Supplementary appendix

In this supplement, we describe the selection process used to determine the final model specification reported in the main paper. As explained in the Methods text, we evaluated different model specifications in terms of a) statistical tests for overall model fit, b) the predictive performance of the model (calibration, discrimination), and c) the clinical face-validity.

The evaluated model specifications are listed in the **Supplementary Table**.

Overall *model fit* was statistically evaluated for all 12 specifications. Statistical tests were inconclusive on the best model, particularly in non-nested model comparisons. In most cases, both models in such comparisons added significant value over the other. The tests did suggest interaction terms to be of limited value (not statistically significant).

To assess *predictive performance*, first, the degree of discrimination was examined for all 12 specifications. Discrimination was similar for specifications 2-3 and 5-8, and superior compared with other model specifications (**Supplementary Table 1**). All these models yielded *C*-statistics within ±0.005 from 0.785 for AN and 0.737 for CRC. Corrections for optimism were not derived for all models, but these were of negligible order for the models included in the main paper. We conclude that, for discrimination, the inclusion of all available information is more important than the form in which it is included (categorical *vs*. continuous, with *vs*. without transformation). In models with categoric F-Hb variables, measured concentrations between 0 μg/g and the *limit of detection* were suggestive of adverse outcomes, despite the potential conflation with 0 μg/g for some participants.

Calibration was evaluated for only four of the models with the highest concordance statistics: specifications 2, 3, 5 and 8. In the absence of suitable objective criteria to compare the calibration curves, we used our subjective judgment to select specifications 2 and 3 as the most appropriate, despite the overestimated risk for participants in the highest riskscore percentile.

Finally, riskscore charts were examined to assess *clinical face validity*. Particularly, we assessed whether these charts demonstrated expected patterns of higher predicted risk for older *vs*. younger adults, for men *vs*. women, and for participants with higher *vs*. lower F-Hb concentrations. Specification 3 was considered the best overall, since the desired patterns were clearly visible. The main exceptions to this were the predictions for those with a F-Hb concentration of 30.0-39.9 μg/g in round 1 , and those with a concentration of 20.0-29.9 in round 2. However, this may be due to unstable coefficients estimates (**Table 2**). The coefficients could be smoothed upon future implementation.

Supplementary Table 1. Evaluated risk prediction model specifications [a]

Model	Explanatory variables (*X*)	Discrimination, *C*-statistic [b]	
		Advanced neoplasia	**Colorectal cancer**
1	(1 + age + male + age×male) × (1 + round1_hb3-10 + round1_hb10-20 +…+ round1_hb40-47 + round2_hb3-10 +…+ round2_hb40-47)	0.767 (0.758-0.794)	0.714 (0.691-0.74)
2	(1 + age + male + age×male) × (1 + round1_hb0-10 + round1_hb10-20 +…+ round1_hb40-47 + round2_hb0-10 +…+ round2_hb40-47)	0.784 (0.775-0.792)	0.738 (0.714-0.759)
3	1 + age + male + round1_hb0-10 + round1_hb10-20 +…+ round1_hb40-47 + round2_hb0-10 +…+ round2_hb40-47	0.784 (0.775-0.776)	0.733 (0.708-0.757)
4	1 + age + male + round1&2_hb0-25 + round1&2_hb25-50 + round1&2_hb50-75 + round1&2_hb75-94	0.767 (0.758-0.797)	0.719 (0.695-0.743)
5	(1 + age + male + age×male) × (1 + round1_hb0 + round2_hb0 + round1_hbvalue + round2_hbvalue + round1_hbvalue^2 + round2_hbvalue^2 + log(round1_hbvalue + 0.5) + log(round2_hbvalue + 0.5) + round1_hbvalue×round2_hbvalue)	0.787 (0.778-0.795)	0.741 (0.719-0.764)
6	1 + age + male + round1_hb0 + round2_hb0 + round1_hbvalue + round2_hbvalue + round1_hbvalue^2 + round2_hbvalue^2 + log(round1_hbvalue + 0.5) + log(round2_hbvalue + 0.5)	0.787 (0.777-0.795)	0.734 (0.710-0.758)
7	1 + age + male + round1_hb0 + round2_hb0 + round1_hbvalue + round2_hbdelta + round1_hbvalue^2 + round2_hbdelta^2	0.787 (0.774-0.794)	0.733 (0.708-0.756)
8	1 + age + male + round1_hb0 + round2_hb0 + round1_hbvalue + round2_hbvalue	0.786 (0.776-0.732)	0.734 (0.709-0.758)
9	1 + age + male + round2_hb0-10 + round2_hb10-20 +…+ round2_hb40-47	0.721 (0.710-0.731)	0.669 (0.643-0.696)
10	1 + age + male + round1_hb0-10 + round1_hb10-20 +…+ round1_hb40-47	0.722 (0.712-0.574)	0.677 (0.653-0.701)
11	1 + age + male	0.565 (0.556-0.740)	0.558 (0.534-0.580)
12	1_round1_hb0-10 + round1_hb10-20 +…+ round1_hb40-47 + round2_hb0-10 +…+ round2_hb40-47	0.771 (0.762-0.780)	0.721 (0.700-0.742)

