Best Practice & Research Clinical Gastroenterology 24 (2010) 427–437

Contents lists available at ScienceDirect

Best Practice & Research Clinical Gastroenterology

5

Simulation of colorectal cancer screening: What we do and do not know and does it matter

Marjolein van Ballegooijen, Assistant Professor [a,*], Rob Boer, Senior Investigator [a], Ann G. Zauber, Associate Professor [b]

[a] Department of Public Health, Erasmus MC, PO Box 2040, 3000 CA Rotterdam, The Netherlands
[b] Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Keywords:
Colorectal neoplasms
Mass screening
Simulation
Modelling
Cost-benefit analysis
Occult blood
Colonoscopy
Sigmoidoscopy

Simulation modelling is increasingly used to inform decision-making on screening, including colorectal cancer screening strategies. The strength of simulation is its ability to handle complexity and to identify the implications of uncertainty in a formal, documented, reproducible and consistent way. Important specific uncertainties concerning colorectal cancer screening are the dwell time of adenomas and the associated sensitivity of the various tests. Concerning these issues, for distal colorectal neoplasia, knowledge has been greatly increased by the recent availability of the once only sigmoidoscopy randomised trial results. Other uncertainties concern the quality of life effects of screening, diagnostic and surveillance colonoscopies, and the true total costs of the various screening modalities in a routine high throughput efficient setting. A limitation of simulation of screening is that complexity leads to lack of insight and understanding into the models used, and therefore a lack of sound criticism, acceptance and use amongst decision makers. Modellers are currently focussing on ways to make models and the implications of assumptions more transparent. Thus it is important to further develop the quality and acceptability of simulation, especially that for colorectal cancer screening.

Introduction

Simulation is increasingly used to inform decision-making for screening strategies. In this chapter we discuss the uncertainties in modelling for informing decisions for colorectal cancer (CRC) screening. [1–7] Simulation complements empirical research, the data of which is required to build

* Corresponding author. Tel.: +31 10 704 4268; fax: +31 10 703 8474.
E-mail address: m.vanballegooijen@erasmusmc.nl (M. van Ballegooijen).

1521-6918/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bpg.2010.07.001

upon the simulation model. However, simulation will have little use for informing decisions about screening strategy without empirical data on screening effectiveness. Once there is a sufficient level of empirical data, simulation can be used to move from the empirical results to optimisation of screening strategies and thus to decisions and guidelines. By using simulation one can define what knowledge is already available in the total body of data. Consequently one can also define what are the knowledge gaps still to be filled. Moreover, sensitivity analyses in simulation modelling can be used to quantify the impact of lack of knowledge. This can also be useful to identify what further empirical research is required.

Whether CRC screening is effective or cost-effective is no longer an issue. Several randomised controlled trials have shown significant CRC mortality reduction for fecal occult blood test (FOBT) screening [8–10] and for sigmoidoscopy screening. [11] The cost-effectiveness of FOBT screening has been demonstrated. [2,12,13] This also holds for flexible sigmoidoscopy, [1,4,6,7] given that the simulations used did not overestimate its effectiveness as demonstrated by the recent endoscopy trials. [11,14] The remaining questions concern what are the optimal strategies in specific situations: which screening test or combinations of tests to use for primary screening and what screening intensity (age range, frequency and cut-off criteria for referral to diagnostic colonoscopy) and surveillance rules are most appropriate. Specifically, is it preferable to offer colonoscopy to everyone in the general population or to precede it with risk selection based on FOBT, flexible sigmoidoscopy, or another screening test? We discuss below what are currently uncertainties that make the answers to these questions not completely clear.

The strength of simulation

Relevant knowledge on CRC and CRC screening is obtained from a variety of independently collected data sources. Data on adenoma prevalence by age is available from various autopsy studies and endoscopy studies. Cancer registry data are generally available for CRC incidence, stage distribution and survival. Cause of death registries provides CRC mortality and all cause mortality rates. Next needed are demographic data on age distribution (current and projected) of the population. This is of importance when using population models, which include a range of birth cohorts. Estimates for participation and detection rates of screening are available from pilot screening programs, regional or national screening programs and opportunistic screening. Last but not least, there are data on impact of screening on incidence and mortality from the randomised trials [8–11,14]. Data on demographic and cancer epidemiology have been collected in several different individual regions to varying degrees. The same goes for autopsy data, participation rates, detection rates, and survival data. For example some countries have data on participation at sigmoidoscopy from their own population whereas others do not have such data. Estimates from neighbouring regions might be used in the absence of specific information for the specific population under study. Data on efficacy of screening from randomised trials are scarce and will be used for evaluation for any country. Simulation can address all these relevant issues simultaneously, and also the underlying mechanisms that connect them. An important strength of simulation, therefore, is that it can combine information from different sources and in one coherent model.

