This is a repository copy of *Calibration and Validation of the Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel), an Individual Patient Simulation Model for Investigation of the Cost-effectiveness of Personalised Screening and Surveillance Strategies*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/171343/

Version: Published Version

Monograph:
Mandrik, O. orcid.org/0000-0003-3755-3031, Thomas, C., Strong, M. et al. (1 more author) (2021) Calibration and Validation of the Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel), an Individual Patient Simulation Model for Investigation of the Cost-effectiveness of Personalised Screening and Surveillance Strategies. Report. ScHARR HEDS Discussion Papers . School of Health and Related Research, University of Sheffield , Sheffield.

© 2021 The Author(s). Article available under the terms of the CC-BY-NC-ND 4.0 licence (https://creativecommons.org/licenses/by-nc-nd/4.0/)

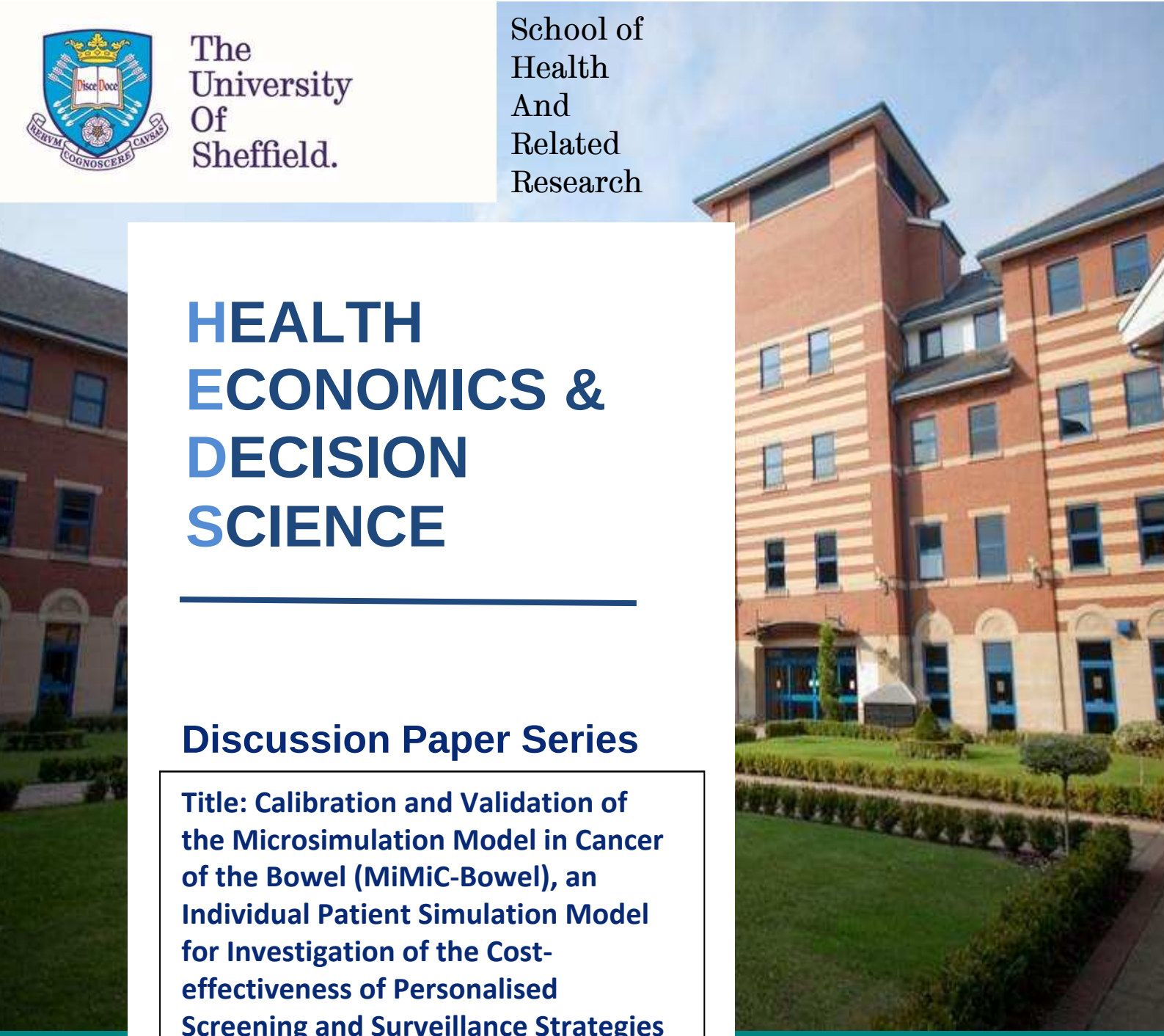
Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

The University Of Sheffield.

School of Health And Related Research



HEALTH ECONOMICS & DECISION SCIENCE

Discussion Paper Series

Title: Calibration and Validation of the Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel), an Individual Patient Simulation Model for Investigation of the Cost-effectiveness of Personalised Screening and Surveillance Strategies

Authors: Olena Mandrik, Chloe Thomas, Mark Strong and Sophie Whyte

Corresponding author: Dr Olena Mandrik, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK. Tel: +44 (0) 114 222 6392 , Email: o.mandrik@sheffield.ac.uk uk

Disclaimer:
This series is intended to promote discussion and to provide information about work in progress.
The views expressed in this series are those of the authors.
Comments are welcome, and should be sent to the corresponding author.

Title: Calibration and Validation of the Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel), an Individual Patient Simulation Model for Investigation of the Cost-effectiveness of Personalised Screening and Surveillance Strategies.

Date: February 2021

Authors: Olena Mandrik, Chloe Thomas, Mark Strong & Sophie Whyte

Corresponding author: Olena (Lena) Mandrik, ScHARR, HEDS, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DAOffice: 1012/13Tel: +44 (0) 114 222 6392
Email: o.mandrik@sheffield.ac.uk

Affiliation of co-authors

Olena Mandrik	School of Health of Related Research, University of Sheffield, United Kingdom ORCID: 0000-0003-3755-3031
Chloe Thomas	School of Health of Related Research, University of Sheffield, United Kingdom ORCID: 0000-0001-8704-3262
Mark Strong	School of Health of Related Research, University of Sheffield, United Kingdom ORCID: 0000-0003-1486-8233
Sophie Whyte	School of Health of Related Research, University of Sheffield, United Kingdom ORCID: 0000-0002-7963-2523

Funding

Calibration and validation analyses were Funded by Research England (Quality-related research funding to Support Evidence Based Policy Making distributed by the University of Sheffield). Development of MiMiC-Bowel was funded by the English National Screening Committee, which is part of Public Health England. The views expressed are those of the authors and not necessarily those of the funding agencies.

Acknowledgement

We are grateful for methodological advice received from Prof. James Chilcott. We are also grateful to Dr. Benjamin Kearns for providing critical feedback on the manuscript.

Contents

Summary

The Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel) incorporates individual cancer risk to inform colorectal cancer (CRC) screening decisions in England. This work reports calibration and cross-validation of the MiMiC-Bowel.

Due to the complex and computationally intensive nature of the model a step-wise calibration approach was taken utilising manual and automated algorithmic fitting. Natural history disease model parameters were estimated via calibration to several data targets including English and adjusted German data relating to previously unscreened persons. The model was cross-validated to four international models.

Incompatibility in calibration data was assessed by analysing CRC incidence, convergence feasibility, and predictions of screening test sensitivity. Data incompatibility was addressed by giving less weight in fitting the parameters to target data with lower reliability and applicability, and adjusting data sets to reflect differences between localities. The MiMiC-Bowel predicted 60% sensitivity of flexible sigmoidoscopy for CRC and 59% for high-risk adenoma. MiMiC-Bowel's predictions of CRC cases that arise from potentially detectable adenomas and impact of perfect polypectomy on risk reduction were within ranges reported by the US models, but the predictions of the sojourn and dwell time were not (29.1 and 5.3 years vs 8-24 and 2-4 years respectively). CRC risk increased less rapidly by age in MiMiC-Bowel than German data suggests, which is supported by population data on pre-screening CRC incidence.

1. Brief description of the MiMiC-Bowel model

The Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel) is a patient-level microsimulation model in R programming language which was developed to simulate the development of colorectal cancer (CRC) over patients' lifetime for the population of England. The model incorporates individual cancer risk and was populated with population characteristics retrieved from the Health Survey for England 2014 [1]. A detailed description of the model is available in the online report [2].

MiMiC-Bowel comprises natural history, symptomatic diagnosis, screening, and surveillance modules. In addition to age and sex, the model incorporates individual risk factors [3]. Each person is assumed to have a normal colorectal epithelium at the age of 30. The population can move through nine health states reflecting the development of CRC (Figure 1). Precancerous conditions were classified as either low-risk adenomas (LRA) or high-risk adenomas (HRA) in line with the surveillance guidelines used in the UK from 2002 to 2019 [4]. CRC diagnosis may occur via symptomatic and chance diagnosis, or through screening. Diagnosis of pre-cancerous lesions in MiMiC-Bowel occurs only via screening.

CRC development via serrated neoplastic pathways was represented in the model as transition from normal epithelium directly to CRC stage A [5]. The transition probability from normal epithelium to CRC stage A was increasing linearly between ages 15 and 100 to reflect the absence of recorded cases for persons younger than 15 years [6]. As no data identified differences in CRC stage distribution by sex, transitions between CRC stages and symptomatic presentation rates (except for persons on CRC stages A and B older than 75 years, to address the lower symptomatic presentation among older population [7]) were assumed to be invariant by age and sex [8 9].

2. Uncertainty of the MiMiC-Bowel predictions

Because of the stochastic nature of the model, the predictions of the model on smaller populations may vary significantly for rare events, such as CRC cases in young persons. To determine the model population size required to ensure stable calibration results, the model was run for a range of

population sizes up to 6.7 million. Standard errors in modelling predictions of CRC incidence among males and females was generated for populations of different sizes (Figure 1). The maximum standard error of CRC incidence among modelled population of less than 0.1 (considered acceptable) was achieved with the population of 1.3 million people.