[a] All models were of the functional form: log(OR_y) ~ Xß. Here y is the dependent variable (yes/no relevant outcome). Evaluated predictors in X include age (years/10); male sex (yes/no); categorical F-Hb variables (round1_ or round2_hbX-Y), categorical summed F-Hb concentrations (round1&2_hbX-Y), continuous F-Hb variables (round1_ or round2_hbvalue), the increase in

F-Hb (round2_hbdelta), log-transformed or squared F-Hb terms, and several interactions (denoted by the × sign). For categorical variables, concentrations were rounded to whole numbers above for ease of notation; the lower bound is included and the upper bound is not, except in hb0-10, where 0 is not included. [b] Not adjusted for overfitting or optimism.

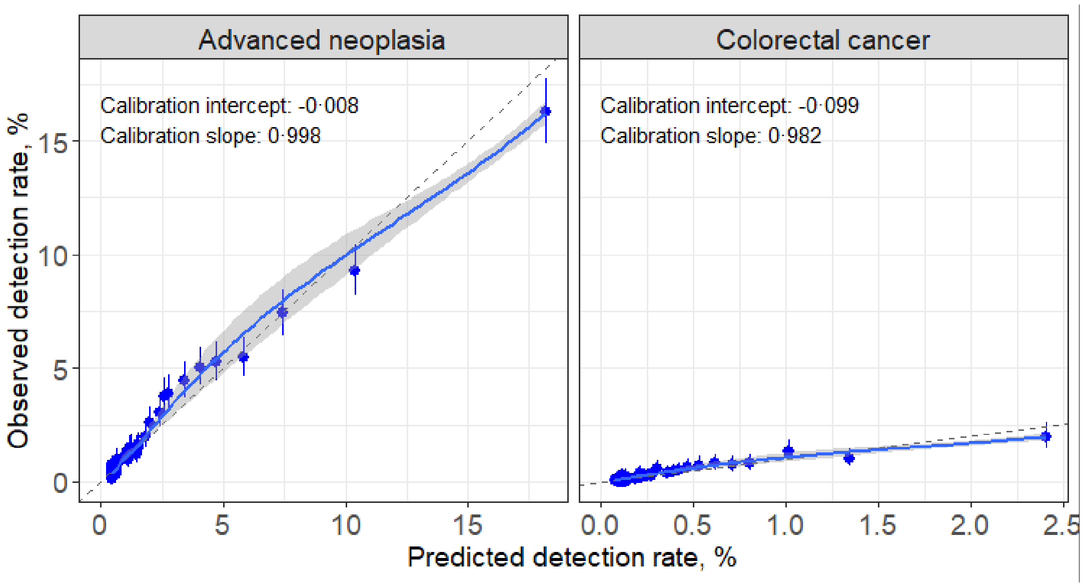
Supplementary Table 2. Study population characteristics and outcomes for external validation