Other strengths of simulation are related to the fact that screening is a complex issue. This is, amongst others, due to the important role of different dwell-time distributions for various steps in the progression of disease, which may moreover depend on age. An extremely important characteristic of simulation is that it can handle this kind of complexity. Another characteristic is its formality, which means that assumptions are identifiable. The basis for each assumption can be questioned. As a result it becomes clear to what extent the assumptions are evidence based and alternatively, when merely on expert opinion. The formal character of simulation also enables reproducibility and consistency between successive analyses, by maintaining assumptions or by selectively and explicitly changing assumptions.

Based on the characteristics discussed above, an important strength of simulation is the ability to identify uncertainties, and to account for them in an explicit, coherent and quantitative way. In a first step uncertainties are identified as part of the data analysis. For parameters that can be measured directly, confidence intervals show the uncertainty. However for parameters which are not directly observable, so called ‘deep parameters’ (for example dwell time, see below), a simulation-based

analysis is applicable. For such an analysis, the study that produced the observed data needs to be simulated. By repeating the simulation with various assumptions for the parameters, one can investigate for which ones the model can reproduce the observed data. In this way, uncertainties can be identified and quantified.

Some of the uncertainties have a relatively wide range, others have a smaller one. This range, however, may not reflect the impact on decisions. For decision analyses, the impact of uncertainties is a key issue. This impact depends on the specific questions and thus may differ between decision analyses. Therefore, as a next step, the impact of the uncertainties on the cost-effectiveness is investigated. In so-called sensitivity analyses, the cost-effectiveness calculations are repeated under alternative assumptions, varying the parameters within the previously identified uncertainty range. The results indicate to what extent the conclusiveness of decision analyses is affected by uncertainties.

A critical point of simulation studies is whether relevant uncertainties are fully acknowledged and whether appropriate sensitivity analyses are included in the analysis. If not, the use of simulation can lead to incorrect decisions. Willingness to acknowledge and recognise our uncertainties is essential for model-based analysis.

Specific uncertainties concerning colorectal cancer screening

Dwell time and test sensitivity

Important differences between endoscopy and FOBT screening are their respective sensitivities for different disease phases. Therefore, sensitivity estimates are important. Moreover, given that endoscopy has a considerably higher sensitivity for adenomas than even sensitive FOB tests, the dwell time of adenomas is most important for how the tests compare. Dwell time and sensitivity are major factors for the effectiveness of screening, screening frequency, and comparisons among various screening strategies.

Dwell time and sensitivity cannot be measured directly and are deep parameters. The reason is that screening is about detecting disease before it gives symptoms. The screen-detectable disease phase is only revealed when screening is applied. Once detected, then the screen-detected cancer (precursor) is typically treated. Consequently it is unclear how long the neoplasia was there when detected and how it would have developed if it would have been ignored. On the other hand, given that a gold-standard test is usually not available, (lack of) sensitivity is also not directly observable. The best direct information on missed cases in individuals with negative screen results comes from studies of different tests in the same asymptomatic individuals, including colonoscopy as the most sensitive test. However even colonoscopy does miss adenomas [15–17] and only minimum estimates are available for miss rates based on colonoscopy repeated in the same individuals [15,16] or in studies where CTC also has been performed. [17] Sessile or flat adenomas may be missed at a considerable rate with any of these procedures.

Deep parameters are similar to indirect comparisons in that both are associated with large uncertainties. However, to discern dwell time and sensitivity as separate parameters is crucial to model the effect of various screening strategies on incidence and mortality. For example, based on the Flexiscope Trial, [11] if we want to estimate the population effect of a lower sensitivity of sigmoidoscopy than that obtained in the trial, then we need to know the correct dwell-time distribution. Only then can we keep the effect of dwell time constant in our projections and estimate the impact of a lower sensitivity.

Another example is that in many trials screening has been repeated at a certain interval, so that the impact of longer intervals between screenings has not been observed. This was the case for the FOBT trials. [8–10] Interval cancers may be detected because of a prior false-negative screen, or as a new neoplasm. The relative importance of these two mechanisms differs by the length of the interval. In the short-term, the role of missed cases (false negatives) is relatively more important. Therefore, for extrapolation of trial results to longer intervals, it again is necessary to discern the effect of dwell times that are short enough to fit within the observed follow-up time from the effect of lack of sensitivity on the other hand.

Next to varying screening intervals, the cut-off points for test positivity may also be varied. Varying the criteria for test positivity will change the sensitivity (and the specificity) but leave the natural

history unchanged. Moreover tests are often developed further and modified, which also affect the test characteristics.