Figure 1. Standard errors in modelling predictions of CRC incidence among males (1a) and females (1b). Each cycle includes a population of 6,700 people

Figure 1a

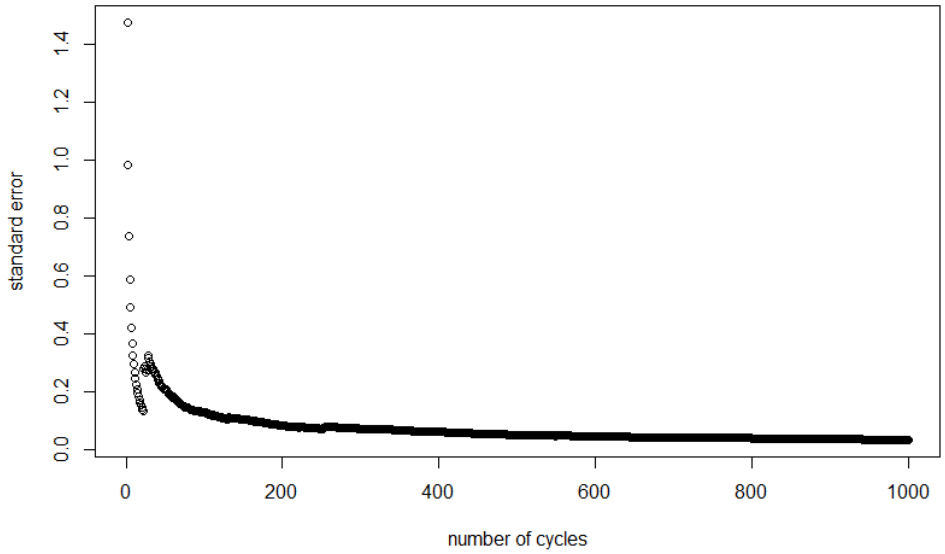
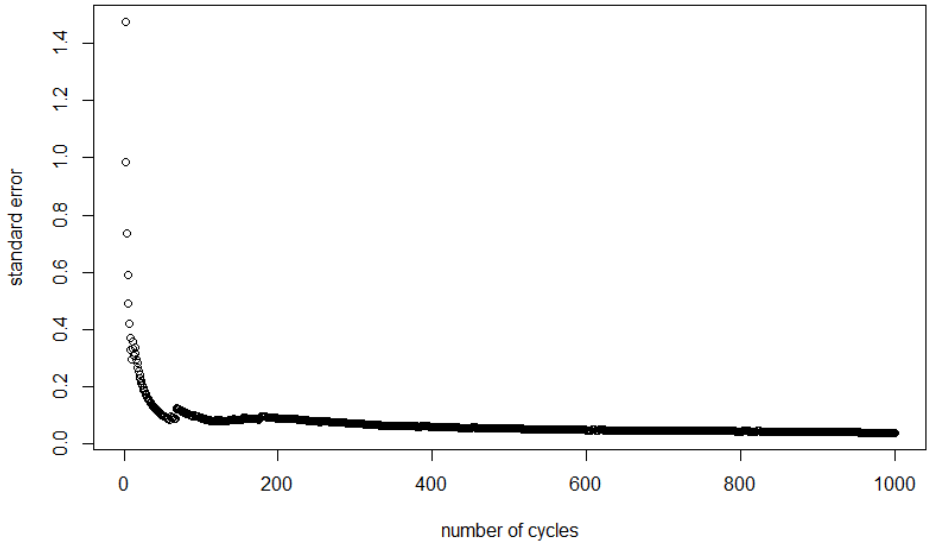


Figure 1b



3. Calibration of the MiMiC-Bowel: the framework

To calibrate the MiMiC-Bowel model we developed the following framework based on a trial-and-error approach:

(1) Selection of the calibration targets and assessment of their compatibility;

(2) Selection of the algorithm and acceptance metrics;

(3) Calibrating the model using the step-wise approach combining a manual search to find an initial set of parameters and Metropolis–Hastings (MH) algorithm to retrieve the distribution of the parameters;

(4) Validating the predictions against expected sensitivity values for screening tests.

(5) Cross-validating the MiMiC-Bowel predictions to other models.

4. Calibration targets of the MiMiC-Bowel

4.1. Calibration targets 1&2: Prevalence of adenomas (unscreened population)

Data on prevalence of adenomas in the UK were not available, so data from comparable settings were considered. Data on advanced/non-advanced lesions from the German colonoscopy screening programme were selected as the most reliable estimates due to the large sample sizes (> 4 million people) and geographical similarity compared to other studies [8 10].

Prevalence of adenomas was estimated as the detection rate of pre-cancerous lesions at screening colonoscopy in a previously unscreened population divided by the estimated sensitivity of screening colonoscopy (0.765 for LRA and a 0.925 for HRA) [11 12]. Further adjustments were made to German data to make it more comparable to the UK, using information from the UK Flexible Sigmoidoscopy Screening randomized controlled trial (UKFSST)[13] as reported in Box 1.

Box 1. Adjustment of low and high risk adenoma data to reflect the prevalence of non-advanced and advanced adenoma

The definition of advanced adenoma used by Brenner et al (2010)[9] differed from the high risk/low risk categorisation used in MiMiC-Bowel. The UKFSS Trial [14] collected information on both advanced adenoma and HRA prevalence following flexible sigmoidoscopy (personal communication with the authors), and was used to convert the prevalence of advanced and non-advanced lesions reported by Brenner et al (2014) [8] to high- and LRA. The proportion of advanced adenomas that are high-risk was 0.81 for males and 0.66 for females. Using this conversion approach for adenoma prevalence based on Brenner colonoscopy screening assumes that these ratios are similar for proximal and distal CRC which may not be the case (unlike colonoscopy, flexible sigmoidoscopy has very low sensitivity to proximal CRC); however, other evidence was not available.

The process of data conversion was the following:

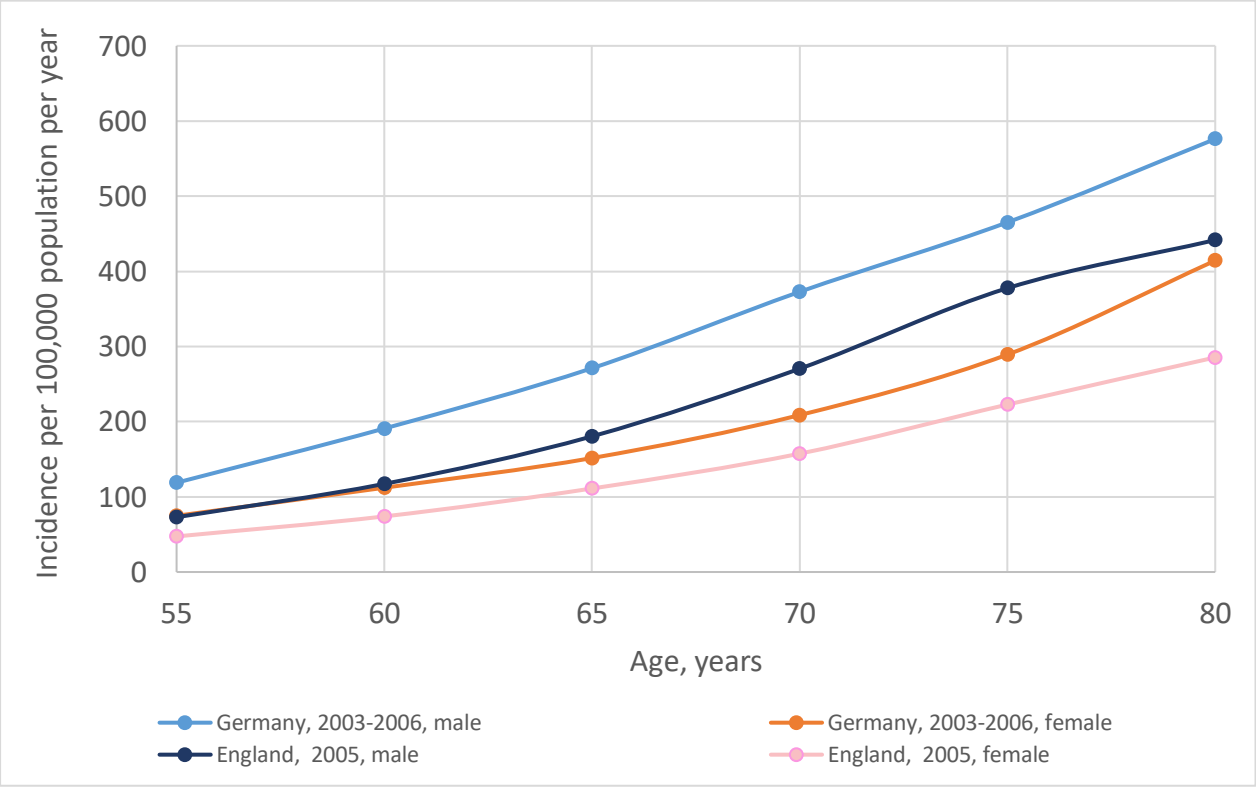
1. Prevalence of advanced and non-advanced adenoma was calculated from detection rate and test sensitivity (0.966 for advanced and 0.765 for non-advanced adenoma).
2. Total adenoma prevalence was calculated as a sum of advanced and non-advanced adenoma.
3. Prevalence of HRA was calculated by multiplying prevalence of advanced adenoma among men by 0.81 and among women by 0.66.
4. Prevalence of LRA was calculated as the difference between total adenoma prevalence and HRA prevalence.

4.2. Calibration targets 3-7: Undiagnosed CRC prevalence total and by stages

Undiagnosed CRC prevalence was defined as the number of persons with undiagnosed CRC in an unscreened population. The detection rate of CRC at screening colonoscopy by age and sex for the 55-73 year-old population in the German national colonoscopy screening registry 2003 to 2007 was used [9]. Undiagnosed CRC prevalence was estimated as CRC detection rate divided by colonoscopy sensitivity to CRC (0.966)[15]. Observing an approximately 10% higher incidence of CRC in Germany compared to England prior to screening implementation (Figure 2) and constrained by the absence of data on undiagnosed CRC in England, we adjusted the prevalence of undiagnosed cancer in Germany downwards to reflect this difference. Firstly, the German prevalence data[9] were multiplied by 0.9 to represent the likely underlying prevalence in England. The data then were also adjusted to take account of model cycle length limitations arising from calculating the outcomes at the end of the annual model cycle in contrast to the target prevalence data that were sourced from colonoscopy screening study inviting the population through the year. This adjustment was required because it

would be expected that German colonoscopy screening would pick up some people who would normally be diagnosed symptomatically by the end of the year anyway, and on average, half of the people due to present symptomatically that year would be picked up by screening before symptomatic diagnosis. These people should already be counted in diagnosed incidence data for that year, and so should be excluded from the undiagnosed prevalence data. To do this, half of the annual CRC incidence in Germany in 2003-2004 was scaled to the target population size and then subtracted from CRC undiagnosed prevalence data in males and females. For the population of screening start age (55-59 years), this conversion due to undiagnosed prevalence required a multiplier of 0.88 for males and 0.85 for females.

Figure 2. Comparison of CRC incidence in Germany and England



Stage distribution of undiagnosed CRC was based on a German multi-centre cohort study in 2003-2010 [16]. It was assumed that stages I-IV applied in Brenner et al (2016) [16] correspond to stages A to D by Duke's classification (Table 1).