	Participants	Advanced neoplasia [b]				Colorectal cancer [b]			
	Number (%)	Number	OR	95%CI	P-value	Number	OR		P-value
All	11 903 (100%)	90	-			24	-		
Sex									
Female	6500 (54.6%)	33	Ref		<.001	9	Ref		0.14
Male	5403 (45.4%)	57	2.1	1.4-3.2		15	2.0	0.9-4.6	
Age, years (mean 60.7±6.7)									
50-54	2603 (21.9%)	16	Ref		0.1	0	Ref		0.003
55-59	2872 (24.1%)	17	1.0	0.5-1.9		2	-		
60-64	2803 (23.5%)	18	1.0	0.5-2.1		8	-		
65-69	2091 (17.6%)	21	1.6	0.9-3.2		7	-		
70-75	1497 (12.6%)	18	2.0	1-3.9		7	-		
Unknown [a]	37 (0.3%)	0	0	-		0	-		
First F-Hb concentration, µg Hb/g faeces									
0	2334 (19.6%)	8	Ref		<.001	2	Ref		0.59
0.1-2.5	1444 (12.1%)	5	1.0	0.3-3.1		2	1.6	0.2-11.5	
2.6-9.9	350 (2.9%)	7	5.9	2.1-16.5		1	3.3	0.3-36.9	
10-46.9	-	-				-			
Missing [a]	7775 (65.3%)	70	2.6	1.3-5.5		19	2.9	0.7-12.3	
Second F-Hb concentration, µg Hb/g faeces									
0	6244 (52.5%)	22	Ref		<.001	8	Ref		0.67
0.1-2.5	1250 (10.5%)	9	2.1	0.9-4.5		2	1.2	0.3-5.9	
2.6-9.9	785 (6.6%)	13	4.8	2.4-9.5		2	2.0	0.4-9.4	
10-46.9	-	-				-			
Missing [a]	3624 (30.4%)	46	3.6	2.2-6.1		12	2.6	1.1-6.3	