Finally, there may be alternative tests that are sensitive for other disease characteristics. This is true for CRC screening, where endoscopy detects adenomas and cancers based on macroscopic morphology, whereas FOBT detects components of bloodshed in the stool. In this case, using results on mortality reduction from a trial with one test to estimate expected mortality reduction with another test (e.g. by adjusting only estimates of sensitivity) is not reliable. A basically different test requires its own randomised controlled trial, especially when it detects much more and earlier disease, as is the case (per screening round) with endoscopy relative to FOBT. It would appear that the adenoma dwell time is relatively more important for endoscopy compared to FOBT screening, since endoscopy is more sensitive for adenomas. However with screening with a highly sensitive FIT every two years or even annually, adenomas are also detected in many of the participating individuals. [18,19] Given that fecal immunochemical test (FIT), for which there are no randomised controlled trials, detects more adenomas than guaiac FOBT, information from endoscopy trials on dwell time of adenomas will be of use for projecting the effectiveness of long-term FOBT screening as well. Again, this requires assessment of the dwell time and sensitivity as separate components of the underlying process that leads to changes in incidence and reduction of mortality.

So, in order to apply the results from a particular trial to alternative strategies or populations, we require understanding the background mechanisms of natural history, screening, and treatment behind the observed results.

The fact that dwell time and sensitivity cannot be observed or measured directly also means that simulation is important to estimate them. Moreover, variation in one assumption will influence variation in another one. Decreasing the sensitivity of a screening test, to some extent, may have the same effect on the projected rate of interval cancers as increasing the dwell time of preclinical disease. Therefore, when calibrating to interval cancers, they must be estimated simultaneously. By varying both, the range in combinations of sensitivity and dwell times that are in agreement with the data are identified. To make this possible, the natural history of the disease, in this case colorectal cancer and its precursors, needs to be described explicitly in the simulation model (Fig. 1).

The important adenomas to detect are those that, without intervention, progress to clinical CRC (before death from another cause). For these clinically relevant adenomas, we need to know how long

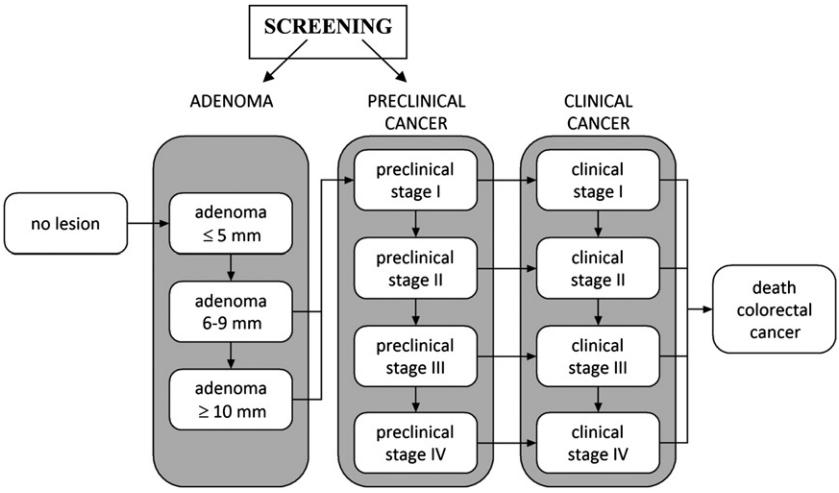

Fig. 1. Graphical representation of natural history of colorectal cancer as modelled by MISCAN model as an example of a published CRC screening simulation model. Screening provides the opportunity to intervene in the natural history of the adenoma carcinoma sequence. Screening can either remove a precancerous lesion (i.e., adenoma), thus moving a person to the 'No lesion' state, or through early cancer detection, which makes an undiagnosed cancer clinically detected at a potentially earlier stage of disease where it is more amenable to treatment.

the process is from adenoma onset to cancer. Average dwell time is not sufficient; we also need the full ranges for the dwell-time distribution, describing what proportion develops in a relatively short period. This dwell time, together with sensitivity, determines what screening interval to use, and at what age to stop screening. With short dwell times, the added value of shorter screening intervals will be large. With long dwell times, this added value is limited and screening in advanced age is relatively ineffective because most of the relevant lesions were already detected at a younger age.

Design of required empirical studies

Cancer incidence after a negative screening test is necessary for the estimation of both dwell time of adenomas and test sensitivity. The maximum information for a given test comes from a randomised trial with a once-only test and a follow-up that is long enough to see the effect dissipating. This is in contrast with trials that include more than one screening round with a specified screening interval. Usually, the negative predictive value for developing clinical cancers is not yet attenuated at the time of the next screening round. As a result, the information on the effectiveness of screening with intervals longer than the one used in the trials is limited. A trial with a once-only endoscopy in the intervention arm, and no screening in the control arm would permit assessment of dwell time of adenomas. The results would show the incidence reduction by year since the screening induction.

