenterologist physicians (lead by BT); they conducted an extensive chart review and handpicked terms that indicates abnormality in CT scans and recommendation for a colonoscopy. Examples of selected indication includes “abnormal scan”, “diverticulitis”, and “unexplained weight loss” etc. Next step is to identify colonoscopy procedures with indications that matched with the terms in the corpus. For the identified procedures, collect patient’s medical record number, date and time of the procedure, procedure indication, and gastroenterologist performing the procedure. Then link imaging orders placed 180 days prior to the procedure date using the extracted metrics. To account for delay in scheduling, follow-

up appointments and urgency of procedure, the 180 days window was considered reasonable.

Secondly, to link radiology reports done after colonoscopy, collect aforementioned metrics for all completed colonoscopy procedures. The follow up appointments could be done by the proceduralist (which could be a gastroenterologist, surgical endoscopist or by a family medicine physician) or by a related specialist including surgeons (for perforations related to the procedure, etc.), or an oncologist, or a hospitalist. Identify imaging orders authorized/ordered by these physicians within the two months of the procedure and link the orders.

3. Application

At University of Arkansas for Medical Sciences (UAMS), the algorithm to link colonoscopy related orders was implemented using Structured Query Language (SQL). UAMS utilizes the EPIC platform (Epic Systems Corp, Verona, WI) for electronic health records. We identified 16,900 colonoscopy procedures performed at UAMS between May 2014 and September 2020.

To link pathology orders related to colonoscopy procedure, we ran the algorithm on 16,900 colonoscopy and 11,182 pathology orders placed at UAMS respectively. The algorithm classified colonoscopy orders in to three categories, colonoscopy procedure resulting in: 1) no pathology order ( $col_{p=0}$ ), 2) one pathology order ( $col_{p=1}$ ), and 3) more than one pathology orders ( $col_{p>1}$ ). Of the 16,900 procedure orders, 5,800 had zero pathology orders, 10,900 had one pathology order, and 200 had more than one pathology orders respectively. The algorithm’s accuracy was evaluated to that of manual review done by two trained data warehouse analyst and a gastroenterologist (BT). A random sample of colonoscopy orders ( $n=400$  [ $col_{p=0}=99$ ,  $col_{p=1}=256$ ,  $col_{p>1}=45$ ],  $N=16,900$ ,  $CI=95\%$ ) was selected. Test for marginal homogeneity (Stuart-Maxwell test,  $k=3$  and  $df=2$ ) between the three mutually exclusive categories ( $col_{p=0}$ ,  $col_{p=1}$ ,  $col_{p>1}$ ) was performed. The Stuart-Maxwell test was selected due to paired nature of the three categories and to test marginal homogeneity for all categories simultaneously. For both algorithm and chart-review, the frequency of the three pathology-order categories found in the sample was computed. The value of Stuart-Maxwell statistic was  $<9.21$  ( $\alpha=0.01$  and  $df=2$ ), inferencing the distribution between the categories for the automated-process is similar to the distribution for the chart review.

To link imaging orders related to colonoscopy procedure, we ran the algorithm on 16,900 colonoscopy and 7,364 imaging orders placed at UAMS respectively. The algorithm identified 3,510 colonoscopy orders that constitute procedure indication and 1,409 follow-up imaging orders respectively. We randomly selected colonoscopy orders 347 ( $N=3,510$ , confidence interval = 95%,  $\epsilon=5\%$ ) and 303 ( $N=1,409$ , confidence interval = 95%,  $\epsilon=5\%$ ) from the two categories. The two analysts validated algorithm’s accuracy to link imaging orders via chart review. The accuracy reported was 96.2% and 98.7% for linking pre and post imaging orders to associated procedure orders.

The results from the algorithm were transformed and stored in a mapping table to link various report types associated to a specific colonoscopy. For example, as show in Table 1, a colonoscopy procedure “Col\_1234” is associated with two pathology orders (“Path\_4001” and “Path\_9901”) and an imaging order (“Img\_54012”).

**Table 1.** Mapping table layout that links various pathology and imaging orders to a specific colonoscopy.

Colonoscopy Order ID	Order Type	Associated Order ID
Col_1234	Patho	Path_4001
Col_1234	Patho	Path_9901
Col_1234	Image	Img_54012
Col_7855	Patho	Path_8091

#### 4. Discussion and Conclusion

A comprehensive study of colonoscopy quality metrics poses an additional challenge of vital procedure details being distributed across multiple unstructured document types, quality of procedure metrics extraction depends on how these documents are linked. We built and tested an automated generalizable EHR workflow that links colonoscopy, pathology and imaging documents based on the vital and most commonly captured order attributes. Irrespective of the EHR software used, the proposed algorithm can be employed to link these document types and is not limited to SQL based implementation.

Applying NLP techniques to extract clinical predictors from the consolidated unstructured procedure documents would facilitate comprehensive and temporal reporting of vital details. This significantly reduces data accessibility time and facilitates clinical and translational endoscopy research. Continuous evaluation of procedure outcomes and, provider and health care facility performance, will reduce the need for repeat procedures and failed detection rates of adenomas. Thereby, reducing cost of procedure and improving quality of care.

**Funding Acknowledgment and Consent:** Patients' data used were obtained under IRB approval (IRB# 262202) at the UAMS. This study was supported in part by the Translational Research Institute (TRI), grant UL1 TR003107 received from the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

#### References

- [1] de Lange T, Halvorsen P, Riegler M. Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy. *World J Gastroenterol.* 2018;24(45):5057-62.
- [2] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians.* 2018;68(6):394-424.
- [3] Anderson JC, Butterly LF. Colonoscopy: quality indicators. *Clin Transl Gastroenterol.* 2015;6(2):e77-e.
- [4] Rex DK, Schoenfeld PS, Cohen J, Pike IM, Adler DG, Fennerty MB, et al. Quality indicators for colonoscopy. *Gastrointestinal endoscopy.* 2015;81(1):31-53.
- [5] Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association : JAMIA.* 2011;18(2):181-6.
- [6] Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *Journal of the American Medical Informatics Association : JAMIA.* 2011;18(5):544-51.
- [7] Mehrotra A, Dellon ES, Schoen RE, Saul M, Bishehsari F, Farmer C, et al. Applying a natural language processing tool to electronic health records to assess performance on colonoscopy quality measures. *Gastrointestinal endoscopy.* 2012;75(6):1233-9.e14.
- [8] Raju GS, Lum PJ, Slack RS, Thirumurthi S, Lynch PM, Miller E, et al. Natural language processing as an alternative to manual reporting of colonoscopy quality metrics. *Gastrointestinal endoscopy.* 2015;82(3):512-9.