Exploring the Unknown and the Unknowable with Simulation Models

Louise B. Russell, PhD

This issue of *Medical Decision Making* (*MDM*) is devoted to simulation modeling in health. The papers were not invited for the issue but are a special collection from the many on modeling that are submitted to *MDM* every year. Those under ''Exploring Model Structure'' deal with various aspects of reality that might be modeled: different ways of modeling preclinical disease, the impact of quality of life on health outcomes, correlations among health outcomes, the balance of benefits and harms over different time horizons, and adapting models to the stage of the decision. Those under ''Calibrating Models'' describe methods for assigning values to model parameters for which no direct evidence is available. And the papers under ''Representing Uncertainty in Models'' discuss methods for incorporating uncertainty in the data into a model and showing its effects on predicted outcomes.

Simulation models are used to predict and compare the consequences of medical interventions. The typical application examines variations of an intervention that have not been, perhaps cannot be, tested in trials. Hlatky[1] recently observed that ''a model is particularly useful in scenarios with multiple clinical options because it is difficult to study several strategies in a single clinical trial.'' Pignone and Ransohoff[2] make the same point in their editorial about colorectal cancer screening models in this issue. The alternative strategies are potentially observable but expensive to observe in terms of resources and elapsed time since it can take years to mount a trial and follow the participants long enough to determine the outcomes of, say, different screening schedules. Models provide decision makers with results much more quickly.

From Rutgers University, New Brunswick, New Jersey.

Address correspondence to Louise B. Russell, PhD, Institute for Health, Rutgers University, 112 Paterson Street, New Brunswick, NJ 08901, USA; telephone: (848) 932-6507; e-mail: lrussell@ifh.rutgers.edu.

DOI: 10.1177/0272989X11412078

Exactly this sort of analysis was conducted to inform the US Preventive Services Task Force's 2009 recommendation for biennial mammography.[3] Six models simulated the lifetime consequences—reduction in breast cancer deaths, life years gained, number of mammograms, false-positive mammograms—of 20 different screening strategies, far more than could be tested in a clinical trial. The strategies were differentiated by screening frequency and the ages at which screening started and stopped. All 6 models supported the conclusion that biennial screening captured most of the benefit of annual screening with far fewer false-positive results.

Screening illustrates a problem common to decision making in health and thus to tools that support decision making—the need to imagine the unknown. To explore the effects of screening, most simulation models include a component, or submodel, that describes the trajectory of the condition before it reaches the point at which it can be detected by current screening tests, the unknown. In this issue, Wever and others[4] explore 2 common screening submodels, stage shift and cure, each of which reflects a different idea about how preclinical disease might act. The stage-shift approach models mortality reductions from screening as due to the discovery of the disease at an earlier stage, which has a lower mortality rate. The cure approach assumes more simply that earlier detection means a higher probability of surviving to die from other causes; those who die of cancer die at the same time they would have died without screening. Both approaches allow for variants; Wever and others[4] describe 2 forms of the stage-shift approach.

In Wever and others analysis,[4] the cure submodel but not the stage-shift submodel was fitted to the trial data used to test both submodels, so it is premature to conclude that the cure approach is superior. But the article raises important questions about how to represent and explore ideas about early disease, important because they affect estimates of screening's benefit under different screening

schedules. In the breast cancer screening analysis, 4 of the 6 models used the stage-shift approach (Mandelblatt and others, Appendix Table 1).[3] Estimates of the benefit retained by biennial screening, compared with annual screening, ranged from 68% to 90%, and the 2 models at the high end of that range were both stage-shift models (Mandelblatt and others, Table 2).[3] But the other 2 stage-shift models produced estimates of 74% and 76%, and estimates of life years gained from screening were not correlated with type of screening submodel (Mandelblatt and others, Appendix Table 4).[3] Thus, again, it is not clear that one approach produces consistently different results from the other.

In principle, it is possible to observe the earlier stages of disease, and it may someday become possible in practice. The unknown will become known. In the meantime, the consequences of screening, whether evaluated by clinical trials or by models, remain subject to uncertainty due in large part to ignorance of the true course of disease. In this issue, van Ballegooijen and others[5] ask, ''Are model differences a limitation, or even a failure, of modeling? We think the differences between our CRC [colorectal cancer] screening models reflect genuine uncertainty because all 3 models provide good fit to observed data. The demonstrated differences indicate areas where additional data are needed to inform models and where, in the absence of data, strong assumptions must be made.''

The practical importance of this lack of information is shown by the unexpected results of more than 2 decades of screening for breast and prostate cancer.[6] The common view when screening began was that cancer was usually fatal and that catching it earlier, before symptoms appeared and while it was still small, would save many lives. After 20 years of widespread screening, the evidence strongly suggests that many cancers are not fatal. Before screening, those cancers often went undiscovered, and the host individuals died of other causes. Today, many of them are discovered—a man's risk of a prostate cancer diagnosis or a woman's risk of a breast cancer diagnosis has almost doubled since 1980, from 1 in 11 to 1 in 6—but the incidence of late-stage cancers and the death rates from prostate and breast cancer have not declined commensurately. Screening programs were based on one idea about early disease. The evidence they have produced supports other ideas, ideas that can be explored through simulation models.

