Development and validation of risk prediction models with plasma biomarkers for colorectal cancer

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Zhe Fang[1], Dong Hang[2,3], Kai Wang[1], Amit Joshi[4], Kana Wu[2], Andrew T. Chan[4,5,6,7], Shuji Ogino[1,7,8], Edward L. Giovannucci[1,2,6], Mingyang Song[1,2,4,5]

Background

- Identification of colorectal cancer (CRC) at an early stage (i.e., advanced polyp) and interrupting the natural history is important to lower incidence and prolong survival[1, 2]. Stratifying the population into risk categories can potentially improve the efficiency of screening and facilitate the development of personalized interventional approaches for CRC.
- Risk prediction models for CRC are not widely used in clinical practice yet. There are a range of existing models mainly based on questionnaire information, which have shown a modest discriminatory ability[3].
- There is an urgent need to examine the added value of plasma biomarkers for CRC prediction. So far, several groups of biomarkers have been implicated in CRC and may aid early detection, including those related to inflammation, metabolic disturbances, sex hormones, and vitamin D.

Objectives

To examine whether adding a set of biomarkers (**CRP, IL6, TNFRSF1B, GDF15, ADIPOQ, LEPR, C-peptide, IGF1, IGFBP3, SHBG, and 25(OH)D**) can improve the performance of risk prediction model for CRC that is only based on questionnaire information (National Cancer Institute Risk Assessment Tool for CRC)

Methods

- **Data source**: Nurses' Health Study, Health Professionals Follow-up Study
- **Inclusion criteria**: had at least one studied biomarker

  Exclusion criteria: a) had any missing studied biomarkers
  b) had any outliers of studied biomarkers

Phase A Model Building

Training Set
HPFS N=736
(253 cases/483 controls)
NHS N=639
(308 cases/331 controls)

Basic Model: NCI model
Men: age, BMI, smoking, vigorous activity, NSAIDs use, family history of CRC, vegetable intake, endoscopy and polyp history
Women: HRT and menopausal status were included instead of smoking

Select biomarkers from 11 biomarker candidates
Using **stepwise selection** with
entry level P=0.15
stay level P=0.2

Phase B Internal Validation

Validation Set
500 bootstrap-generated random samples with the sample size as training set

Developed Model: NCI + biomarkers model

- **Model Evaluation**

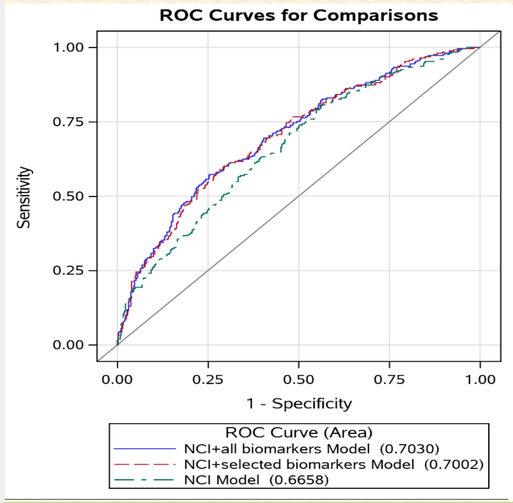
a) **Discrimination** → C-statistics
b) The improvement in the C-statistics → Mann-Whitney U test
c) The improvement in the **reclassification** → IDI, NRI

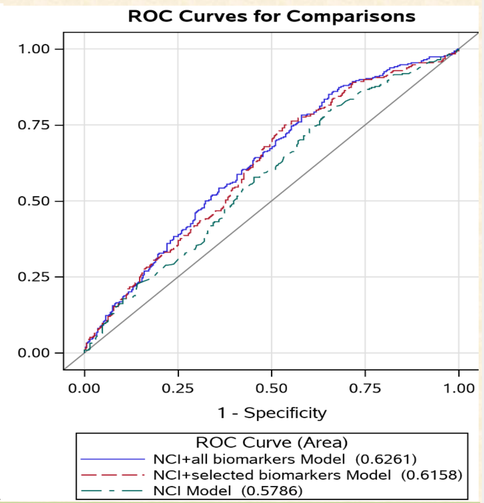
- **Secondary analysis**

a) Investigating whether adding a genetic risk score (GRS) for CRC into NCI + biomarkers model can further improve model performance
b) Utilizing genotype data from the Genetic and Epidemiology of CRC Consortium (GECCO) among participants in our study with available GRS

$$GRS = \left(\sum_{i=1}^{39} \beta_i \times SNP_i\right) \times \left(39 / \sum_{i=1}^{39} \beta_i\right)$$

Key Findings





Model	Men (HPFS)		Women (NHS)	
	Training dataset	Validation dataset	Training dataset	Validation dataset
NCI	0.67(0.62-0.71)	0.69(0.65-0.72)	0.58(0.54-0.63)	0.61(0.56-0.65)
NCI + stepwise selected biomarkers	0.70(0.66-0.74)	0.73(0.69-0.77)	0.62(0.57-0.66)	0.66(0.61-0.70)
P value	0.008	/	0.06	/
NCI + all biomarkers	0.70(0.66-0.74)	0.73(0.70-0.77)	0.63(0.58-0.67)	0.66(0.62-0.70)
P value	0.53	/	0.30	/

For men, **GDF15, ADIPOQ, TNFRSF1B, and SHBG** were selected into the model;
For women, **IGF1 and IGFBP3** were selected.

	Men (HPFS)			Women (NHS)		
	NCI Model	NCI + selected biomarkers Model		NCI Model	NCI + selected biomarkers Model	
	Values	Values	P value	Values	Values	P value
Category-free NRI (%)	Ref	35.7(20.7,50.7)	<0.0001	Ref	25.2(9.8,40.6) (13.0-43.7)	0.001
IDI (95% CI)	Ref	0.036 (0.021,0.050)	<0.0001	Ref	0.019(0.009,0.030) (0.011-0.034)	0.0004

Model	Men (HPFS)	Women (NHS)
NCI + GRS	0.70 (0.66-0.74)	0.58 (0.53-0.63)
NCI + biomarkers	0.71 (0.67-0.76)	0.59 (0.54-0.64)
NCI + biomarkers + GRS	0.73 (0.69-0.77)	0.60 (0.55-0.65)
P value [a]	0.06	0.36

Conclusions

- Adding a set of plasma biomarkers to the risk factor-based NCI model improved the discriminatory accuracy and reclassification.

- **Public Health Relevance**

a) Knowledge about individuals' risk has been shown to affect screening behaviors over time and be utilized to select screening modalities.
b) The components of biomarkers that can be easily assessed in clinical practice and embedded within the electronic health record system allow the risk prediction to be linked to the clinical, further advancing the uptake and efficiency of screening.

- **Future directions**

a) External validation of risk models incorporating plasma biomarkers for CRC
b) How risk models for CRC stratify risk within the general population
c) The impact of risk models on screening programs in certain scenarios where each person begins screening when their individual risk of CRC reaches a predefined threshold

References

1. Lindholm E, Brevinge H, Haglind E. Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. British Journal of Surgery. 2008;95(8):1029-36.
2. Mandel JS, Church TR, Bond JH, et al. The effect of fecal occult-blood screening on the incidence of colorectal cancer. New England Journal of Medicine. 2000;343(22):1603-7.
3. Smith T, Muller DC, Moons KGM, et al. Comparison of prognostic models to predict the occurrence of colorectal cancer in asymptomatic individuals: a systematic literature review and external validation in the EPIC and UK Biobank prospective cohort studies. Gut. 2019;68(4):672-83.