Development of a Multivariable Prediction Model to Identify Patients Unlikely to Complete a Colonoscopy Following an Abnormal FIT Test in Community Clinics

Amanda Petrik (✉ amanda.f.petrik@kpchr.org)  
Kaiser Permanente Northwest  https://orcid.org/0000-0002-8891-6042

Erin Keast  
Center for Health Research, Kaiser Permanente Northwest

Eric Johnson  
Center for Health Research, Kaiser Permanente Northwest

David H. Smith  
Center for Health Research, Kaiser Permanente Northwest

Gloria D. Coronado  
Center for Health Research, Kaiser Permanente Northwest

Research article

Keywords: Colorectal cancer screening, fecal immunochemical test, colonoscopy, multivariable prediction model, predictive analytics, precision medicine, follow-up colonoscopy

DOI: https://doi.org/10.21203/rs.3.rs-23641/v3

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: Colorectal cancer (CRC) is the 2nd leading cancer killer in the US. The Strategies and Opportunities to STOP Colon Cancer in Priority Populations (STOP CRC) project aimed to increase CRC screening among patients in Federally Qualified Health Centers (FQHCs) through a mailed fecal immunochemical test (FIT) outreach program. However, rates of completion of the follow-up colonoscopy following an abnormal FIT remain low. We developed a multivariable prediction model using data available in the electronic health record to assess the probability of patients obtaining a colonoscopy following an abnormal FIT test.

Methods: To assess the probability of obtaining a colonoscopy, we used Cox regression to develop a risk prediction model among a retrospective cohort of patients with an abnormal FIT result.

Results: Of 1596 patients with an abnormal FIT result, 556 (34.8%) had a recorded colonoscopy within 6 months. The model shows adequate separation of patients across risk levels for non-adherence to follow-up colonoscopy (bootstrap-corrected C-statistic > 0.63). The refined model included 8 variables: age, race, insurance, GINI income inequality, long term anticoagulant use, receipt of a flu vaccine in the past year, frequency of missed clinic appointments, and clinic site. Probability of obtaining a follow-up colonoscopy within 6 months varied across quintiles; patients in the lowest quintile had an estimated 18% chance, whereas patients in the top quintile had a greater than 55% chance of obtaining a follow-up colonoscopy.

Conclusions: Knowing who is unlikely to follow-up on an abnormal FIT test could help identify patients who need an early intervention aimed at completing a follow-up colonoscopy.

Trial registry: This trial was registered at ClinicalTrials.gov (NCT01742065) on December 5, 2012. The protocol is available.

Background

Colorectal cancer (CRC) is the 2nd leading cancer killer in the United States. Mailed fecal immunochemical testing (FIT) outreach programs can effectively increase CRC screening rates among underserved populations (1-3). The Strategies and Opportunities to STOP Colon Cancer in Priority Populations (STOP CRC) project aimed to increase CRC screening among patients in Federally Qualified Health Centers (FQHCs) through mailed FIT outreach (4). FIT looks for hidden blood in the stool, which may be a sign of polyps or cancer in the colon or rectum. An abnormal test means that blood was found in the stool. For these patients, a follow-up colonoscopy is recommended, yet rates of completion of the follow-up colonoscopy remain low (5-7). Among patients who receive care in community clinics in the United States, follow-up colonoscopy rates are consistently as low as 50% (7-10). Liss and Chubak have identified rates of achieving a follow-up colonoscopy after an abnormal FIT test of 54% and 50% respectively in community health center patients (11, 12). This is of concern because delaying a follow-up colonoscopy up to 12 months following an abnormal fecal test is associated with increased cancer diagnoses and advanced cancer stage at the time of diagnosis (13).
Barriers can inhibit patients’ ability to complete this follow-up colonoscopy. Patient-level barriers to completing a follow-up colonoscopy may include fear of results, inability to take time off of work, the cost of preparation supplies or the colonoscopy, inability to complete adequate bowel prep, difficulty finding a driver on the day of the procedure, having competing health concerns, and lack of understanding that the procedure was necessary (7, 10, 14, 15).

Provider and system-level barriers may include limited colonoscopy capacity, failure to refer the patient to the specialist or schedule the procedure, failure to communicate expectations about the procedure or preparation for the procedure, and lack of adequate workflows to complete the referral (6-8). Interventions like patient navigation, where a navigator helps address these barriers to screening, can close the gaps and improve follow-up rates (15, 16).

While it is likely cost-prohibitive to “navigate” all patients with an abnormal FIT, stratifying the patients in greatest need of navigation could target resources to close gaps in screening. The use of the electronic health record (EHR) to identify patients at risk for failure to follow-up on abnormal screening, who may be candidates for personalized interventions, may improve the precision of healthcare delivery (17). Therefore, we aimed to develop a multivariable prediction model using patient level data only available in the EHR to identify patients who are unlikely to undergo colonoscopy following an abnormal FIT test. We hypothesized that we could accurately predict which patients have a low probability of obtaining a colonoscopy.

