CRC-AIM recalibration

To initially assess the performance of the model, we used the posterior means of the 23 calibrated parameters from CRC-SPIN that are described in Rutter et al.[1] We compared the natural history outputs of CRC-AIM with publicly available outcomes from CRC-SPIN, which have been described in multiple publications over time. Through this process, we observed a high degree of comparability in almost every natural history output between both models. However, we want to describe some of the discrepancies we observed that required us to recalibrate CRC-AIM. Here, we describe these dissimilarities and the recalibration process in greater detail.

Adenoma growth parameters

We observed a discrepancy between CRC-AIM and CRC-SPIN in the distribution of the most-advanced adenoma size by age (**Figure 1**). Specifically, at age 40, CRC-AIM reported only 4% of adenomas were ≥10 mm, compared to 8% for CRC-SPIN. Because adenomas only begin to generate after age 20, we hypothesized that certain adenomas were growing too slowly after generation. We discovered colon-based adenomas were growing too slowly—for adenomas generated at ages less than 40, in our implementation of the model, it was improbable for them to become ≥10 mm by the age of 40.

However, the adenoma size distribution at age 65 was similar for CRC-AIM and CRC-SPIN, including for adenomas ≥10 mm (**Table 1**).[2] Aside from differences in model implementation, there are at least two possible explanations for this discrepancy. One possibility is that after age 40, there is enough time for the adenomas to grow to match

the size distribution from CRC-SPIN. Another possibility is that CRC-SPIN could have been recalibrated after this publication and the impact of a potential recalibration has not been fully described.

Transition probabilities by location/sex

We also observed that CRC-AIM was generating lower values for overall SEER cancer incidence compared to CRC-SPIN (**Figure 2A**), which also resulted in an underestimation of overall cumulative cancer risk (**Figure 2B**).[2] Overall CRC incidence is a direct function of age-stratified CRC incidence by location and cancer site. In the prototype version of CRC-AIM, age-stratified cancer incidence was only slightly under-generated for colon cancers in males and females and rectal cancers in females—approximately 4-13% below expected results (**Table 2**). The primary reason for under-generated overall CRC incidence was rectal cancers in males, which were consistently 30% below expected results (**Table 2**).

We considered multiple potential explanations for this discrepancy. Our first consideration was sojourn time (ST), because CRC incidence is partly a function of ST: changing sojourn time will shift cancer incidence when all other model parameters are kept equal. However, the prototype version of CRC-AIM replicated the overall ST for CRC-SPIN[2] (see **Table 3** for the CRC-AIM prototype), which was unsurprising because ST parameters for preclinical cancers are explicitly defined by formulas and therefore are tightly controlled. Additionally, ST is neither age- nor sex-specific in both CRC-SPIN[1] and CRC-AIM, and while the rectal cancer incidence in males was greatly underestimated, the rectal cancer incidence in females was relatively comparable

between models. Therefore, we concluded that ST did not likely contribute to this discrepancy.

Next, we considered other components that contribute to CRC incidence—namely, adenoma generation (both sex- and age-specific), adenoma growth, and adenoma transition parameters. Adenoma generation was comparable between CRC-SPIN and CRC-AIM.[2] Although we had experienced discrepancies in the adenoma growth by location (see “Adenoma growth parameters”), adenoma growth is not sex-specific. In addition, the differences in CRC incidence for rectal cancer incidence in males occurred across all age groups, not just in individuals over 40 years old.

The only remaining parameters were those that enabled adenomas to transition to preclinical cancer—specifically, the adenoma size (γ_1) and age of adenoma initiation (γ_2) parameters. These parameters are highly sensitive, with minor variations having a large impact on cancer incidence. The minor yet consistent differences for age-stratified colon cancer incidence in males and females and rectal cancer incidence in females could simply have been due to the lack of precision in the published estimates of the gamma parameters estimates (reported only to the thousandths place) or due to differences in the way the programs are implemented. Using the published parameter estimates, we found that simple plots of the CDF for adenomas transitioning to preclinical cancer also indicated that rectal cancers in males would be underestimated. Those curves revealed that the rectal-male curves lie underneath the other location/sex curves across multiple ages of adenoma initiation (**Figure 3**), demonstrating that these parameters were contributing to the discrepancy in CRC incidence.

