Accuracy of Referring Provider and Endoscopist Impressions of Colonoscopy Indication

Mariam Naveed, MD; Meredith Clary, MD; Chul Ahn, PhD; Nisa Kubiliun, MD; Deepak Agrawal, MD, MPH, MBA; Byron Cryer, MD; Caitlin Murphy, PhD, MPH; and Amit G. Singal, MD, MS

Abstract
Background: Referring provider and endoscopist impressions of colonoscopy indication are used for clinical care, reimbursement, and quality reporting decisions; however, the accuracy of these impressions is unknown. This study assessed the sensitivity, specificity, positive and negative predictive value, and overall accuracy of methods to classify colonoscopy indication, including referring provider impression, endoscopist impression, and administrative algorithm compared with gold standard chart review. Methods: We randomly sampled 400 patients undergoing a colonoscopy at a Veterans Affairs health system between January 2010 and December 2010. Referring provider and endoscopist impressions of colonoscopy indication were compared with gold-standard chart review. Indications were classified into 4 mutually exclusive categories: diagnostic, surveillance, high-risk screening, or average-risk screening. Results: Of 400 colonoscopies, 26% were performed for average-risk screening, 7% for high-risk screening, 26% for surveillance, and 41% for diagnostic indications. Accuracy of referring provider and endoscopist impressions of colonoscopy indication were 87% and 84%, respectively, which were significantly higher than that of the administrative algorithm (45%; P < .001 for both). There was substantial agreement between endoscopist and referring provider impressions (κ = 0.76). All 3 methods showed high sensitivity (>90%) for determining screening (vs nonscreening) indication, but specificity of the administrative algorithm was lower (40.3%) compared with referring provider (93.7%) and endoscopist (84.0%) impressions. Accuracy of endoscopist, but not referring provider, impression was lower in patients with a family history of colon cancer than in those without (65% vs 84%; P = .001). Conclusions: Referring provider and endoscopist impressions of colonoscopy indication are both accurate and may be useful data to incorporate into algorithms classifying colonoscopy indication.

doi:10.6004/jnccn.2017.0123

Background
Over the past decade, colonoscopy use has increased dramatically in the United States. This is partly related to increasing uptake of colorectal cancer (CRC) screening examinations, but colonoscopy is often performed for other reasons, including surveillance of those with prior adenomas and for diagnostic evaluation of patients with gastrointestinal symptoms. In fact, diagnostic and surveillance colonoscopies account for most endoscopy examinations performed in the United States.1 Accurately classifying colonoscopy indication is crucial, with implications for clinical care, healthcare quality reporting, and policy reform.2 Ability to distinguish between screening colonoscopies and examinations performed for other reasons is also important for research examining predictors of screening uptake and adherence and impact of screening on patient
outcomes. Several administrative or claims data-based algorithms have been developed to classify colonoscopy indication; however, none have demonstrated simultaneously high sensitivity and specificity in discriminating between screening and other colonoscopies. Sensitivity and specificity range from 29% to 84% and from 58% to 93%, respectively. Natural language processing (NLP) algorithms have also been tested, but these require further refinement and validation before widespread dissemination.

Meanwhile, referring provider and/or endoscopist impression of colonoscopy indication is often used for several important clinical decisions. First, the indication listed by the referring provider drives procedural urgency and timing of colonoscopy scheduling. In many centers, particularly those with limited endoscopic capacity, patients referred for diagnostic colonoscopy, including evaluation of positive fecal immunochemical testing, are prioritized over those referred for average-risk screening. The indication listed by the endoscopist can also have implications for reimbursement and patient copays, given that insurance companies are required to provide screening—but not surveillance or diagnostic examinations—without a copayment. Endoscopist impression also governs the recommended interval for repeat colonoscopy, with shorter intervals typically required in high-risk screening or surveillance examinations compared with average-risk screening examinations.

Despite the central role of the referring provider and endoscopist in the CRC screening process, surprisingly scant data exist regarding accuracy of provider impressions of colonoscopy indication. Thus, the goal of our study was to characterize the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy of referring providers and endoscopists in determining indication for colonoscopy, and to identify correlates of accuracy.