Knowing who may be at risk for not adhering to recommendations for a follow-up colonoscopy after an abnormal FIT test could help providers and clinics identify patients in need of early interventions (including patient navigation) aimed at completing a colonoscopy. Precision delivery of interventions to those most likely to benefit might optimize patient outcomes and enhance opportunities to sustain successful interventions in low-resource settings.

**Methods**

To predict each patient’s probability of obtaining a colonoscopy, we developed a risk prediction model using data from patients with an abnormal fecal test at the 26 STOP CRC clinics. We followed guidelines set forth by the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnoses statement (18, 19). This model was designed to be put into practice at community clinics using data available in the EHR. Our objective was to predict patients who may benefit from interventions to complete recommended follow-up.

*Setting and Participants*

This retrospective analysis used data from the STOP CRC project, and included eligible patients who have returned a FIT with an abnormal result during the study period. OCHIN, formerly the Oregon Community Health Information Network, is a nonprofit health care controlled network that provides a centrally hosted EHR for primary care clinics. The STOP CRC project included 26 clinics in Oregon and California that
served as the setting for this cohort. This project was approved by the Kaiser Permanente Northwest IRB (Protocol #4364). Clinics operated in diverse settings, were diverse in size, and were part of 8 health centers.

To be eligible in the STOP CRC study, patients had to have been 50-75 years old and not up to date with CRC screening including fecal testing in past 11 months or colonoscopy in past 9 years. Patients were excluded from STOP CRC if they had co-morbid conditions that would make screening inappropriate, such as a history of CRC, colectomy, or dialysis. Our complete inclusion criteria are described elsewhere (20). We then assembled a complete retrospective cohort of STOP CRC patients who subsequently completed CRC screening by FIT test and obtained an “abnormal” result. All patients with at least one abnormal FIT result from February 4, 2014 through February 28, 2016 were identified (n=1723). For patients with more than one abnormal test result, the date of the most recent result was time zero, the start of follow-up.

Outcome and duration of follow-up

The outcome measure of interest was whether a patient received a colonoscopy within 6 months of receiving their abnormal FIT test result. For the Cox model, the outcome was determined if a colonoscopy was completed within in 180 days following the abnormal FIT test. Patients were not censored for loss to follow-up, as community clinics do not track membership. Completed colonoscopies were determined through procedure codes in the EHR.

Predictor characteristics

Statistical Analysis

We evaluated the characteristics predicting follow up colonoscopy using a Cox proportional hazards model and a logistic model in SAS® System Software. We fit a full model of patients with complete data, and used a step-down process to manually remove the weakest characteristics one covariate at a time to simplify the model so that the final model retained at least 90% of the variation explained of the full model.

For the final model, we calculated the mean observed risk of completing the colonoscopy and plotted mean observed and predicted risks in quintiles using risk predictiveness curves that showed the distribution of predicted risks of completing the colonoscopy.(22) Discrimination was measured by a bootstrap corrected C-statistic. Variation explained was measured with an $R^2$ statistic. The Cox regression coefficients were then translated into a simplified points -based risk scoring system to improve use in the clinical setting. A higher number of points mean a higher likelihood of completing a follow up colonoscopy (23). This allows clinicians to translate the model into practice without calculating the regression equation exactly. Table 3 shows the expected and observed probability of completing a colonoscopy within 6 months of an abnormal fit test by points. The points assignment reflects the
Table 3. Expected and observed probability of completing a colonoscopy within 6 months of abnormal FIT test, by points score

| Points Score* | Expected Probability, % | Observed probability, % | (95% CI) |
|---------------|-------------------------|-------------------------|---------|
| 140-149       | 17.30%                  | 18.00%                  | (7.0-29.0) |
| 150-159       | 19.20%                  | 14.50%                  | (6.4-22.6) |
| 160-169       | 21.30%                  | 18.10%                  | (10.2-26.0) |
| 170-179       | 23.60%                  | 27.90%                  | (19.8-35.9) |
| 180-189       | 26.10%                  | 23.60%                  | (16.1-31.1) |
| 190-199       | 28.80%                  | 28.80%                  | (21.5-36.0) |
| 200-209       | 31.80%                  | 33.80%                  | (25.9-41.8) |
| 210-219       | 34.90%                  | 40.70%                  | (32.3-49.1) |
| 220-229       | 38.30%                  | 34.60%                  | (26.4-42.8) |
| 230-239       | 41.90%                  | 39.60%                  | (30.2-49.1) |
| 240-249       | 45.70%                  | 45.20%                  | (34.9-55.5) |
| 250-259       | 49.60%                  | 54.70%                  | (43.1-66.2) |
| 260-269       | 53.70%                  | 54.30%                  | (44.0-64.5) |
| 270-279       | 57.90%                  | 46.40%                  | (33.0-59.9) |
| 280-289       | 62.20%                  | 66.70%                  | (51.2-82.1) |
| 290-299       | 66.50%                  | 58.30%                  | (37.1-79.6) |

