Clarifying Differences in Natural History between Models of Screening: The Case of Colorectal Cancer

Marjolein van Ballegooijen, MD, PhD, Carolyn M. Rutter, PhD, Amy B. Knudsen, PhD, Ann G. Zauber, PhD, James E. Savarino, PhD,† Iris Lansdorp-Vogelaar, PhD, Rob Boer, PhD, Eric J. Feuer, PhD, J. Dik F. Habbema, PhD, Karen M. Kuntz, ScD

Background. *Microsimulation models are important decision support tools for screening. However, their complexity makes them difficult to understand and limits realization of their full potential. Therefore, it is important to develop documentation that clarifies their structure and assumptions. The authors demonstrate this problem and explore a solution for natural history using 3 independently developed colorectal cancer screening models.* ***Methods***. *The authors first project effectiveness and cost-effectiveness of colonoscopy screening for the 3 models (CRC-SPIN, SimCRC, and MISCAN). Next, they provide a conventional presentation of each model, including information on structure and parameter values. Finally, they report the simulated reduction in clinical cancer incidence following a one-time complete removal of adenomas and preclinical cancers for each model. They call this new measure the maximum clinical incidence reduction (MCLIR).* ***Results***. *Projected effectiveness varies widely across models. For example, estimated mortality reduction for colonoscopy screening every 10 years from age 50 to 80 years, with surveillance in adenoma patients, ranges from 65% to 92%. Given only conventional information, it is difficult to explain these differences, largely because differences in structure make parameter values incomparable. In contrast, the MCLIR clearly shows the impact of model differences on the key feature of natural history, the dwell time of preclinical disease. Dwell times vary from 8 to 25 years across models and help explain differences in projected screening effectiveness.* ***Conclusions***. *The authors propose a new measure, the MCLIR, which summarizes the implications for predicted screening effectiveness of differences in natural history assumptions. Including the MCLIR in the standard description of a screening model would improve the transparency of these models.*
Key words: *microsimulation; natural history; colorectal cancer; screening models; standard model output.* ***(Med Decis Making 2011;31:540–549)***

Received 18 July 2010 from the Department of Public Health, Erasmus MC, Rotterdam, the Netherlands (MvB, IL-V, RB, JDFH); Group Health Research Institute, Seattle, Washington (CMR, JES); Institute for Technology Assessment, Massachusetts General Hospital, Boston (ABK); Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York (AGZ); Division of Cancer Control and Population Sciences, National Cancer Institute, Washington, DC (EJF); and Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis (KMK). This work was supported by the National Institutes of Health/National Cancer Institute (U01-CA-088204, U01-CA-097427, U01-CA-097426, and U01-CA-115953). Revision accepted for publication 4 April 2011.
†Deceased.

Address correspondence to Marjolein van Ballegooijen, MD, PhD, Department of Public Health, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands; telephone: +31 10 703 8450; fax: +31 10 703 8451; e-mail: m.vanballegooijen@erasmusmc.nl.

DOI: 10.1177/0272989X11408915

Microsimulation modeling is a widely used tool for decision analyses examining the benefits of cancer screening. Although clinical trials show the effectiveness of interventions for only a few screening strategies, for a limited follow-up period, and for a specific population, models enable the extrapolation of information gained from clinical trials to alternative screening strategies and in a broader population. Several recent studies used models to evaluate the effectiveness and cost-effectiveness of screening for colorectal cancer (CRC)[1–5] and for other cancers.[6–9]

Models are complex, and it takes a major effort from stakeholders, such as clinicians and policy makers, to understand the implications of underlying assumptions and why different models produce different answers. Even individuals with

an understanding of model structure and parameter values find it difficult to grasp the full implications of these assumptions. Developing clear and simple model documentation that demonstrates the impact of assumptions is therefore critical.

In this article, we explore additional ways of documenting the natural history assumptions of screening models. Natural history assumptions are typically used in combination with assumptions about screening to determine the effect of screening on disease outcomes. We focus particularly on assumptions that affect the dwell time of the disease, the preclinical phase when disease is biologically present, but yet does not yield clinical symptoms.[10] Dwell time represents the time frame in which the disease could be caught by screening and therefore is a key factor for the potential impact of screening.

To allow model comparisons, a great deal of effort has been put toward improving the description of *model structure and model inputs*. Although all inputs of a model should always be reported (e.g., in an online appendix),[6] assumptions on dwell time and their implications remain difficult to assess even with this documentation. Another level of documentation has been proposed,[11,12] focusing on specific model *predictions*. We used this approach to describe the impact of model assumptions about the dwell time of preclinical disease on screening effectiveness. Specifically, we explored one possible prediction measure, the maximum clinical incidence reduction (MCLIR). This measure represents the reduction in incidence of clinically diagnosed disease following a one-time complete removal of preclinical disease. As an example, we applied this method to 3 independently developed CRC screening models.