Sojourn time

We observed that the mean sojourn times (STs) of colon- and rectal-based cancers for CRC-SPIN (averaged to 1.9 years for 2010[3] and 1.6 years for 2011[4]) was noticeably shorter than the published ST of MISCAN (3.0 years[4]) and SimCRC (4.0 years[4]).[2] This discrepancy prompted a more thorough analysis into modeling ST, including estimates based on real-world evidence and model validation studies.

One study by Zheng and Rutter estimated sojourn time in a study of 42,079 patients who received fecal occult blood tests (FOBT).[5] The estimated mean sojourn time (in years) for colon and rectal cancers, respectively, was determined to be 3.86 and 3.35 for 45- to 54-year-olds, 3.78 and 2.24 for 55- to 64-year-olds, and 2.70 and 2.10 for 65- to 74-year-olds,[5] all values larger than the mean STs for CRC-SPIN. Another study evaluated the accuracy of the CISNET model outputs against outcomes from the United Kingdom Flexible Sigmoidoscopy Screening (UKFSS) trial.[6] Notably, although CRC-SPIN did not accurately predict the number of screen-detected cancers, the two models with the longer sojourn times (MISCAN and SimCRC) did. Based on this observation, the authors concluded that the mean ST is probably between 1.6 and 4.0 years and that the actual value is on the higher end of that estimated range. Finally, Brenner et al used data from a German national screening database to estimate sex- and age-stratified mean sojourn times, which were determined to be between 4.5 and 5.8 years for the subgroups assessed.[7]

We determined that this evidence provided sufficient justification to extend the sojourn time for CRC-AIM beyond that of CRC-SPIN, whose STs for colon and rectal cancers are 1.9 and 2.7 years, respectively. Therefore, we conservatively extended the

average sojourn by one year each (to 2.9 and 3.7 years, respectively) but maintained the standard deviation ($\mu_c*\tau_c$, $\mu_r*\tau_r$) of the distributions equivalent to CRC-SPIN.[2]

We chose to be conservative in extending the sojourn time because preclinical cancer detection depends on both the interval length of a screening test and the ST length. If the ST length is shorter than a screening interval, then there is a possibility of no chance to detect a preclinical cancer since the development from preclinical to clinical CRC could fall entirely in the window between screening tests.[4] However, if the ST is double that of a screening interval, then there are two chances for a screening test can detect the preclinical cancer. In this way, for cancers with a longer ST, a greater overall screening benefit could be conferred to patients who are screened with a test with a longer screening interval. We did not want to bias CRC-AIM to favor screening tests with longer intervals. We hypothesize that CRC-SPIN's sojourn time will be updated through recalibration but the recalibrated parameters are unknown.

Recalibration methodology

After updating the model to include the evidence-based sojourn times as described above, we conducted a sequential recalibration of the discrepant parameters, first updating the adenoma growth parameters and then updating the transition to preclinical cancer parameters. In general, we used space-filling Latin hypercube sampling across the published posterior distribution estimates (Bayesian credible intervals) for the parameters.[3] For adenoma growth parameters, we restricted combinations of β_1/β_2 to limit the probability of an adenoma reaching 10 mm within 10 years from 0.0001 to 0.25, similar to the CRC-SPIN v2.x recalibration.[8]

For the adenoma growth recalibration targets, we used a combination of the adenoma number and size distribution by age for CRC-SPIN and the adenoma size distribution for the most advanced adenoma for CRC-SPIN.[2] For the transition probability recalibration targets, we used the 1975-1979 SEER and CRC-SPIN v1.0 observed cancer rates as lower and upper limit optimization targets.