*Probabilities calculated for point ranges with at least 20 patients

Results

Of 11,622 patients with a completed fecal test, 1,723 (14.8%) were abnormal, and 699 (40.6%) of those had a subsequent completed colonoscopy in their EHR record within 12 months (Figure 1). However, only 597 (34.6%) of those patients had record of a completed a colonoscopy within 6 months of their abnormal FIT test. For this analysis, one small clinic system was excluded due to low numbers of patients with abnormal FIT results (n=13). We also only included patients with non-missing data for all predictors (n=1,596). Of the 1,596 patients included in the final model, 34.8% (n=556) had recorded completed colonoscopies within 6 months.

Table 1 illustrates all baseline characteristics for the entire cohort and the subgroup that had a recorded completed colonoscopy within 6 months. Overall, patients were typically white (83.3%), aged 50-64 (81.5%) and had a low rate of preventive screenings: flu shots (14.3%); prior CRC screening (38.3%). Only eight variables were retained for the final model as they contributed to the explained variation in risk.
Table 1. Characteristics at baseline for all patients and patients with a colonoscopy

|                          | All Patients | With Colonoscopy | Univariate | Likelihood Ratio | p-value |
|--------------------------|--------------|------------------|------------|------------------|---------|
|                          | N (% of all) | N (% row)        | HR         |                  |         |
| All                      | 1596 100.00% | 556 34.80%       |            |                  | 0.0040  |
| **Age**                  |              |                  |            |                  |         |
| 50-54                    | 498 31.20%   | 200 40.20%       | ref        |                  |         |
| 55-59                    | 425 26.60%   | 156 36.70%       | 0.91       |                  |         |
| 60-64                    | 377 23.60%   | 122 32.40%       | 0.76       |                  |         |
| 65-69                    | 202 12.70%   | 62 30.70%        | 0.71       |                  |         |
| 70-75                    | 94 5.90%     | 16 17.00%        | 0.37       |                  |         |
| **Sex**                  |              |                  |            |                  | 0.4032  |
| Male                     | 757 47.40%   | 275 36.30%       | ref        |                  |         |
| Female                   | 839 52.60%   | 281 33.50%       | 0.9        |                  |         |
| **BMI**                  |              |                  |            |                  | 0.1467  |
| <24                      | 420 26.30%   | 143 34.10%       | ref        |                  |         |
| 25-29                    | 453 28.40%   | 149 32.90%       | 0.98       |                  |         |
| 30-34                    | 349 21.90%   | 137 39.30%       | 1.22       |                  |         |
| 35-39                    | 209 13.10%   | 65 31.10%        | 0.89       |                  |         |
| 40+                      | 165 10.30%   | 62 37.60%        | 1.15       |                  |         |
| **Language**             |              |                  |            |                  | 0.0780  |
| Non-English              | 312 19.60%   | 95 30.50%        | ref        |                  |         |
| English                  | 1284 80.50%  | 461 35.90%       | 1.26       |                  |         |
| **Race**                 |              |                  |            |                  | 0.0599  |
| Non-White                | 266 16.70%   | 72 27.10%        | ref        |                  |         |
| White                    | 1330 83.30%  | 484 36.40%       | 1.48       |                  |         |
| **Ethnicity**            |              |                  |            |                  | 0.0270  |
| Non-Hispanic             | 1445 90.50%  | 496 34.30%       | ref        |                  |         |
| Hispanic                 | 151 9.50%    | 60 39.70%        | 1.21       |                  |         |
| **Insurance**            |              |                  |            |                  | 0.5652  |
| Uninsured                | 265 16.60%   | 86 32.50%        | ref        |                  |         |
| Medicaid                 | 748 46.90%   | 282 37.70%       | 1.18       |                  |         |
| Medicare                 | 435 27.30%   | 136 31.30%       | 0.94       |                  |         |
| **Commercial**           | 148 9.30%    | 52 35.10%        | 1.1        |                  |         |
| Tobacco Use              |              |                  |            |                  | 0.9812  |
| Never/Quit               | 1153 72.20%  | 393 34.10%       | ref        |                  |         |
| Current User             | 443 27.80%   | 163 36.80%       | 1.11       |                  |         |
| Percent of Census Tract with College Degree | | | | | 0.0697 |
| 4.9 - 14.6               | 346 21.70%   | 125 36.10%       | ref        |                  |         |
| 14.7 - 19.9              | 337 21.10%   | 96 28.50%        | 0.73       |                  |         |
| 19.9 - 25.7              | 324 20.30%   | 125 38.60%       | 1.08       |                  |         |
| 26.0 - 36.8              | 282 17.70%   | 94 33.30%        | 0.88       |                  |         |
| 36.9 - 77.7              | 307 19.20%   | 116 37.80%       | 1.04       |                  |         |
| Percent of Census Tract Households below FPL | | | | | 0.2315 |
| 2.7 - 11.4               | 283 17.70%   | 105 37.10%       | ref        |                  |         |
| Income Range | Census Tract Median Household Income | $14,000 - $36,000 | $36,000 - $41,000 | $41,000 - $47,000 | $47,000 - $56,000 | $56,000 - $149,000 |
|--------------|-------------------------------------|--------------------|------------------|--------------------|------------------|-------------------|
| 11.4 - 14.8  | 288 18.10% 100 34.70% 0.91         | 331 20.70% 101 30.50% ref |
| 14.9 - 19.4  | 309 19.40% 105 34.00% 0.89         | 330 20.70% 125 37.90% 1.32 |
| 19.5 - 25.8  | 333 20.90% 131 39.30% 1.05         | 286 17.90% 101 35.30% 1.18 |
| 26.1 - 53.9  | 383 24.00% 115 30.00% 0.75         |

