Computer Disease Simulation Models: Integrating Evidence for Health Policy

Carolyn M. Rutter, PhD[1,2], **Amy B. Knudsen, PhD**[3], and **Pari V. Pandharipande, MD, MPH**[4]

[1] Group Health Research Institute, Seattle, WA

[2] Department of Biostatistics, University of Washington, Seattle, WA

[3] Institute for Technology Assessment, Department of Radiology, Massachusetts General Hospital, Boston, MA

[4] Abdominal Imaging and Institute for Technology Assessment, Department of Radiology, Massachusetts General Hospital, Boston, MA

Abstract

Computer disease simulation models are increasingly being used to evaluate and inform healthcare decisions across medical disciplines. The aim of researchers who develop these models is to integrate and synthesize short-term outcomes and results from multiple sources to predict the long-term clinical outcomes and costs of different healthcare strategies. Policy makers, in turn, can use the predictions generated by disease models together with other evidence to make decisions related to healthcare practices and resource utilization. Models are particularly useful when the existing evidence does not yield obvious answers or does not provide answers to the questions of greatest interest, such as questions about the relative cost effectiveness of different practices. In this review we focus on models used to inform decisions about imaging technology, discussing the role of disease models for health policy development and providing a foundation for understanding the basic principles of disease modeling. We draw from the collective computed tomographic colonography (CTC) modeling experience, reviewing 10 published investigations of the clinical effectiveness and cost-effectiveness of CTC relative to colonoscopy. We discuss the implications of different modeling assumptions and difficulties that may be encountered when evaluating the quality of models. Finally, we underscore the importance of forging stronger collaborations between researchers who develop disease models and radiologists, in order to ensure that policy-level models accurately represent the experience of everyday clinical practices.

1.0 Introduction

Computer disease simulation models are increasingly being used to integrate and synthesize short-term evidence about diagnostic tests and procedures to project the long-term clinical

Correspondence: Carolyn M. Rutter, PhD, Biostatistics Unit, Group Health Research Institute, 1730 Minor Avenue, Ste 1600, Seattle, WA 98101, Phone: 206-287-2190, Fax: 206-287-2017, rutter.c@ghc.org.

and cost outcomes that are typically of interest to policy makers. A recent example of the use of modeling to inform policy decisions is the high-profile evaluation of mammography screening intervals for the 2009 US Preventive Services Task Force breast cancer screening recommendations (1–5). Models incorporate evidence on a variety of elements that impact the effectiveness of a particular technology or procedure, including disease incidence and mortality, operating characteristics of screening and diagnostic tests, patient acceptance of the tests, and treatment effectiveness

For example, the Centers for Medicare and Medicaid Services recently considered predictions from colorectal cancer models together with an independent review of the evidence on colorectal cancer screening that was commissioned by the Agency for Healthcare Research and Quality (6) when deciding if computed tomographic colonography (CTC) should be a reimbursable method of colorectal cancer screening among Medicare enrollees. The modeling study, conducted by an international team of public health researchers, cancer experts, and biostatisticians, used three independently developed models to predict health outcomes and lifetime cancer-related costs associated with a program of CTC screening compared to colonoscopy screening, sigmoidoscopy screening, fecal occult blood testing, and no colorectal cancer screening (7). Each of the three models demonstrated that the predicted life-years gained (vs. no colorectal cancer screening) from CTC screening among Medicare enrollees were only slightly less than the life-years gained from colonoscopy screening. They found that CTC screening could be a cost-effective option for colorectal cancer screening of the Medicare population if the cost of a CTC was substantially lower than a colonoscopy or if the availability of CTC screening would attract people who would otherwise not undergo colorectal cancer screening. This analysis of CTC motivates our discussion about the uses of disease models, the types of disease models, and how they are developed and applied in comparative-effectiveness and cost-effectiveness research.

2.0 Role of Disease Simulation Models in Health Policy for Radiologic Imaging

Researchers seeking information about the impact of imaging technologies on clinical outcomes encounter a variety of types of clinical evidence. However, each has its limitations. These limitations can often be overcome by integrating data with a computer disease simulation model. These models are particularly useful when some short-term outcomes are available about a technology, but evidence of the impact on long-term health outcomes is lacking.

Most studies of imaging tests focus on test accuracy, that is, how well a test can identify disease. While accuracy studies are key to understanding one-time test performance, they provide no information about clinical outcomes (8). Randomized trials are the gold standard study design for assessing outcomes. However, they are usually not a feasible method for addressing health policy questions because of their costs and long time horizon for completion, and because of the rapidly evolving nature of healthcare technologies. This is especially true for imaging technologies. In addition, randomized trials may be ethically difficult, or even impossible, to carry out once an imaging test is widely accepted as the standard of care, as is the case with mammography, or when the benefits are widely perceive to outweigh the risks, as in the case of coronary CT angiography (9). Finally, long-term randomized trials are often not possible because of funding limitations. As a result, randomized trials generally provide information about shorter-term outcomes such as the severity of disease detected (e.g., early or late stage cancer) or 1- to 5-year survival outcomes. Trials that focus on long-term disease outcomes following imaging are rare. One example is the Prostate, Lung, Colorectal, and Ovarian (PLCO) Screening Trial, which

randomized 38,349 men and 39,115 women to receive either routine medical care or active screening. In addition to colorectal and prostate cancer screening, the active screening arm included lung cancer screening with a baseline chest x-ray followed by annual chest x-rays for three years and, among women, ovarian cancer screening with transvaginal ultrasound and CA125 (10,11). In spite of the high value of randomized trials like PLCO, the results are not available for guiding policy now: the trial began in 1993 and mortality results are anticipated in 2015.