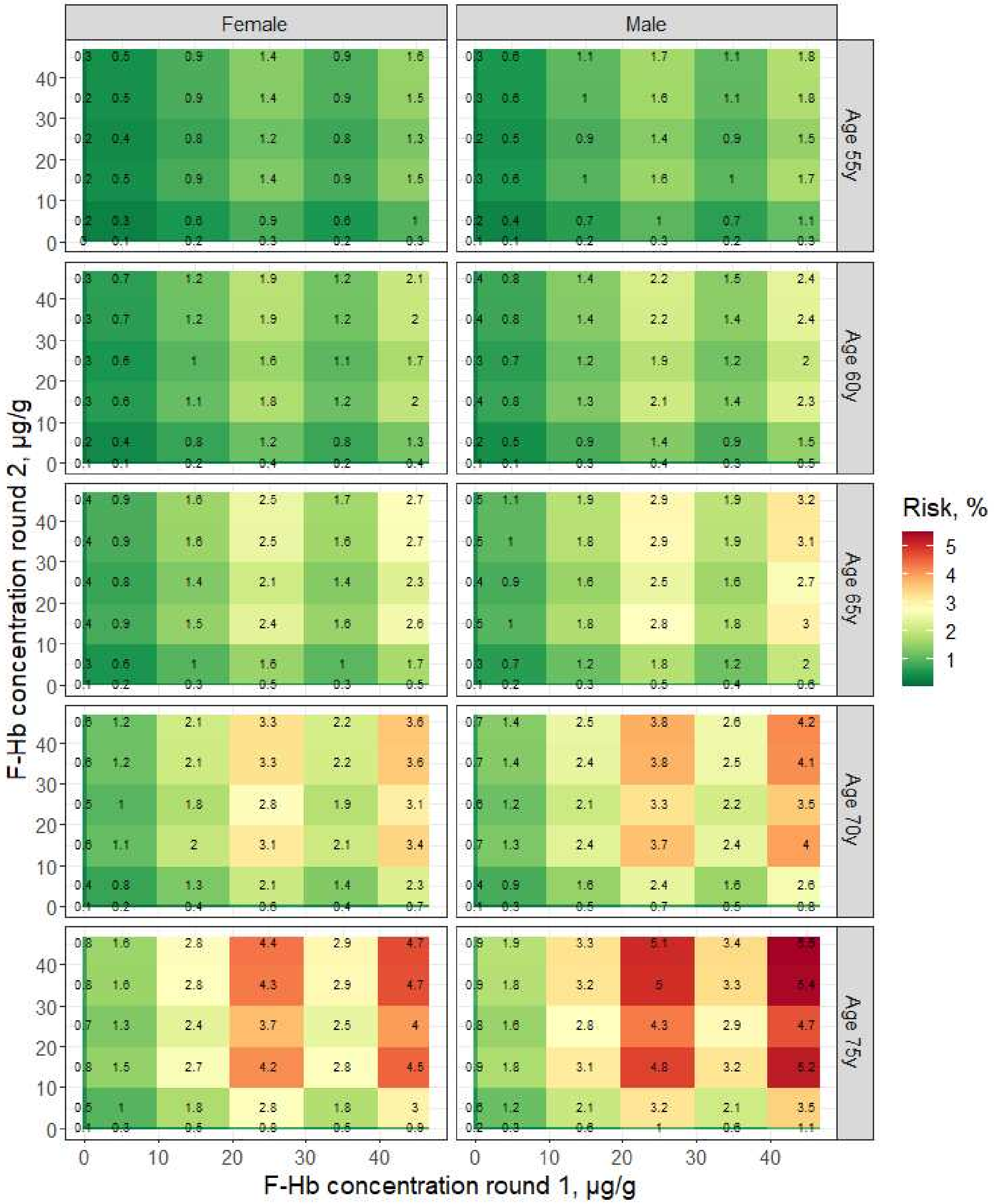
Abbreviations: F-Hb = faecal haemoglobin; OR = odds ratio. [a] Values were imputed using multiples imputation. [b] Observed among participants with a F-Hb ≥47 µg Hb/g faeces in Round 3.

Supplementary Figure 1. Observed vs. predicted FIT screening outcomes.

Blue dots represent observed detection rates with 95%CIs for each riskscore percentile; the blue line is a fitted Loess curve with 95% confidence bounds (grey area). Adequate calibration is indicated by overlap of the grey area with the diagonal (predicted=observed). A calibration intercept and slope close to 0 and 1, respectively, further confirm adequate calibration.