| Census Tract Median Household Income | 0.6530 |
|-------------------------------------|--------|
| $14,000 - $36,000                   | 331 20.70% 101 30.50% ref |
| $36,000 - $41,000                   | 330 20.70% 125 37.90% 1.32 |
| $41,000 - $47,000                   | 353 22.10% 117 33.10% 1.11 |
| $47,000 - $56,000                   | 286 17.90% 101 35.30% 1.18 |
| $56,000 - $149,000                  | 296 18.60% 112 37.80% 1.33 |

| Census Tract Unemployment | 0.0009 |
|---------------------------|--------|
| 2.6-8.1                   | 323 20.20% 132 40.90% ref |
| 8.2-10.2                  | 285 17.90% 83 29.10% 0.65 |
| 10.2-12.7                 | 293 18.40% 88 30.00% 0.68 |
| 12.7-15                   | 397 24.90% 146 36.80% 0.87 |
| 15.0-32.4                 | 298 18.70% 107 35.90% 0.84 |

| Census Tract Population Density (People per square mile of land area) | 0.2521 |
|-----------------------------------------------------------------------|--------|
| 0.8 - 174                                                              | 238 14.90% 96 40.30% ref |
| 176 - 1,571                                                            | 217 13.60% 62 28.60% 0.67 |
| 1,574 - 3,770                                                          | 289 18.10% 83 28.70% 0.64 |
| 3,781 - 6,576                                                          | 358 22.40% 144 40.20% 0.98 |
| 6,593 - 26,873                                                         | 494 31.00% 171 34.60% 0.8 |

| Census Tract GINI Income Inequality | 0.4162 |
|-------------------------------------|--------|
| 0.27 - 0.38                         | 329 20.60% 102 31.00% ref |
| 0.38 - 0.41                         | 326 20.40% 115 35.30% 1.15 |
| 0.41 - 0.43                         | 326 20.40% 122 37.40% 1.24 |
| 0.43 - 0.47                         | 256 16.00% 97 37.90% 1.26 |
| 0.47 - 0.82                         | 359 22.50% 120 33.40% 1.09 |

| Low access Census Tract at 1/2 mile for urban areas or 5 miles for rural areas | 0.8152 |
|-------------------------------------------------------------------------------|--------|
| No                                                                             | 309 19.40% 121 39.20% ref |
| Yes                                                                            | 1287 80.60% 435 33.80% 0.81 |

| Emergency Room Visits per 1,000 Medicare Enrollees (County) | 0.7264 |
|-------------------------------------------------------------|--------|
| 0                                                           | 356 22.30% 107 30.10% ref |
| 1                                                           | 914 57.30% 337 36.90% 1.24 |
| 2+                                                          | 326 20.40% 112 34.40% 1.16 |

| Urban/Rural County Cluster (10-50k population) | 0.7809 |
|-----------------------------------------------|--------|
| Rural (<10K population)                       | 246 15.40% 97 39.40% 1.39 |
| Urban (50k+ population)                       | 1074 67.30% 375 34.90% 1.16 |