Methods

Study Population
We randomly sampled 400 adult (age >18 years) patients undergoing colonoscopy at the North Texas VA Healthcare System between January 1, 2010, and December 30, 2010. During this period, colonoscopy was performed for all indications, including average-risk screening, high-risk screening, surveillance, and diagnostic evaluation. Patients were identified using an endoscopic electronic medical reporting system, ProVation MD (ProVation Medical, Minneapolis, MN), which collects endoscopic data from all patients undergoing colonoscopy. We excluded patients with a history of inflammatory bowel disease and an existing diagnosis of hereditary colon cancer syndrome (based on chart review), given the high probability of misclassifying colonoscopy indication in these patients. The study was approved by the North Texas VA Institutional Review Board.

Data Collection
We considered 4 methods to classify colonoscopy indication: referring provider impression, endoscopist impression, administrative algorithm, and electronic medical record (EMR) gold standard. Referring provider impression was recorded from the colonoscopy referral order and progress notes from the corresponding clinic visit or telephone encounter. Endoscopist impression was recorded from the procedure and preprocedure history and physical notes from the day of colonoscopy. The administrative algorithm used a validated set of ICD-9 codes from the periprocedural period to determine indication, as described elsewhere. Two authors (M.N., M.C.) reviewed patient EMRs for the gold standard indication based on history lists (eg, family history of colon cancer), progress notes (eg, gastrointestinal symptoms), laboratory data (eg, iron deficiency anemia), radiologic data (eg, bowel thickening on imaging), and prior pathology and colonoscopy data (eg, history of adenomas). Family history of colon cancer denoting high-risk screening was defined as one first-degree relative with CRC or advanced adenoma diagnosed before 60 years of age, or 2 first-degree relatives with CRC or advanced adenomas diagnosed at any age. Indications were classified into 1 of 4 mutually exclusive categories: diagnostic (ie, performed for signs or symptoms, such as iron deficiency anemia, change in bowel habits, or hematochezia), surveillance (ie, personal history of polyps or CRC), high-risk screening (ie, increased risk for CRC, such as family history of CRC), or average-risk screening. When more than one category was possible, we used the following trumping

© JNCCN—Journal of the National Comprehensive Cancer Network | Volume 15 Number 7 | July 2017
algorithm based on perceived urgency of requiring endoscopic evaluation:

\[ \text{diagnostic} > \text{surveillance} > \text{high-risk screening} > \text{screening} \]

**Statistical Analysis**

We evaluated accuracy of colonoscopy indication as determined by the referring provider impression, endoscopist impression, and administrative algorithm compared with the gold standard EMR (Aim 1). Accuracy was calculated as the ratio of correctly identified colonoscopy indications to total number of colonoscopies. Agreement between referring provider and endoscopist impression was measured using the \( \kappa \) statistic.

To evaluate validity of colonoscopy indication (screening vs nonscreening colonoscopy) by classification method (Aim 2), we compared referring provider impression versus gold standard, endoscopist impression versus gold standard, and administrative algorithm versus gold standard. We calculated point estimates and 95% confidence intervals for the following validity measures: sensitivity, specificity, PPV, and NPV.

We used Tisnado’s criteria for evaluating validity measures: ≥0.9 indicates excellent agreement, ≥0.8 to <0.9 indicates good agreement, ≥0.7 to <0.8 indicates fair agreement, and <0.7 indicates poor agreement.

Fisher exact test (categorical) or Student t test (continuous) were used to identify correlates of accuracy for referring provider and endoscopist impression. Potential correlates included age, sex, race, family history of CRC, and referring provider specialty. Statistical significance was defined as \( P < .05 \) for all analyses. All data analyses were conducted using SAS 9.3 (SAS Institute, Cary, NC).

**Results**

**Patient Characteristics**

Baseline characteristics of the patient cohort are shown in Table 1. Median age was 60 years (range, 26–89 years), and most (89%) were male. Most patients (65%) were white, 30% were black, and 5% were Hispanic. Colonoscopies were performed for several indications as determined by gold standard EMR: 103 (26%) for average-risk screening, 29 (7%) high-risk screening, 163 (41%) surveillance, and 105 (26%) for diagnostic purposes.