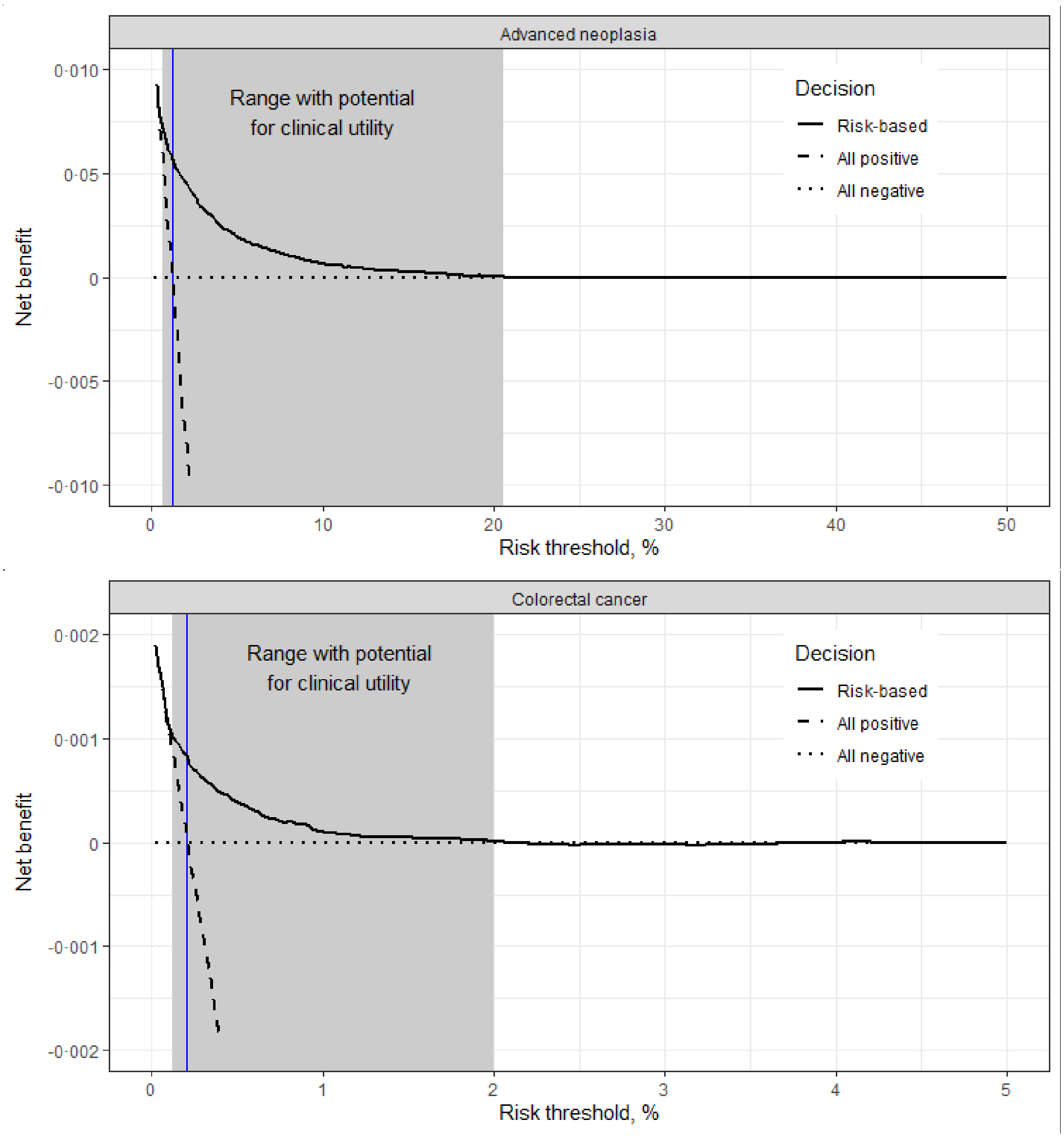


Supplementary Figure 2. Riskscore chart for future colorectal cancer.



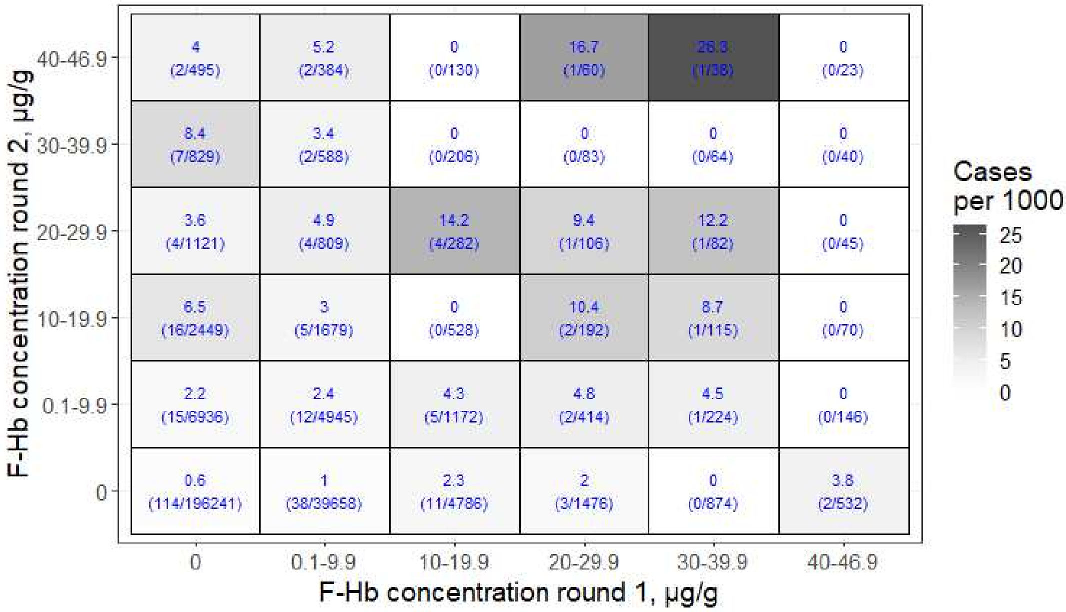
Supplementary Figure 3. Decision curve analysis of the potential benefit for risk-stratification in faecal immunochemical test screening.