| Charlson Comorbidity                          | 0.7870 |
|------------------------------------------------|--------|
|                                      | Count | 95% CI       | ref Count | 95% CI       | Reference Value |
|--------------------------------------|-------|--------------|-----------|--------------|-----------------|
| Asthma/COPD dx in 2 years prior to index |       |              |           |              | 0.1816          |
| No                                   | 1122  | 70.30%       | 404       | 60.00%       | ref             |
| Yes                                  | 474   | 29.70%       | 152       | 40.00%       | 0.87            |
| Diabetes dx in 2 years prior to index |       |              |           |              | 0.2072          |
| No                                   | 881   | 55.20%       | 322       | 60.00%       | ref             |
| Yes                                  | 715   | 44.80%       | 234       | 40.00%       | 0.86            |
| Severe mental illness                |       |              |           |              | 0.7889          |
| No                                   | 1455  | 91.20%       | 504       | 80.00%       | ref             |
| Yes                                  | 141   | 8.80%        | 52        | 20.00%       | 1.09            |
| Mood disorder (Depression, Bipolar) dx in 2 years prior to index |       |              |           |              | 0.6492          |
| No                                   | 1006  | 63.00%       | 342       | 60.00%       | ref             |
| Yes                                  | 590   | 37.00%       | 214       | 40.00%       | 1.1             |
| Substance/alcohol abuse dx in 2 years prior to index |       |              |           |              | 0.6928          |
| No                                   | 1264  | 79.90%       | 434       | 80.00%       | ref             |
| Yes                                  | 332   | 20.80%       | 122       | 20.00%       | 1.14            |
| Long term anticoagulant use          |       |              |           |              | 0.0353          |
| No                                   | 1545  | 96.80%       | 546       | 95.00%       | ref             |
| Yes                                  | 51    | 3.20%        | 10        | 5.00%        | 0.5             |
| Blood in Stool prior to abnormal FIT |       |              |           |              | 0.3026          |
| No                                   | 1538  | 96.40%       | 538       | 95.00%       | ref             |
| Yes                                  | 58    | 3.60%        | 18        | 5.00%        | 0.86            |
| Hemorrhoid/Anal Fissure prior to abnormal FIT |       |              |           |              | 0.3546          |
| No                                   | 1514  | 94.90%       | 526       | 95.00%       | ref             |
| Yes                                  | 82    | 5.10%        | 30        | 5.00%        | 1.08            |
| Prior CRC screening                  |       |              |           |              | 0.2966          |
| No                                   | 985   | 61.70%       | 362       | 60.00%       | ref             |
| Yes                                  | 611   | 38.30%       | 194       | 40.00%       | 0.82            |
| Flu shot within 1 year of index date |       |              |           |              | 0.0000          |
| No                                   | 1368  | 85.70%       | 452       | 85.00%       | ref             |
| Yes                                  | 228   | 14.30%       | 104       | 15.00%       | 1.57            |
| Number of outpatient encounters in year prior to index date |       |              |           |              | 0.3248          |
| 0                                    | 203   | 12.70%       | 85        | 11.90%       | ref             |
| 1                                    | 173   | 10.80%       | 52        | 10.10%       | 0.65            |
| 2                                    | 196   | 12.30%       | 59        | 10.10%       | 0.64            |
| 3                                    | 209   | 13.10%       | 80        | 13.20%       | 0.86            |
The eight characteristics retained in the final Cox regression model included age, race, insurance, GINI income inequality, long term anticoagulant use, receipt of a flu vaccine in the past year, frequency of missed clinic appointments and clinic size (Table 2.). No notable differences were determined when the model was run for men and women separately, so therefore we combined men and women to develop one model. Table 2 also shows hazard ratios, confidence intervals, and number of risk points assigned to each characteristic. The hazard ratios and risk score points for the final prediction model indicated that health center, age, long term anti-coagulant use, and receipt of a flu vaccine in the past year were the variables with highest points assigned in the model.
| Variable                                  | Hazard Ratio | (95% CI)     | Likelihood ratio p-value | Points |
|-------------------------------------------|--------------|---------------|--------------------------|--------|
| **Age**                                   |              |               |                          |        |
| 50-54                                     | ref          |               |                          | 83     |
| 55-59                                     | 0.92         | (0.74 - 1.13) |                          | 76     |
| 60-64                                     | 0.76         | (0.61 - 0.96) |                          | 60     |
| 65-69                                     | 0.76         | (0.55 - 1.04) |                          | 59     |
| 70-75                                     | 0.38         | (0.22 - 0.65) |                          | 0      |
| **Race**                                  |              |               |                          | 0.0019 |
| Non-White                                 | ref          |               |                          | 0      |
| White                                     | 1.48         | (1.14 - 1.91) |                          | 34     |
| **Insurance**                             |              |               |                          | 0.5174 |
| Uninsured                                 | ref          |               |                          | 3      |
| Medicaid                                  | 1.15         | (0.90 - 1.48) |                          | 15     |
| Medicare                                  | 1.03         | (0.77 - 1.38) |                          | 5      |
| Commercial                                | 0.97         | (0.67 - 1.40) |                          | 0      |
| **Census Tract GINI Income Inequality**   |              |               |                          | 0.4446 |
| 0.27 - 0.38                               | ref          |               |                          | 0      |
| 0.38 - 0.41                               | 1.14         | (0.87 - 1.49) |                          | 11     |
| 0.41 - 0.43                               | 1.17         | (0.90 - 1.53) |                          | 14     |
| 0.43 - 0.47                               | 1.25         | (0.94 - 1.66) |                          | 19     |
| 0.47 - 0.82                               | 1.27         | (0.97 - 1.67) |                          | 21     |
| **Long term anticoagulant use**           |              |               |                          | 0.0315 |
| No                                        | ref          |               |                          | 54     |
| Yes                                       | 0.54         | (0.29 - 1.01) |                          | 0      |
| **Flu shot within 1 year of index date**  |              |               |                          | 0.0001 |
| No                                        | ref          |               |                          | 0      |
| Yes                                       | 1.59         | (1.28 - 1.98) |                          | 40     |
| **Count of no-show encounters in year prior to index date** |           |               |                          | 0.0151 |
| 0                                         | ref          |               |                          | 31     |
The mean predicted risk of completion of colonoscopy was 34.8%, and the model was able to accurately predict the patients who were least likely to receive a follow-up colonoscopy (lowest two quintiles, 15.9% and 28.5% respectively). Likelihood of obtaining a follow-up colonoscopy within 6 months varied across quintiles: patients with the highest predicted risk of non-adherence (bottom quintile) had an estimated 16% chance of obtaining a colonoscopy; whereas, patients with the lowest predicted risk of non-adherence (top quintile) had a greater than 55% chance of obtaining a follow-up colonoscopy. Figure 2 shows the predictiveness curve for colonoscopy completion. The open circles are the observed proportions (o) and the line represents the predicted probability of colonoscopy completion.