Table 1. Undiagnosed CRC prevalence, target data used in model calibration

Sex and Age	Population	Prevalence				
		Total	stage A	stage B	stage C	stage D
Male, 55-59	236028	1166	595	198	315	58
Male, 60-64	272832	2045	1043	348	552	102
Male, 65-69	281400	2778	1417	472	750	139
Male, 70-75	170073	2506	1278	426	677	125
Female, 55-59	351716	863	440	147	233	43
Female, 60-64	354559	1293	659	220	349	65
Female, 65-69	330965	1729	882	294	467	86
Female, 70-75	187580	1528	779	260	413	76

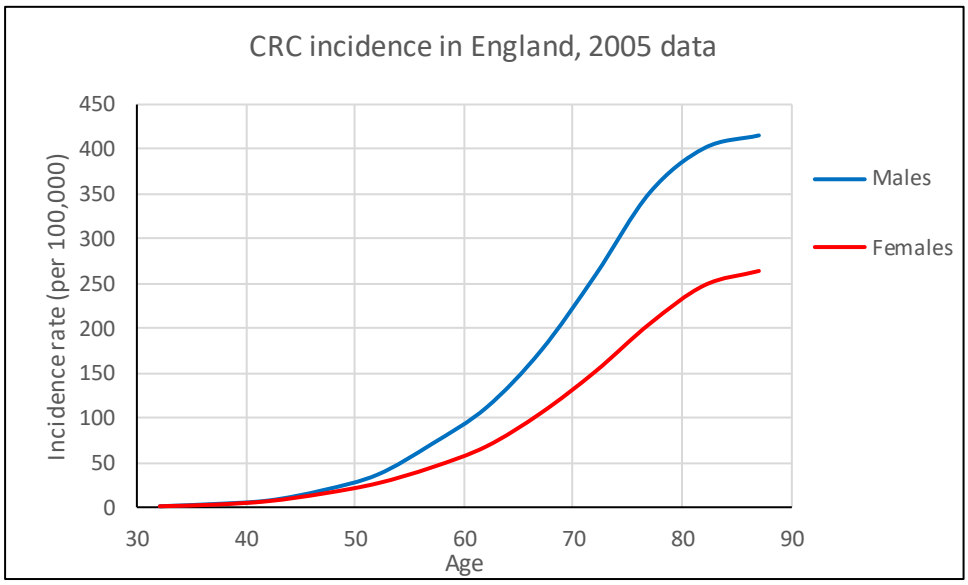
4.3. Calibration targets 8-12: Total CRC incidence, developed through serrated pathways, and by stages

English data from 2005 Cancer Registration Statistics were used together with census data for England (2005) [6 17](Table 2, Figure 3) to assess CRC incidence in the absence of screening. For both males and females 15% of CRC cases were assumed to develop through the serrated neoplastic pathway [5 18].

Table 2. CRC incidence calibration target data [6 17]

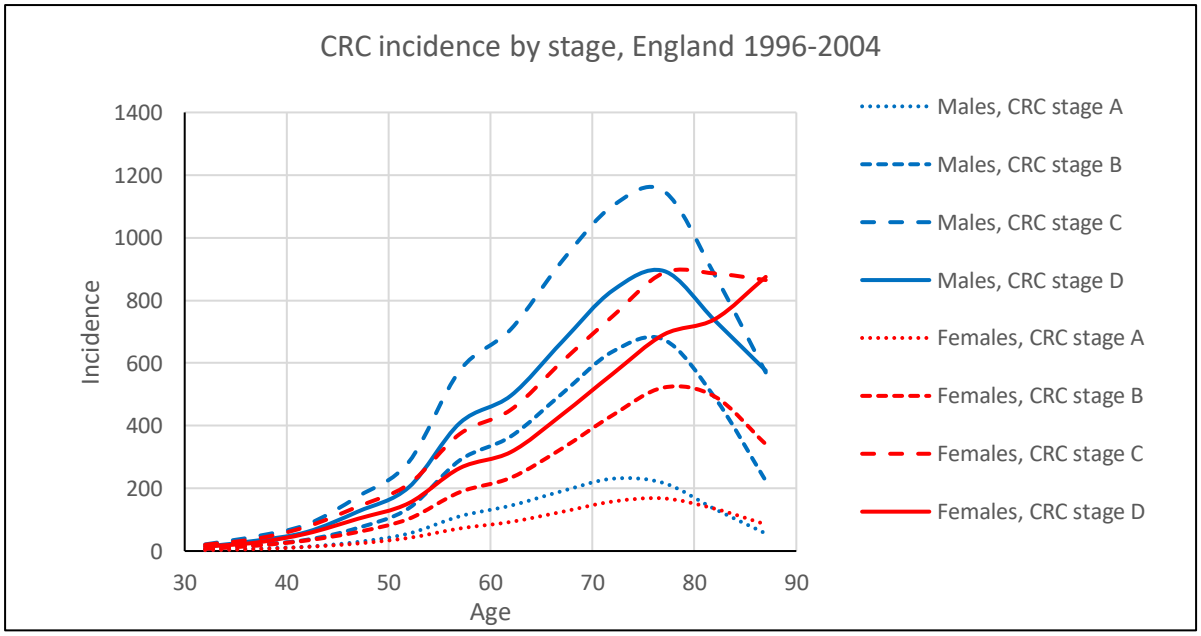
Age	Males				Female			
	All CRC	Population size	Incidence rate (per 100,000)	Incidence rate	All CRC	Population size	Incidence rate (per 100,000)	Incidence rate
30-34	45	2,087,828	2	0.002%	43	2,117,737	2	0.002%
35-39	107	2,305,591	5	0.005%	89	2,339,424	4	0.004%
40-44	196	2,284,554	9	0.009%	181	2,331,691	8	0.008%
45-49	404	2,011,847	20	0.020%	330	2,040,492	16	0.016%
50-54	681	1,820,534	37	0.037%	512	1,852,733	28	0.028%
55-59	1,387	1,936,082	72	0.072%	899	1,978,944	45	0.045%
60-64	1,715	1,528,329	112	0.112%	1,094	1,590,038	69	0.069%
65-69	2,285	1,302,570	175	0.175%	1,491	1,407,541	106	0.106%
70-74	2,802	1,079,412	260	0.260%	1,908	1,253,751	152	0.152%
75-79	2,936	835,292	351	0.351%	2,269	1,104,514	205	0.205%
80-84	2,241	559,108	401	0.401%	2,254	908,552	248	0.248%
85+	1,426	343,698	415	0.415%	2,166	820,546	264	0.264%

Figure 3. CRC incidence used in MiMiC-Bowel model calibration by sex



To include CRC incidence by stage we used the data on stage distribution for the population aged 0 to 100 diagnosed between 1996 and 2004 in England. A review of sex-related differences in CRC incidence in the UK suggested no clear difference by sex for stage I/II or for more advanced stages [19]. The same stage distribution for CRC incidence was applied for males and females to obtain sex specific target data (Figure 4).

Figure 4. CRC incidence by stage



4.4. Compatibility of calibration targets

The data selection process considered multiple characteristics of the sources of calibration targets, including the quality of the reporting of the data and their representativeness to the general population of the country, population sample size, healthcare and population comparability. The compatibility of datasets used in the calibration process was explored through analysis of:

(1) Incidence in the studied population (England) and the population for sourced data (Germany);

(2) Consistency in the direction of change in calibrated parameters to fit each of the multiple databases separately;

(3) Compatibility of the data used in the calibration (prevalence of lesions retrieved from German sources) against other data sources that were not directly used to calibrate the model (predicted faecal immunochemical test [FIT] test sensitivity in England).

5. Algorithm and acceptance metrics

5.1. Goodness-of-fit metric

The calibration process aims to obtain a parameter set with a good fit to each of the calibration targets (12 for each sex). This is achieved by minimising the total sum of squared errors (SSE):

$$Total\ SSE\ = \left(\sum_{i=1}^{\#\ data\ targets} \frac{N\ of\ data\ points_i}{weight_{\ i} \times\ variance_i} \times SSE_i \right) + \left(\sum_{j=1}^{\#\ priors} \frac{prior_j}{weight_j} \right)$$

Where i –the target data set by sex (eg CRC A incidence among females), SSEi – sum of squared errors for each data set i, N of data $points_i$ - number of age groups for which data were reported in the data set i, j – contribution of a weighted prior j to the total SSE, $weight_i$ and $weight_j$ are weightings assigned to each target data set i and each prior j (a measure of how relevant the data are for the model) (Table 3).

Table 3. Weightings, variances and priors used in the calibration

i	Target Data Set i	Weight_i	N of observation points_i	Variance_i	
				Males	Females
1	LRA prevalence	0.5	5	42,280,670	26,101,070
2	HRA prevalence	0.5	5	3,472,459	1,232,347
3	Undiagnosed CRC	0.5	4	17,581	6,967
4	Undiagnosed CRC A	0.5	4	4,574	1,812
5	Undiagnosed CRC B	0.5	4	508	201
6	Undiagnosed CRC C	0.5	4	1,282	508
7	Undiagnosed CRC D	0.5	4	44	17
8	CRC incidence	1	12	30,863	20,882
9	CRC A incidence	0.5	12	176	99
10	CRC B incidence	0.5	12	1,484	929
11	CRC C incidence	0.5	12	4,901	3,301
12	CRC D incidence	0.5	12	2,957	2,259
13	CRC incidence developed through serrated pathways*	1	1	236,926	157,673

The priors	Weight	Mean, (95% CI)		$Alpha, Beta, mC$**	
		Males	Females	Males	Females
LRA prevalence, age≥80	0.2	28% (24-32%)	19% (17-22%)	128.3, 335.2, 271.8	143.1, 595, 361.1
HRA prevalence, age≥80	0.2	8%, (7 -9%)	5% (4.5-6%)	163.4, 1878.2, 408.4	168.5, 3065.8, 451.7

Abbreviations: CRC - colorectal cancer, LR – low-risk, HR – high-risk.

* The variance calculation for CRC incidence developed through serrated pathways was based on assumption of high uncertainty in the expected values, assumed being between 12% and 18%.

** Calculation of the priors components (alpha, beta, and mC) was based on mean prevalence of lesions in age group 75-79 years old and standard distribution assuming the prevalence in age group older than 80 years be within 15% of the prevalence in age group 75-59 years. The calculation of priors using Alpha, beta, and mC components was the following:

$$The\ prior \quad = -[(\alpha - 1)log(prevalence) + (\beta - 1)log(1 - prevalence) - mC]$$

The objective function component variance$_i$ was calculated as the SSE when comparing the target data to 3% error on target data. Hence, the variances scale the SSE's for each data set relative to the number of observations in the data set:

$$variance_i = \sum_{n=1}^{\#\ age\ groups} \left(Occurrence_{n,i} \times 3\%\right)^2$$

Where $variance_i$ - variance for the target data set i, $occurrence_{n,i}$ – number of occurrences in the target data set i in each age group n; n – age group for which occurrence is reported.

This ensures that the objective function will result in a model which fits to all the data sets and avoids domination by the data set with the largest numbers.

Data on the prevalence of LRA and HRA was not available for persons aged 80 and over with the model predicting more than 30% of prevalence events in persons aged 80 comparing to 75 years (what was considered as implausible). Thus, the calibration applied a prior that the prevalence in this age group would be within 15% of that observed for persons aged 75-79 years. The weightings, variances and calculation of contribution of the priors to total SSE are detailed in Table 3.