**Accuracy for Colonoscopy Indication**

The overall accuracy of referring provider and endoscopist impression of indication was 87% (347/400) and 84% (335/400), respectively, compared with gold standard (Table 2). There was substantial agreement between referring provider and endoscopist impression (\( \kappa \), 0.76). Both referring provider and endoscopist impression had a higher accuracy than the administrative algorithm (45%; 178/400; \( P < .001 \) for both). Although the administrative algorithm had high accuracy for average-risk screening (95.2%), accuracy was substantially lower for diagnostic (47.6%) and surveillance (18.4%) colonoscopies (see supplemental eTable 1, available with this article at JNCCN.org). There was poor agreement between the administrative algorithm and referring provider (\( \kappa \), 0.23) and endoscopist (\( \kappa \), 0.26) impressions.

Table 3 shows the point estimates and 95% confidence intervals for the 4 validity measures (sensitivity, specificity, PPV, and NPV) when comparing referring provider impression, endoscopist impression, and the administrative algorithm versus the EMR gold standard for determining screening versus nonscreening indication. All 3 methods showed high sensitivity (>90%), but specificity of the administrative algorithm was lower (40.3%) compared with both referring provider (93.7%) and endoscopist (84.0%) impressions. Although the specificity of the referring provider impression was higher than endoscopist impression, this was offset by lower sensitivity (91.7% vs 98.5%).
pattern for NPV and PPV, wherein all 3 methods had high NPV (>90%), but the administrative algorithm had a much lower PPV (43.1% vs 87.7% for referring provider and 75.1% for endoscopist).

**Correlates for Accuracy of Colonoscopy Indication**

The accuracy of endoscopist impression was significantly lower in patients with a family history of colon cancer compared with patients with no family history (65% vs 84%; P=.001). In contrast, accuracy of the referring provider impression of indication did not differ among patients with (85%) and without (87%) a family history of colon cancer (P=.81). No difference in accuracy was seen between endoscopist and referring provider impression by patient age or sex.

**Discussion**

To the best of our knowledge, this study is the first to evaluate the accuracy of referring provider and endoscopist impression to classify colonoscopy indication. This is a critically important question because colonoscopy indication has implications for procedural urgency, reimbursement and patient copayments, appropriate screening/surveillance intervals, and healthcare quality reporting. We found both referring provider and endoscopist impression were highly accurate in classifying colonoscopy, with a high agreement.

The higher accuracy of provider impression for determining indication compared with administrative algorithms in our study highlights the need for improved algorithms. Prior administrative algorithms derived from the VA population have demonstrated limited accuracy, particularly when externally validated, given differences in coding practices between institutions.4,5 We found both referring provider and endoscopist impressions had significantly higher accuracy than a prior administrative algorithm. Although administrative algorithms can leverage available EMR data to address important epidemiologic questions in large populations, they are prone to misclassification given possible missing and/or inaccurate data in the EMR.12 Misclassification may be particularly likely for colonoscopy indication given receipt of colonoscopy with polypectomy at outside institutions, incomplete capture of CRC family history, and inaccurate coding for gastrointestinal symptoms.

Given a trade-off between sensitivity and specificity, administrative algorithms often prioritize one measure over another, highlighting the importance of characterizing sensitivity, specificity, PPV, and NPV in addition to overall accuracy.12 Although all 3 algorithms in our study had high sensitivity and NPV, referring provider and endoscopist impressions had higher specificity and PPV than the administrative

### Table 2. Referring Provider and Endoscopist Impression of Colonoscopy Indication

<table>
<thead>
<tr>
<th>Perspective of Colonoscopy Indication</th>
<th>Average-Risk Screening</th>
<th>High-Risk Screening</th>
<th>Surveillance</th>
<th>Diagnostic</th>
<th>Overall Accuracy (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold standard</td>
<td>103</td>
<td>29</td>
<td>163</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>Referring provider</td>
<td>103</td>
<td>35</td>
<td>154</td>
<td>108</td>
<td>86.8% (83.0–89.9)</td>
</tr>
<tr>
<td>Endoscopist</td>
<td>131</td>
<td>42</td>
<td>124</td>
<td>103</td>
<td>83.8% (79.8–87.2%)</td>
</tr>
<tr>
<td>Administrative algorithm</td>
<td>281</td>
<td>0</td>
<td>38</td>
<td>81</td>
<td>44.5% (39.6–49.5)</td>
</tr>
</tbody>
</table>