Risk score points can be assigned to a patient to determine their risk of completing a colonoscopy. For example, we can score a patient who is on Medicaid (15 points), white (34 points), 54 years old (83 points), receives his care at health center 3 (100 points), has not missed appointments (31 points), has received a flu shot (40 points), isn't on anticoagulants (54 points) and lives in an area with low income inequality (21 points). His total point count is 378, which predicts that he has an 81% probability of completing a colonoscopy, compared to the 35% likelihood of the average patient (data not shown).

The model showed modest separation of patients across risk levels for non-adherence to follow-up colonoscopy (C-statistic>0.66, bootstrap-corrected C-statistic>0.63) and excellent calibration or high agreement between observed and predicted risk. The R2 statistic, derived from the D-statistic, showed only 14% of the variation in outcome was explained in this model (R2 (95% CI) =14.03 (10.17-18.18), D (95% CI)=0.83 (0.69-0.96)). A logistic regression, predicting the completion of a colonoscopy, showed similar results for non-adherence to follow-up colonoscopy (C-statistic=0.66, bootstrap-corrected C-statistic>0.64).

Discussion
This model was created to identify patients at the greatest need for targeted interventions, such as patient navigation, to complete the screening process for CRC. We recognize that the performance of the model has limitations. The C-statistic, while suboptimal, shows adequate separation of patients across risk levels for non-adherence to follow-up colonoscopy, yet the $R^2$ indicates the discrimination and calibration could be further improved. However, focusing efforts for improving follow-up colonoscopy among patients in the lowest quintiles could provide value in the population most in need for understanding the importance of follow-up. Identifying the barriers among these patients and targeting interventions could produce improvements. When putting the model into practice, targeting the lowest probability groups could result in the greatest improvements.

Both patient and system level barriers were used in the final model, indicating the importance of recognizing multilevel barriers in adherence to colonoscopy following an abnormal FIT. The strongest predictors were age, health center, anticoagulant use, and flu shot vaccination. The youngest patients were more likely to obtain colonoscopies, which may be explained by newer entry into screening eligibility and they are healthier. However, recognizing lower likelihood of follow-up screening among older patients could help in efforts to close the gap in colonoscopies. Patients’ likelihood of receiving colonoscopy varied by health center, and this is a complex variable that could represent a variety of systems and patient level factors. While the system level factors impacting colonoscopy completion can include access to colonoscopies, location, and community characteristics, the referral process, scheduling, wait lists and capacity, it is also affected by many patient level barriers reflected in the health system like transportation barriers, inability to take time off work and mistrust of the system. The health system variable is important in this model and should be further examined.