Models also let us explore aspects of disease and its interactions with interventions, which can never be observed, the unknowable. In this issue, Erenay and others[7] model the history of metachronous colorectal cancer. The progress of metachronous cancers cannot be observed because they are removed, thus censored, as soon as they are discovered. Yet, best practice for surveillance depends on understanding their course. Erenay and others[7] started with a conceptual model of disease, defining parameters that describe current thinking about its natural history and setting ''biologically plausible'' ranges for their unknown values or, where data were lacking, ranges that ''would not contradict clinical intuition.'' Within those ranges, they identified the best-fitting values for the parameters, using as their guide the correspondence between the model's predictions and Mayo Clinic data on the 5-year incidence of first metachronous polyps and metachronous cancers and the 5-year mortality rate from such cancers.

Also in this issue, Whyte and others[8] report another calibration exercise, again of colorectal cancer (but not metachronous cancers) and again building on clinical characterizations of ''the adenoma–carcinoma sequence, which is how most CRC is thought to arise.'' They use a stage-shift approach to model the benefits of screening and report not only the best-fitting values for the parameters but also the 95% credible intervals for use in probabilistic sensitivity analysis; the values reported are those that produce the best fit to the incidence of polyps and cancer before and after the start, in 2007, of a national program of screening for colorectal cancer in England. The 95% credible intervals underscore the scientific uncertainty about the process being modeled. Although the model's projections of incidence and mortality fit the observed data well (Figures 6–8), the lower and upper bounds of the intervals for some of the parameters differ by a factor of 10 (Whyte and others, Table 1).

Defining parameters that depict different ideas about how the world works and calibrating values for them is one way models can help us explore the unknowable by testing the implications of our ideas against what is known. In a very different context, Dyson[9] used a simulation model to test ideas about how biological life first began, a process that humans can never hope to observe since it preceded us by eons. The plausibility of his preferred ideas was supported because the model worked best for parameter values that were consistent with what is known about current biological life and with the widely accepted hypothesis that early life must have been simpler.

Calibrated parameter values are tested by the fit between model predictions based on those values

and observed outcomes, such as disease incidence and mortality. But models that fit the same data equally well can differ considerably in their predicted outcomes and implications for medical care. This fact serves as a reminder that the data themselves have a wide range of uncertainty: The 95% confidence intervals around observed outcomes are often wide even at a population level. In addition, when much is unknown, the number of possible parameters and the best-fitting values for them are not uniquely determined. As van Ballegooijen and others[5] observe about the differences among their 3 colorectal cancer screening models, ''A model with long dwell time combined with low [test] sensitivity can project similar effectiveness to a model with short dwell time and high sensitivity.''

In this issue, Jackson and others[10] propose to include different ideas about how the world works—alternative ''model structures''—as parameters within a single model, in principle nesting all alternatives within a larger global model. The crucial step is to ''define [a] global model with parameters defining structures.''[10(Figure 1)] Each structural parameter can then be assigned a distribution of possible values, informed by data if data are available, by expert opinion if data are not available. The implications of those values for the results can be explored through probabilistic sensitivity analysis. The stage-shift and cure approaches could, for example, be nested in the same overall model. The stage-shift submodel could be used for some projections and contrasted with other projections based on the cure rate approach. In an individual-level model, both submodels could be used together, with individuals assigned to a submodel according to some probability, if it were thought that reality might be a mixture of the two.

Nesting all possibilities in a single model requires that analysts and clinical experts have enough knowledge and imagination to think of all the important alternatives. Jackson and others[10] note that ''appropriate characterization of structural uncertainties is essential for determining both the cost-effectiveness of interventions and the value of conducting further research. . . . Subject matter experts should be involved early in the modeling process, so that realistic structural alternatives are considered.'' The caveat here is that what is modeled must first be imagined. This will depend on the state of the science. What has not yet been imagined cannot be modeled, no matter how many subject matter experts are consulted. For immediate decisions, ''strong [and probably wrong] assumptions are required.''[5]

Over the longer term, models can help us explore the unknown, refine our ideas about it, and identify the evidence and trials necessary to move toward a better understanding of disease. How can we best use this capacity of models? Can modelers develop ways to explore the unknown and unknowable more completely and efficiently? There are algorithms for exploring parameter space. Can algorithms be developed to help explore model space—which is the same as idea space—so that we can conceptualize more clearly the things we cannot observe directly, and may never be able to observe directly, about the course of disease and treatment?

REFERENCES

1. Hlatky MA. What we can learn from a decision model: comment on ''Cost-effectiveness of adding magnetic resonance imaging to rheumatoid arthritis management.'' Arch Intern Med. 2011; 171:667–8.

2. Pignone M, Ransohoff DF. Cross-model comparisons to improve the value of modeling: the case of colorectal cancer screening. Med Decis Making. 2011;31(4):524–526.

3. Mandelblatt JS, Cronin KA, Bailey S, et al, for the Breast Cancer Working Group of the Cancer Intervention and Surveillance Modeling Network (CISNET). Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Ann Intern Med. 2009;151:738–47.

4. Wever EM, Draisma G, Heijnsdijk EAM, de Koning HJ. How does early detection by screening affect disease progression? Modeling estimated benefits in prostate cancer screening. Med Decis Making. 2011;31(4):550–558.

5. 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

6. Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. JAMA. 2009;302:1685–92.

7. Erenay FS, Alagoz O, Banerjee R, Cima RR. Estimating the unknown parameters of the natural history of metachronous colorectal cancer using discrete-event simulation. Med Decis Making. 2011;31(4):611–624.

8. Whyte S, Walsh C, Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. Med Decis Making. 2011;31(4):625–641.

9. Dyson F. Origins of Life. Cambridge, UK: Cambridge University Press; 1985.

10. Jackson CH, Bojke L, Thompson SG, Claxton K, Sharples LD. A framework for addressing structural uncertainty in decision models. Med Decis Making. 2011;31(4):662–674.