The measure we propose in this article, the MCLIR, is a summary measure. For more detailed results, especially for subgroups of individuals with and without adenomas or cancer at the time of the hypothetical one-time complete disease removal, we refer to the analyses presented in a companion article published in this issue of *Medical Decision Making*.[13]

METHODS

We demonstrate the limitations associated with model description using the 3 CRC models. Each of these models was developed as part of the National Cancer Institute's (NCI's) Cancer Intervention and Surveillance Modeling Network (CISNET). To enhance the transparency of the models, we created structured model descriptions, referred to as ''model profiles.'' These profiles are posted online (http://cisnet.cancer.gov/profiles/). Despite the detail provided in these model profiles, the implications of differences in natural history assumptions only became apparent to us, collaborating modelers, after carrying out specific model predictions. We use these models to demonstrate how our proposed measure of MCLIR adds value to the description of the model structure.

The 3 models are Microsimulation Screening Analysis (MISCAN), from the Erasmus University Medical Centre, Department of Public Health; Colorectal Cancer Simulated Population model for Incidence and Natural history (CRC-SPIN), from the Group Health Research Institute, Seattle, Washington; and Simulation Model of Colorectal Cancer (SimCRC), from the University of Minnesota and Massachusetts General Hospital. Although independently developed, the 3 microsimulation models have much in common. Each model describes the natural history of CRC in individuals using a structure that builds on the adenocarcinoma sequence.[14] Individuals are simulated one by one from a young age to death. At age 20 years, individuals have no adenomas. From there onward, they are at risk of developing new adenomas in the colorectum; the risk depends on age. Multiple adenomas may form within an individual, and in the absence of screening, adenomas grow in size and may eventually transform into preclinical invasive CRC. These preclinical cancers may later become clinically detected through presentation of symptoms. Once a cancer becomes clinically detected, a stage-specific survival function determines whether and when the simulated individual dies of CRC. Simulated individuals are also at risk of dying from other causes. As a consequence, even an individual diagnosed with incurable CRC may die from another cause before dying from the CRC.

The MISCAN and SimCRC models do not explicitly define the starting size of an adenoma. But because the models were calibrated to adenoma prevalence data (described below), this implies that simulated adenomas are initiated when they are macroscopically detectable. The CRC-SPIN model initiates adenomas at 1 mm, explicitly assuming that this is the earliest point of detection.

The 3 models were calibrated to the same CRC incidence and adenoma prevalence data. CRC incidence was calibrated based on Surveillance,

Epidemiology, and End-Results (SEER) CRC incidence rates in 1975 to 1979 because this represents CRC incidence in the United States when there was little or no CRC screening.[15] Adenoma prevalence rates were based on autopsy data that reported adenoma information by age and sex.[16]

In the Results section, we first provide the projected effectiveness and cost-effectiveness of colonoscopy screening for the 3 models, showing how they differ. Next, we provide a conventional description of each model, including information that would usually be published with a decision analysis. Finally, for each model, we provide the background incidence (no intervention) by age and the simulated MCLIR.

The Population and the Comparator Situation

For all 3 models, the simulated population represents the general US population that is 25 years and older. By calibrating the models without intervention to SEER CRC incidence rates in 1975 to 1979, we assumed this would be the incidence level if there were no screening for CRC. None of the models explicitly simulated incidental detection of symptomless disease; it was included in the observed 1975 to 1979 CRC incidence to which the models were calibrated for the incidence without screening.

The Effectiveness and Cost-Effectiveness of Screening

We examined the effectiveness and cost-effectiveness of screening as predicted by each model for colonoscopy screening every 10 years beginning at age 50 and ending at age 80 years.[17,18] Under this scenario, polypectomy was followed by colonoscopy surveillance. We assumed that the sensitivity of colonoscopy was 75% for small adenomas (<5 mm), 85% for medium adenomas (6−9 mm), and 95% for large adenomas (10+ mm) and cancers. We also assumed that colonoscopies are complete (reach the cecum) in 95% of the procedures, and the reach in the remaining 5% is distributed evenly over the colorectum. When an adenoma is detected, it is completely removed. When preclinical cancer is detected, the models simulate a stage-specific survival that (stochastically) replicates survival as observed in SEER (SEER 1997–2001[15]). All models use the same stage-specific survival distributions for screen- and clinically detected cancers. Finally, all models assumed 1 fatal complication per 10,000 screening or surveillance colonoscopies in which at least 1 adenoma or cancer is detected and removed.

We use life years gained to measure effectiveness and the projected number of colonoscopies as our cost, with cost-effectiveness measured by the number of colonoscopies per 1000 life years gained. We also estimate the projected incidence and mortality reductions relative to no screening.

Conventional Model Descriptions

We provide descriptions of each natural history model, drawn from their CISNET profiles and published articles, at a level of detail that would normally be accepted in a published article.