Observational studies are a common first approach for estimating the effect of imaging tests on disease outcomes because they use existing data and can be carried out at relatively low cost. Observational studies often include a broader patient population than randomized trials, providing important additional information, even when results from randomized trials are available (12). Observational studies can also provide information about longer-term outcomes following imaging and information about practices that are "off protocol". However, observational studies are prone to multiple biases (13–15). Although analyses can adjust for these biases, they can only account for known potential biases that were measured in the study. Accordingly, the results from observational studies alone may not provide sufficient evidence to inform health policy decision makers of the effectiveness of a technology.

Computer disease models provide a method for combining results from all available studies to fill gaps in existing knowledge. Disease models can be used to extrapolate results from accuracy studies to predict longer-term clinical outcomes. For example, the Digital Mammography Imaging Screening Trial (DMIST), a large study assessing the accuracy of film and digital mammography for breast cancer screening, found that the overall diagnostic accuracy of the two modalities was similar, but that digital mammography was superior in certain patient subgroups (16). Using a model of the natural history of breast cancer that incorporated the DMIST data on test performance, Tosteson et al. (3) projected the long-term health outcomes and incremental cost-effectiveness of alternative breast cancer screening strategies, including all-film and all-digital strategies and strategies that targeted use of digital mammography based on age and breast density. They found that all-digital mammography screening was no more effective, and in some cases it was less effective, than targeted use of digital mammography in terms of quality-adjusted life years. The authors note that their modeling analysis shows "how a shift to all-digital mammography screening may result in health gains for younger women (especially those with dense breasts), possibly at the expense of older women (especially those with nondense breasts)."

Disease models can also be used to simulate comparisons that were not evaluated in a trial. For example, McMahon and colleagues (17) estimated the effectiveness of helical CT for lung cancer screening among current and former smokers in the single-arm Mayo CT screening study. Using de-identified data from the study participants, they simulated the study outcomes, as well as predicted outcomes for a hypothetical control group that did not undergo helical CT screening. They concluded that while helical CT screening could reduce lung cancer-specific mortality, it would have little impact on overall mortality because of competing causes of death in current and former smokers.

For newly developed or rapidly evolving imaging technologies, studies of long-term outcomes are not possible because of insufficient observation time. In this context, disease models provide a tool for real-time evaluation of a new or evolving technology by combining information about the natural history of disease with the most recent results describing the sensitivity and specificity of the technology for disease detection.

Models can also be used to explore and identify aspects of a diagnostic test or intervention that have the largest impact on effectiveness and cost-effectiveness. For example, because CTC is less invasive than colonoscopy, it has the potential to draw more people into colorectal cancer screening. Models can be used to evaluate the impact of potential increased adherence attributable to this feature of CTC on the effectiveness and cost-effectiveness of CTC screening. Models can also assess the value of obtaining more information on uncertain aspects of the clinical decision, thereby helping to prioritize future research efforts (18). For example, using a model of colorectal cancer, Hassan and colleagues (19) determined that having additional information on screening uptake rates with less-invasive screening tests such as CTC compared with colonoscopy would be of greatest value in identifying an optimal colorectal cancer screening strategy.

Models are typically used to predict effectiveness measures such as life-years gained per treated individual, or life-years gained per imaging exam (20). Model predictions often, but not always, incorporate costs. These include the costs of the imaging test, as well as those of subsequent interventions or treatments and associated adverse events (e.g., mammogram costs plus the costs of treating breast cancer that would not have been diagnosed without the imaging test). The net cost or savings associated with an imaging strategy are calculated by comparison to the costs of a no-intervention scenario. Cost-effectiveness analysis combines life-years gained with cost predictions, providing policy makers with a framework for identifying the interventions that yield the greatest health benefit given resource constraints (21).

3.0 Basics of Disease Models

Disease models provide a framework for integrating and synthesizing information from multiple sources, including accuracy studies, randomized trials, and observational studies, that describe various aspects of the disease of interest and relevant technologies. Models are structured with a pre-specified set of health states (e.g. living without the disease of interest, living with preclinical disease, living with diagnosed disease while undergoing treatment, etc.) that describe the relevant aspects of the disease process. Modelers integrate information from different studies by using data to choose probabilities of transitioning through these disease states. The transitions can be specified as fixed probabilities or they may be based on mathematical formulae that allow the probability of transition to depend on characteristics of simulated individuals (e.g., age, sex, disease risk factors). The goal of this process of "model calibration" is to build a model that precisely reproduces the existing data, so later model predictions are as accurate as possible. In selecting the data for model calibration, modelers can choose to disregard unreliable or poorly designed studies and give greater weight to more rigorous and larger studies.

Computer disease models generally combine a natural history model that describes the disease process in the absence of screening with an intervention model to estimate the effectiveness of a diagnostic procedure or other medical intervention. Compared to natural history models, intervention models are relatively simple because they are based on observable outcomes (test sensitivity and specificity) and known or recommended intervention schedules. Intervention models often assume ideal or overly simplified circumstances. For example, models may assume no variability in test performance across individuals, even though such variability undoubtedly occurs because of both patient- and practitioner-specific factors (22,23). In addition, models may incorporate assumptions of perfect adherence with protocols (e.g., screening at pre-specified intervals), or equal adherence with all strategies under consideration (e.g., colonoscopy and CTC). These simplifying assumptions may be necessary because of a lack of data on adherence with repeated tests, or may be made to allow strategies to be compared in a head-to-head manner.

Disease models are categorized into three broad classes with increasing degrees of complexity: decision tree models, cohort models, and microsimulation models. All models make assumptions. In general, simpler disease models make different, but not fewer, assumptions about the disease processes than more complex models.