Parameters were then optimized using a Kriging model approach. The resulting recalibrated parameters are reported in the main manuscript, and the outputs from recalibration are also reported.

References

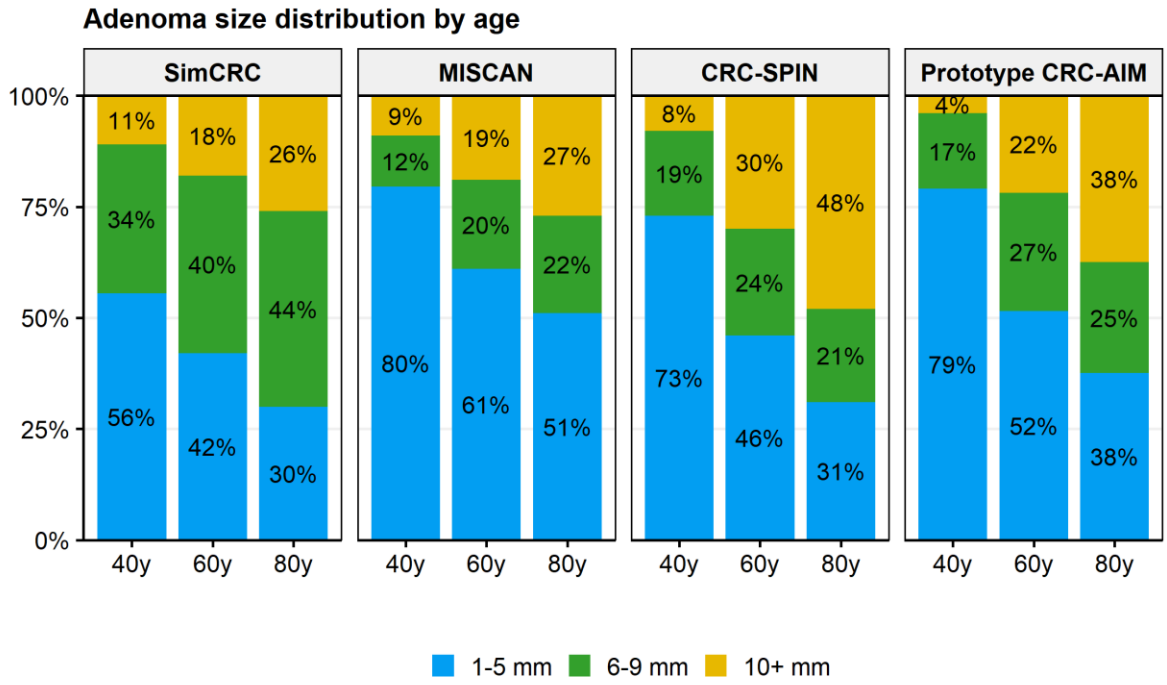
1. Rutter CM, Miglioretti DL, Savarino JE. Bayesian Calibration of Microsimulation Models. *J Am Stat Assoc.* 2009;104(488):1338-1350.
2. Piscitello A, Saoud L, Matney M, et al. Description and Validation of the Colorectal Cancer and Adenoma Incidence & Mortality (CRC-AIM) Microsimulation Model. *bioRxiv.* 2020.
3. Rutter CM, Savarino JE. An evidence-based microsimulation model for colorectal cancer: validation and application. *Cancer Epidemiol Biomarkers Prev.* 2010;19(8):1992-2002.
4. Kuntz KM, Lansdorp-Vogelaar I, Rutter CM, et al. A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. *Med Decis Making.* 2011;31(4):530-539.
5. Zheng W, Rutter CM. Estimated mean sojourn time associated with hemoccult SENSA for detection of proximal and distal colorectal cancer. *Cancer Epidemiol Biomarkers Prev.* 2012;21(10):1722-1730.
6. Rutter CM, Knudsen AB, Marsh TL, et al. Validation of Models Used to Inform Colorectal Cancer Screening Guidelines: Accuracy and Implications. *Med Decis Making.* 2016;36(5):604-614.
7. Brenner H, Altenhofen L, Katalinic A, Lansdorp-Vogelaar I, Hoffmeister M. Sojourn time of preclinical colorectal cancer by sex and age: estimates from the German national screening colonoscopy database. *Am J Epidemiol.* 2011;174(10):1140-1146.