Maximum Clinical Incidence Reduction (MCLIR)

The prediction measure MCLIR is defined as the simulated incidence reduction in the remaining years until, for example, age 80 years, after complete removal of all prevalent detectable preclinical disease, relative to the background (no intervention) disease incidence, and is reported as a percentage. To estimate the MCLIR in CRC, we simulated 2 cohorts. The first had no intervention and was used to provide background incidence. The second simulated cohort had preclinical disease (adenomas and preclinical cancers) completely removed (100% sensitivity, 100% cure) at a given age.

For the example presented here, we placed the complete removal of disease at age 65 years, in the middle of the 50- to 80-year age range often recommended for average-risk individuals. The resulting MCLIR provides a direct measure of the impact of the natural history of disease after age 65 years because we have eliminated variations due to test sensitivity, repeated screening, and surveillance. Models projected CRC incidence after the intervention, which included only clinically diagnosed cases since there was no further screening (and the detected preclinical cancers at age 65 years were not included). When calculating CRC incidence, individuals were removed from the risk set (censored) once they transitioned to clinically diagnosed CRC or died from other causes.

Several simulation issues need clarification because they affect the MCLIR. First, it must be clear where detectability starts (in the case presented: at the onset of small adenomas as defined in our example models). Next, the composition of the population addressed must be determined (e.g., the distribution of gender and categories of risk factors). Finally, it must be clear whether and how the model accounts for incidental detection of asymptomatic

(preclinical) disease (e.g., by imaging procedures for other diseases).

Dwell time is often modeled in an implicit way, for example, by specifying the proportions of individuals who transition in and out of disease states based on probabilities that vary by age and other factors. But even when time to progression is a direct model input, implied differences in dwell time across models can be difficult to determine because of structural differences. The MCLIR, however, relies only on model outputs. The definition of the simulation run is relatively simple (no missed cases and complete cure of cancers and precancerous lesions that are prevalent at the time of the one-time, perfect intervention), and the output considered is driven by dwell time (''what percent of the clinical cancers at each age had a dwell time shorter than *x* years?''). It shows the potential of screening if its effectiveness were strictly limited by the length of the preclinical phase: Even the perfect test could not detect the disease earlier because detectable disease was not yet in existence.

Moving the age of adenoma removal from age 65 years to another age will affect the MCLIR depending on how dwell time parameters vary by age. In the event of such dependency, and if detailed documentation is required, presenting the MCLIR for different ages would provide additional information. However, there is a limitation in the amount of useful information that can be displayed. For a standard manuscript, we would propose the ''mid-screening-age-only'' approach presented here.

We calculated the MCLIR for 1-year periods, for 5-year periods, and for the period from the intervention up to age 80 years (in this case, a 15-year period). To calculate the MCLIR for an *x*-year period, we cumulated the incidence reduction percentages for each year in that period and divided the result by the number of years. For the reduction between 5 and 10 years after disease removal at age 65, we use the following notation: $MCLIR_{65}^{5-10}$. We propose this notation as a standard way to report the MCLIR.

RESULTS

Effectiveness and Cost-Effectiveness Results

Simulation of colonoscopy screening every 10 years from age 50 to age 80 years shows differences in projected effectiveness, with CRC-SPIN and SimCRC projecting greater effectiveness than MISCAN (Table 1). The incidence reduction varied from 52% to 91% and mortality reduction from 65% to 92%, with SimCRC and CRC-SPIN predicting the largest reductions. The life years gained (from 207 to 327 per 1000 individuals) and the number of colonoscopies per life year gained (from 19 to 13) varied accordingly: fewer life years gained and more colonoscopies per life year gained in MISCAN. In general, the results of CRC-SPIN and SimCRC are close to each other.

Table 1 Model Outcomes for Colonoscopy Screening Every 10 Years from Age 50 to 80 Years, with Surveillance in Adenoma Patients

	MISCAN	CRC-SPIN	SimCRC
Incidence[a] reduction, %	52	91	82
Mortality[a] reduction, %	65	92	84
Life years gained, No.[b]	207	260	327
Screening colonoscopies, No.[b]	2288	2580	2574
Surveillance colonoscopies, No.[b]	1715	1341	1609
Total colonoscopies, No.[b]	4002	3921	4184
CRC incidence cases prevented, No.[b]	30	56	54
CRC mortality cases prevented, No.[b]	19	25	30
Colonoscopies per incidence case prevented, No.	135	70	77
Colonoscopies per life year gained, No.	19	15	13

CRC, colorectal cancer; MISCAN, Microsimulation Screening Analysis; CRC-SPIN, Colorectal Cancer Simulated Population model for Incidence and Natural history; SimCRC, Simulation Model of Colorectal Cancer.
[a]25+ years of age.
[b]Per one thousand 25-year-old individual's lifetime.