Decision tree models are relatively simple and are used to track outcomes associated with different courses of action (24,25). Outcomes can include disease status, 5-year survival, number of procedures, and costs. At each branching point, the tree specifies the chance of each subsequent outcome occurring, for example whether a positive result is a true or false positive finding. Alternative courses of action are compared by calculating the expected value of the outcome from each (i.e., multiplying the value assigned to each potential outcome by the probability that each occurs). A key limitation of decision trees is that they do not explicitly incorporate time and therefore are not suited to evaluation of repeated events, such as imaging implemented within a screening program.

Cohort models allow greater complexity than decision trees by describing the movement of groups of individuals through disease states over time. Cohort models allow movement between health states (e.g., "well", "sick", "dead") at fixed time intervals (e.g., one year) and are typically specified as Markov models: the probability of the next transition depends only on the current state and is independent of the prior history (i.e., how one got to the state) (26). Markov health states are, as a result, commonly described as "memory-less." This means that if, for example, only one health state of metastatic disease is designated in a cancer model, then once patients develop metastatic disease, patients who originally had low-grade primary cancers become indistinguishable from those who had high-grade primary cancers. Cohort models can be programmed in a spreadsheet or in a standard decision analysis software program (e.g., TreeAge Pro). However, these models may be impractical if "memory" of past events is important, or if the number of unique health states simulated becomes intractable (27).

Microsimulation models are more flexible than cohort models because they simulate the movement of individuals through disease states (28). Microsimulation models do not require the memory-less property of Markov models and allow events in continuous time rather than requiring transitions based on specified cycle lengths. Microsimulation models also allow specification of complex disease states that vary by individual and tumor characteristics, and these can be described by continuous variables (e.g., body mass index, or tumor size in mm), rather than categorically, as is often the case in cohort models. The price of greater flexibility is increased complexity and simulation times. Additionally, the complexity of microsimulation models typically requires customized programming using a flexible programming language such as C++, R, or Delphi.

There is a natural tension between model simplicity and model complexity, as we demonstrate in the following section comparing models used to evaluate CTC. The appropriate degree of model complexity depends on a variety of factors, including modeling goals and available information. Complex disease models are generally developed as part of a larger research agenda focusing on specific disease screening or treatment. These models are developed specifically for repeated use including assessment of interventions that may become available after initial model development. In contrast, simpler models may be developed to address a single question.

4.0 Example: Models Used to Evaluate CTC vs. Colonoscopy

As an example of the diversity of disease models, the multiple factors they may consider, and the ways in which they can be applied, we compared models used to evaluate both CTC and colonoscopy for colorectal cancer screening. We searched the published literature and

identified 10 unique models that have evaluated these technologies. Key model characteristics, discussed below, are summarized in Table 1. When models have been used in multiple publications examining CTC, we describe the publication with the most detailed model description.

Three of the models (29–31) evaluated colorectal cancer screening with CTC or with colonoscopy only, while the other seven models evaluated CTC, colonoscopy, and other screening modalities (e.g., fecal occult blood testing, flexible sigmoidoscopy, and strategies that combined the two tests) (7,32–35). One model was a decision tree (29), six were cohort models (30–35), and three were microsimulation models (MISCAN, CRC-SPIN, and SimCRC, used in a joint publication) (7). The models varied both in terms of the general level of complexity and the specific assumptions made. Most models specified a similar overall disease process, with eight (7,29,30,32,34,35) of the 10 models assuming that all cancers arose through the adenoma-carcinoma sequence (36,37). Two models also allowed for de novo cancers (31,33). Three of the models (29,31,32) allowed hyperplastic polyps, though it was unclear if they allowed co-occurrence of adenomas and hyperplastic polyps. All treated the hyperplastic polyps as "nuisance" polyps that induced unnecessary procedures; none allowed progression to cancer through this pathway (38).

The simulation of adenoma onset by age is an essential part of a colorectal cancer natural history model. While the malignant potential of adenomas of various sizes is not fully known, an incident adenoma represents a potential starting point for disease. The modelers uniformly based adenoma incidence on adenoma prevalence data, either from autopsy studies (7,31,34,39)(MISCAN, SimCRC), screening studies (29), or a combination of screening and autopsy data (7) (CRC-SPIN). However, even when modelers used the same type of studies for adenoma prevalence (e.g., autopsy studies alone), they did not necessarily use the same specific studies. Most modelers who used adenoma prevalence data from autopsy studies selected only one or two autopsy studies from at least 14 possible data sources (40). Thus the different CTC models integrated different evidence, which may contribute to the differences in their predictions.

While all of the models incorporated evidence on adenoma prevalence, the level of detail used to describe the occurrence of adenomas within individuals differed greatly. Conceptually, all models included an adenoma state, which included all individuals with at least one adenoma. The three microsimulation models (7) described this state in the most detail, simulating the size and location of multiple adenomas that may simultaneously occur. Most cohort models simulated at most one adenoma or polyp per individual (the most advanced) at any point in time and did not simulate the location of the adenoma or polyp in the colorectal tract. Models that do not allow synchronous adenomas, with some missed and some identified, need to adjust imaging characteristics that are reported at the lesion-level to sensitivities and specificities at the person-level to avoid underestimating the likelihood of detecting a person with at least one adenoma.

The models also differed with respect to whether and how they accounted for adenoma location. Only the three microsimulation models (7) and one cohort model (34) accounted for adenoma location. The microsimulation models simulated six locations (i.e., cecum; ascending, transverse, descending, or sigmoid colon; or rectum) and the cohort model simulated two (i.e., proximal colon vs. distal colon and rectum). Although not the focus of this review, the implications of not explicitly simulating adenoma location are important to consider when evaluating models that assess sigmoidoscopy screening strategies because sigmoidoscopy only detects lesions in the rectum or sigmoid colon.

