Development of Risk Prediction Models Combining Routine EHR Data for Colorectal Cancer Screening Referral Decisions

Jennifer A Cooper1, Prof. Tom Marshall2, Dr. Ronan Ryan2, Dr. Nick Parsons1, Dr. Chris Stinton1, Dr. Sian Taylor-Phillips1
1Department of Health Sciences, Gibbet Hill Campus, The University of Warwick, Coventry, CV4 7AL, UK
2Institute of Applied Health Research, University of Birmingham, West Midlands, B15 2TT, UK

METHAB Theme: eHealth and mHealth for diagnosis, prognosis and monitoring

Background

• Risk prediction models incorporating both faecal occult blood test (FOBT) results and other colorectal cancer (CRC) risk factors have demonstrated increased sensitivity than FOBT alone.1,2
• Electronic primary care records have a rich level of data including; symptoms, diagnoses, lab test results and prescriptions which may add a further dimension to a risk based prediction model.
• The BCSS receives data for its participants from the NHS Spine which houses demographic information drawn from GP records. There is capacity to draw further data from the Spine/GP records to improve screening referral decisions (Figure 1).
• This study aimed to determine:
  (i) the availability of GP data for key predictors of CRC in the screening population. 
  (ii) whether this additional information can inform more accurate screening referral decisions in future risk prediction models.

Results

• Multivariable model - The Cox Regression model including FOBT results (n=98,303, 1197 events) had 13 predictors and 2 interactions.
• The model for negative FOBT patients (n=95,792, 587 events) is given in Table 1.
• Predictors retained in both models: age, sex, smoking, MCV, family history GI cancer, previous polyps, abdominal pain, diarrhoea, flatulence & change in bowel habit.
• Model performance - Table 2 gives the optimism adjusted performance metrics.
• Results for the negative FOBT model: Distribution of the LP (Figure 3), discrimination also assessed by analyzing separation between KM curves for 4 risk groups (Figure 4).
• Baseline survival after shrinkage (0.988) was estimated to give absolute risk probabilities for each individual (Equation 1) (Van Houwelingen’s heuristic shrinkage).
• A Calibration plot is given for deciles of risk (Figure 5).
• Gompertz parametric model best fit for AIC (7497.7), cumulative hazard and KM plots.

Table 1: Variable Odds Coefficient (95% CI) P Value (P>|z|)
Smoking Status (baseline never smoked): 0.285
Ex-smoker: 0.258
Current smoker: 0.953
IBS: 0.901
Previous polyps: 1.225
Abdominal pain: 0.944
Ovarian polyps: 0.000
Metabolic syndrome: 0.004
Family history GI cancer: 0.015
Pattern of smoking: 0.000
Smokers: 0.000
Ovarian polyps: 0.000
Abdominal pain: 0.000
Age at FOBT: 1.089
Continuous: 0.000
Change in bowel habit: 0.004

Table 2: Optimism adjusted performance metrics

Model including FOBT results

<table>
<thead>
<tr>
<th>Statistic</th>
<th>C statistic</th>
<th>c-Slope</th>
<th>D statistic</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOBT patients</td>
<td>0.850</td>
<td>0.991</td>
<td>2.298</td>
<td>0.558</td>
</tr>
<tr>
<td>FOBT patients</td>
<td>0.650</td>
<td>0.944</td>
<td>0.836</td>
<td>0.144</td>
</tr>
</tbody>
</table>

Figure 1: Data schematic of the BCSS and interlinks with other data services: BCSS: Bowel Cancer Screening System.

Figure 2: Rate of electronic BCSP notifications received. Blue line: Actual rate of BCSP notifications. Red line: Expected rate of notifications. Green line: LOSAT to assess the overall trend. AEB data: 1st September 2012.

Figure 3: Distribution of linear predictor (Mean 0.15, SD 0.45).

Figure 4: Separation of Kaplan-Meier curves for 4 risk groups based on the linear predictor.

Conclusions

• This research has identified predictors which could be considered for inclusion in a future risk adjusted screening model.
• Additional data could be drawn onto the BCSS to contribute to a referral algorithm.
• Calculating individual risk can help referral decisions as well as patients and screening practitioners make a more informed choice.
• Strengths: Large sample size, internal validation to quantify optimism, development of the AEB date for data quality assurance.
• Limitations Sample size reduced due to complete variables (not MAR), recording of cancer could be enhanced using cancer registry data, relies on recorded symptoms.
• Future Research Similar analyses with the FIT & SNOMED codes. Inclusion of more lab results, and screening adherence.

References


Figure 3: Kaplan-Meier survival estimates

Figure 4: Calibration plot for negative FOBT model. Table 1: Cox Regression model for negative FOBT patients (age centred 66.77, df 39, EPV = 15.1). Table 2: Optimism adjusted performance for both models. Equation 1: Equation for absolute risk prediction.

Data Collection - Large database of electronic primary care records (The Health Improvement Network, THIN).
• Cohort derived in part by identifying an Acceptable Electronic BCSP (AEB) date for each practice in England for participants aged 60-74.
• Risk Factor/Predictor Information – >30 clinical predictors were developed using Read code, drug code lists and AHD strategies. Combined outcome was CRC and polyps.
• Completeness of variables and univariable analysis was investigated for this cohort.
• Statistical Analysis – A prediction model combining the FOBT with other clinical predictors was developed using Cox Regression and multivariable fractional polynomials with backwards elimination (‘mfp’ Function in Stata, p<0.05).
• For internal validation, optimism adjusted performance metrics were determined using bootstrapping. Absolute risk predictions generated by estimating the baseline survival.
• Analysis was repeated for negative FOBT patients to assess whether other predictors could still warrant screening referral despite a negative result.

Figure 5: Calibration plot for negative FOBT model. Table 1: Cox Regression model for negative FOBT patients (age centred 66.77, df 39, EPV = 15.1). Table 2: Optimism adjusted performance for both models. Equation 1: Equation for absolute risk prediction.

Figure 5: Calibration plot for negative FOBT model. Table 1: Cox Regression model for negative FOBT patients (age centred 66.77, df 39, EPV = 15.1). Table 2: Optimism adjusted performance for both models. Equation 1: Equation for absolute risk prediction.