Cancers after a negative test either come from adenomas that had their onset after the baseline screening, or from adenomas or preclinical cancers that have been missed at baseline. High sensitivity combined with a short dwell time is to some extent compatible with the same results as a lower sensitivity with a longer dwell time. Distinction can be made based on the fact that missed cases on average do give earlier interval cancers than new cases. However, the distinction will be limited by the fact that the shape of the dwell time distribution, and therefore the fraction of rapidly developing adenomas, is also unknown. Given their mutual unidentifiability, dwell time and sensitivity have to be estimated from the trial data simultaneously. The different combinations of test sensitivities and dwell-time distributions that can reproduce the data determine the remaining uncertainty. This is the range of uncertainty that must be explored in a sensitivity analysis.

Empirical studies available, on their way, planned, or not expected

We are at a special time in colorectal cancer screening evaluation when several empirical trials have just been published and we still have to incorporate these studies into our knowledge base for modelling. [11,14]

The first large-scale randomised endoscopy study with incidence and mortality as endpoints was the Norwegian Colorectal Cancer Prevention (NORCCAP) study. [14] This trial had three arms, one with a once-only baseline sigmoidoscopy, one in which the baseline sigmoidoscopy was combined with a FIT and one with usual care. The number of individuals invited for sigmoidoscopy was 13,823 and 8846 (63%) participated (two intervention arms, no preselection of the study population). The follow-up time was 7 years for incidence and 6 years for mortality. Very recently, the results of another large-scale randomised endoscopy trial (Flexiscope) have been published. [11] In this trial there was a once-only sigmoidoscopy arm and a usual care arm. The study population was considerably larger than in the earlier Norwegian trial (57,237 invited for sigmoidoscopy, 40,674 (71%) participated after preselection of the study population) and the follow-up longer (11 years). Both trials were performed without a specific active CRC screening program in the usual care arm.

As already noted, a trial with a once-only test and a long follow-up provides the best information on dwell time and sensitivity. The results give invaluable information on the effects that can be expected from endoscopy screening at different intervals.

Simulation models must definitely be validated with these data which are the first evidence from randomised controlled trials on the efficacy of endoscopic screening and thus provide very relevant new information on CRC and CRC screening. To the extent that models do not reproduce the observed data on incidence and mortality, the models will need to be calibrated to these new data.

For this validation and calibration it is necessary to adjust the simulations to the specific characteristics of the trial. The adjustments concern the age range of the study population, screening uptake,

reach of the sigmoidoscopy, incidence, stage distribution, survival and (by consistency) mortality in the non-attendees and in the control group. One would also want, if relevant, to account for whether there has been a preselection of individuals willing to participate in the trial before inclusion in the randomisation, as in the UK trial, versus direct randomisation of the complete eligible population, as in the Norwegian trial. This would concern other effects than an increased participation rate (which is reproduced in the simulation and thus accounted for) and possibly a higher or lower total risk in the population included in the randomisation (which is expressed and thus accounted for in the incidence and mortality in the control group). In principle, there are all kinds of possible effects. For instance, a higher or lower a priori risk could be associated with slower or faster developing disease. Slower developing disease could be associated with a larger efficacy, due to higher screen-detectability. It could also be associated with a smaller efficacy, due to better survival already without screening. However, given the fact that any such possible effects of preselection take place before randomisation, the study does not give information on such effects. It does make it interesting to compare the risk in the control group to the expected risk based on cancer registry data from the background population, but how this affects the observed efficacy remains speculative. Clearly, a trial without preselection does not imply this kind of uncertainty. After validation and calibration, when the model is used for projections, a realistically expected participation rate should be used, which most probably would be lower than the rate observed after preselection in that same population. Thus far, none of the CRC modelling groups has published model validation or calibration to the sigmoidoscopy trials.

