**Name of Journal:** *Artificial Intelligence in Gastroenterology*

**Manuscript NO:** 56295

**Manuscript Type:** ORIGINAL ARTICLE

***Retrospective Study***

**Machine learning better predicts colonoscopy duration**

Podboy AJ *et al*. Machine learning and colonoscopy duration

Alexander Joseph Podboy, David Scheinker

**Alexander Joseph Podboy,** Division of Gastroenterology and Hepatology, Stanford University School of Medicine, Stanford, CA 94305, United States

**David Scheinker,** Department of Management Science and Engineering, Stanford University School of Engineering, Stanford, CA 94305, United States

**David Scheinker,** Department of Preoperative Services, Lucile Packard Children's Hospital Stanford, Stanford, CA 94304, United States

**Author contributions:** All authors equally contributed to this paper, in regards to conception and design of the study, literature review and analysis, drafting and critical revision and editing, and final approval of the final version.

**Corresponding author: Alexander Joseph Podboy, MD, Academic Fellow,** **Doctor,** Division of Gastroenterology and Hepatology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305, United States. [apodboy@stanford.edu](mailto:apodboy@stanford.edu)

**Received:** April 23, 2020

**Revised:** June 15, 2020

**Accepted:** June 17, 2020

**Published online:** July 28, 2020

**Abstract**

BACKGROUND

The use of machine learning (ML) to predict colonoscopy procedure duration has not been examined.

AIM

To assess if ML and data available at the time a colonoscopy procedure is scheduled could be used to estimate procedure duration more accurately than the current practice.

METHODS

Total 40168 colonoscopies from the Clinical Outcomes Research Initiative database were collected. ML models predicting procedure duration were developed using data available at time of scheduling. The top performing model was compared against historical practice. Models were evaluated based on accuracy (prediction – actual time) ± 5, 10, and 15 min.

RESULTS

ML outperformed historical practice with 77.1% to 68.9%, 87.3% to 79.6%, and 92.1% to 86.8% accuracy at 5, 10 and 15 min thresholds.

CONCLUSION

The use of ML to estimate colonoscopy procedure duration may lead to more accurate scheduling.

**Key words:** Machine Learning; Colonoscopy; Endoscopy; Artifical intelligence; Practice outcomes; Operations

**Citation:** Podboy AJ, Scheinker D. Machine learning better predicts colonoscopy duration. Artif Intell Gastroenterol 2020; 1(1): 30-36

**URL:** https://www.wjgnet.com/2644-3236/full/v1/i1/30.htm

**DOI:** https://dx.doi.org/10.35712/aig.v1.i1.30

**Core tip:** Machine learning has been utilized to predict surgical procedure duration and enhance operating room proficiency, however its usefulness for predicting colonoscopy procedure duration has not been examined. In determination of procedure duration by machine learning outperformed historical practice.

**INTRODUCTION**

Current colonoscopy scheduling models utilize either historical averages or predetermined time allotments (usually 30-45 min). Scheduling has not evolved to incorporate patient information, case complexity, procedure environment, or operator proficiency. Failure to assess for these variables can lead to significant misjudgments of procedural duration. These errors can result in both under- and overutilization of endoscopy room time leading to increased cost, misappropriation of endoscopy resources, delays, and decreases to patient and provider satisfaction[1].Machine learning (ML) has been utilized to predict surgical procedure duration and enhance operating room proficiency, however its usefulness for predicting colonoscopy procedure duration has not been examined[2,3].

Our aim was to assess if ML and data available at the time a colonoscopy procedure is scheduled could be used to estimate procedure duration more accurately than the current practice.

**MATERIALS AND METHODS**

The Clinical Outcomes Research Initiative (CORIv.4) database was queried for all colonoscopies with complete procedural duration times from 2008-2014 following approval from our institutional review board.