8. Rutter C, Ozik J, DeYoreo M, Collier N. Microsimulation Model Calibration using Incremental Mixture Approximate Bayesian Computation. *arXiv.* 2018.

9. Zauber A, Knudsen AB, Rutter CM, Lansdorp-Vogelaar I, Kuntz KM. Technical Report: Evaluating the benefits and harms of colorectal cancer screening strategies: a collaborative modeling approach. 2015; AHRQ Publication No. 14-05203-EF-2:https://www.uspreventiveservicestaskforce.org/Home/GetFile/1/16540/cisnet-draft-modeling-report/pdf.

10. Knudsen AB, Lansdorp-Vogelaar I, Rutter CM, et al. Cost-effectiveness of computed tomographic colonography screening for colorectal cancer in the medicare population. *J Natl Cancer Inst.* 2010;102(16):1238-1252.

159 **Figure 1. Comparison of the distribution of the most-advanced adenoma size by**
160 **age between CISNET models (SimCRC, MISCAN and CRC-SPIN) and a prototype**
161 **version of CRC-AIM.** Size distribution is evaluated into 40, 60, and 80 years. The CRC-
162 AIM prototype was prior to recalibration. Data adapted from Zauber et al.[9]



163

Figure 2. Comparison of colorectal cancer (CRC) incidence and cumulative CRC risk between CISNET models (SimCRC, MISCAN and CRC-SPIN) and a prototype version of CRC-AIM. (A) CRC incidence by age and model and (B) cumulative probability of developing CRC. The CRC-AIM prototype was prior to recalibration. Data adapted from Zauber et al.[9]