Meanwhile, the results of one more trial, SCORE [20] which is very similar to the UK trial, is anticipated with 17,148 invited for sigmoidoscopy, and 9911 (58%) participated in having sigmoidoscopy. [20] Finally, in the USA the PLCO, with repeated sigmoidoscopy and a randomised usual care arm, is ongoing. [21]

Sigmoidoscopy trials give limited information concerning proximal CRC. To assume that dwell time and endoscopy (colonoscopy) sensitivity in the proximal colon are the same as the dwell time and endoscopy (sigmoidoscopy) sensitivity in the distal part, is not trivial. Case control study results, which have a weaker design than randomised trials for evaluating screening effectiveness, suggest that interval cancer rates after colonoscopy are higher proximally than distally. [22,23]

With these issues in mind, the multicountry Nordic-European Initiative on Colorectal Cancer trial (NordICC) randomised colonoscopy trial was started in 2009 with a planned enrolment of 22,000 randomised trail to colonoscopy and 44,000 to usual care. [24] So far two countries have started in this trial, both with a colonoscopy versus a usual care arm: a large one in Poland (5000 colonoscopies planned), and a smaller one in the Netherlands (2500 colonoscopies planned) under an assumption of 50% adherence with colonoscopy. In both countries, so far there is no screening program active in the usual care arm. Norway will start this trial soon. Funding for this trial in additional countries is being sought in order to increase its power. Increasingly, a usual care arm will include FOBT screening, since the introduction of FOBT programs is taking place or planned in several countries. In the analyses, one can account for FOBT screening both in the control and in the intervention arm. Of course, the more screening in the control arm, the less power there is to show effectiveness of colonoscopy screening. Also, the magnitude of the effect will be more uncertain.

Meanwhile, also for proximal colon cancer, the main clinical question is whether to screen primarily with endoscopy or with FOBT. The proximal colon can be evaluated either by sigmoidoscopy (by performing colonoscopy in individuals who have any or significant adenomas detected in the distal colon), by FOBT (guaiac or immunochemical), by a combination of sigmoidoscopy and FOBT, or by primarily screening with colonoscopy. Given these alternatives, the information obtained by contrasting colonoscopy with any FOBT in randomised trials will be most relevant and useful, either to show difference in effectiveness or to show how close they are in reducing mortality. Therefore trials that randomised between colonoscopy and FIT are considered and planned in Europe (Barcelona) http://clinicaltrials.gov/ct2/show/NCT00906997 for colorectal cancer screening in average-risk population: immunochemical fecal occult blood testing versus colonoscopy) and US centres in the Veterans Administration setting.

Quality of life effects

Effective screening results in several positive outcomes, including reduction in the number of advanced cancers and their associated treatments, fewer life-years lost, and the loss of quality of life is prevented.

On the other hand, screening requires colonoscopies and polypectomies, which are associated with negative effects on quality of life. The frequencies of these events are much higher than for the number of cancer cases and cancer deaths prevented. More importantly, the frequency of these procedures will differ between endoscopy and FOBT screening. Therefore, the estimates for the loss in quality of life associated with endoscopy and adenectomy, and how this relates with the number of life-years gained, could make a difference when comparing FOBT to endoscopy screening for net benefits.

There is methodology (for example the standard gamble) for translating descriptive measurement of health-related quality of life effects into utilities (a weighing index for the quality of life between zero for death and one for perfect health) and thus to produce QALYs (quality-adjusted life-years). [25] But there are two unresolved issues in how to value these negative side effects. First, there is a lack of data describing the impact of colonoscopy and other CRC screening tests on health-related quality of life. Such impact would include mental discomfort and worry or anxiety, and physical discomfort surrounding screening. Secondly, even if such data are available, estimates for utilities are not as precise compared to the small effects that are associated with e.g. preparing for colonoscopy, undergoing colonoscopy, receiving a positive FOBT result, and becoming an adenoma patient (causing exposure to a perceived increased risk). However, even small differences in estimates can influence decisions because of the large numbers, and not accounting for small effects represent a strong assumption.

Further research is needed in this area, both empirical and methodological. Meanwhile, sensitivity analyses will probably always play an important role to show the possible impact of quality of life effects on how screening strategies compare. Such sensitivity analyses are certainly advisable where CRC screening is concerned, including to also report results in terms of life-years gained, without quality of life correction. This clearly shows the influence of any assumed correction for quality of life.

Costs

There is uncertainty about the costs of both screening and of cancer treatment. Screening induces large costs whereas treatment is associated with potential savings. The choice between tests will be most affected by the unit costs of the tests. Risk selection based on FOBT or sigmoidoscopy preceding diagnostic colonoscopy will have its own cost, and will affect the number of colonoscopies. Difference in incidence and mortality reduction between tests will influence treatment savings. How to estimate costs depends on the perspective of the Cost Effective Analysis (CEA). If the question is posed by the health care third-party payer, one probably would want to use reimbursement rates. Most CEA's take the societal perspective or at least an approximation. In that case, opportunity costs are the standard. For practical reasons, reimbursement rates are also often used in societal analyses. However, reimbursement rates may be based on the specific clinical setting. The costs of e.g. colonoscopies in a screening setting may be lower than clinical colonoscopies. Therefore it is necessary to measure costs in the specific screening setting. Important items to account for are the number of screening colonoscopies per working day, and the level of training and education required to perform the procedure in a screening setting. The same kind of cost assessments must be made for FOBT or FIT. For example, the US Medicare reimbursement rate of $4.54 for guaiac FOBT probably is not sufficient to cover laboratory personnel who perform the test, which consists of a non-automated procedure. For test kits that have to be purchased from manufacturers, it may be relevant to negotiate the price based on the large numbers required in population-based screening, and use the result in the CEA. Manufacturers may use the CEA's to set their price at the highest acceptable level. This however is an undesired way of using CEA results.