### Table 3. Performance Characteristics of Perspective to Determine Screening Indication of Colonoscopy

<table>
<thead>
<tr>
<th>Perspective of Colonoscopy Indication</th>
<th>Screening</th>
<th>Nonscreening</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referring provider</td>
<td>138</td>
<td>262</td>
<td>91.7% (85.6–95.8)</td>
<td>92.7% (90.0–96.3)</td>
<td>87.7% (81.0–92.7)</td>
<td>95.8% (92.6–97.9)</td>
</tr>
<tr>
<td>Endoscopist</td>
<td>173</td>
<td>227</td>
<td>98.5% (94.6–99.8)</td>
<td>84.0% (79.0–88.1)</td>
<td>75.1% (68.0–81.4)</td>
<td>99.1% (96.9–99.9)</td>
</tr>
<tr>
<td>Administrative algorithm</td>
<td>281</td>
<td>119</td>
<td>91.7% (85.6–95.8)</td>
<td>40.3% (34.4–46.4)</td>
<td>43.1% (37.2–49.1)</td>
<td>90.8% (84.1–95.3)</td>
</tr>
</tbody>
</table>

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.
algorithm. Specificity is the most important measure for classifying outcomes, whereas PPV is most important for identifying a cohort of persons with a condition of interest, such as those undergoing screening colonoscopy. A recent algorithm developed by Lee et al., using commonly available electronic medical data sources, demonstrated high sensitivity, specificity, and PPV for accurately classifying colonoscopy indications. However, the algorithm has only been validated in a single healthcare setting, and external validation studies are needed. If externally validated with similar performance in other health systems, this administrative algorithm may allow routine accurate capture of colonoscopy indication.

Overall, our data suggest referring provider and/or endoscopist impressions are readily available data that can be used as part of future algorithms. Although these data are available in most EMR systems, they are rarely discrete and therefore not easily extractable. These data instead require manual chart review, which can be both time-consuming and expensive. NLP, a field in computer science in which the computer is trained to “read” and extract meaningful information from free-text reports, may allow for a more accurate mechanism for obtaining these data from free text fields in the EMR and provide a quality measure assessment of the indication. Although NLP may offer a long-term solution to determine colonoscopy indication, further refinement and validation of NLP-based algorithms is still needed.

Although accurate, referring providers and endoscopists misclassify indication in approximately 15% of cases; therefore, algorithms that rely solely on these modalities should be approached with caution. Additional clinical data would be needed to accurately determine procedure indication; however, the burden and costs of additional data collection must be weighed against any gains in accuracy. One potential target for data collection is family history of CRC, because this was associated with endoscopist impression accuracy in our study. In fact, endoscopist impression was incorrect in one-third of cases with a positive family history. Prior studies have shown that verification of colon cancer family history can be easily missed unless time is taken to carefully ask the patient and to review the EMR. It is clear that efforts must be made to improve collection of family history of colon cancer given its importance in discriminating high-risk versus average-risk screening and its impact on screening interval recommendations.

Our findings should be interpreted in light of study limitations. This study was performed in a single healthcare setting with a limited sample of patients and may not be generalized to other practice settings. However, the VA offered unique advantages specific to our research question. The VA is an integrated healthcare system with a comprehensive EMR, providing access to referral orders, progress notes, laboratory results, endoscopy reports, and pathology data to determine an accurate gold standard indication. A second limitation of our study is that the authors were not blinded to the referring providers’ or endoscopists’ impressions when ascertaining the gold standard indication. Third, given the retrospective nature of the study, there is the potential for unmeasured confounders and missing data that could have led to misclassification of the gold standard indication. For example, family history is not accurately or consistently collected during clinic visits and reported family history of CRC in the general population can often be inaccurate.

Conclusions

Although it is clear colonoscopy indication is important, the best approach to accurately determine indication is less evident. There are several available algorithms, each using a different method and each with its own limitations. This study demonstrated that referring providers and endoscopists may offer a readily available and highly accurate source for determining colonoscopy indication.

References

Accuracy of Colonoscopy Indication


See JNCCN.org for supplemental online content.