The models varied in their description of adenoma malignancy potential. As shown in Table 1, several models based malignancy potential on transitions between clinically-relevant size categories: less than 10 mm vs. 10 mm or larger (31); 1–5 mm, 6–9 mm, and 10 mm or larger (7,33,34). One model specified a functional form for adenoma growth, resulting in continuous adenoma size (7) (CRC-SPIN). One model did not describe adenoma size (39). Two models described high- and low-risk adenomas, with high-risk adenomas defined as 10 mm or larger or less than 10 mm with villous features or high-grade dysplasia, and low-risk adenomas defined as all others. Adenoma size is important because accuracy is linked to lesion size for both CTC and colonoscopy. One of the models describing high- and low-risk adenomas based test sensitivity on size (32), but the other used adenoma risk category as a proxy for size (35) when specifying the accuracy of CTC and colonoscopy. Models that describe a single lesion describe the largest lesion and require person-level test accuracy, whereas models that allow multiple lesions of varying size may use lesion-level test accuracy. Two models ignored small (<6 mm) adenomas (29,34), assuming that these would not progress to cancer within the timeframe of the study (29) or would not be detected or removed at screening (34); the latter contradicts typical colonoscopy screening practices.

While there was much variability in the approach to modeling adenomas, the models were quite similar in the data used to simulate cancer incidence rates. All but one model used colorectal cancer incidence rates from cancer registries (7,30–35); the decision-tree model considered a 3-year time horizon and its developers used data from observational studies to inform 3-year cancer risk. Cancer incidence data for model calibration should be selected to reflect the natural history of the disease; for colorectal cancer, that means incidence prior to the dissemination of screening. Rates during post-screening periods may be elevated due to detection of indolent cancer that would not otherwise have been clinically detected. However, colorectal cancer incidence rates have fallen over time, partially due to detection and removal of adenomas through screening (41). Most models that used registry data did not specify the specific time periods of the obtained data (30,31,33–35), although the three microsimulation models (7) and one cohort model (32) specified the use of registry data collected before the use of colorectal cancer screening tests became more common (i.e., late 1970s to mid-1980s).

While the models used to compare CTC and colonoscopy were broadly similar, each used a somewhat different structure: using different data to describe adenoma prevalence, different categories for adenoma malignancy potential, and potentially different data sources for all elements. In the following section, we compare the findings from these diverse studies.

4.1 Predicted Effectiveness of CTC Relative to Colonoscopy

All of the models we examined used expected number of years of life as the primary measure of the effectiveness of screening for colorectal cancer using either colonoscopy or CTC. Consistent with recommendations on the conduct of cost-effectiveness analyses (21,42), future years of life were discounted to their present value in all models, although different discount rates were used. As shown in Table 1, the models evaluated the screening strategies in somewhat different populations. Most models focused on individuals from age 50 to death, though the three microsimulation models focused on the Medicare population of individuals 65 and older. The decision-tree analysis (29) estimated outcomes over a 3-year period following the initial screening exam, corresponding with the shortest interval for a follow-up exam. Many modelers reported results from sensitivity analyses, which explored the impact of their assumptions on results. Unless otherwise noted, we refer to the findings from each study's from primary, or "base case", assumptions. Vijan and colleagues(32) evaluated both 3D and 2D CTC. We focus exclusively on their results for 3D CTC because they found that a screening program based on 3D CTC was both less costly and more effective than screening based on 2D CTC (that is 3D CTC dominated 2D CTC).

All of the cohort and microsimulation models we reviewed found that colonoscopy screening every 10 years was more effective in terms of discounted years of life than CTC screening every 5 (7,35) or 10 years (30–34). The analysis by Vijan and colleagues (32) found that colonoscopy was more effective than CTC every 10 years, but *less* effective than CTC every 5 years. In addition, Teleford and colleagues (35) only reported discounted quality-adjusted years of life, which accounts for the fact that a year of life in perfect health is valued more than a year of life in poor health (20). The decision-tree model also found that colonoscopy was more effective over the 3-year period that it considered. As shown in Table 2, the assumed sensitivities and specificities of CTC and colonoscopy varied across studies, reflecting differences in the accuracy studies used to inform the models. Most models incorporated data suggesting that colonoscopy was more sensitive than CTC for detection of all lesions. Lee and colleagues (34) assumed that CTC was more sensitive than colonoscopy for detection of cancers and large (>10 mm) adenomas, but still found colonoscopy to be a more effective screening strategy than CTC.

4.2 Predicted Costs and Cost-Effectiveness of CTC Relative to Colonoscopy

There was some diversity in findings related to the costs and cost-effectiveness of colonoscopy and CTC screening (Table 2). Six models found that colonoscopy screening was less costly than CTC screening (7,30,31,33,34) so that colonoscopy was dominant, being both more effective and less costly than CTC. The analysis by Vijan and colleagues(32) found that colonoscopy every 10 years dominated 3D CTC every 10 years while 3D CTC every 5 years was more effective but also more costly than colonoscopy every 10 years. Three cohort models found that colonoscopy screening was more costly than CTC (30,33,34). Compared to the other models, the relative unit cost of colonoscopy versus CTC in these three models was much higher. They were 1.5 times higher for colonoscopy without polypectomy versus CTC, compared to at most 1.2 times higher for models finding colonoscopy to be less expensive. Only two of these studies (33,34) found CTC to have a more favorable incremental cost-effectiveness ratio than colonoscopy. Compared to other models, these two studies assumed a lower accuracy for CTC detection of small adenomas that are unlikely to lead to colorectal but increase costs because of referral to colonoscopy (shown in Table 2). In addition, Lee and colleagues (34) assumed that CTC was more accurate than colonoscopy for detection of cancers and adenomas 10 mm or larger. Other modelers explored the impact of CTC sensitivity on their findings with inconsistent results. Sonnenberg and colleagues(30) found that the cost-effectiveness of CTC increased with increasing sensitivity, but their model did not explicitly describe adenoma size. Ladabaum and colleagues(31) found that decreasing CTC sensitivity made it a less effective and more costly test. Telford and colleauges(35) found that increased sensitivity for *advanced adenomas* increased effectiveness and reduced costs for CTC. These reports suggest that when determining cost-effectiveness, assumptions about CTC accuracy to detect adenomas with high malignant potential are more important than assumptions about the overall accuracy or CTC, pointing to the importance of the underlying natural history model for colorectal cancer, and, in particular, assumptions about the rate at which adenomas transition to cancer.