Conventional Natural History Model Descriptions

The following natural history descriptions were retrieved from recent articles.[5,19,20] Model parameters are described in Tables 2 to 4. For additional detailed information concerning the respective models, we refer to the CISNET model profiler (http://cisnet.cancer.gov/profiles/).

The MISCAN model. The MISCAN-Colon natural history parameters are presented in Table 2. A person-specific risk index is generated for each individual in the simulated population. Subsequently, adenomas are generated in the population according to this person-specific risk index and an age-specific

Table 2 Conventional Model Description: Basic Structure of the Microsimulation Screening Analysis (MISCAN)–Colon Natural History Model[5]

Model Parameter	Value
a) Distribution of risk for adenomas over the general population	Gamma distributed, mean 1, variance 2
b) Probability that a new adenoma is progressive	Dependent on age at onset: 0–45 years: linearly increasing from 0 to 3% 45–65 years: linearly increasing from 3% to 17% 65–100 years: linearly decreasing from 17% to 13%
c) Mean duration of development of progressive adenomas to clinical cancer	20 years (without competing other-cause mortality)
d) Mean duration of preclinical cancer	3.6 years
e) Mean duration of adenoma	16.4 years
f) Percentage of nonprogressive adenomas that do not develop beyond 6 to 9 mm	50%
g) Percentage of nonprogressive adenomas that become 10 mm or larger	50%
h) Percentage of cancers that develop from a 6- to 9-mm adenoma and from a 10+-mm adenoma	30% of cancer develops from 6 to 9 mm, 70% from 10+ mm

incidence rate of adenomas. This results in no adenomas for most persons and 1 or more adenomas for others. Adenomas can progress in size from small (1−5 mm) to medium (6−9 mm) to large (10+ mm). Most adenomas will never develop into cancer (nonprogressive adenomas), but some (progressive adenomas) may eventually become malignant, transforming from medium or large size to a preclinical stage I cancer. The cancer may then progress from stage I to stage IV. In every stage, there is a probability of the cancer being diagnosed because of symptoms.

The CRC-SPIN model. The CRC-SPIN natural history is described in Table 3. The risk of an observable adenoma ($\geq$1 mm) is modeled using a nonhomogeneous Poisson model that allows adenoma risk to depend on gender and to increase with age. The log-risk of adenoma occurrence varies across individuals and has a Normal distribution. Each adenoma is stochastically assigned a time to reach 10 mm, based on a type II extreme value distribution. Adenoma size is modeled continuously in time using a Janoschek growth curve model,[21,22] with growth rates determined by a transformation of the time to 10 mm. CRC-SPIN allows direct transition to CRC from adenomas of any size. The probability that an adenoma transitions to preclinical cancer increases with size and the age of the person at adenoma initiation. The time from preclinical cancer to clinical cancer is modeled using a lognormal distribution. Adenoma growth, the probability of transition to preclinical cancer, and dwell time all depend on adenoma location (colon or rectum). Once a cancer becomes clinically detectable, it is stochastically assigned a size and stage at clinical detection. The CRC-SPIN model was developed to examine CRC screening and therefore projects CRC incidence and mortality through age 85.

The SimCRC model. The SimCRC natural history is described in Table 4. A simulated individual is stochastically assigned a person-specific risk index. Over time, each person is at risk for forming 1 or more adenomas, based on his or her risk index, age, and gender. Each adenoma may grow in size from small ($\leq$5 mm) to medium (6−9 mm) to large ($\geq$10 mm). Medium-size and large adenomas may directly transit to preclinical colorectal cancer. Preclinical cancers may progress in stage (I–IV) and may be detected by the presence of symptoms, becoming a clinical case.

The SimCRC model allows for heterogeneity in growth and progression rates across multiple adenomas within an individual. Although all adenomas have the potential to develop into colorectal cancer, most will not do so in an individual's lifetime. The likelihood of adenoma growth and progression to colorectal cancer is allowed to vary by location in the colorectal tract (i.e., proximal colon v. distal colon v. rectum).

Dwell time in the 3 models. The dwell time from adenoma onset to clinical cancer in each model is influenced by all the input parameters listed in the respective Tables 2 to 4, with the exception of the

Table 3 Conventional Model Description: Basic Structure of the CRC-SPIN Natural History Model[19]

Adenoma Risk: Nonhomogeneous Poisson Process

Log-risk for the ith individual =

$$\alpha_{0i}+\alpha_1 sex_i+\sum_{k=1}^{4}\delta(A_k<age_i(t)\le A_{k+1})\left\{age_i(t)\alpha_{2k}+\sum_{j=2}^{k}A_j(\alpha_{2j-1}-\alpha_{2j})\right\}$$