Patients indicated as users of anti-coagulant medications were less likely to obtain a colonoscopy following an abnormal FIT. Removing a patient from an anti-coagulant may be required prior to a colonoscopy and could be a deterrent for completing the test. Patients who were vaccinated for the flu in the past year were more likely to complete colonoscopy, indicating a trend of compliance for recommended preventative care. Patients with Medicaid insurance, a free or low-cost health coverage for low-income patients, were more likely to complete the colonoscopy. This, which may indicate program effectiveness or the effect of removing cost barriers. The GINI Income Inequality Ratio is a statistical measure of income inequality where a measure of 1 indicates total inequality and a measure of 0 indicates total equality. This analysis shows that living in an area of inequity (closer to 1.0) decreases one's chances of completing a colonoscopy. Finally, patients who were non-white were less likely to complete a colonoscopy. Addressing issues among non-white populations through interventions is necessary to close the gap in disparities in CRC screening.

We believe this is the first model to predict likelihood of follow-up after an abnormal fecal test. Further research is needed to test the effectiveness of interventions for patients who have a low and moderate probability of completing follow-up colonoscopy.

**Limitations**
There are known deficiencies in capturing completed colonoscopies and referrals in the EHR, and especially in the community clinic setting where most colonoscopies are referred to outside providers or specialties without a direct link to the EHR. Therefore, updating records relies on clinic processes. While analyzing chart abstracted colonoscopies could be the gold standard, it only explains why patients may or may not have colonoscopies recognized by the physician in obscure data points in the EHR (24). This population is a primarily FQHC population and is therefore not generalizable to patients who obtain care in other types of clinics or healthcare settings. Further, this population is primarily in Oregon and Northern California, indicating regional limitations to generalization. Other populations and settings may not have EHR records to capture the predictors in our model. The health center variable is highly collinear with the other variables except age. System-level predictors may matter more than patient-level predictors and these models may need to be validated for each population and setting where they will be put into practice. We sought to develop a model that will transport to other health systems, clinics, and populations. The validity of the analysis would be increased with external validation and could support widespread use.

Conclusions

Understanding the differences in patients who are more likely to complete colonoscopy, can lead to tailored outreach to patients in need of interventions. Doing so will target resources, reduce disparities, and save lives.

List Of Abbreviations

CRC Colorectal cancer
ED emergency department
EHR electronic health record
FIT fecal immunochemical test
FQHC Federally Qualified Health Center
STOP CRC Strategies and Opportunities to STOP Colon Cancer in Priority Populations

Declarations

Ethics approval and consent to participate

This research project was approved by the Kaiser Permanente Northwest IRB (KPNW IRB, Protocol #4364). Participating clinic systems and collaborating organizations ceded human subjects review authority to the KPNW IRB. A waiver of informed consent was granted for this project.
Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number UH3CA188640. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The study sponsor had no role in study design; collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication.

Authors' Contributions

All authors contributed significantly to this manuscript. EK developed datasets and analyzed all data, EJ and DS directed, interpreted and lead the development of the risk model, GC acted as principal investigator, and provided context and application to the clinical settings, and AP participated in all of the above and led the writing of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests at this time.

Consent to publish

Not applicable

Acknowledgements

Not applicable

References

1. Gupta S, Halm EA, Rockey DC, Hammons M, Koch M, Carter E, et al. Comparative effectiveness of fecal immunochemical test outreach, colonoscopy outreach, and usual care for boosting colorectal cancer screening among the underserved: a randomized clinical trial. JAMA internal medicine. 2013;173(18):1725-32.

2. Coronado GD, Golovaty I, Longton G, Levy L, Jimenez R. Effectiveness of a clinic-based colorectal cancer screening promotion program for underserved Hispanics. Cancer. 2011;117(8):1745-54.

3. Singal AG, Gupta S, Tiro JA, Skinner CS, McCallister K, Sanders JM, et al. Outreach invitations for FIT and colonoscopy improve colorectal cancer screening rates: A randomized controlled trial in a safety-net health system. Cancer. 2016;122(3):456-63.
4. Coronado GD, Burdick T, Petrik A, Kapka T, Retecki S, Green B. Using an Automated Data-driven, EHR-Embedded Program for Mailing FIT kits: Lessons from the STOP CRC Pilot Study. J Gen Pract (Los Angel). 2014;2.

5. Oluloro A, Petrik AF, Turner A, Kapka T, Rivelli J, Carney PA, et al. Timeliness of Colonoscopy After Abnormal Fecal Test Results in a Safety Net Practice. Journal of community health. 2016;41(4):864-70.

6. Martin J, Halm EA, Tiro JA, Merchant Z, Balasubramanian BA, McCallister K, et al. Reasons for Lack of Diagnostic Colonoscopy After Positive Result on Fecal Immunochemical Test in a Safety-Net Health System. Am J Med. 2017;130(1):93.e1-.e7.