In decision curve analysis, the net benefit is directly related to the choice of risk threshold (no interpretable unit). The idea is that the chosen risk threshold exposes how screening participants or policy makers weigh false-positive *vs.* false-negative outcomes (p:1-p).[15] Risk-stratified screening or follow-up (solid black line) may add value over uninformed strategies when the associated net benefit exceeds that of treating everyone as *high-risk* (dashed line) or *low risk* (dotted line). In our case, there is potential for clinical utility for risk thresholds of 0.6-20.6% for advanced neoplasia, and 0.1-2.0% for cancer, which includes the average detection rate of those outcomes (blue line) within the study population as also highlighted in **Figure 4**.



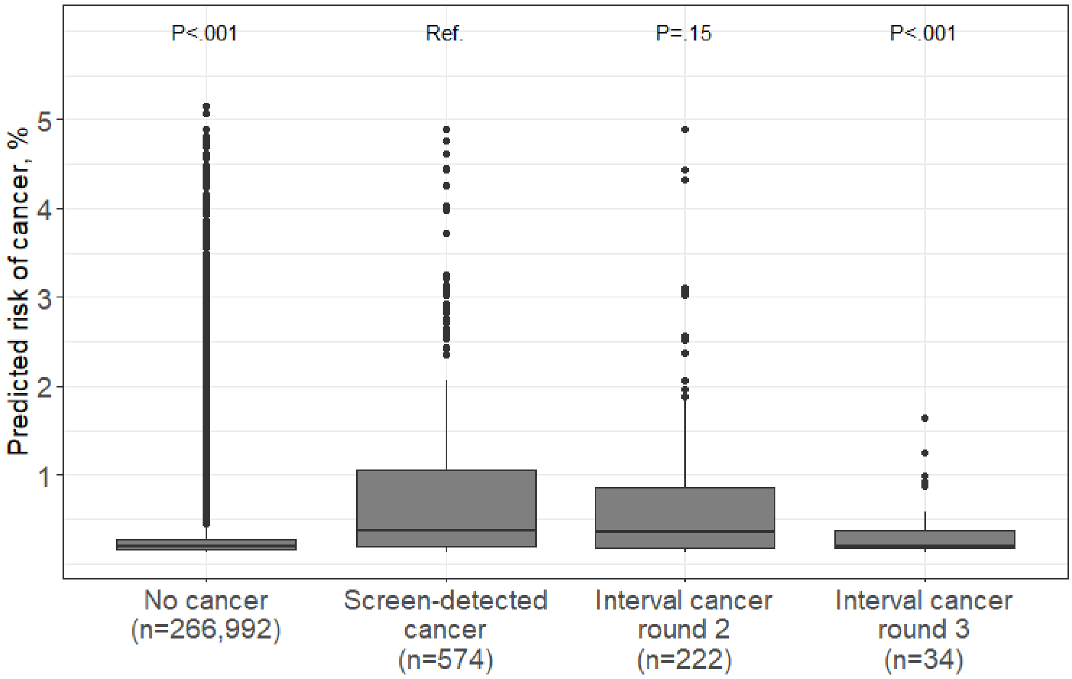
Supplementary Figure 4. Interval colorectal cancer by F-Hb concentration.

Tiles show the proportion of FIT participants with interval colorectal cancers by measured F-Hb concentration in round 1 and 2. Labels provide exact proportions as well as case counts and population denominators (in parentheses).