The CORI database is a national central repository of endoscopic procedures from a physician network of academic, community and veteran administration hospitals/practices. The details of the repository can be found in previous publications[4]. ML models were trained on variables with < 20% missing values and variables available prior to the procedure. Procedures with duration < 5 and > 280 min were excluded. All statistical analyses were performed in R-studio version 3.5.3 (Boston, Massachusetts).80% of the cases were used for training data and the remaining 20% used to compare the performance of these models. To reduce skew in the data, the target variable (procedural duration), was logarithmically transformed in line with previous publications[3,5].

Following established methodology[3,5,6], several models were tuned to predict procedure-time duration using cross-validation. The various models included random forest, gradient boosting machine, least absolute shrinkage and selection operator or LASSO, and extreme gradient boosting models (xgboost). The best performing model was selected based on lowest root mean squared error of the model and trained using historical data (2008-2013) to predict ‘current’ data (2014). Predictions derived from the best performing model were compared with the current standard of using historical means. Models were evaluated based on accuracy (prediction – actual time) within thresholds of 5, 10, and 15 min to account for operational considerations.

**RESULTS**

Total 40168 colonoscopies from 75 different sites from 2008 to 2014 with procedural duration information were obtained. 32136 (80%) of the cases were used for training the algorithm with the remaining 8032 (20%) used compare the performance of these models. A total of five patient (age, gender, race, ASA class, pediatric status), eight provider (endoscopist ID, degree of performing provider, degree year of performing provider, specialty of provider, gender and race/ethnicity of the provider, fellow involvement) and twelve procedure specific [(procedure year, procedure order, site ID, site type (University *vs* Community), location of procedure/facility type, duration of procedure, primary indication of procedure, depth intended of the procedure, sedation type used, state, and region)] variables were all selected for model analysis and training.

Table 1 demonstrates background characteristics of the final cohort. The best performing machine learning algorithm was the xgboost model. Figure 1 depicts the final models accuracy. The percentages of procedures for which the xgboost and the historical models generated forecasts within the 5, 10 and 15 min threshold were 77.1% *vs* 68.9%, 87.3% *vs* 79.6%, and 92.1% *vs* 86.8% (*P* < 0.001). The most important features of the model were: patient age, procedure year, and the degree year of provider year (Figure 2).

**DISCUSSION**

We demonstrated that machine learning predicts colonoscopy procedure duration more accurately than the currently accepted standard practice and the improvement was greater as the tolerance for error decreased.

Our results mimic similar applications of machine learning algorithms. Bartek *et al*[6] compared the standard practice of using average historical procedure duration and surgeon estimates of procedural duration compared to predictions derived from a machine learning model. Using a 10% accuracy threshold, the machine learning algorithm outperformed both traditional practices (39% ML *vs* 32% surgeon derived and 30% historical means). In an analysis of feature importance, the authors noted that fundamental case information, such as mean duration of the last ten procedures, was the most important predictive feature, with patient health metrics having a smaller total impact. However, our results suggest that patient specific factors may play a greater role in determining colonoscopy procedure duration. While again provider and procedural factors demonstrated high importance, patient specific factors (such as age, female sex) factored substantially into our model’s final predictions.

There are several strengths to our analysis. A large number of colonoscopies from a national repository of endoscopic procedures composed of a wide array of procedures, patients, and providers from an assortment of practice environments were analyzed. Inclusion of a national database increases generalizability by limiting regional or practice related biases.

However, there are several limitations to our analysis. Procedure reporting to the CORI database is voluntary and there may be an inherent selection bias in which easier colonoscopies were more likely to be reported to the database. This is supported by the relatively low overall procedural duration in our cohort. While the effects of a longer average procedure duration on our model are unknown, we anticipate more resiliency to increased error in the ML model compared to historical means, further enhancing the overall accuracy of the model compared to traditional practice.

While the algorithm was successful, it largely represents a rudimentary proof of concept option. Several variables that have been associated with difficult or lengthy colonoscopies in previous reports[7] were either not available or too incomplete in this current data set to allow for analysis. Inclusion of variables associated with difficult colonoscopies including body mass index, previous abdominal or pelvic surgeries, bowel habits, weight, height *etc*. would potentially improve the models accuracy.