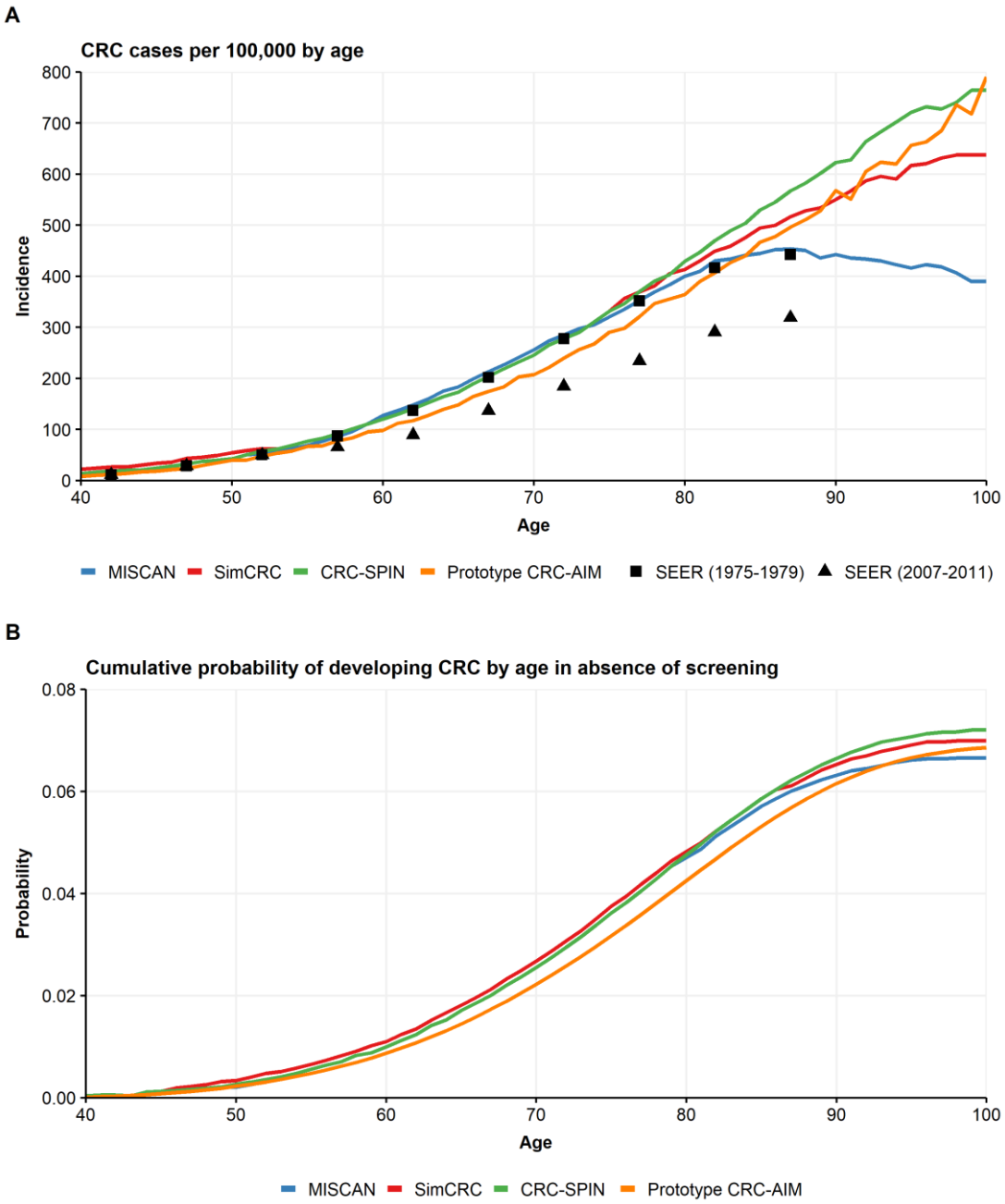


Figure 3. Cumulative transition probability for an adenoma to become preclinical cancer as a function of adenoma size. The transition probability is visualized at three different ages at adenoma initiation (25, 45, and 65 years), stratified by sex and adenoma location. The vertical red line indicates the maximum adenoma size in the model (50 mm).