- Baseline log-risk, α_{0i}, is Normally distributed, mean Λ, standard deviation σ
- $\delta(\cdot)$ is an indicator function with $\delta(x) = 1$ when x is true and $\delta(x) = 0$ otherwise
- $age_i(t)$ is the ith individual's age at time t
- $A_1 = 20$, $A_2 = 50$, $A_3 = 60$, $A_4 = 70$, $A_5 = \infty$ (effectively 100 years old)

Calibrated parameters: Λ, σ, α_1, α_{21}, α_{22}, α_{23}, and α_{24}

Adenoma Growth: Janoshek Growth Curve

$d_{ij}(t) = d_\infty - (d_\infty - d_0)\exp(-\lambda_{ij}t)$

- $d_{ij}(t)$ is the maximum diameter of the jth adenoma in the ith individual at time t after initiation.
- $d_0 = 1$ mm, minimal detectable adenoma size
- $d_\infty = 50$ mm, maximum adenoma size
- Time to reach 10 mm: $-\ln((d_\infty - 10)/(d_\infty - d_0))/\gamma$

Time to Reach 10 mm: Type 2 Extreme Value Distribution

Adenomas in the colon: distribution parameterized by β_{1c} and β_{2c}

Adenomas in the rectum: distribution parameterized by β_{1r} and β_{2r}

Calibrated parameters: β_{1c}, β_{2c}, β_{1r}, and β_{2r}

Transition to Preclinical Cancer: Normal Cumulative Distribution

Probability of transition, male colon	$\Phi(\{\ln(\gamma_{1cm}size)+\gamma_{2cm}(a-50)\}/\gamma_3)$
Probability of transition, male rectum	$\Phi(\{\ln(\gamma_{1rm}size)+\gamma_{2rm}(a-50)\}/\gamma_3)$
Probability of transition, female colon	$\Phi(\{\ln(\gamma_{1cf}size)+\gamma_{2cf}(a-50)\}/\gamma_3)$
Probability of transition, female rectum	$\Phi(\{\ln(\gamma_{1rf}size)+\gamma_{2rf}(a-50)\}/\gamma_3)$

Where $\Phi(\cdot)$ is the standard Normal cumulative distribution function, *size* is adenoma size in mm, and a is age at adenoma initiation.

Calibrated parameters: γ_{1cm}, γ_{2cm}, γ_{1rm}, γ_{2rm}, γ_{1cf}, γ_{2cf}, γ_{1rf}, γ_{2rf}, and γ_3

Sojourn Time: Lognormal Distribution

Preclinical colon cancer, lognormal with mean μ_c, standard deviation $\tau_c\mu_c$

Preclinical rectal cancer, lognormal with mean μ_r, standard deviation $\tau_r\mu_r$

Calibrated parameters: μ_c, τ_c, μ_r, and τ_r

CRC-SPIN, Colorectal Cancer Simulated Population model for Incidence and Natural history.

parameters on (fractions of) nonprogressive adenomas in MISCAN (items b, f, and g in Table 2).

Maximum Clinical Incidence Reduction (MCLIR)

Table 5b presents the MCLIR after disease removal at age 65 years for each model in the successive 5-year periods from 65 years onward. SimCRC and CRC-SPIN models projected large $MCLIR65^{0-15}$ values for the 15 years until age 80 years (90% and 88%, respectively). The MISCAN model projected a $MCLIR65^{0-15}$ of 51%. The differences show that the assumptions of the first 2 models imply a longer preclinical phase, because more adenomas and polyps are caught in the preclinical stage at age 65 years and are prevented from progressing to clinical cancer. Figure 1 shows these results as the percentage of the background incidence (100% – MCLIR) by year. This figure demonstrates that the MISCAN model projected that 62% of cancers in individuals 75 years old developed within 10 years (namely, after age 65 years), whereas the CRC-SPIN and SimCRC models projected that only 4% and 9% of cancers in individuals 75 years developed within 10 years, respectively. These differences show why CRC-SPIN and SimCRC project greater effectiveness and more favorable cost-effectiveness of screening than MISCAN.

As shown in Table 5a, the differences in MCLIR between CRC-SPIN and SimCRC, on one hand, and MISCAN, on the other cannot, be explained by differences in background incidence since these rates are very similar. This combination of similar incidence and different dwell times is not surprising: A given incidence curve can be reproduced by

Table 4 Conventional Model Description: Basic Structure of the SimCRC Natural History Model