5.2. Search strategy and acceptance criteria

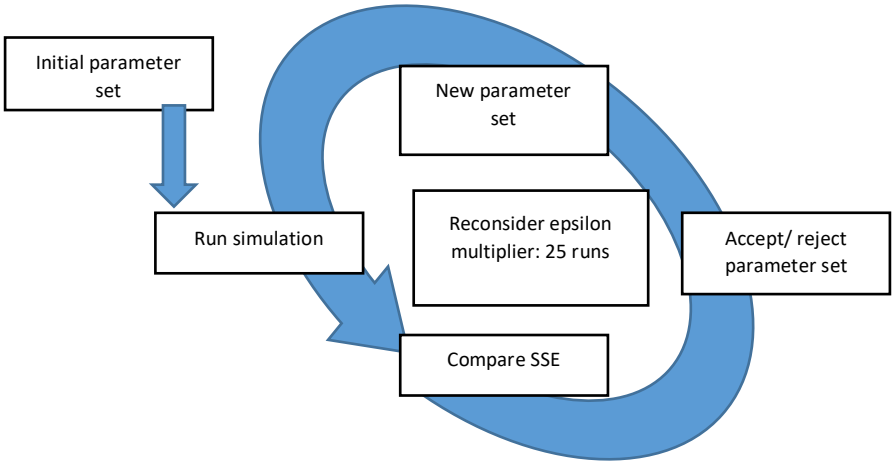
The calibration used the MH algorithm [20] (Figure 5) to estimate the posterior probability distributions of model parameters. An increment, epsilon (initially set at 10% of each parameter's value), was used for each parameter to determine the maximum step size for each iteration of the algorithm. The new parameter set values are obtained by adding a random sample from a uniform distribution (-epsilon, epsilon) to the current parameter values. As the algorithm converges on the solution parameter set it is efficient to reduce the maximum step size. This tuning was achieved by using an epsilon multiplier, which was initially set to one and subsequently decreased by 20% if during the last 25 calibration cycles two or less parameter sets were accepted.

The proposal parameter set in calibration was always accepted if the proposal parameter set had lower total SSE than the current parameter set (p = 1). The probability that the proposed parameter set was accepted when it results in higher SSE than with the current parameter set was:

$$p = e^{(-0.5\times(Total\ SSE(proposed\ set) - Total\ SSE(current\ set)))}$$

Where p – probability of accepting the proposal parameter if the SSE with proposed parameter set is higher than with the current parameter set.

Figure 5. Metropolis–Hastings algorithm to calibrate of model parameters



The probability of accepting a proposal set which increases the objective function by more than 5 units is less than 10%.

5.3. Step-wise calibration approach

As the model is complex (requiring calibration of 47 parameters) and computationally intensive to run (around 26 minutes for one calibration run, i.e. for population of 1.3 million people), an approach of fitting to all data sets simultaneously was found to be unfeasible. Hence a step-wise calibration approach (Table 4) was adopted in this project to enable fit to the multiple data sets by finding a good estimation for the initial set of calibration parameters. The way in which the calibration process was split into consequential steps was based on the nature of CRC development (Figure 6) and the modelling assumptions. Step one applied manual adjustment of the parameter to fit CRC incidence developed via serrated pathways since only one value required adjustment. The purpose of manual adjustment of parameters at step 2 was to retrieve a good starting parameter set to be used in MH-algorithm (step 3), and so the retrieved values of the parameters were not fixed in the following calibration steps. Steps one, three, four, and five resulted in fixing different sets of calibrated parameters to their final values. Step 6 used the final parameter values retrieved from the previous steps (i.e. used a starting parameter set with good fit to all target databases) to retrieve the distributions for all 47 calibrated parameters for probabilistic sensitivity analysis.

Table 4. Step-wise calibration process of MiMiC-Bowel

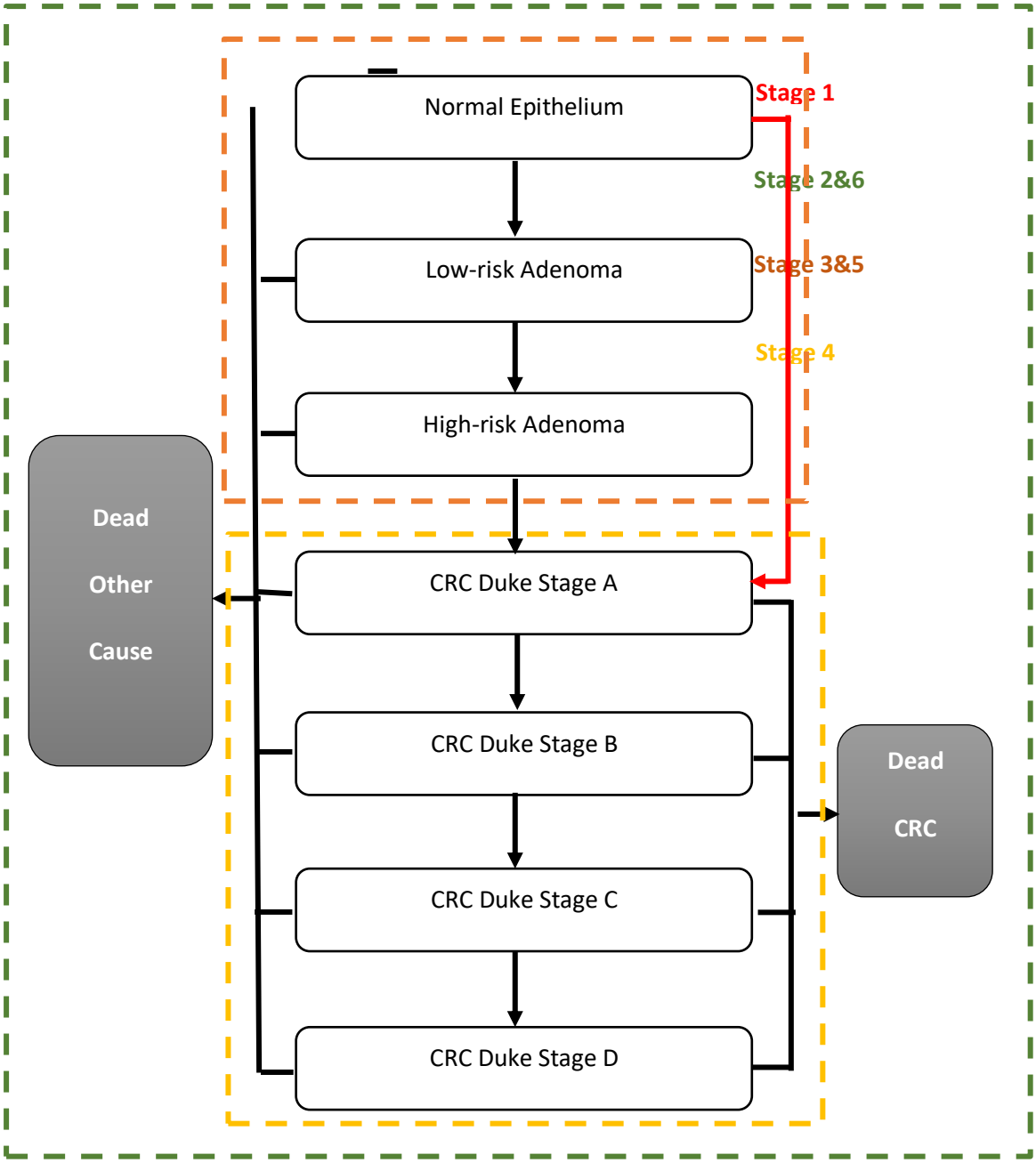
Step	Parameters calibrated	Target data set(s)[1]	Approach	N of parameters
1	Transition from normal epithelium to CRC A	CRC incidence via serrated pathway	Manual[2]	1
2	NHD parameters	Adenoma prevalence, CRC incidence	Manual[2]	43
3	Pre-cancer NHD parameters for males	Adenomas prevalence, CRC incidence, and undiagnosed CRC (males)	MH-algorithm	18
4	Cancer NHD parameters (transitions between cancer stages, symptomatic presentation rate, probability of death from undiagnosed CRC D, decrement of symptomatic presentation among people older than 75 years old)	CRC incidence, undiagnosed CRC and CRC by stages (males)	MH-algorithm	9
5	Pre-cancer NHD parameters for females	Adenoma prevalence, CRC incidence, and undiagnosed CRC (females)	MH-algorithm	18
6	All NHD parameters: pre-cancer and cancer NHD parameters, symptomatic presentation rates, probability of death from undiagnosed CRC D, decrement of symptomatic presentation among people older than 75 years old	All calibration targets: Adenomas prevalence, CRC incidence, undiagnosed CRC, incidence and undiagnosed CRC by stages (males and females)	MH-algorithm	47

[1] See sub-sections on calibration targets explaining how they were derived

[2] Manually adjusting the parameters to fit calibration targets

Abbreviations: CRC – colorectal cancer, MH – Metropolis-Hasting algorithm, NHD – natural history of disease.

Figure 6. Structure of the CRC natural history model health states and calibration steps



6. Calibration outcomes

6.1. CRC natural history model parameter estimates obtained via calibration

The applied calibration framework allowed retrieval of the model parameters (Table 5 and Figure 7). Transition probabilities differed by sex being higher for males transitioning from normal epithelium to LRA and LRA to HRA up to age 63 while being higher for females of all ages transitioning from HRA to CRC.