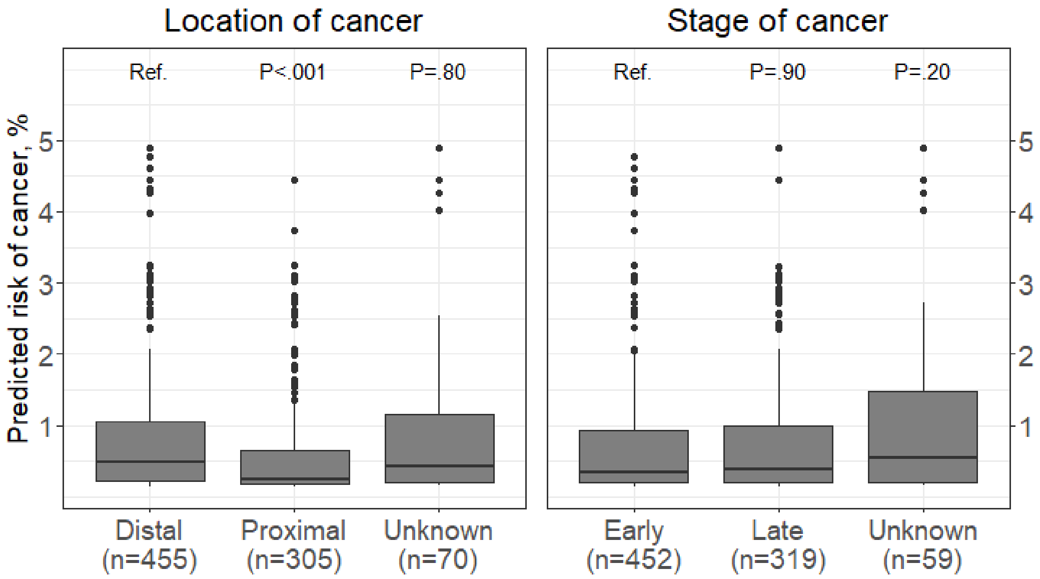
Supplementary Figure 5. Predicted risk of cancer by type of outcome.

The boxplots represent the distribution of riskscores for participants by outcome category in the prediction model for CRC. Reported P-values are from a pairwise Wilcoxon test to examine subgroup differences in predicted CRC risk.



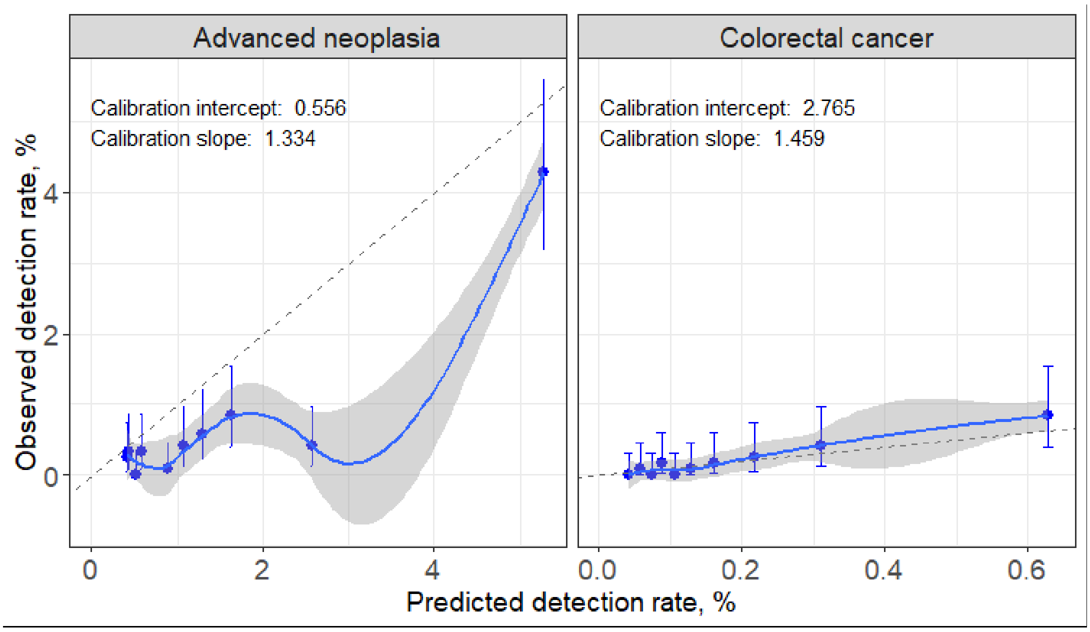
Supplementary Figure 6. Predicted risk of cancer by type of outcome.