Adenoma risk index: Drawn for each individual from a normal distribution truncated to nonnegative values with a mean of 1 and a variance ranging from 0.020 to 0.021.
Adenoma growth index: Drawn for each adenoma from a normal distribution truncated to nonnegative values with a mean of 1 and a variance ranging from 0.392 to 0.438.
Adenoma growth: Location- and sex-specific annual probability that adenoma of a given size category transitions to the next size category, ranging from 0.038 to 0.603 (adenoma <6 mm) and 0.009 to 0.368 (adenoma 6−9 mm).
Transition to preclinical CRC: Sex- and location-specific probability that a 10+ mm adenoma transitions to a stage I preclinical CRC. Modeled as a logistic function of age with intercept ranging from −5.812 to −4.479, coefficient on age ranging from 0.008 to 0.058, and coefficient on the square of age ranging from -4.5×10^{-5} to -7.0×10^{-6}.
Progression of preclinical (i.e., undiagnosed) CRC: Location-, sex-, and stage-specific annual probability that a preclinical CRC progresses to the next stage, ranging from 0.241 to 0.388 (stage I), 0.275 to 0.654 (stage II), and 0.509 to 0.769 (stage III).
Symptom detection: Location-, sex-, and stage-specific annual probability that a preclinical CRC is detected by symptoms, ranging from 0.030 to 0.192 (stage I), 0.204 to 0.523 (stage II), 0.520 to 0.751 (stage III), and 0.910 to 0.974 (stage IV).

CRC, colorectal cancer; SimCRC, Simulation Model of Colorectal Cancer.

very different dwell times by adjusting the (unobservable) age of onset of adenomas that make it to clinical cancers. Given the similarity in age trend of the background incidence, the differences in MCLIR are determined entirely by dwell time assumptions.

DISCUSSION

A complete description of the structure and inputs of a model is necessary but limited in the insights that are provided for natural history models. In correspondence with the idea of developing standard outputs to describe models, we propose to include the projected MCLIR as a prediction measure in the description of screening models. By comparing the MCLIRs between models, the implications of differences in natural history and, more specifically, the models' implicitly assumed length of the preclinical disease phase become apparent. The MCLIR should be relatively easy to calculate with any screening model. Because it is based on model output instead of input, it is uniformly applicable regardless of the type of model. We propose this approach as a general way to describe models that are used to estimate the effectiveness or cost-effectiveness of screening strategies.

Closely related output measures to express the impact of dwell time are lead time (restricted to disease that would progress to clinical cancer without intervention) and dwell time itself (from disease onset to clinical cancer). In MISCAN, CRC-SPIN, and SimCRC, dwell time on average was 8, 25, and 21 years, respectively.[13] Given that dwell time is the main driver of the MCLIR, it clearly gives the same type of information but arranged differently. The MCLIR is presented by time since complete removal of disease (e.g., age 65 years). The corresponding way to present dwell time would be its distribution by time since disease onset at, for example, age 65 years. The difference between the metrics is their relation to clinical incidence by age, given that at each age, the incident cases represent a difficult-to-grasp mix of shorter and longer dwell times. This relation is not straightforward for the dwell time distribution, whereas it is part of the metric of the MCLIR. Dwell time is closer to the inputs of the model; the MCLIR is closer to effects of screening that could be observed. Even though lead time, like the MCLIR, is bound to an age of intervention, it lacks straightforward relation to the cancer incidence by age. In addition, the lead time distribution (as opposed to average lead time) is difficult if not impossible to output for many models that do not produce paired life histories with and without intervention (e.g., models built with the TreeAge program).

We chose to present the MCLIR for age 65 years because this is in the middle of the age range (50–80 years) when individuals are often recommended to get CRC screening. As pointed out in the Methods section, it may be useful to present the MCLIR for different ages (e.g., the $MCLIR_{55}$, $MCLIR_{65}$, and $MCLIR_{75}$) if dwell time assumptions differ by age. Similarly, if dwell time depends on disease characteristics (such as location of disease, e.g., colon cancer v. rectal cancer) or other patient characteristics than age (such as gender or race), then it would be valuable to present the MCLIR for each of these groups.

Are model differences a limitation, or even a failure, of modeling? We think the differences

Table 5 Simulated Outcomes of 3 Models Representing the 25 and Older US General Population: (a) The Background CRC Incidence by Age Group per 100,000 Person Years and (b) the Maximum Clinical Incidence Reduction ($MCLIR_{65}$) after Complete Removal of Preclinical Cancer and Adenomas at Age 65 for 5-Year Periods and until Age 80, by Age Group and by Period Since Complete Removal

Age, years	50–54	55–59	60–64	65–69	70–74	75–79	80–84	65–79
(a)				**Background Incidence**				
MISCAN	51	88	139	208	286	370	440	279
CRC-SPIN	52	85	127	180	247	327	424	243
SimCRC	61	96	142	202	277	357	424	271
(b)				**$MCLIR_{65}$**				
Time Since Complete Removal, y				0–5	5–10	10–15	15–20	0–15
MISCAN, %				84	55	30	17	51
CRC-SPIN, %				100	99	92	77	90
SimCRC, %				100	97	86	74	88

MISCAN, Microsimulation Screening Analysis; CRC-SPIN, Colorectal Cancer Simulated Population model for Incidence and Natural history; SimCRC, Simulation Model of Colorectal Cancer.