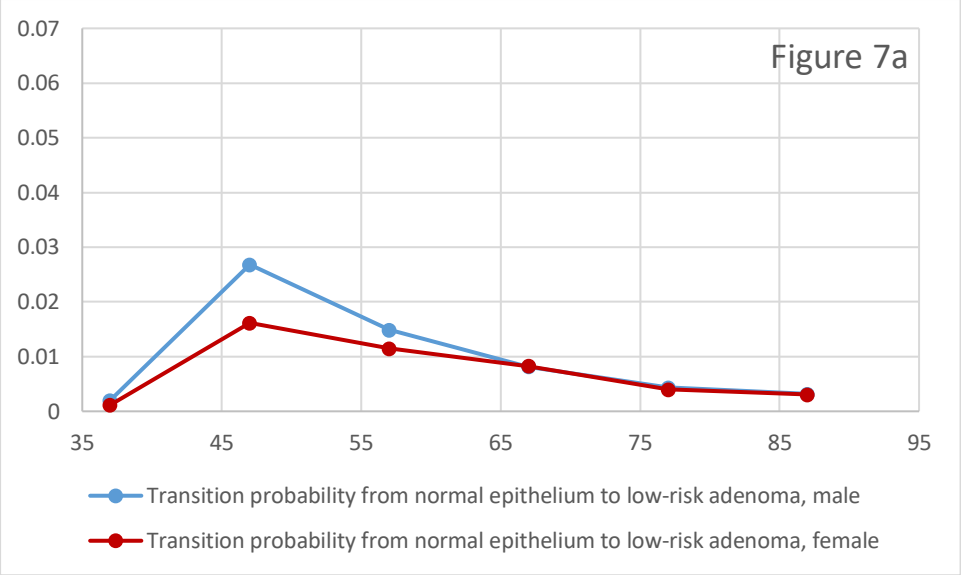
Table 5. Calibrated parameter values

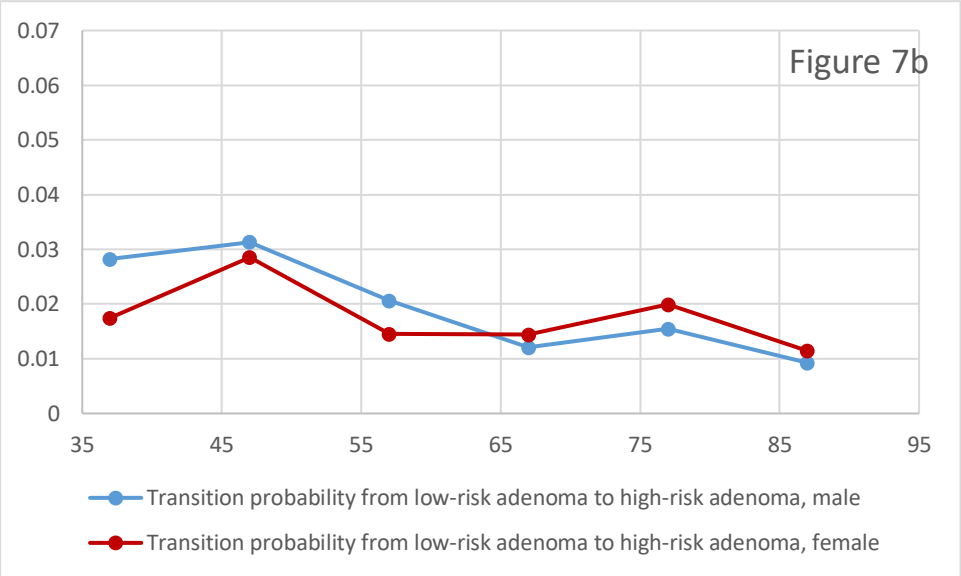
Parameter	Best-fit value	Mean Posterior Estimate (95% Credible Interval Percentiles)
Transition probability from normal epithelium to LRA, male, 37 y.o	0.0020	0.00196 (0.00196, 0.00197)
Transition probability from normal epithelium to LRA, male, 47 y.o.	0.0268	0.02714 (0.02711, 0.02717)
Transition probability from normal epithelium to LRA, male, 57 y.o.	0.0149	0.01498 (0.01497, 0.01500)
Transition probability from normal epithelium to LRA, male, 67 y.o.	0.0082	0.00808 (0.00807, 0.00809)
Transition probability from normal epithelium to LRA, male, 77 y.o.	0.0044	0.00432 (0.00431, 0.00433)
Transition probability from normal epithelium to LRA, male, 87 y.o.	0.0032	0.00324 (0.00323, 0.00324)
Transition probability from LRA to HRA, male, 37 y.o.	0.0282	0.02818 (0.02815, 0.02821)
Transition probability from LRA to HRA, male, 47 y.o.	0.0313	0.03184 (0.03180, 0.03188)
Transition probability from LRA to HRA, male, 57 y.o.	0.0206	0.02069 (0.02067, 0.02072)
Transition probability from LRA to HRA, male, 67 y.o.	0.0121	0.01224 (0.01222, 0.01226)
Transition probability from LRA to HRA, male, 77 y.o.	0.0155	0.01515 (0.01511, 0.01518)
Transition probability from LRA to HRA, male, 87 y.o.	0.0093	0.00924 (0.00923, 0.00925)
Transition probability from HRA to CRC, male, 37 y.o.	0.0092	0.00916 (0.00915, 0.00918)
Transition probability from HRA to CRC, male, 47 y.o.	0.0163	0.01618(0.01616, 0.01620)
Transition probability from HRA to CRC, male, 57 y.o.	0.0181	0.01739 (0.01737, 0.01741)
Transition probability from HRA to CRC, male, 67 y.o.	0.0284	0.02759 (0.02752, 0.02765)
Transition probability from HRA to CRC, male, 77 y.o.	0.0502	0.05120 (0.05112, 0.05128)
Transition probability from HRA to CRC, male, 87 y.o.	0.0352	0.03584 (0.03580, 0.03588)
Transition probability from normal epithelium to LRA, female, 37 y.o.	0.0012	0.00115 (0.00115, 0.00115)
Transition probability from normal epithelium to LRA, female, 47 y.o.	0.0162	0.01638 (0.01636, 0.01640)
Transition probability from normal epithelium to LRA, female, 57 y.o.	0.0115	0.01145 (0.01144, 0.01147)
Transition probability from normal epithelium to LRA, female, 67 y.o.	0.0083	0.00845 (0.00843, 0.00846)
Transition probability from normal epithelium to LRA, female, 77 y.o.	0.0040	0.00389 (0.00388, 0.00389)
Transition probability from normal epithelium to LRA, female, 87 y.o.	0.0030	0.00307 (0.00307, 0.00307)
Transition probability from LRA to HRA, female, 37 y.o.	0.0175	0.01831 (0.01827, 0.01834)
Transition probability from LRA to HRA, female, 47 y.o.	0.0285	0.02838 (0.02834, 0.02841)
Transition probability from LRA to HRA, female, 57 y.o.	0.0145	0.01427(0.01425, 0.01428)
Transition probability from LRA to HRA, female, 67 y.o.	0.0144	0.01416(0.01414, 0.01418)
Transition probability from LRA to HRA, female, 77 y.o.	0.0199	0.01937(0.01935, 0.01940)
Transition probability from LRA to HRA, female, 87 y.o.	0.0114	0.01179 (0.01177, 0.01181)
Transition probability from HRA to CRC, female, 37 y.o.	0.0047	0.00477 (0.00477, 0.00478)
Transition probability from HRA to CRC, female, 47 y.o.	0.0208	0.02032(0.02029, 0.02035)
Transition probability from HRA to CRC, female, 57 y.o.	0.0272	0.02673 (0.02669, 0.02676)
Transition probability from HRA to CRC, female, 67 y.o.	0.0359	0.03647 (0.03642, 0.03653)
Transition probability from HRA to CRC, female, 77 y.o.	0.0650	0.06384 (0.06371, 0.06398)
Transition probability from HRA to CRC, female, 87 y.o.	0.0531	0.05283(0.05276, 0.05289)
Transition probability from normal epithelium to CRC at age 100 (linear gradient with the transition set to zero at age 15)	0.0006	0.00056 (0.00056, 0.00056)
Decrement in symptomatic cancer A,B presentation among people older than 75 years	0.0361	0.03715 (0.037103, 0.03720)
Probability to die undiagnosed at stage D	0.0400	0.04029 (0.04026, 0.04032)
Transition probability from CRC stage A to stage B	0.2932	0.28784 (0.28745, 0.28822)
Transition probability from CRC stage B to stage C	0.5539	0.56207 (0.56104, 0.56310)

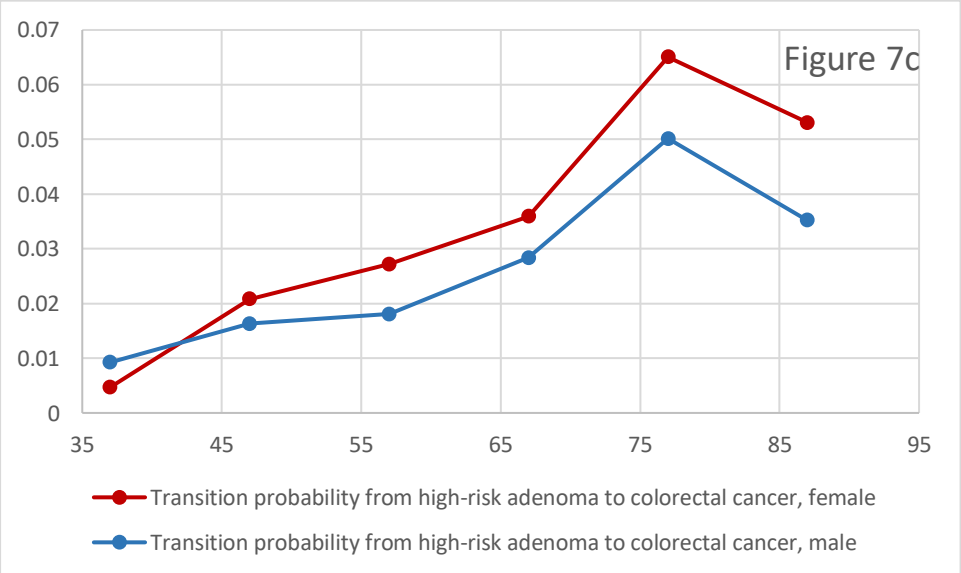
Transition probability from CRC stage C to stage D	0.3500	0.36705 (0.36655, 0.36756)
Probability of being symptomatic, CRC stage A	0.0203	0.02025 (0.02022, 0.02029)
Probability of being symptomatic, CRC stage B	0.1429	0.12981 (0.12965, 0.12996)
Probability of being symptomatic, CRC stage C	0.2741	0.26771 (0.26738, 0.26804)
Probability of being symptomatic among those who die diagnosed, CRC stage D	0.2500	0.20386 (0.20359, 0.20413)

y.o. – years old.

Figure 7. Calibrated transition probabilities of CRC development by age







6.2. Fit of the model to target data sets

The model parameters derived via the calibration process resulted in a good fit between model outcomes and target prevalence of adenomas, CRC incidence, undiagnosed CRC, and CRC incidence by stages (Figure 8). For incidence of CRC, model predictions were within target data confidence intervals for women of all ages except 75-85 year olds (0.7-6.6% difference in the means), and for men from 45 to 80 years old. The prevalence of HRA, which precedes CRC development, was within confidence intervals for all females and males.