Another issue is the rapidly increasing costs for CRC treatment in advanced stages. This increase may get to the point that CRC screening becomes cost-saving. [26] Even though the fact that CRC screening is cost-effective would be enough to implement it, a cost-saving situation would add to decisions for further efforts to disseminate screening, e.g. by installing population-based programs free of charge, possibly with personal invitations to participate. The increase in colorectal cancer treatment costs is due to the increasing use of new generations of high-cost chemotherapy drugs. [27] In a rapidly changing situation, it is likely that available data, which are always at least one but often several years behind, are not up to date. Outdated data can cause serious underestimation of CRC treatment costs. Alternatively,

calculating the chemotherapy costs based on protocols could overestimate costs, given that in practise not all individuals with CRC will receive the full regiment of drugs, depending on comorbidity and how individuals are insured. Therefore it is important to create methods for retrieval of up to date treatment cost data. Now that data registrations are usually computerised, this in principle is possible.

Discussion

Other uncertainties

We have discussed uncertainty concerning dwell time of adenomas and the sensitivity of various tests (especially concerning proximal disease), unit costs of screening test and of treatment costs. There are, of course, other uncertainties. One issue is regression of adenomas. The consequence of regression would be that more non-progressive adenomas are detected at repeat screening and colonoscopy surveillance after adenectomy, and that more individuals with an on average lower risk are referred to (further) surveillance. [28] Another concern is the uncertainty regarding dependency between test results, in particular dependency between false-negative results. In case of such dependency, the probability for a false-negative result is higher after a previous false-negative test result than on average. In such a situation, an adenoma or preclinical cancer that has been missed, has an increased probability to be missed again in the next screening round. This could play a role in FOBT screening for non-bleeding neoplasias, or in endoscopy screening, when a neoplasia is flat, localised behind a colon fold, or in a difficult to reach part of the colon, at least in this particular individual. Especially for the more sensitive immunological FOBT, for which there are no randomised controlled trial data and that is performed every one or two years, unaccounted for dependency would cause an overestimate for the program sensitivity.

The role of microsimulation

The role of uncertainties is independent of the type of simulation model applied. This also holds for the strengths of simulation and how it should be performed with care. Meanwhile, simulation models used in colorectal cancer screening evaluation are often microsimulation models. In microsimulation, at least at some level, entities are simulated one-by-one. For the simulation of screening, this is typically the case for individual life histories. By simulating enough (thousands or millions) individuals, a realistic population is simulated. In some models, adenomas are also microsimulated, whereas in other models the occurrence of multiple adenomas is summarised into one disease process per individual. Microsimulation models are stochastic. Occurrences are governed by probabilities. Random drawings from probability distributions determine what the next event is and when it takes place. This explains the ability of microsimulation to handle complexity while being flexible at the same time. These characteristics, together with the ability to handle many kinds of probability distributions, make microsimulation especially advantageous for screening evaluation and cost-effectiveness analysis.

Joint analyses of independent trials

When there are results from different randomised controlled trials to inform estimation of deep parameters, one would want the estimates to be consistent with all of the trials, and to use the combined power of the studies. However, a joint analysis is hampered by differences between the trials, especially when the trials have been designed and performed independently from each other. Relevant differences may concern the background epidemiology and age range of the study population, the screening interval, the cut-off for referral to diagnostic colonoscopy, the participation rate, the risk selection in participators (observable by comparing the risk in the non-participators to the control group), and whether non-participators are included in follow-up rounds or not. Also, the results are sometimes only available in a specific published tabulation following specific definitions. With simulation, it is possible to simulate every trial separately, accounting for all its peculiarities, and deriving the results in the same way as the observed results also per trial. By using joint natural history and possibly also sensitivity parameters, these can be estimated by minimising the differences between simulated and observed results for all the considered trials simultaneously. [29] It is even

possible to combine trials with different tests. For example, given that FOBT and endoscopy trials both contain information on dwell time of preclinical cancers, one would want to account for both types of trials when estimating this dwell time, so that the natural history is consistent with all available data. So, using simulation, different types of trials can be investigated in a joint analysis, making it potentially very useful for optimising data analysis.