Although 8 out of 10 models found CTC to be either dominated by colonoscopy or have a less-attractive incremental cost-effectiveness ratio than colonoscopy, several noted that CTC could be cost-effective at a lower per-exam cost (7,30–32). For example, the microsimulation models found that CTC could be cost-effective compared to colonoscopy and other screening strategies if exams cost 22% to 41% of the cost of a colonoscopy (without polypectomy).

One limitation of these studies is that while many models (30,32–35) considered less than perfect adherence to screening recommendations, all assumed the same degree of adherence

to colonoscopy and CTC screening. Few explored the impact of differential adherence to colonoscopy and the less invasive CTC exam. Sonnenberg and colleagues(30) provide graphs of the effect of adherence on cost per life-years gained, but did not provide information on the impact of adherence on life-years gained alone. In sensitivity analyses, Hassan and colleagues (33) found that CTC was equally effective and less costly than colonoscopy if the colonoscopy adherence rate was 60% vs. 65% for CTC, which is a less than 10% improvement in relative adherence. Microsimulation model results (7) also suggested that CTC could be cost-effective compared to colonoscopy and other screening modalities if its availability would result in otherwise unscreened individuals being screened; however, relative adherence with CTC vs. other modalities would need to be at least 25% higher for this to be the case.

5.0 Model Evaluation

Evaluating computer disease simulation models is difficult because of their complexity, as illustrated by the models comparing CTC and colonoscopy. Therefore, it is incumbent upon modelers to enable assessment of their models by providing a clear description of the assumptions and data sources used to develop their models. That is, modelers must strive to make their models "transparent".

Model transparency often focuses on clearly describing the model structure and inputs. The structure of the natural history model is a basic assumption, often based on a combination of published findings and expert opinion. The assumed model structure is critical, because it provides a framework for combining information used to calibrate the model. However, the implications of using a particular model structure may not always be apparent. The interplay between model structure and calibration data can further confuse model evaluation. For example, because individuals with more than one adenoma have more opportunity to be detected, assumptions about the number of adenomas within individuals affect implied person-level detection rates.

Communicating all the assumptions and implications of a model is challenging. However, the models evaluating CTC vs. colonoscopy provide examples of how these features might be presented. Model predictions are both the results from a model and an important, but underused, avenue for model evaluation (43). For example, when models are calibrated, selecting probabilities and parameters to reproduce known or expected results, the selected parameters may be published but calibration results are rarely included. Thus, readers are given little or no information about how well a model was able to fit different calibration points. Lee and colleagues (34) described their model's fit to data on adenoma prevalence and provided an online appendix with information about the model's fit to data on colorectal cancer incidence, mortality, and stage at diagnosis. Such goodness-of-fit results have the potential to provide an intuitive sense of the model to end users. Another example of the use of model predictions to aid in model evaluation pertains to adenoma "dwell time", the time from the onset of an adenoma of detectable size to clinically detected colorectal cancer. Dwell time is an important component of all colorectal cancer models as it represents the period of potential disease prevention through polypectomy or mortality reduction through earlier cancer diagnosis and treatment initiation. However dwell time is typically an output rather than an input of a disease simulation model, thus it is not easy to compare models on this aspect. Providing model predictions of the mean dwell time may provide a more intuitive indication of a model's structure and assumptions. Assumptions about dwell time are particularly important when comparing tests with different screening intervals. For example, Vijan and colleagues(32) found that shorter dwell times improved the cost-effectiveness of CTC carried out every 5 years relative to colonoscopy every 10 years.

Between-model comparisons provide another method for assessing structural assumptions on model predictions. The Cancer Incidence and Surveillance Modeling Network (CISNET), a National Cancer Institute-funded program that began in 2000 that has supported the development of multiple disease models for five cancer sites, has performed several comparative modeling analyses (1,4,7,44–47). These CISNET analyses, which compare independently developed models with different structures and assumptions, shed light on the impact and importance of the different assumptions. However, CISNET is an exception. Direct comparison of models is uncommon for modelers funded through different mechanisms.

Finally, model validation provides an important, but often unavailable, method for estimating the accuracy of model predictions. Models are validated by testing their ability to predict outcomes that were not used for model calibration. Since the data needed for model development are often sparse, setting aside data for later model validation is often not feasible. For this reason, validation often requires new data that arise from studies published after model development is complete. For example, one of the microsimulation models used to evaluate CTC (CRC-SPIN) was later validated using a case-control study examining the effect of colonoscopy on colorectal cancer mortality (43). While validation often requires several assumptions related to both the population and the intervention model, it remains the best "test" of a model's predictive ability.

6.0 Discussion

Policy makers need to gather, synthesize, and interpret information to make decisions. If presented with multiple sources of information (e.g., results from primary studies), then each decision-maker carries out his or her own internal synthesis. Disease models offer a way to integrate and synthesize information using the structure of the natural history and intervention models. Disease models range in complexity from relatively simple decision trees, to more detailed cohort models and complex microsimulation models. Simpler models are relatively easy to construct and quickly implement, and may be a natural first step towards developing more complex models. Cohort models can be readily developed within a four- or even two-year funding cycle, offering a natural adjunct to projects focusing on accuracy of imaging tests or short-term outcomes from testing. Microsimulation models require longer-term support for their development. This investment is returned in a detailed model that can be applied to a wide range of problems.