Figure 8. Predicted modelling and target outcomes

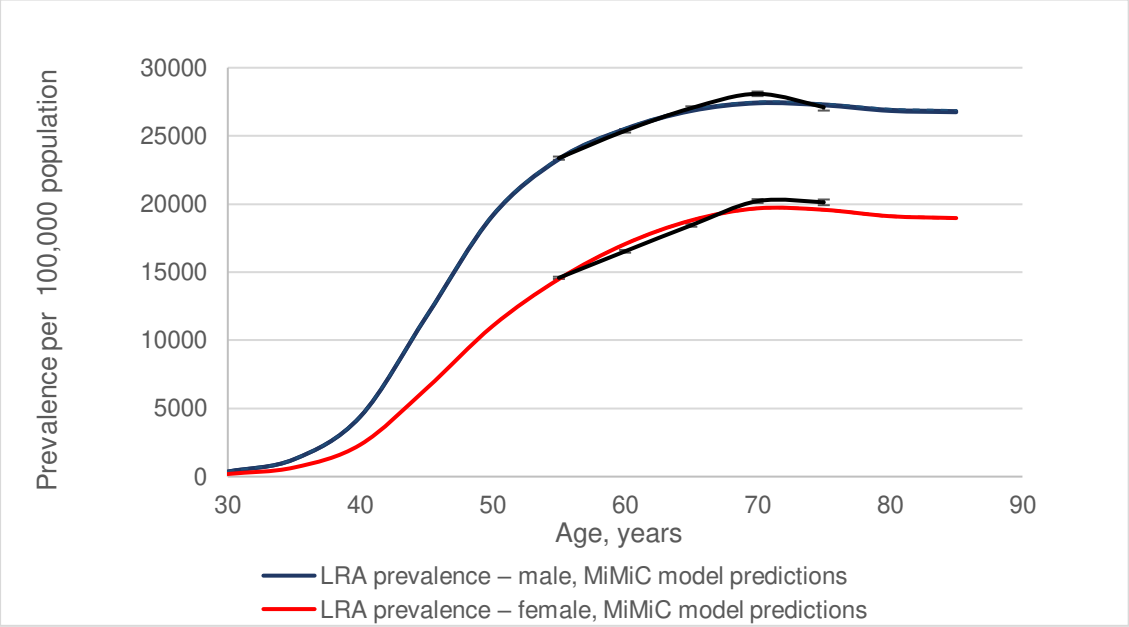


Figure 8a. Prevalence of LRA predicted by the model to target data

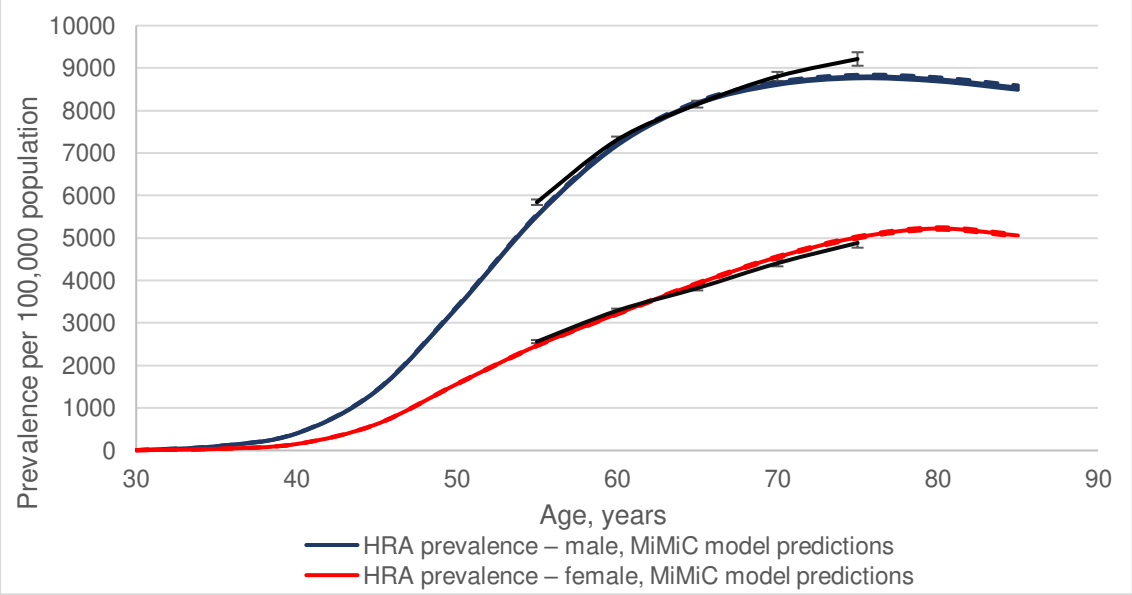


Figure 8b. Prevalence of HRA predicted by the model to target data

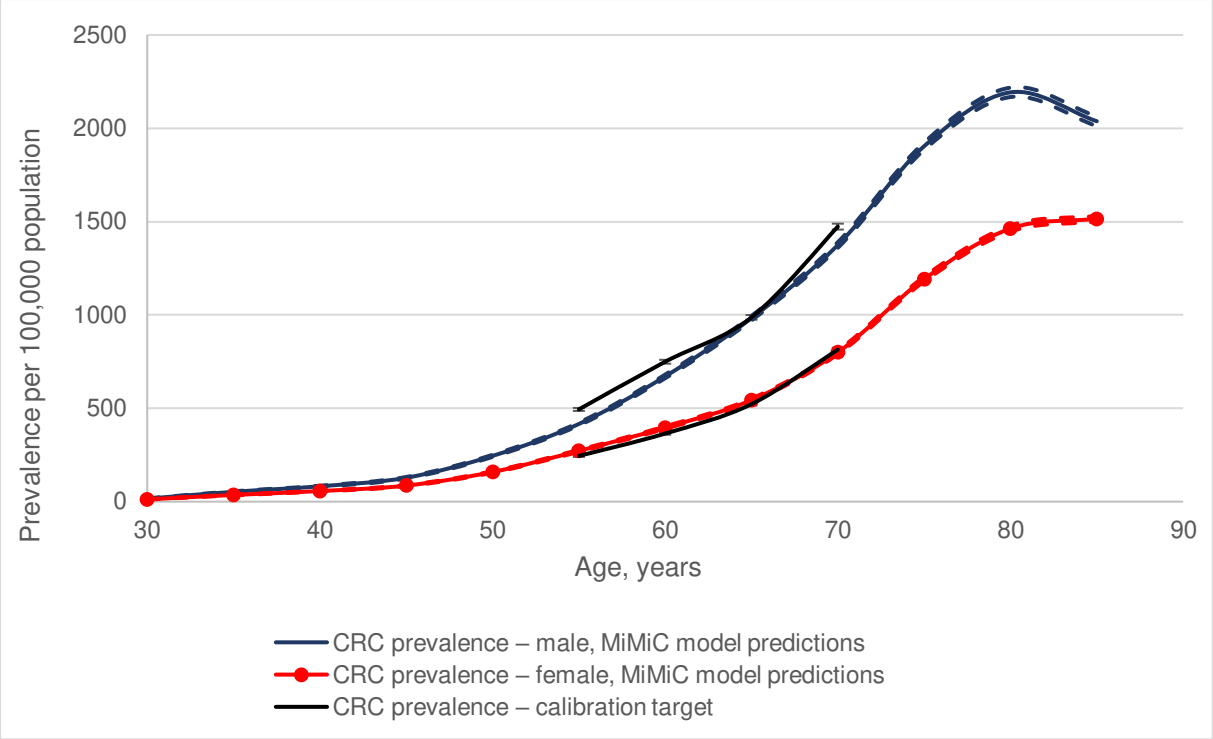


Figure 8c. Prevalence of CRC predicted by the model to target data

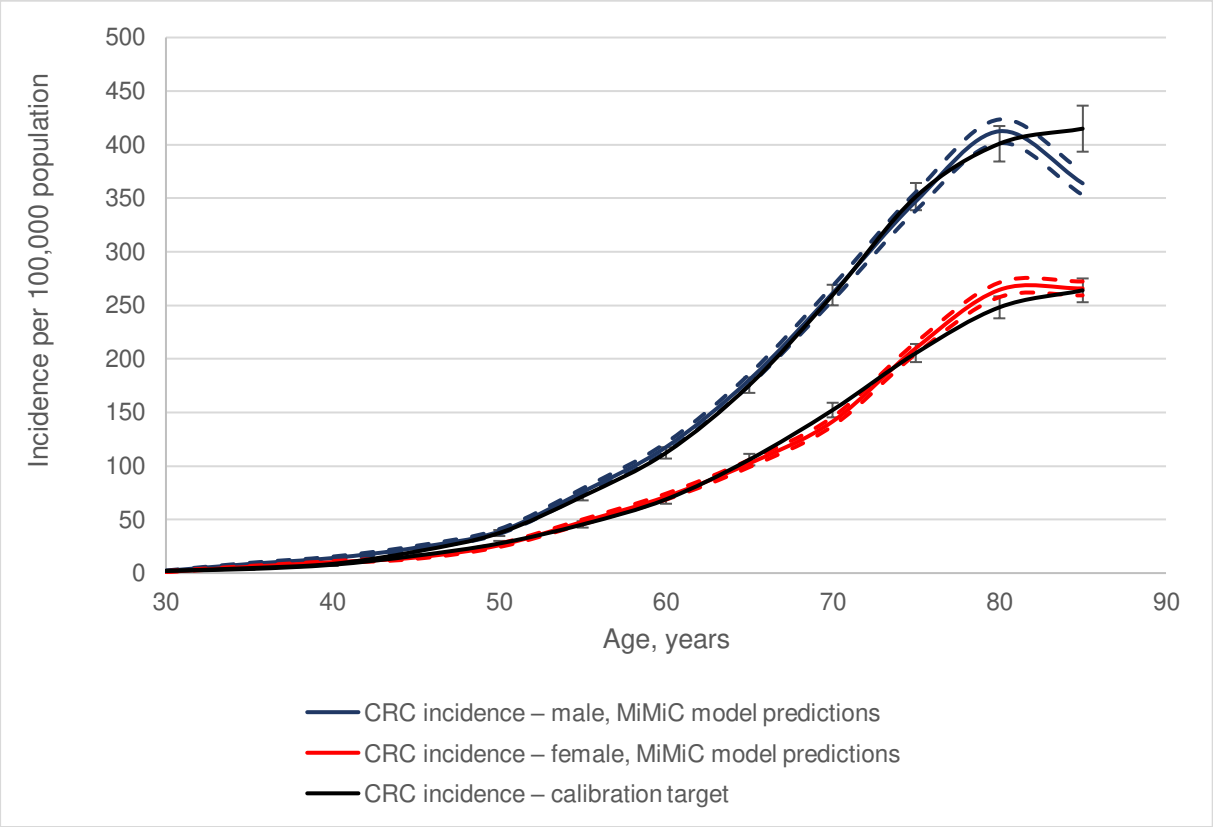


Figure 8d. Incidence of CRC predicted by the model to target data

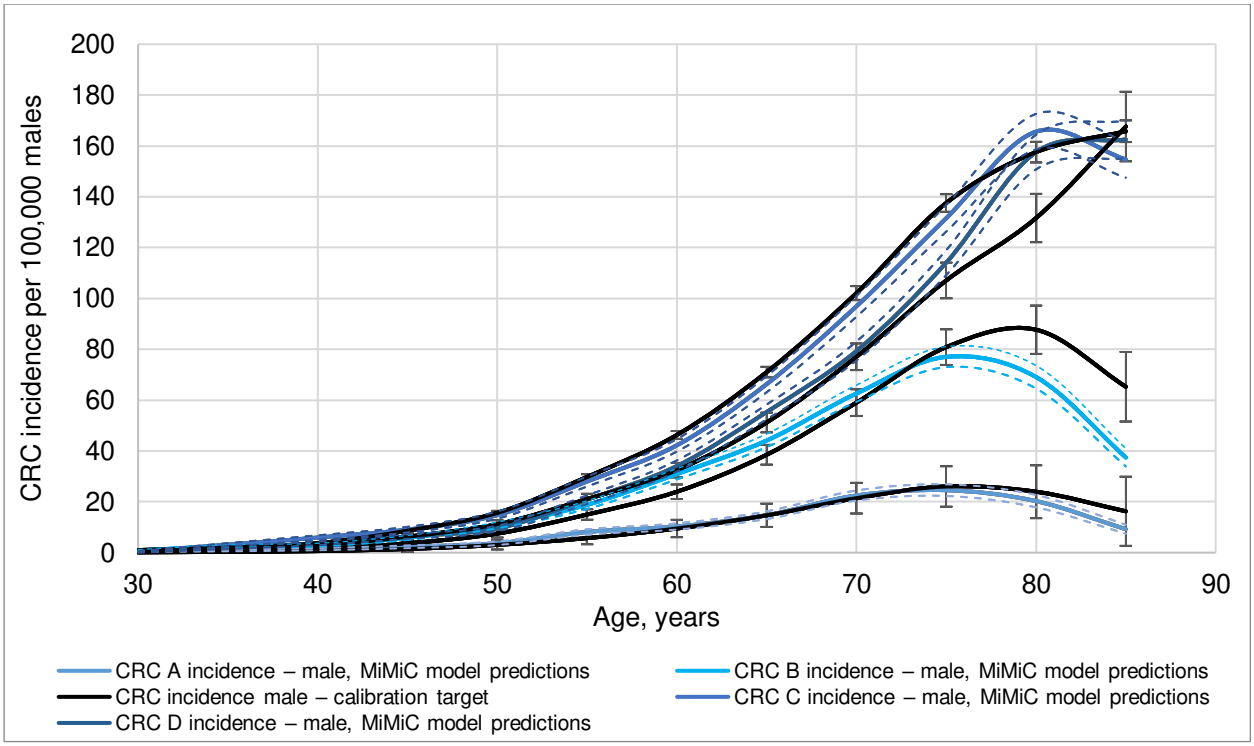


Figure 8e. Incidence of CRC among males predicted by the model to target data

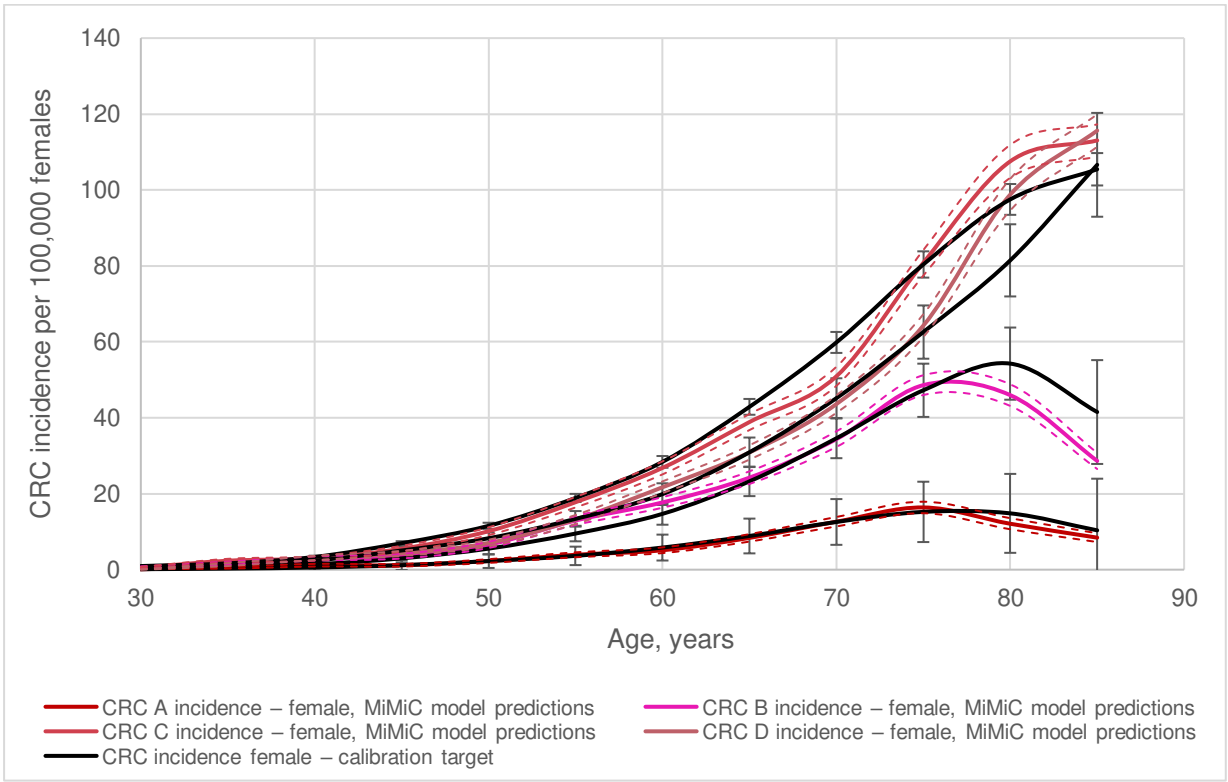


Figure 8f. Incidence of CRC among females predicted by the model to target data

6. Validation of MiMiC-Bowel

The validation of MiMiC-Bowel included comparisons of the following predictions: (a) sensitivity of screening tests (flexible sigmoidoscopy [FS] and FIT20) to plausible sensitivity values (external validity of the model); (b) unobserved outcomes (e.g. time of tumour progression) to predictions from Cancer Intervention and Surveillance Modeling Network (CISNET) models [21 22]; and (c) risks of lesions to estimations in a German model by Brenner et al (2014) [8].

6.1. External validity of screening sensitivity estimated using MiMiC-bowel

For FIT20, validation assumed upper bound of sensitivity to CRC of 60% [23]. The process of calculation of upper bounds of sensitivity rates of FS in population aged 55 years is reported in Box 2; this resulted to the upper bound for sensitivity of FS in population screening was 72% for CRC and 69% for HRA.

Box 2. Steps for calculation of sensitivity of FS used as upper bounds in validation

1. Sensitivity of FS was retrieved from the literature:[12 24] in distal colon – 97% for CRC and 93% for advanced adenoma, and in proximal colon – 21% for both CRC and advanced adenoma.
2. Proportion of proximal and distal lesions in population aged 55 years diagnosed with CRC was assessed from UKFSST data:[25] distal colon/rectum represents 79% of lesions in males and 71% in females.
3. Sensitivity of FS to CRC and advanced adenoma was calculated by weighting to the proportions of proximal and distal lesions: sensitivity to proximal lesions* proportion of proximal lesions + sensitivity to distal lesions *proportion of distal lesions. This resulted in an estimate of the upper bound of sensitivity in the trial (UKFSST) of 78% for CRC and 75% for HRA.
4. The predicted sensitivity of FS in UKFSST was adjusted to a ratio of lesions detection rate in the national bowel cancer screening programme (BCSP) and UKFSST. The comparative performance of BCSP and UKFSST was calculated using the detection rates of CRC (0.17% in BCSP and 0.24% in UKFSST) and the detection rates of HRA (2.20% in BCSP and 2.32% in UKFSST). The division of detection rate of lesions in BCSP (2.37%) and the detection rate of lesions in UKFSST (2.56%), resulted to a multiplier of 92%.

Sensitivity of tests was calculated by dividing detection rates in BCSP (Box 2) and FIT pilot by prevalence of lesions estimated with the target calibration datasets.

Using the BCSP in England as a target, the calibrated model predicted the sensitivity of FS at 60% for CRC and 59% for HRA for 55-year olds, what is below the calculated upper bound values. The predicted sensitivity of FIT20 for population aged 55 years was 48% for CRC and 32% for HRA, what is also below the upper bound values reported in the literature.

6.1. Cross-validity of MiMiC- Bowel model to CISNET models

The first part of the cross-validation was conducted through comparison of the CRC disease natural histories between three models of the CISNET [21 22]. Since no confidence interval was reported for the CISNET model predictions, cross-validation assumed fitting the confidence interval or outcomes generated by MiMiC-Bowel to ranges reported by three CISNET models. Considering lack of information whether cross-validation of CISNET models considered all adenomas as detectable (or only HRA), [21 22] the validation task retrieved predictions of MiMiC-Bowel on all- and high-risk adenomas separately.