The limitations of simulation

The strengths of formal simulation are also its enemy: complexity. Complexity of simulation models leads to lack of insight and understanding and contributes to mistakes, lack of sound criticism, acceptance and use. On the other hand, simulation offers a unique opportunity for more rational decision-making in cancer screening, so it is important to overcome the limitations.

There are in principle two ways to improve the situation regarding complexity: decrease complexity, or find ways to explain it. Models should for practical reasons not include issues that are irrelevant for the question addressed. So in a way, they should be as simple as possible. But models must also include all relevant issues. In the case of screening, the lowest level of complexity is still quite complex. Therefore, it is questionable whether complexity can be decreased at any level, other than for educational or communicative purposes. Nevertheless, this is a field for further investigation.

On the other hand, modellers are currently working to make modelling more transparent. One of the important opportunities to do this is offered by the US National Cancer Institute, which funds the Cancer Intervention and Surveillance Modeling Network (CISNET), a consortium of modelling research groups. (http://cisnet.cancer.gov/). One of the aims of CISNET is to bring modelling to a higher level by having several groups (per cancer site) collaborate with each other and then to publish the structure and inputs of their models on-line in a standardised way. This is important for documentation and reproducibility. Another approach is to perform base cases, in which different model inputs are standardised to compare outputs, and thus to clarify different views on screening issues. It becomes increasingly clear that to develop a set of standardised outputs aimed at describing models is helpful in presenting key model characteristics. Input description in principle will be as complex as the model itself, and is difficult to standardise on a more detailed level because of different model structures. The value of standardised output is largely independent of the underlying model and is expected to add to clarifying models conceptually.

Conclusion

In conclusion, simulation offers a powerful method for the complexity of evaluation of screening. The challenge is to further develop the quality and acceptability of the work. In terms of uncertainties regarding colorectal cancer screening, much progress has been achieved by the availability of the sigmoidoscopy trial results. Further work needs to be done concerning the screening effect in proximal disease, quality of life effects of screening and societal costs of both screening and cancer treatment.

Practice points

- Simulation modelling is a complex tool which enables optimal use of knowledge from randomised controlled trials and other empirical studies to evaluate alternative screening strategies in different screening situations
- However to receive the benefits of simulation modelling to inform health policy, there must be a willingness for public policy professionals to invest time to understand the key issues of simulating (cancer) screening. Given the fact that these are primarily also the key issues of cancer screening itself, and given the utility of simulation, this investment should pay back. It would also greatly help if the modellers addressed these issues in a more standardised and transparent way.

Research agenda

- Validation of our CRC screening models with respect to the efficacy observed in the new sigmoidoscopy RTC's by defining the corresponding combinations of dwell time and test sensitivity along this dwell time
- Estimating the differences in screening efficacy in right- and left-sided disease and what is the reason behind such differences; to this end, colonoscopy arms are required in RCT's
- To measure the true total costs of routine high throughput efficient screening and screening follow-up practise. Retrieve up to date chemotherapy and other CRC cancer treatment costs representing average clinical practise.

Conflict of interest

None.

Acknowledgements

Drs. van Ballegooijen, Boer, and Zauber are supported in part by Cancer Intervention and Surveillance Modeling Network grants from the National Cancer Institute: (U01-CA-097426). In addition Drs. van Ballegooijen and Zauber are also supported in part by U01-CA-115953, and Drs. van Ballegooijen also in part by the Dutch Ministry of Health, Health Care Prevention Program–Implementation (ZonMw project number 120720011).