7. Issaka RB, Singh MH, Oshima SM, Laleau VJ, Rachocki CD, Chen EH, et al. Inadequate Utilization of Diagnostic Colonoscopy Following Abnormal FIT Results in an Integrated Safety-Net System. The American journal of gastroenterology. 2017;112(2):375-82.

8. Yu C, Skootsky S, Grossman M, Garner OB, Betlachin A, Esrailian E, et al. A Multi-Level Fit-Based Quality Improvement Initiative to Improve Colorectal Cancer Screening in a Managed Care Population. Clinical and translational gastroenterology. 2018;9(8):177.

9. Singal AG, Gupta S, Skinner CS, Ahn C, Santini NO, Agrawal D, et al. Effect of Colonoscopy Outreach vs Fecal Immunochemical Test Outreach on Colorectal Cancer Screening Completion: A Randomized Clinical Trial. JAMA. 2017;318(9):806-15.

10. Jetelina KK, Yudkin JS, Miller S, Berry E, Lieberman A, Gupta S, et al. Patient-Reported Barriers to Completing a Diagnostic Colonoscopy Following Abnormal Fecal Immunochemical Test Among Uninsured Patients. Journal of general internal medicine. 2019.

11. Liss DT, Brown T, Lee JY, Altergott M, Buchanan DR, Newland A, et al. Diagnostic colonoscopy following a positive fecal occult blood test in community health center patients. Cancer Causes Control. 2016;27(7):881-7.

12. Chubak J, Garcia MP, Burnett-Hartman AN, Zheng Y, Corley DA, Halm EA, et al. Time to Colonoscopy after Positive Fecal Blood Test in Four U.S. Health Care Systems. Cancer Epidemiol Biomarkers Prev. 2016;25(2):344-50.

13. Meester RG, Zauber AG, Doubeni CA, Jensen CD, Quinn VP, Helfand M, et al. Consequences of Increasing Time to Colonoscopy Examination After Positive Result From Fecal Colorectal Cancer Screening Test. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2016;14(10):1445-51.e8.

14. Lamanna A, Sheaffer H, Guerra C, Kochman M. Colorectal Cancer Screening Navigation for the Underserved: Experience of an Urban Program. Gastroenterol Hepatol (N Y). 2016;12(9):547-51.

15. Green BB, Anderson ML, Wang CY, Vernon SW, Chubak J, Meenan RT, et al. Results of Nurse Navigator Follow-up After Positive Colorectal Cancer Screening Test: A Randomized Trial. Journal of the American Board of Family Medicine: JABFM. 2014;27(6):789-95.

16. Escoffery C, Fernandez ME, Vernon SW, Liang S, Maxwell AE, Allen JD, et al. Patient Navigation in a Colorectal Cancer Screening Program. Journal of Public Health Management & Practice.
17. Parikh RB, Kakad M, Bates DW. Integrating Predictive Analytics Into High-Value Care: The Dawn of Precision Delivery. JAMA. 2016;315(7):651-2.

18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Medicine. 2015;13(1):1.

19. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine. 2015;162(1):W1-73.

20. Coronado GD, Vollmer WM, Petrik A, Taplin SH, Burdick TE, Meenan RT, et al. Strategies and Opportunities to STOP Colon Cancer in Priority Populations: design of a cluster-randomized pragmatic trial. Contemp Clin Trials. 2014;38(2):344-9.

21. Mehta SJ, Jensen CD, Quinn VP, Schottinger JE, Zauber AG, Meester R, et al. Race/Ethnicity and Adoption of a Population Health Management Approach to Colorectal Cancer Screening in a Community-Based Healthcare System. Journal of general internal medicine. 2016;31(11):1323-30.

22. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, et al. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol. 2008;167(3):362-8.

23. Harrell F. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer-Verlag; 2001.

24. Petrik A, Green BB, Vollmer WM, Le T, Bachman B, Keast E, et al. The Validation of Electronic Health Records in Accurately Identifying Patients Eligible for Colorectal Cancer Screening in Safety Net Clinics. Family Practice. 2016;33(6):639-43.

Figures
**Figure 1**

Risk Model and Patient Population from STOP CRC

- **Total Population**
  - Patients with a FIT test completed 2/4/2014 - 2/28/16
  - $N = 11,622$

- **Model Population**
  - Patients excluded from the total population who were seen at clinics too small to contribute and who had missing data

  - Patients with an abnormal FIT test included in the model
    - $N = 1,596$

  - Patients with a colonoscopy within 6 months of an abnormal FIT test
    - $N = 556$

  - Patients with a colonoscopy within 12 months of an abnormal FIT test
    - $N = 1,723$

  - Patients with a colonoscopy within 6 months of an abnormal FIT test
    - $N = 597$
Figure 2

Observed and Predicted Probability Colonoscopy Completion

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- APPENDIXSTOPriskscore20171017rd.doc
- TRIPODChecklistv220190731.docx