The use of an algorithm trained on prospectively collected data with greater provider, environmental, patient, and procedural information may lead to improvements in colonoscopy procedure scheduling. Such improvements may contribute to improved efficiency, patient and provider satisfaction, and reduced costs. Further study is necessary to examine the implications of the deployment of such a model in a clinical setting, and assess if such models can be used in other gastrointestinal procedures.

**ARTICLE HIGHLIGHTS**

***Research background***

The usefulness of machine learning (ML) for predicting colonoscopy procedure duration has not been examined.

***Research motivation***

A ML algorithm trained on endoscopic data derived from the Clinical Outcomes Research Initiative database predicted colonoscopy procedure duration more accurately than the currently accepted standard practice and the improvement was greater as the tolerance for error decreased.

***Research objectives***

The aim of this study was to assess if ML and data available at the time a colonoscopy procedure is scheduled could be used to estimate procedure duration more accurately than the current practice.

***Research methods***

Total 40168 colonoscopies were collected. ML models predicting procedure duration were developed using data available at time of scheduling. The top performing model was compared against historical practice.

***Research results***

ML outperformed historical practice with 77.1% to 68.9%, 87.3% to 79.6%, and 92.1% to 86.8% accuracy at 5, 10 and 15 min thresholds, and the most important features of the model were: patient age, procedure year, and the degree year of provider year.

***Research conclusions***

The use of ML to estimate colonoscopy procedure duration may lead to more accurate scheduling.

***Research perspectives***

Further study is necessary to examine the implications of the deployment of such a model in a clinical setting, and assess if such models can be used in other gastrointestinal procedures.

**REFERENCES**

1 **Almeida R**, Paterson WG, Craig N, Hookey L. A Patient Flow Analysis: Identification of Process Inefficiencies and Workflow Metrics at an Ambulatory Endoscopy Unit. *Can J Gastroenterol Hepatol* 2016; **2016**: 2574076 [PMID: 27446830 DOI: 10.1155/2016/2574076]

2 **Stepaniak PS**, Heij C, Mannaerts GH, de Quelerij M, de Vries G. Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. *Anesth Analg* 2009; **109**: 1232-1245 [PMID: 19762753 DOI: 10.1213/ANE.0b013e3181b5de07]

3 **Master N,** Zhou Z, Miller D, Scheinker D, Bambos N, Glynn P. Improving predictions of pediatric surgical durations with supervised learning. *Int J Data Sci Anal* 2017; **4**: 33-52 [DOI: 10.1007/s41060-017-0055-0]

4 **Holub JL**, Morris C, Fagnan LJ, Logan JR, Michaels LC, Lieberman DA. Quality of Colonoscopy Performed in Rural Practice: Experience From the Clinical Outcomes Research Initiative and the Oregon Rural Practice-Based Research Network. *J Rural Health* 2018; **34 Suppl 1**: s75-s83 [PMID: 28045200 DOI: 10.1111/jrh.12231]

5 **Scheinker D**, Valencia A, Rodriguez F. Identification of Factors Associated With Variation in US County-Level Obesity Prevalence Rates Using Epidemiologic vs Machine Learning Models. *JAMA Netw Open* 2019; **2**: e192884 [PMID: 31026030 DOI: 10.1001/jamanetworkopen.2019.2884]

6 **Bartek MA**, Saxena RC, Solomon S, Fong CT, Behara LD, Venigandla R, Velagapudi K, Lang JD, Nair BG. Improving Operating Room Efficiency: Machine Learning Approach to Predict Case-Time Duration. *J Am Coll Surg* 2019; **229**: 346-354.e3 [PMID: 31310851 DOI: 10.1016/j.jamcollsurg.2019.05.029]

7 **Anderson JC**, Messina CR, Cohn W, Gottfried E, Ingber S, Bernstein G, Coman E, Polito J. Factors predictive of difficult colonoscopy. *Gastrointest Endosc* 2001; **54**: 558-562 [PMID: 11677470 DOI: 10.1067/mge.2001.118950]

**Footnotes**

**Institutional review board statement:** This research was deemed by the institutional review board of Stanford University.