Unlike the CISNET models, MiMiC-Bowel predicted considerably slower growth of adenomas among older populations (Table 6). If all adenomas are considered to be detectable, then the predicted values of the proportion of adenomas developed within 10- and 20-years of CRC diagnosis in CISNET models were generally higher than values predicted by MiMiC-Bowel. If only HRA are considered 100% detectable, MiMiC-Bowel predictions were slightly higher than those of CRC-SPIN and SimCRC models, though still lower than the predictions of the MISCAN model [22].

Table 6. CRC cases that arise from potentially detectable adenomas: model predictions for MiMiC-Bowel and CISNET models[21 22]

Age at CRC diagnosis	MISCAN	CRC-SPIN	SimCRC	MiMiC-Bowel, all adenoma*	MiMiC-Bowel, HRA*
Adenomas developed within 10 years of CRC diagnosis (%)					
55	72	3	10	18	48

65	67	4	9	3	28
75	62	4	9	1	23
Percent of adenomas developed within 20 years of cancer diagnosis (%)					
65	94	24	39	47	77
75	92	25	37	16	59
85	89	28	33	8	53

*The 95% confidence interval is narrower than 0.1 unit point and so is not reported here

For the percentage of adenomas developing into CRC over time, MiMiC-Bowel predictions around sojourn and dwell time were closer to CRC-SPIN and SimCRC models than to MISCAN [22] (Table 7); both dwell time and sojourn time were however, higher in MiMiC-Bowel. The CISNET models predicted the cumulative 20-year incidence rate for a 55-year old cohort under a condition of no other cause mortality among the US population (3.2-3.6%), which was higher than those predicted by MiMiC-Bowel in England (Table 8): 2.62%, 95%CI (2.61-2.63) among males and 1.55%, 95%CI (1.50-1.51) among females with the 2005 population cohort. MiMiC-Bowel predicted the 20-year incidence as almost 10-fold higher among persons with adenomas than with normal epithelium at age 55 years: 5.1% vs 0.6%.

Table 7. Model predicted sojourn and dwell time for MiMiC-Bowel and CISNET models[21 22]

Summary measure	MISCAN	CRC-SPIN	SimCRC	MiMiC – Bowel (with the best-fit parameters)	MiMiC – Bowel (mean)*
Adenoma dwell time (from LRA incidence to CRC A)	7.6	24.2	21.2	29.1	29.1
Sojourn time (from preclinical CRC A to CRC diagnosis)	3.0	1.6	4.0	5.3	5.4
Overall dwell time (from LRA to CRC diagnosis)	10.6	25.8	25.2	34.4	34.5

CRC – colorectal cancer; LRA – low-risk adenoma

*The 95% confidence interval is narrower than 0.1 unit point and so is not reported here

Table 8. Cumulative colorectal cancer incidence for 55-year-old cohort

Predicted outcomes	CISNET models, range[22]	MiMiC -Bowel model		
		Persons	Males	Females
20-year cancer incidence for 55-year old cohort	3.2-3.6%	2% (5.1% for those with adenoma at 55, 0.6% with NE at age 55)	2.6% (6% for males with adenoma at 55, 0.7% with NE at age 55)	1.5% (4.2% for females with adenoma at age 55, 0.5% with NE at age 55)
20-year cancer incidence for 55- year old cohort with polypectomy at age 55 years	0.4 – 2.7%	0.9%	1.0%	0.7%
20-year cancer incidence for the whole cohort of 55 year olds	Not reported	0.7%	0.8%	0.6%

Reduction in 20-year cumulative cancer risk, %	30% (MISCAN), 87-89% (CRC-SPIN, SimCRC)	65%	71%	60%

CRC – colorectal cancer, NE – normal epithelium, y.o – year-old

Predictions of 20-year CRC incidence for a 55-year-old cohort with perfect polypectomy at age 55 years, resulted in decreased incidence compared to CISNET models: 0.9%, 95%CI (0.91 – 0.93) in MiMiC-Bowel vs 0.4-2.7%. This reduced 20- year incidence in the whole population (including those who had normal epithelium anyway) to 0.7% (0.6% in females and 0.8% in males). The perfect polypectomy resulted in a 71% reduction in 20-y cumulative cancer risk in MiMiC-Bowel, - within the range 30-89% predicted by CISNET models [22]. The ratio of 20-year CRC risks for people with lesions at 55 and for people with normal epithelium at 55 years was comparable to the MISCAN model (8.5 in MIMIC-bowel vs 7 in MISCAN), while it was much higher in the other two CISNET models (29 in SimCRC and 75 in CRC-SPIN)[22], which could be related to definition of detectable lesions (all lesions vs only HRA). Perfect polypectomy decreased the ratio of the 20-year CRC risks to 1.33 in MiMiC-Bowel. A ratio higher than 1 is expected given that MiMiC-Bowel is based on individual level risks and people who develop adenomas by age 55 years will tend to be higher risk and so more likely to develop CRC in the future than people who did not develop adenomas by age 55 years.

