Cross-Model Comparisons to Improve the Value of Modeling: The Case of Colorectal Cancer Screening

Michael Pignone, MD, MPH, David F. Ransohoff, MD

Policy makers need rigorous and transparent methods to compare and evaluate different health care strategies. To compare strategies for the same condition, head-to-head trials are often impractical because they require large sample sizes and years to conduct, and they are not well suited for comparing multiple different options or measuring long-term outcomes. Modeling may offer advantages for conducting this type of comparative effectiveness research because of its ability to compare multiple strategies, including different combinations of interventions and different time intervals. In addition, cost-effectiveness modeling can examine incremental efficiency (cost per unit of health gained compared with the next best strategy), thus providing information directly relevant to decision makers.

Modeling, however, has several important limitations: It requires using multiple types of evidence (about biology or clinical behavior) drawn from a variety of sources, and model results can be compromised by inaccurate inputs or assumptions. Because modeling is inherently complex, it can be difficult to understand, explain, and critically evaluate. For these reasons, it may not be accepted or used by decision makers uncomfortable with such details.

Addressing these limitations, particularly improving transparency and ''accessibility'' (or ''understandability''), might increase the value of modeling, as suggested by recent guidelines for best practices in modeling.[1,2] Following these guidelines can increase the value of a model by explicit identification and estimation of model inputs, examination of the effect of different assumptions on results through sensitivity analyses (including cross-model comparisons), and identification of simpler, more easily interpretable ways of presenting methods and results for end users, including policy makers. One other way to increase a model's value and appropriate use is to show the extent to which ''deep'' assumptions about disease natural history can affect outcomes. Assessing such assumptions is particularly challenging because the biology or natural history of health conditions may be particularly difficult to observe directly.

Colorectal cancer (CRC) screening offers an excellent case study to illustrate the use of modeling to help inform preventive care policy. First, several randomized controlled trials have directly demonstrated that screening for CRC (fecal occult blood testing [FOBT] and sigmoidoscopy) can be effective in reducing CRC incidence and mortality compared with no screening.[3,4] Although other screening methods (e.g., colonoscopy, radiological testing, newer FOBT) are available, they have as yet not been evaluated in rigorous randomized trials (and such trials are so expensive and cumbersome that few are likely to ever be done). In addition, no trials have compared different screening methods head to head. Because available screening tests differ considerably in important details (e.g., frequency, invasiveness, preparation required), the decision about which strategy or strategies to employ is complex. Finally, many aspects of CRC and CRC screening are well understood, allowing modelers to accurately represent many key parameters.

With this wealth of evidence, choices, and clinical importance, CRC screening has been the subject of multiple well-developed models[5] that have generally found screening (by any of the commonly endorsed screening strategies) to be effective compared with

From the University of North Carolina, Chapel Hill.

Address correspondence to Michael Pignone, MD, MPH, University of North Carolina, 5039 Old Clinic, Chapel Hill, NC 27599; e-mail: pignone@med.unc.edu.

DOI: 10.1177/0272989X11412195

no screening; however, they have not reached consistent conclusions about the incremental effects of different screening strategies.[5,6] Perhaps the most rigorous modeling effort for CRC screening has been conducted by the Cancer Intervention and Surveillance Modeling Network (CISNET)] collaboration, a group of previously independent modelers who have come together to apply best practices in modeling, rigorously evaluate different models for CRC screening, and use this information to help policy makers.[7]

In this issue of *Medical Decision Making* (*MDM*), 2 articles from the CISNET modelers examine differences across 3 CRC screening models from their collaboration (Microsimulation Screening Analysis [MISCAN], Colorectal Cancer Simulated Population model for Incidence and Natural history [CRC-SPIN], and Simulation Model of Colorectal Cancer [SimCRC]).[8,9] These articles examine how the natural history of CRC has been simulated in each model, based on CRC incidence data and cross-sectional data on adenoma prevalence, and how choices or assumptions about natural history affect the calculated efficacy of screening. The principal difference between the models is that MISCAN assumes that some adenomas are nonprogressive, whereas CRC-SPIN and SimCRC do not. As a result, the average dwell times (the amount of time required for adenomas to develop into cancers and for cancers to become symptomatic and ultimately lethal) are quite different across models. Kuntz and colleagues[8] noted important differences in dwell times across the 3 models: The average dwell time in MISCAN is 10.6 years v. 25.2 and 25.8 years for SimCRC and CRC-SPIN, respectively. Nearly all of this difference was related to the time involved in adenoma progression.