The boxplots represent the distribution of prediction riskscores for CRC patients by location and stage of diagnosed CRC. Proximal location was defined as proximal to the splenic flexure. Early stage was defined as stage I or II. Reported P-values are from a pairwise Wilcoxon test to examine subgroup differences in predicted CRC risk.



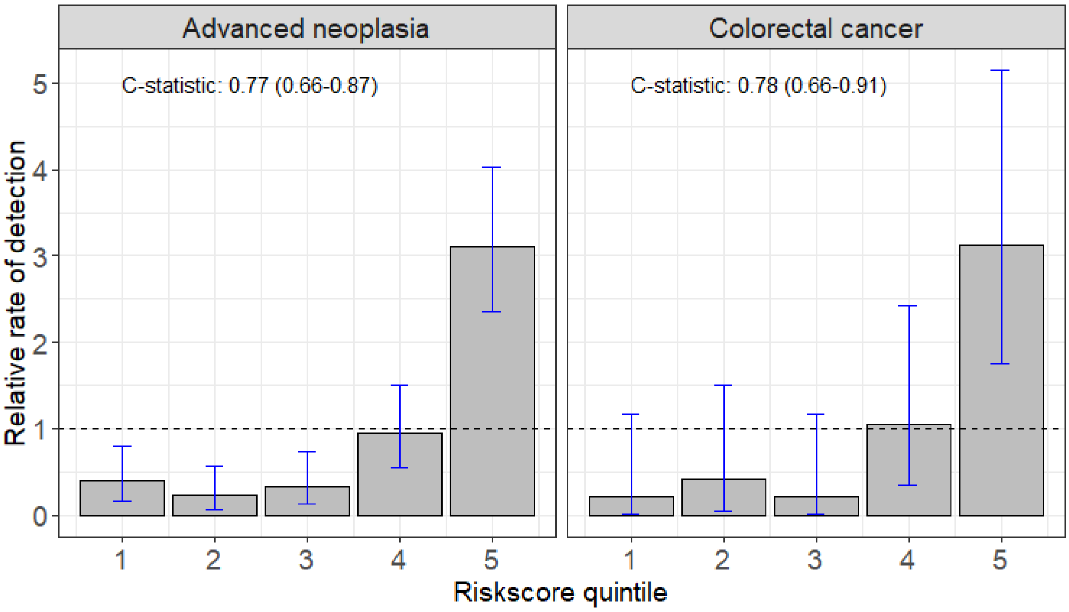
Supplementary Figure 7. Observed vs. predicted FIT screening outcomes in an external population.

This analysis was performed for external validation in an independent screening population. Blue dots represent observed detection rates with 95%CIs for each riskscore percentile; the blue line is a fitted Loess curve with 95% confidence bounds (grey area). Adequate calibration is indicated by overlap of the grey area with the diagonal (predicted=observed). A calibration intercept and slope close to 0 and 1, respectively, further confirm adequate calibration.



Supplementary Figure 8. Risk stratification in an external population.

This analysis was performed for external validation in an independent screening population. The x-axis plots population subgroups rank-ordered by riskscore (quintiles). The y-axis plots observed outcomes relative to the total study population.


Advanced neoplasia
Colorectal cancer
C-statistic: 0.77 (0.66-0.87)
C-statistic: 0.78 (0.66-0.91)
Relative rate of detection
Riskscore quintile