6.2. Cross-validity of MiMiC- Bowel to modelled predictions in Germany

The second part of cross validation included comparing risks of lesions (adenoma, clinically manifest CRC, all CRC cases) up to various ages for men and women free of neoplasm at screening colonoscopy by age, estimated in a German model by Brenner et al (2014) assuming 100% sensitivity for colonoscopy [8]. The prevalence of HRA predicted by MiMiC-Bowel was converted to advanced adenoma rate [13].

For both men and women, MiMiC-Bowel predicted slightly higher risk of CRC incidence among those with negative colonoscopy at defined age than reported by Brenner et al (2014) [8]; this risk though increased with time less rapidly in MiMiC-Bowel than in the German model, resulting in

higher predictions at age 80 years in Germany than England, for each age (55-70 year old) at negative colonoscopy (Table 9).

Table 9. Risk of clinically manifest CRC up to various ages for men and women free of neoplasm at screening colonoscopy by age

Age at negative colonoscopy (years)	Men Risk of clinically manifest CRC (%) up to age (years)				Women Risk of clinically manifest CRC (%) up to age (years)			
	65	70	75	80	65	70	75	80
Brenner et al (2014) [8]								
55	0.09	0.44	1.26	2.48	0.04	0.25	0.82	1.86
60		0.09	0.49	1.26		0.06	0.34	1.03
65			0.11	0.51			0.09	0.44
70				0.10				0.10
MiMiC-Bowel (predictions with the best-fit parameters)								
55	0.18	0.39	0.68	1.0	0.16	0.34	0.60	1.0
60		0.19	0.38	0.61		0.16	0.35	0.64
65			0.20	0.37			0.18	0.38
70				0.19				0.19
MiMiC-Bowel (mean and 95% confidence interval)								
55	0.18 (0.17-0.18)	0.36 (0.35 - 0.37)	0.64 (0.62-0.65)	0.99 (0.97-1.0)	0.15*	0.33 (0.32-0.33)	0.59 (0.58-0.59)	1.0 (0.9-1.0)
60		0.18 (0.17-0.18)	0.34 (0.34-0.35)	0.57 (0.56-0.57)		0.17*	0.34*	0.60 (0.59-0.61)
65			0.18 (0.17-0.18)	0.33 (0.32-0.34)			0.18 (0.17-0.18)	0.35 (0.35-0.36)

70	0.19 (0.18-0.19)	0.18*

CRC – colorectal cancer

*The 95% confidence interval is narrower than 0.01 unit point and so is not reported here

The 10-year risk of all CRC neoplasms and the risk of advanced neoplasms for men and women free of neoplasm at screening colonoscopy (and not undergoing any further CRC screening) was higher in all age groups in the German modelled population. Both cancer prevalence and incidence though were higher in MiMiC-Bowel (Table 10), which shows that in MiMiC-Bowel for a population aged 55 years and older pre-cancer transitions were lower, while adenoma to CRC transitions were higher than in the German model. These differences may be explained by model structures: Brenner's Markov model utilises a simpler five-state structure excluding serrated pathways [8].

Table 10. 10-year risk (%) of colorectal neoplasms for population free of neoplasm at screening colonoscopy

Age at negative colonoscopy	Men				Women			
	Any neoplasm	Advanced neoplasm	Cancer	Clinically manifest cancer	Any neoplasm	Advanced neoplasm	Cancer	Clinically manifest cancer
Brenner et al (2014) model [8]								
55	20.4	3.4	0.28	0.09	13.3	2	0.14	0.04
60	19.9	3.2	0.31	0.09	14	2.1	0.2	0.06
65	19.1	3.2	0.4	0.11	14.2	2.6	0.3	0.09
70	16	2.5	0.33	0.1	12.3	2.1	0.31	0.1
MiMiC – Bowel model (predictions with the best-fit parameters)								
55	11.3	1.81	0.39	0.18	8.4	0.82	0.32	0.16
60	8.4	1.10	0.38	0.19	6.2	0.71	0.35	0.17
65	6.1	0.80	0.38	0.2	5.1	0.55	0.39	0.18
70	4.3	0.59	0.35	0.19	3.7	0.39	0.39	0.19

MiMiC – Bowel model (mean and 95% confidence interval)								
55	11.3*	1.4 (1.3-1.4)	0.35 (0.34-0.35)	0.17*	8.4*	0.79 (0.78-0.79)	0.31*	0.15*
60	8.3*	0.79 (0.78-0.79)	0.35 (0.34-0.35)	0.18 (0.17-0.18)	7.2 (7.1-7.2)	0.69 (0.68-0.69)	0.33*	0.17*
65	6.0 (5.9-6.0)	0.56 (0.55-0.56)	0.36 (0.35-0.36)	0.18 (0.17-0.18)	5.7 (5.6-5.7)	0.61*	0.36*	0.17*
70	4.3 (4.2-4.3)	0.40 (0.39-0.40)	0.36 (0.35-0.37)	0.19 (0.18-0.19)	3.9*	0.46*	0.36*	0.18*

References

1. Health Survey for England 2014: NHS Digital; 2014 [Available from: https://data.gov.uk/dataset/health_survey_for_england.
2. Thomas C, Mandrik, O. and Whyte, S. Development of the Microsimulation Model in Cancer of the Bowel (MiMiC-Bowel), an Individual Patient Simulation Model for Investigation of the Cost-effectiveness of Personalised Screening and Surveillance Strategies, 2020.
3. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019;51(1):76-87. doi: 10.1038/s41588-018-0286-6 [published Online First: 2018/12/05]
4. Atkin WS, Saunders BP. Surveillance guidelines after removal of colorectal adenomatous polyps. *Gut* 2002;51 Suppl 5:V6-9. doi: 10.1136/gut.51.suppl_5.v6 [published Online First: 2002/09/11]
5. Greuter MJ, Xu XM, Lew JB, et al. Modeling the Adenoma and Serrated pathway to Colorectal CAncer (ASCCA). *Risk Anal* 2014;34(5):889-910. doi: 10.1111/risa.12137 [published Online First: 2013/11/01]
6. Cancer Registration Statistics, England, 2005.
7. Starfield B. Threads and yarns: weaving the tapestry of comorbidity. *Ann Fam Med* 2006;4(2):101-3. doi: 10.1370/afm.524 [published Online First: 2006/03/30]
8. Brenner H, Altenhofen L, Stock C, et al. Incidence of colorectal adenomas: birth cohort analysis among 4.3 million participants of screening colonoscopy. *Cancer Epidemiol Biomarkers Prev* 2014;23(9):1920-7. doi: 10.1158/1055-9965.Epi-14-0367 [published Online First: 2014/07/12]
9. Brenner H, Altenhofen L, Hoffmeister M. Sex, age, and birth cohort effects in colorectal neoplasms: a cohort analysis. *Ann Intern Med* 2010;152(11):697-703. doi: 10.7326/0003-4819-152-11-201006010-00002 [published Online First: 2010/06/02]
10. Bretthauer M, Kaminski MF, Loberg M, et al. Population-Based Colonoscopy Screening for Colorectal Cancer: A Randomized Clinical Trial. *JAMA Intern Med* 2016;176(7):894-902. doi: 10.1001/jamainternmed.2016.0960 [published Online First: 2016/05/24]
11. van Rijn AF, Dekker E, Kleibeuker JH. [Screening the population for colorectal cancer: the background to a number of pilot studies in the Netherlands]. *Ned Tijdschr Geneeskd* 2006;150(50):2739-44. [published Online First: 2007/01/18]
12. Martin-Lopez JE, Beltran-Calvo C, Rodriguez-Lopez R, et al. Comparison of the accuracy of CT colonography and colonoscopy in the diagnosis of colorectal cancer. *Colorectal Dis* 2014;16(3):O82-9. doi: 10.1111/codi.12506 [published Online First: 2013/12/05]
13. Atkin W, Wooldrage K, Parkin DM, et al. Long term effects of once-only flexible sigmoidoscopy screening after 17 years of follow-up: the UK Flexible Sigmoidoscopy Screening randomised controlled trial. *Lancet* 2017;389(10076):1299-311. doi: 10.1016/s0140-6736(17)30396-3 [published Online First: 2017/02/27]
14. Atkin W, Wooldrage K, Parkin DM, et al. Long term effects of once-only flexible sigmoidoscopy screening after 17 years of follow-up: the UK Flexible Sigmoidoscopy Screening randomised controlled trial. *The Lancet* 2017;389(10076):1299-311. doi: 10.1016/s0140-6736(17)30396-3
15. Bressler B, Paszat LF, Chen Z, et al. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. *Gastroenterology* 2007;132(1):96-102. doi: 10.1053/j.gastro.2006.10.027 [published Online First: 2007/01/24]
16. Brenner H, Jansen L, Ulrich A, et al. Survival of patients with symptom- and screening-detected colorectal cancer. *Oncotarget* 2016;7(28):44695-704. doi: 10.18632/oncotarget.9412 [published Online First: 2016/05/24]
17. Census. Office for National Statistics. : Office for National Statistics. , 2005.

18. East JE, Atkin WS, Bateman AC, et al. British Society of Gastroenterology position statement on serrated polyps in the colon and rectum. *Gut* 2017;66(7):1181-96. doi: 10.1136/gutjnl-2017-314005 [published Online First: 2017/04/30]
19. White A, Ironmonger L, Steele RJC, et al. A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK. *BMC Cancer* 2018;18(1):906. doi: 10.1186/s12885-018-4786-7 [published Online First: 2018/09/22]
20. Chib S, Greenberg E. Understanding the Metropolis-Hastings Algorithm. *The American Statistician* 1995;49(4):327-35. doi: 10.1080/00031305.1995.10476177
21. 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):604-14. doi: 10.1177/0272989x15622642 [published Online First: 2016/01/10]
22. 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):530-9. doi: 10.1177/0272989x11408730 [published Online First: 2011/06/16]
23. Niedermaier T, Tikk K, Gies A, et al. Sensitivity of Fecal Immunochemical Test for Colorectal Cancer Detection Differs According to Stage and Location. *Clin Gastroenterol Hepatol* 2020 doi: 10.1016/j.cgh.2020.01.025 [published Online First: 2020/01/29]
24. Castro I, Estevez P, Cubiella J, et al. Diagnostic performance of fecal immunochemical test and sigmoidoscopy for advanced right-sided colorectal neoplasms. *Dig Dis Sci* 2015;60(5):1424-32. doi: 10.1007/s10620-014-3434-6 [published Online First: 2014/11/20]
25. Brown JP, Wooldrage K, Kralj-Hans I, et al. Effect of once-only flexible sigmoidoscopy screening on the outcomes of subsequent faecal occult blood test screening. *J Med Screen* 2019;26(1):11-18. doi: 10.1177/0969141318785654 [published Online First: 2018/10/05]