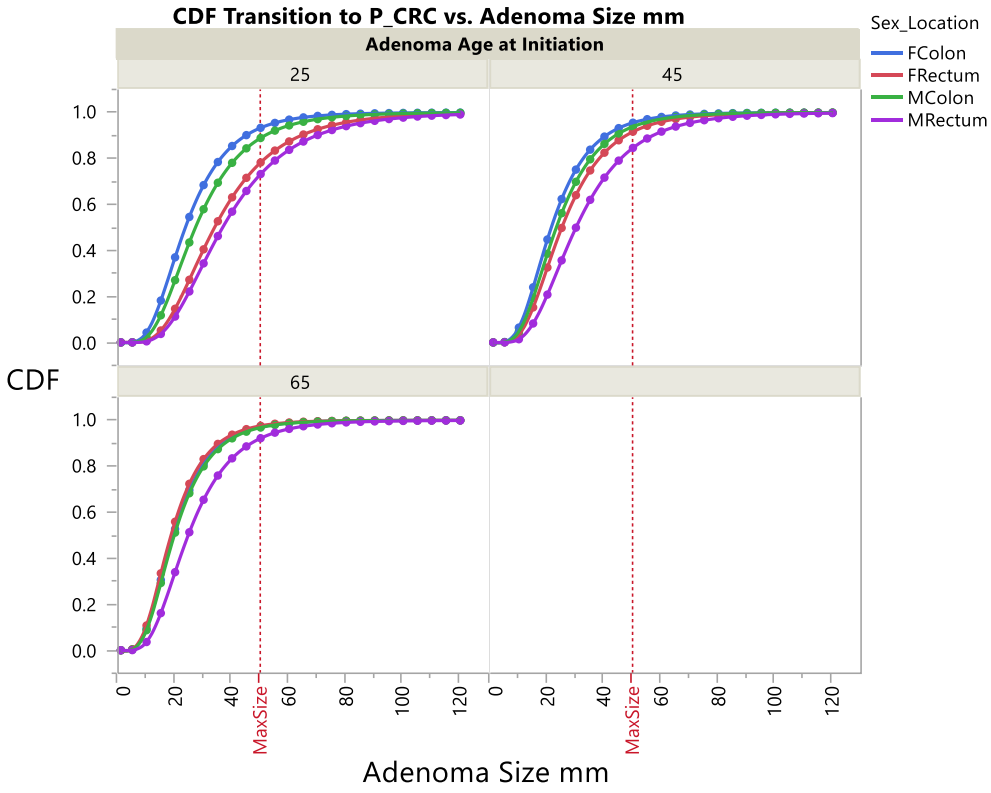


Table 1. Detailed adenoma statistics at age 65 for CISNET models (MISCAN, SimCRC, and CRC-SPIN) and a prototype version of CRC-AIM. The CRC-AIM prototype was prior to recalibration. Data adapted from Knudsen et al.[10]

*The table from Knudsen et al[10] labels this category as “1-10 mm”.

Outcome		MISCAN	SimCRC	CRC-SPIN	Prototype CRC-AIM
Adenoma prevalence, age 65		39.80%	37.20%	30.70%	not calculated
Number of adenomas per 1000 by site and size at age 65 y					
Proximal Colon	1-5 mm	121.2	171.7	190.2	200.2
	6-9 mm	69.9	186.2	67.8	75.1
	≥10 mm*	61.8	23.9	40.8	43.6
Distal Colon	1-5 mm	134.4	124.2	124.5	131.0
	6-9 mm	77.4	18.2	44.4	49.4
	≥10 mm	68.4	41.6	26.7	28.5
Rectum	1-5 mm	133.5	8.7	14.1	15.5
	6-9 mm	76.8	16.0	9.1	9.4
	≥10 mm	68.1	15.8	20.2	22.6

Table 2. Comparison to estimates of clinical cancers per 100,000 among SEER 1975-1979 estimates published in Rutter et al, CRC-SPIN, and a prototype version of CRC-AIM. The CRC-AIM prototype was prior to recalibration. Data adapted from Rutter et al.[1]

Location, Sex	Age (years)	SEER 1975-1979, per Rutter 2009	CRC-SPIN (Rutter 2009)	Prototype CRC-AIM	Difference between CRC-SPIN and Prototype CRC-AIM (%)
Colon, Female	20-49	4.8	4.4	not calculated	NA
	50-59	43.3	45.5	41.19	-9.48%
	60-69	100.7	99.3	94.63	-4.70%
	70-84	216.7	207.3	198.75	-4.12%
Rectum, Female	20-49	1.87	2.0	not calculated	NA
	50-59	20.4	18.6	18.45	-0.80%
	60-69	42.5	39.7	41.44	4.38%
	70-84	73.9	82.1	83.58	1.80%
Colon, Male	20-49	4.51	4.2	not calculated	NA
	50-59	45.9	50.6	43.85	-13.35%
	60-69	121.4	120.0	111.88	-6.76%
	70-84	268.4	263.4	257.15	-2.37%
Rectum, Male	20-49	2.3	3.2	not calculated	NA
	50-59	30.0	29.8	21.23	-28.76%
	60-69	71.4	63.2	42.97	-32.00%
	70-84	128.0	123.8	84.83	-31.48%

188 **Table 3. Dwell time summary statistics for a prototype version of CRC-AIM.** The

189 CRC-AIM prototype was prior to recalibration. IQR = interquartile range

Measure	Statistic	Prototype CRC-AIM
Adenoma dwell time, y	Mean	26.7
	Median	25
	IQR	17-35
Sojourn time (preclinical cancer dwell time), y	Mean	2.1
	Median	1.6
	IQR	1.0-2.7
Overall dwell time (full adenoma-carcinoma sequence), y	Mean	28.9
	Median	27.2
	IQR	18.9-37.3

190