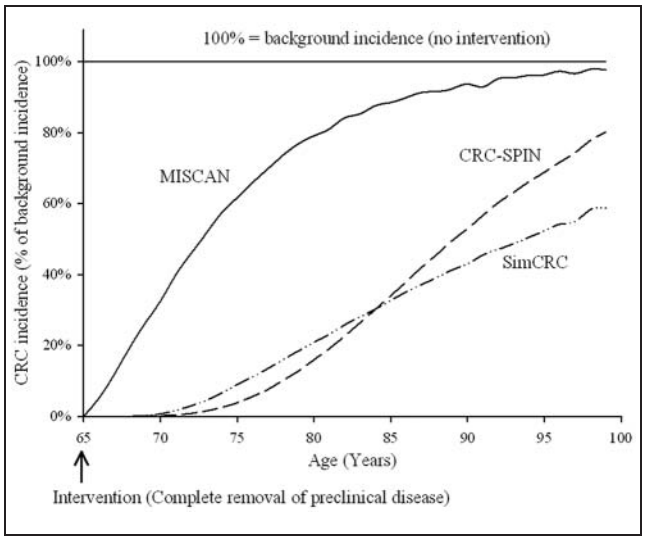

Figure 1 Colorectal cancer (CRC) incidence after complete removal of adenomas and preclinical cancers prevalent at age 65 years, as a percentage of the background incidence (100% – $MCLIR_{65}$); calculations for 1 cohort of individuals with 3 models representing the US general population: MISCAN, CRC-SPIN, and SimCRC. MCLIR, maximum clinical incidence reduction; MISCAN, Microsimulation Screening Analysis; CRC-SPIN, Colorectal Cancer Simulated Population model for Incidence and Natural history; SimCRC, Simulation Model of Colorectal Cancer.

between our CRC screening models reflect genuine uncertainty because all 3 models provide good fit to observed data such as CRC incidence and adenoma prevalence rates.[13] The demonstrated differences indicate areas where additional data are needed to inform models and where, in the absence of data, strong assumptions must be made. Only when more relevant data have become available will it become clear which model is more accurate. Very recently, the incidence and mortality end point results of a large randomized controlled study that investigated once-only sigmoidoscopy have become available.[23] This study's 11-year follow-up contains strong information on dwell time in combination with endoscopy sensitivity, at least for the distal colon. We expect that after the 3 model groups have calibrated their models to these new data, the MCLIRs will differ substantially less. Remaining uncertainty will concern the proximal part of the colon that is not reached by the sigmoidoscopy. In the meantime, the way to handle uncertainties is to perform sensitivity analyses to investigate the robustness of the results for the uncertainties.

In this article, we presented CRC screening models. Our approach, however, is relevant for any screening model, including those for nonneoplastic disease. The reason is that screening by definition presumes a detectable preclinical phase before disease becomes symptomatic. The duration of this phase always is an important determinant of the potential of screening. Although the concept is generalizable, specific issues may need attention when the MCLIR is applied to other diseases. One issue is that, unlike for CRC, the possibility of ''incidental'' detection of asymptomatic disease (e.g., breast or lung cancer on a computed tomography examination for unrelated indication) may be important. As

pointed out in the Methods section, the absence or presence of incidental detection when simulating the MCLIR should be specified since the MCLIR assumes no further screening. The best comparison of MCLIRs between models will be made when they handle incidental detection in a similar way.

In the models presented, the simulation of natural history begins at the onset of detectable disease (i.e., the onset of small adenomas). In some models, the simulation of natural history may begin before the disease is in a detectable state. Given that nondetectable disease is not relevant for the effectiveness of screening, we included detectability in the definition of the MCLIR. If the onset of detectable disease is not defined in a model, the MCLIR will automatically simulate the removal of any and all preclinical disease. In that case, however, it becomes more important to also present the clinical incidence reduction (CLIR) after removal that is incomplete due to assumed realistic lack of sensitivity (see next paragraph).

The MCLIR addresses the modeling of the limitation imposed by the natural history (the limitation for screening effectiveness). To describe the limitation imposed by (lack of) test sensitivity is a logical next step. Interestingly, one could use the same method used for the MCLIR by presenting the CLIR after screening with the base case sensitivity. Because of the high sensitivity, the CLIR after colonoscopy was only slightly lower than the MCLIR for our models (results not shown). The difference between the MCLIR and the CLIR after removal of disease detected with a fecal occult blood test (FOBT) of course will be much larger. Also, dwell time of detectable preclinical disease and sensitivity of the test under evaluation are to some extent interchangeable when calibrating models to data on screening effectiveness. A model with long dwell time combined with low sensitivity can project similar effectiveness to a model with short dwell time and high sensitivity. Typically, these models will show different MCLIRs but similar CLIRs after detection with estimated sensitivity.