A common criticism regarding the use of disease models to aid decision-making is that because models are prediction tools, their results are, by nature, more uncertain than those directly derived from patient data. However, while they are indeed predictive, disease models are developed by integrating directly observed data and they synthesize information from a broad range of studies. Improving the transparency and interdisciplinary involvement in model development will ensure that model assumptions and structure are more broadly understood. In addition, working toward methods that produce standardized and clinically relevant, relatable model predictions will allow more direct model comparisons, easing communication of model assumptions and their implications.

Models are especially useful when no long-term data are available to inform decision making. In these cases, models can provide robust predictions that cannot be reasonably achieved through other means. However, model predictions should always be considered in the context of all other available evidence. For example, the Medicare Evidence Development and Coverage Advisory Committee (MedCAC) considers a range of evidence and explicitly considers the quality of evidence before it gives its coverage recommendation to the Centers for Medicare and Medicaid Services (48). When considering coverage for

CTC screening, MedCAC considered evidence on the operating characteristics and risks of different screening tests for colorectal cancer based on a systematic review (6) and testimony from several experts in the fields of cancer screening, gastroenterology, and radiology, in addition to microsimulation model predictions (49)

The use of computer disease models to guide decision-making is a natural result of the increasing amount and diversity of information available and these models are an important tool for estimating the comparative effectiveness of different imaging studies. Models are only as good as the current state of knowledge, and like other technologies they can and should be continually improved. One avenue for improvement is increased collaboration between radiologists and modelers. Radiologist can provide critical expertise during model development, guiding selection of key studies and information for calibration. In turn, modelers could advise radiologists on how to design efficient studies that will also yield data to fill knowledge gaps in faced when modeling disease processes. At study completion, these collaborations will result in the best possible predictions of long-term outcomes from short-term trial data.

Acknowledgments

References

1. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005; 353(17):1784–92. [PubMed: 16251534]

2. Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective cost-effectiveness analysis of screening mammography. J Natl Cancer Inst. 2006; 98(11):774–82. [PubMed: 16757702]

3. Tosteson AN, Stout NK, Fryback DG, et al. Cost-effectiveness of digital mammography breast cancer screening. Ann Intern Med. 2008; 148(1):1–10. [PubMed: 18166758]

4. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Ann Intern Med. 2009; 151(10):738–47. [PubMed: 19920274]

5. Lee JM, McMahon PM, Kong CY, et al. Cost-effectiveness of Breast MR Imaging and Screen-Film Mammography for Screening BRCA1 Gene Mutation Carriers. Radiology. 2010; 254(3):793. [PubMed: 20177093]

6. Whitlock EP, Lin JS, Liles E, Beil TL, Fu R. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2008; 149(9):638–58. [PubMed: 18838718]

7. Knudsen AB, Lansdorp-Vogelaar I, Rutter CM, et al. Cost-effectiveness of computed tomographic colonography screening for colorectal cancer in the medicare population. J Natl Cancer Inst. 2010; 102(16):1238–52. [PubMed: 20664028]

8. Gatsonis C. The promise and realities of comparative effectiveness research. Stat Med. 2010; 29(19):1977–81. discussion 96–7. [PubMed: 20683886]

9. Douglas PS, Budoff M, Tunis S, Woodard PK, Justman RA, Honigberg R. A new era for cardiovascular imaging? Implications of the revoked national coverage decision for CT angiography on future imaging reimbursement. JACC Cardiovascular imaging. 2008; 1(3):398–403. [PubMed: 19356455]

10. Partridge E, Kreimer AR, Greenlee RT, et al. Results from four rounds of ovarian cancer screening in a randomized trial. Obstet Gynecol. 2009; 113(4):775–82. [PubMed: 19305319]

11. Hocking WG, Hu P, Oken MM, et al. Lung Cancer Screening in the Randomized Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Journal of the National Cancer Institute. 2010; 102(10):722–31. [PubMed: 20442215]

12. Fleurence RL, Naci H, Jansen JP. The Critical Role Of Observational Evidence In Comparative Effectiveness Research. HAF Health Affairs. 2010; 29(10):1826–33.

13. Berger ML, Mamdani M, Atkins D, Johnson ML. Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report-Part I. Value in Health. 2009; 12(8):1044–52. [PubMed: 19793072]

14. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good Research Practices for Comparative Effectiveness Research: Approaches to Mitigate Bias and Confounding in the Design of Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report-Part II. Value in Health. 2009; 12(8):1053–61. [PubMed: 19744292]

15. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report-Part III. Value in Health. 2009; 12(8):1062–73. [PubMed: 19793071]

16. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005; 353(17):1773–83. [PubMed: 16169887]

17. McMahon PM, Kong CY, Johnson BE, et al. Estimating Long-term Effectiveness of Lung Cancer Screening in the Mayo CT Screening Study1. Radiology. 2008; 248(1):278–87. [PubMed: 18458247]

18. Claxton K, Sculpher M. Using value of information analysis to prioritise health research: some lessons from recent UK experience. Pharmacoeconomics. 2006; 24(11):1055–68. [PubMed: 17067191]

19. Hassan C, Di Giulio E, Hunink MGM, et al. Value-of-information analysis to guide future research in colorectal cancer screening. Radiology. 2009; 253(3):745–52. [PubMed: 19789242]

20. Gold, MR. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.

21. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med. 1977; 296(13):716–21. [PubMed: 402576]

22. Elmore JG, Jackson SL, Miglioretti DL, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. Radiology Radiology. 2009; 253(3):641–51.

23. Johnson CD, Chen MH, Toledano AY, et al. Accuracy of CT colonography for detection of large adenomas and cancers. N Engl J Med. 2008; 359(12):1207–17. [PubMed: 18799557]

24. Drummond, MF.; Drummond, MF. Mfteeohcp. Methods for the economic evaluation of health care programmes. 3. Oxford; New York: Oxford University Press; 2005.