References

[1] Loeve F, Brown ML, Boer R, van Ballegooijen M, van Oortmarssen GJ, Habbema JD. Endoscopic colorectal cancer screening: a cost-saving analysis. J Natl Cancer Inst Apr 5, 2000;92(7):557–63.
[2] Pignone M, Saha S, Hoerger T, Mandelblatt J. Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med Jul 16, 2002;137(2):96–104.
[3] Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. Ann Intern Med Nov 4, 2008; 149(9):659–69.
[4] Zauber AG, Knudsen AB, Rutter CM, Lansdorp-Vogelaar I, Savarino JE, van Ballegooijen M, et al. Cost-effectiveness of CT colonography to screen for colorectal cancer. Available from: http://www1.cms.hhs.gov/mcd/viewtechassess.asp?from2=viewtechassess.asp&where=index&tid=58&; 2009.
[5] van Ballegooijen M, Habbema JDF, Boer R, Zauber AG, Brown ML. Report to the Agency for Healthcare Research and Quality: a comparison of the cost-effectiveness of fecal occult blood tests with different test characteristics in the context of annual screening in the Medicare population. Available from: http://www/cms.hhs.gov/coverage/download/id87.zip; August 2003.
[6] Knudsen AB, Lansdorp-Vogelaar I, Rutter CM, Savarino JE, Van Ballegooijen M, Kuntz KM, et al. Cost-effectiveness of CT colonography screening for colorectal cancer in the Medicare population. J Natl Cancer Inst (in press).
[7] Lansdorp-Vogelaar I, Kuntz KM, Knudsen AB, Wilschut JA, Zauber AG, Van Ballegooijen M. Stool DNA testing to screen for colorectal cancer in the Medicare population: a cost-effectiveness analysis. Ann Int Med, (in press).
[8] Mandel J, Bond J, Church T, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med May 13, 1993;328(19):1365–71.
[9] Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet Nov 30, 1996;348(9040):1472–7.
[10] Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet Nov 30, 1996;348(9040):1467–71.
[11] Atkin WS, Edwards R, Kralj-Hans I, Wooldrage K, Hart AR, Northover JM, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet May 8, 2010;375(9726):1624–33.
[12] Whynes DK, Neilson AR, Walker AR, Hardcastle JD. Faecal occult blood screening for colorectal cancer: is it cost-effective? Health Econ Feb 1998;7(1):21–9.
[13] Gyrd-Hansen D, Sogaard J, Kronborg O. Colorectal cancer screening: efficiency and effectiveness. Health Econ Feb 1998;7 (1):9–20.
[14] Hoff G, Grotmol T, Skovlund E, Bretthauer M. Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. Br Med J 2009;338:b1846.

[15] Hixson LJ, Fennerty MB, Sampliner RE, Garewal HS. Prospective blinded trial of the colonoscopic miss-rate of large colorectal polyps. Gastrointest Endosc Mar–Apr 1991;37(2):125–7.
[16] Rex DK, Cutler CS, Lemmel GT, Rahmani EY, Clark DW, Helper DJ, et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. Gastroenterology Jan 1997;112(1):24–8.
[17] Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. N Engl J Med Dec 4, 2003;349(23):2191–200.
[18] Castiglione G, Zappa M, Grazzini G, Mazzotta A, Biagini M, Salvadori P, et al. Immunochemical vs guaiac faecal occult blood tests in a population-based screening programme for colorectal cancer. Br J Cancer Jul 1996;74(1):141–4.
[19] Hol L, Wilschut JA, van Ballegooijen M, van Vuuren AJ, van der Valk H, Reijerink JG, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer Apr 7, 2009;100(7):1103–10.
[20] Segnan N, Senore C, Andreoni B, Aste H, Bonelli L, Crosta C, et al. Baseline findings of the Italian multicenter randomized controlled trial of "once-only sigmoidoscopy"–SCORE. J Natl Cancer Inst Dec 4, 2002;94(23):1763–72.
[21] Gohagan JK, Prorok PC, Kramer BS, Hayes RB, Cornett JE. The prostate, lung, colorectal, and ovarian cancer screening trial of the National Cancer Institute. Cancer 1995;75:1869–73.
[22] Baxter N, Goldwasser M, Paszat L, Saskin R, Urbach D, Rabeneck L. A population-based case control study of the effectiveness of colonoscopy for prevention of death from CRC. Supplement to gastroenterology, digestive disease week; 2007. Washington, DC.
[23] Brenner H, Hoffmeister M, Arndt V, Stegmaier C, Altenhofen L, Haug U. Protection from right- and left-sided colorectal neoplasms after colonoscopy: population-based study. J Natl Cancer Inst Jan 20, 2010;102(2):89–95.
[24] Bretthauer M. The capsule and colorectal-cancer screening–the crux of the matter. N Engl J Med Jul 16, 2009;361(3):300–1.
[25] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ Mar 2002;21(2):271–92.
[26] Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Habbema JD, Kuipers EJ. Effect of rising chemotherapy costs on the cost savings of colorectal cancer screening. J Natl Cancer Inst; Sep 24, 2009.
[27] Schrag D. The price tag on progress–chemotherapy for colorectal cancer. N Engl J Med Jul 22, 2004;351(4):317–9.
[28] Loeve F, Boer R, Zauber AG, van Ballegooijen M, van Oortmassen GJ, Winawer SJ, et al. National Polyp Study data: evidence for regression of adenomas. Int J Cancer Sep 10, 2004;111(4):633–9.
[29] Lansdorp-Vogelaar I, van Ballegooijen M, Boer R, Zauber A, Habbema JD. A novel hypothesis on the sensitivity of the fecal occult blood test: results of a joint analysis of 3 randomized controlled trials. Cancer Jun 1, 2009;115(11):2410–9.