**Informed consent statement:** The informed consent was waived.

**Conflict-of-interest statement:** Scheinker D serves as an advisor to Carta Healthcare - a healthcare analytics company. No other potential conflicts of interest or financial support to disclose.

**Data sharing statement:** No additional data are available.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

**Manuscript source:** Unsolicited manuscript

**Peer-review started:** April 23, 2020

**First decision:** June 4, 2020

**Article in press:** June 17, 2020

**Specialty type:** Gastroenterology and hepatology

**Country/Territory of origin:** United States

**Peer-review report’s scientific quality classification**

Grade A (Excellent): 0

Grade B (Very good): 0

Grade C (Good): C

Grade D (Fair): 0

Grade E (Poor): 0

**P-Reviewer:** Mohamed SY **S-Editor:** Wang JL **L-Editor: A**  **E-Editor:** Ma YJ

**Figure Legends**

**Figure 1 Accuracy of machine learning model *vs* historical average.**

**Figure 2 Feature importance of machine learning model.**

**Table 1 Cohort background characteristics**

|  |  |  |
| --- | --- | --- |
| **Summary statistics** | **Characteristics** | **Data** |
| **Demographic information** | | |
| Total patients | 40168 |  |
| mean age | 58.95 |  |
| Sex | Female | 17682 |
| Male | 22485 (56.0%) |
| ASA Class | I | 7071 |
| II | 27699 |
| III | 5237 |
| IV | 158 |
| V | 3 |
| Race | Caucasian | 32031 |
| Hispanic | 2219 |
| Black | 2193 |
| Asian | 1140 |
| Native American | 679 |
| Other | 1906 |
| **Procedural information** | | |
| Median procedure year | 2012 | (2008-2014) |
| Total # of Sites | 75 |  |
| Fellow involved | 3575 |  |
| Indication for  procedure | Average risk screening | 12687 |
| Surveillance of adenomatous polyps | 8213 |
| Hematochezia | 3795 |
| High risk  screening | 3272 |
| Anemia | 1508 |
| Diarrhea | 1469 |
| Other | 9224 |
| Procedure order | 1st | 37864 |
| 2nd | 2056 |
| Other | 248 |
| mean duration of procedure | 23.4 min |  |
| Depth intended | Cecum | 31745 |
| Terminal Ileum | 6798 |
| Ascending colon | 570 |
| Ileum | 424 |
| Anastomosis site | 447 |
| Other | 163 |
| Location of the procedure | Hospital endoscopy suite | 15589 |
| Ambulatory surgery center | 14730 |
| unknown | 5739 |
| Office | 2501 |
| Endoscopy suite | 1450 |
| ICU | 88 |
| Region | North Central | 3490 |
| Northeast | 11156 |
| Northwest | 12329 |
| South Central | 776 |
| South East | 1466 |
| South West | 10947 |
| Site type | Community | 25133 |
| HMO | 1000 |
| University | 5676 |
| VA | 8359 |
| Sedation | None | 241 |
| Moderate/ Conscious sedation | 28009 |
| ‘Deep’ Sedation | 7289 |
| General Anesthesia | 2510 |
| Anxiolytic Sedation | 78 |
| **Provider information** | | |
| Gender of provider | Female | 9881 |
| Male | 30287 |
| Median degree year of provider | 1989 | (1962-2009) |
| Degree of performing  provider | DO | 1253 |
| MD | 38851 |
| PA | 64 |
| Provider specialty | Gastroenterology | 33059 |
| Surgery | 2976 |
| Colorectal Surgery | 995 |
| Internal Medicine | 1589 |
| Family medicine | 581 |
| Other | 968 |
| Ethnicity of provider | Hispanic | 419 |
| Non-hispanic | 37148 |

ICU: Intensive care unit; HMO: Health maintenance organization.