25. Detsky AS, Naglie G, Krahn MD, Naimark D, Redelmeier DA. Primer on medical decision analysis: Part 1--Getting started. Medical decision making: an international journal of the Society for Medical Decision Making. 1997; 17(2)

26. Beck J, Pauker S. The Markov process in medical prognosis. Med Decis Making. 1983; 3:419–58. [PubMed: 6668990]

27. Alagoz O, Hsu H, Schaefer A, Roberts M. Markov Decision Processes: A Tool for Sequential Decision Making under Uncertainty. Medical Decision Making. 2010; 30(4):474–83. [PubMed: 20044582]

28. Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic Microsimulation Models for Health Outcomes: A Review. Med Decis Making. 2010

29. Heitman SJ, Manns BJ, Hilsden RJ, Fong A, Dean S, Romagnuolo J. Cost-effectiveness of computerized tomographic colonography versus colonoscopy for colorectal cancer screening.

CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne. 2005; 173(8):877–81.

30. Sonnenberg A, Delco F, Bauerfeind P. Is Virtual Colonoscopy a Cost-Effective Option to Screen for Colorectal Cancer? Am J Gastroenterol. 1999; 94(8):2268–74. [PubMed: 10445561]

31. Ladabaum U, Song K, Fendrick AM. Colorectal neoplasia screening with virtual colonoscopy: when, at what cost, and with what national impact? Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association. 2004; 2(7):554–63. [PubMed: 15224279]

32. Vijan S, Hwang I, Inadomi J, et al. The Cost-Effectiveness of CT Colonography in Screening for Colorectal Neoplasia. Am J Gastroenterol. 2007; 102(2):380–90. [PubMed: 17156139]

33. Hassan C, Zullo A, Laghi A, et al. Colon cancer prevention in Italy: Cost-effectiveness analysis with CT colonography and endoscopy. Dig Liver Dis. 2007; 39(3):242–50. [PubMed: 17112797]

34. Lee D, Muston D, Sweet A, Cunningham C, Slater A, Lock K. Cost Effectiveness of CT Colonography for UK NHS Colorectal Cancer Screening of Asymptomatic Adults Aged 60–69 Years. Applied Health Economics and Health Policy. 2010; v8(n3):141–54. [PubMed: 20369905]

35. Telford JJ, Levy AR, Sambrook JC, Zou D, Enns RA. The cost-effectiveness of screening for colorectal cancer. CMAJ. 2010; 182(12):1307. [PubMed: 20624866]

36. Muto T, Bussey HJ, Morson BC. The evolution of cancer of the colon and rectum. Cancer. 1975; 36(6):2251–70. [PubMed: 1203876]

37. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. Br J Surg. 2002; 89(7):845–60. [PubMed: 12081733]

38. O'Brien MJ. Hyperplastic and serrated polyps of the colorectum. Gastroenterol Clin N Am. 2007; 36:947–68.

39. Sonnenberg A, Delco F, Inadomi JM. Cost-effectiveness of colonoscopy in screening for colorectal cancer. Ann Intern Med. 2000; 133(8):573–84. [PubMed: 11033584]

40. Rutter CM, Yu O, Miglioretti DL. A hierarchical non-homogenous Poisson model for meta-analysis of adenoma counts. Stat Med. 2007; 26(1):98–109. [PubMed: 16372387]

41. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin. 2010; 60(5):277–300. [PubMed: 20610543]

42. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. JAMA. 1996; 276(15):1253–8. [PubMed: 8849754]

43. Rutter CM, Savarino JE. An evidence-based microsimulation model for colorectal cancer. Cancer Epidemiol Biomarkers Prev. 2010 August; 19(2010):1992–2002. [PubMed: 20647403]

44. 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. Annals of Internal Medicine. 2010; 153(6):368–77. [PubMed: 20855801]

45. 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. 2008; 149(9):659–69. [PubMed: 18838717]

46. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. JNCI Journal of the National Cancer Institute. 2009

47. Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the U.S. prostate cancer mortality decline. Cancer Causes & Control. 2008; 19(2):175–81. [PubMed: 18027095]

48. Normand SL, McNeil BJ. What is evidence? Stat Med. 2010; 29(19):1985–8. discussion 96–7. [PubMed: 20683888]

49. U.S. Department of Health & Human Services. MedCAC Meeting, Computed Tomography Colonography (CTC). [Accessed 11/2/2010]. Available at: http://www.cms.gov/mcd/viewmcac.asp?from2=viewmcac.asp&where=index&mid=45&

50. Ladabaum U, Song K. Projected national impact of colorectal cancer screening on clinical and economic outcomes and health services demand. Gastroenterology. 2005; 129(4):1151–62. [PubMed: 16230069]

51. 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. [PubMed: 17455218]

52. Hassan C, Pickhardt PJ, Laghi A, et al. Computed tomographic colonography to screen for colorectal cancer, extracolonic cancer, and aortic aneurysm: model simulation with cost-effectiveness analysis. Arch Intern Med. 2008; 168(7):696–705. [PubMed: 18413551]

53. Pickhardt PJ, Hassan C, Laghi A, Kim DH. CT colonography to screen for colorectal cancer and aortic aneurysm in the Medicare population: Cost-effectiveness analysis. AJR Gastrointestinal Imaging. 2009; 192:1–9.

54. Regge D, Hassan C, Pickhardt PJ, et al. Impact of computer-aided detection on the cost-effectiveness of CT colonography. Radiology. 2009; 250(2):488–97. [PubMed: 19188317]

55. Hassan C, Laghi A, Pickhardt PJ, et al. Projected impact of colorectal cancer screening with computerized tomographic colonography on current radiological capacity in Europe. Aliment Pharmacol Ther. 2008; 27:366–74. [PubMed: 18005247]

56. Lansdorp-Vogelaar I, Ballegooijen Mv, Zauber AG, Boer R, Wilschut J, Habbema JDF. At what costs will screening with CT colonography be competitive? A cost-effectiveness approach. International Journal of Cancer. 2009; 124(5):1161–8.