These differences, derived from the ''deep assumption'' about natural history, have important implications for the evaluation of CRC screening strategies. More than 90% of cancers in SimCRC and CRC-SPIN developed from adenomas present for more than 10 years; in contrast, MISCAN found that most cancers developed from adenomas arising within 10 years, which means they would not have been detected and removed if a colonoscopy had been done 10 years previously. Accordingly, van Ballegooijen and colleagues[9] demonstrated that the predicted efficacy of colonoscopy screening every 10 years for the 3 models differed considerably: MISCAN predicted relative reductions in CRC incidence and mortality of 52% and 65%, respectively. Corresponding values were 91% and 92% for CRC-SPIN and 82% and 84% for SimCRC.

The large differences between MISCAN and the other 2 models may have important implications for clinicians and policy makers. If the average dwell time for polyps is over 20 years, then screening methods that accurately detect and lead to removal of polyps present at age 60, for example, are likely to have nearly as much efficacy as methods that employ additional screening tests at regular intervals after age 60. This would favor devoting resources to tests with high single-application sensitivity, such as colonoscopy, over tests with lower single-application sensitivity that are repeated at more frequent intervals, such as FOBT. On the other hand, if some (or many) polyps grow rapidly into cancers, then screening methods with shorter intervals between tests would be relatively favored over those with longer intervals. Although the CISNET modelers did not find major differences between MISCAN and SimCRC in terms of optimal testing intervals when examining life years gained per number of colonoscopies required, it is possible that such differences would emerge in analyses comparing different strategies with regard to cost per life year gained. Thus, it is important to resolve, through other research, which pattern of natural history is most accurate.

One indirect source of evidence about natural history is the analysis of interval cancers in the polyp prevention studies, which show that a substantial number of cancers appear relatively soon after a colonoscopy that removes all identified adenomas. Martínez and colleagues[10] pooled data from 8 studies of a total of 9167 adults who had colonoscopy and polypectomy. In 48 months (median time) following initial polypectomy, 11.8% had advanced adenomas and 0.6% invasive cancer. These results are somewhat higher than would be expected based on the average dwell times predicted in the SimCRC and SPIN-CRC models and hence lend support to the model of natural history used in MISCAN. However, it is possible that some adenomas and cancers detected in follow-up were present but missed on the initial colonoscopy, rather than arising de novo, so the data do not directly resolve the ''deep assumption'' problem.

To try to understand the different natural history assumptions, the CISNET modelers propose to examine data from ongoing trials of sigmoidoscopy screening[3] whose detailed findings, when published, with respect to adenoma detection and cancer incidence after sigmoidoscopy (and polypectomy) may offer additional insight into the growth pattern of adenomas and cancers. Specifically, if the trials show a low rate of interval CRC or advanced adenoma, it may support the ''long dwell time'' hypothesis. Conversely, if many adenomas or cancers are detected in the 10 years

following sigmoidoscopy, the MISCAN model's heterogeneous natural history assumption (with some rapidly growing adenomas) may be favored.

CRC screening is a good case study to examine the potential for modeling to guide future research and inform health policy. The CISNET collaboration is a particularly good example of how to systematically address differences across models and use them to pose new research questions (either for new studies or for additional analyses of existing data). However, despite these methodological advances, policy makers are only beginning to have experience in using modeling results to help guide health policy.[11] To take best advantage of modeling, policy makers must have core skills in interpreting modeling results, understanding how modeling addresses uncertainty, and evaluating whether the residual uncertainty is likely to affect the net benefit of a given option compared with the alternative. Modelers can facilitate widespread use of models in health policy making by focusing on clear presentation of their results and on the key assumptions that underlie them.

REFERENCES

1. Weinstein MC, O'Brien B, Hornberger J, et al; ISPOR Task Force on Good Research Practices—Modeling Studies. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value Health. 2003;6(1):9–17. PMID: 12535234

2. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. Pharmacoeconomics. 2006;24(4):355–71. PMID: 16605282

3. Atkin WS, Edwards R, Kralj-Hans I, et al; UK Flexible Sigmoidoscopy Trial Investigators. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet. 2010;375(9726):1624–33. PMID: 20430429

4. 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. PMID: 18838718

5. 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. 2002;137(2):96–104. PMID: 12118964

6. Pignone M, Saha S, Hoerger T, Lohr KN, Teutsch S, Mandelblatt J. Challenges in systematic reviews of economic analyses. Ann Intern Med. 2005;142(12, Pt 2):1073–9. PMID: 15968032

7. 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. PMID: 18838717

8. Kuntz KM, Lansdorp-Vogelaar I, Rutter CM, et al. A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. Med Decis Making. 2011;31(4):530–539.

9. van Ballegooijen M, Rutter CM, Knudsen AB, et al. Clarifying differences in natural history between models of screening: the case of colorectal cancer. Med Decis Making. 2011;31(4):540–549.

10. Martínez ME, Baron JA, Lieberman DA, et al. A pooled analysis of advanced colorectal neoplasia diagnoses after colonoscopic polypectomy. Gastroenterology. 2009;136(3):832–41. PMID: 19171141

11. U.S. Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2008;149(9):627–37. PMID: 18838716