Suggestions for using projections to describe models date back to the 1990s.[11] We build on those suggestions by proposing a standardized metric for comparing models. Projections cannot replace input description. The latter is necessary for reproducibility. However, projections do have added value. Where needed, projections can convey the implications of assumptions in a concise manner. Projections, combined with a restricted input description, are suitable to be included in the main text of a journal, whereas complete lists of input parameters can be placed in a supplementary (online) document.

In conclusion, adding the simulated maximum clinical incidence reduction (MCLIR) after complete removal of precursor disease is a simple way to clarify the impact of natural history in (CRC) screening models. It would be worthwhile to include such a measure in all screening modeling papers. We have described how to calculate the MCLIR and proposed standard notation for reporting it.

REFERENCES

1. Song K, Fendrick AM, Ladabaum U. Fecal DNA testing compared with conventional colorectal cancer screening methods: a decision analysis. Gastroenterology. 2004;126(5):1270–9.

2. Pickhardt PJ, Hassan C, Laghi A, Zullo A, Kim DH, Morini S. Cost-effectiveness of colorectal cancer screening with computed tomography colonography: the impact of not reporting diminutive lesions. Cancer. 2007;109(11):2213–21.

3. Vijan S, Hwang EW, Hofer TP, Hayward RA. Which colon cancer screening test? A comparison of costs, effectiveness, and compliance. Am J Med. 2001;111(8):593–601.

4. Knudsen AB, Lansdorp-Vogelaar I, Rutter CM, et al. Cost-effectiveness of CT colonography screening for colorectal cancer in the Medicare population. J Natl Cancer Institue. 2010;102:16.

5. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Boer R, Wilschut J, Habbema JD. At what costs will screening with CT colonography be competitive? A cost-effectiveness approach. Int J Cancer. 2009;124:1161–8.

6. Goldie SJ, Kim JJ, Wright TC. Cost-effectiveness of human papillomavirus DNA testing for cervical cancer screening in women aged 30 years or more. Obstet Gynecol. 2004;103:619–31.

7. Tosteson AN, Stout NK, Fryback DG, et al. Cost-effectiveness of digital mammography breast cancer screening. Ann Intern Med. 2008;148(1):1–10.

8. Kobayashi T, Goto R, Ito K, Mitsumori K. Prostate cancer screening strategies with re-screening interval determined by individual baseline prostate-specific antigen values are cost-effective. Eur J Surg Oncol. 2007;33(6):783–9.

9. McMahon PM, Kong CY, Johnson BE, et al. Estimating long-term effectiveness of lung cancer screening in the Mayo CT screening study. Radiology. 2008;248(1):278–87.

10. Herman CR, Gill HK, Eng J, Fajardo LL. Screening for preclinical disease: test and disease characteristics. AJR Am J Roentgenol. 2002;179(4):825–31.

11. Siegel JE, Weinstein MC, Torrance GW. Reporting cost-effectiveness studies and results. In: Gold MR, Siegel JE, Russel LB, Weinstein MC, eds. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p 276–303.

12. Habbema JDF, Schechter CB, Cronin KA, Clarke LD, Feuer EJ. Modeling cancer natural history, epidemiology, and control: reflections on the CISNET breast group experience. J Natl Cancer Inst Monogr. 2006;36:122–6.

13. 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; 31:530–539.

14. Vogelstein B, Fearon ER, Hamilton SR, et al. Genetic alterations during colorectal-tumor development. N Engl J Med. 1988; 319(9):525–32.

15. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov), SEER*Stat Database: Incidence—SEER 9 Regs Public-Use, Nov 2003 Sub (1973-2001). Frederick, MD: National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch; 2004 (released April 2004, based on the November 2003 submission).

16. Rutter CM, Miglioretti DL, Yu OA. Hierarchical non-homogenous Poisson model for meta-analysis of adenoma counts. Stat Med. 2007;26:98–109.

17. American Cancer Society Colorectal Cancer Advisory Group. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology. 2008;134:1570–95.

18. Winawer SJ, Zauber AG, Fletcher RH, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. CA Cancer J Clin. 2006;56:143–59; quiz 184–5.

19. Rutter CM, Savarino JE. An evidence-based microsimulation model for colorectal cancer: validation and application. Cancer Epidemiol Biomarkers Prev. 2010;19:1992–2002.

20. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, Ballegooijen M van, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. Ann Intern Med. 2008;149:659–69.

21. Janoschek A. Das reaktionskinetische Grundgesetz und seine Beziehungen zum Wachstums- und Ertragsgesetz. Stat Vjschr. 1957;10:25–37.

22. Gille U, Salomon FV. Brain growth in mallards, Pekin and Muscovy ducks (Anatidae). J Zool. 2000;252:399–404.

23. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomized controlled trial. Lancet. 2010; 375(9726):1624–33.