Table 1

Ten studies that used computer simulation models to compare computed tomographic colonography (CTC) and colonoscopy

Study	Model Type	Study Population & Time Horizon	De novo Cancers	Location	Multiple Adenomas	Malignant Potential
Heitman et al (29)	**Decision tree**	**Canadian 50-year-old males to age 53 years**	**No**	**No**	**No**	**Adenoma Size: 6–9 mm vs.10+ mm**
Sonnenberg et al (30)	Markov cohort	US 50-year-olds to death	No	No	No	Not Modeled
Ladabaum et al[E] (31)	Markov cohort	US 50-year-olds to death	Yes	No	No	Adenoma Size: <10 mm vs.10+ mm
Vijan et al (32)	Markov cohort	US 50-year-olds to death	No	No	No	**Low Risk:** 1–5 mm or 6–9 mm without high-risk histology[D] **High Risk:** 10+ mm or <10 mm (1–5 mm or 6–9 mm) with high-risk histology[D]
Hassan et al[F] (33)	Markov cohort	Italian 50-year-olds to age 80 years	Yes	No	No	Adenoma Size: 1–5 mm, 6–9 mm, 10+ mm
Lee et al (34)	Markov cohort	British 50-year-olds to death	No	Yes[A]	Yes[C]	Adenoma Size: 6–9 mm, 10+ mm
Telford et al (35)	Markov cohort	Canadian 50-year-olds to death	No	No	No	**Low Risk:** <9 mm without high-risk histology[D] **High Risk:** 9+ mm or <9 mm with high-risk histology[D]
Knudsen et al (7)						
MISCAN[G]	Microsimulation	US 65-year-olds to death	No	Yes[B]	Yes	Adenoma Size: 1–5 mm, 6–9 mm, 10+ mm
CRC-SPIN	Microsimulation	US 65-year-olds to death	No	Yes[B]	Yes	Adenoma Size: Continuous size in mm
SimCRC	Microsimulation	US 65-year-olds to death	No	Yes[B]	Yes	Adenoma Size: 1–5 mm, 6–9 mm, 10+ mm

[A] Two locations modeled (distal colon and rectum versus proximal colon)

[B] Six locations modeled: cecum; ascending, transverse, descending, and sigmoid colon; rectum

[C] One adenoma may occur in each of two locations

[D] High-risk histology is present when an adenoma has advanced dysplasia or villous features.

[E] Additional publication on CTC with this model: (50)

[F] Additional publications on CTC with this model: (51–55)

[G] Additional publication on CTC with this model:(56)

Table 2

Conclusions about relative costs and of computed tomographic colonography (CTC) and colonoscopy (COL) and assumptions about test sensitivity, specificity and costs for 10 disease simulation models. Unless noted, sensitivity estimates are per-person and costs are in US dollars.

Study	Sensitivity for Adenoma by Size/Risk								Specificity		Per Exam Cost			Least Costly Screening Program
	1–5 mm/Low-Risk		6–9 mm/Low-Risk		≥10 mm/High-risk		Cancer							
	CTC	COL	CTC	COL	CTC	COL	CTC	COL	CTC	COL	COL	COL with polypectomy	CTC	
Heitman et al (29)	*	*	0.61	0.94	0.71	0.96	†	†	0.84	1.0	$547[I]	$668[I]	$445[I]	COL
Sonnenberg et al (30)	0.80	0.95	0.80	0.95	0.80	0.95	0.80	0.95	0.94	1.0	$728	$1,139	$478	CTC[A]
Ladabaum et al (31)	0.70	0.85	0.70	0.85	0.75	0.90	0.95	0.95	0.85	1.0	$820	$1,200	$820	COL
Vijan et al (32)[B]	0.46	0.85	0.83	0.85	0.91	0.95	‡	‡	0.91	1.0	$653	$831	$559	COL
Hassan et al (33)	0.48	0.80	0.70	0.85	0.85	0.90	0.95	0.95	0.86	0.90	€148.2	€228.6	€100.9	CTC
Lee et al (34)	*	*	0.653	0.924[C] 0.950[D]	0.899	0.859[C] 0.883[D]	0.899	0.859[C] 0.883[D]	0.88	1.0	£488[G]	£488[G]	£128	CTC
Telford et al (35)	0.50[E]	0.92[E]	0.50[E]	0.92[E]	0.78[F]	0.97[F]	0.89	0.93	0.91	1.0	$424[I]	$609[I]	$590[I]	COL
Knudsen et al (7)[J]	0[K]	0.75	0.84	0.85	0.92	0.95	0.92	0.95	0.80	0.90	$498[H]	$649[H]	$488[H]	COL

* Adenomas <6 mm not modeled

† Not reported, assumed no prevalent colorectal cancer

‡ Not reported

[A] COL screening was more cost-effective than CTC screening

[B] CTC was evaluated with both 2- and 3-dimensional reading displays. Sensitivity, specificity and costs shown are for 3-dimensional CTC, which dominated 2-dimensional CTC.

[C] Adenomas in the proximal colon

[D] Adenomas in the distal colon

[E] Low-risk adenomas, without high-risk histology

[F] High-risk adenomas, includes lesions <10 mm with high-risk histology

[G] Separate costs for COL with and without polypectomy not specified

[H] Costs from the perspective of the Centers for Medicare and Medicaid Services

[I] Canadian dollars

[J] Sensitivity estimates are per lesion

[K] Analysis assumed a 6 mm referral threshold for a follow-up colonoscopy, thus the effective sensitivity for 1–5mm adenomas was 0