Many applications of mathematical modeling in environmental health represent fusions ... of the structured modeling approach of the physical scientist and the statistical approach of the biologist.

Mathematical Modeling in Environmental Health

In both recent and forthcoming issues of *EHP*, an increasing number of papers and news articles feature applications of statistical and mathematical modeling in environmental health. My comments focus on applications of mathematical models, which have become increasingly popular as tools for organizing and conducting the analysis of complex problems, partially because of easy access to sophisticated software and computational power. Risk assessment applications provide a good example where the model serves as a platform for the integration of the fate and transport, exposure, and response data central to risk estimation.

Generally, the information needed to undertake model-based analysis is of three sorts (*1*). The first is the set of causal hypotheses that describe current understanding of how different processes and variables are interrelated. This information provides the structure of the model. Second, simulation models can incorporate independent information on the range of values of the model's parameters, because many parameters of models based on physical, chemical, or biological processes have a clear experimental interpretation. Values of these parameters are often reported in the literature of the various scientific specialties that underpin the integrated analyses common to environmental health. Finally, such structured models can incorporate data on observed patterns of behavior characteristic of the particular system under analysis.

Although all this seems straightforward enough, an inherent challenge relates to the relative weight one places on the different types of information contained in these three categories. Mathematics has long been the language of engineering and the physical sciences, where basic physical laws form the foundations of analysis as well as of models studied by computational techniques. Hence, those of us who come from this tradition tend to place high value on the causal linkages implicit in model structure and parameterization. Problems in biology, on the other hand, have been a major motivation for the development of the descriptive and empiric approaches of statistical analysis. In this tradition little emphasis is placed on *a priori* model structure, and the goal is to summarize the observed data in an efficient and useful manner.

Both mathematical and statistical modeling have become common tools in risk assessments. On the mathematical side, the tendency to emphasize structural information in model development leads to large and complex models. Beck et al. (*2*) have commented that "there is a natural tendency to rely on the complexity of the model as a form of insurance against the unknown. For, if everything of conceivable relevance has been included in a model, how can its predictions possibly be wrong?" Clearly, model-based predictions can be quite wrong, which has led to well-founded concerns over model validity (*3*). But can complexity be the culprit if each element of the model is based on well-conducted independent studies? The crux of the issue is that, in general, the more complex the model, the wider the variation of the output variables that it can produce with plausible parameter values. So predictions that make sense can be selected only with reference to past behavior observed in the real system, a process sometimes called calibration. The next difficulty is that model outputs that match past behavior, either qualitatively or quantitatively, can be produced by many combinations of plausible parameter values (*1*). This complicates the prediction task as the dimension of the parameter space increases. There are good reasons to be concerned about the complexity of models.
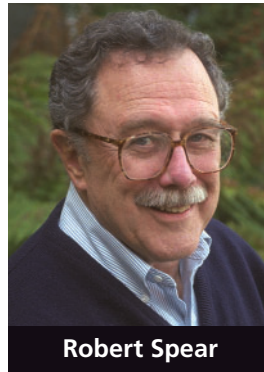
Robert Spear

The biological sciences have come much more recently to mechanistic models. Indeed, Levin et al. (*4*) open their broad-ranging discussion of modeling in biology with the observation that "Mathematical and computational approaches to biological questions, a marginal activity a short time ago, are now recognized as providing some of the most powerful tools in learning about nature." Are any lessons from the use of models in biology relevant to environmental health applications? A recent example of "learning about nature" is the report of Neutel et al. (*5*) on the stability of food webs in ecology. In applications of this sort, the model serves as an integrated and explicit set of hypotheses of how the system works. A plausible model is one that is *phenomenologically* consistent with observed data. Often qualitative system properties, like stability in the food web example, are the focus of attention. The hypotheses expressed by the model can be refuted wholly or partly when its predictions are shown to be inconsistent with the observed behavior of the natural system. Generally in applications of this sort, the model is explanatory rather than predictive. Plausible structure is inferred from observed data in the statistical tradition of biology. Quantitative predictions of the future are generally avoided.

Clearly, many applications of mathematical modeling in environmental health represent fusions, at least implicitly, of the structured modeling approach of the physical scientist and the statistical approach of the biologist. Recognizing this mixed mode offers some strategic guidance in applications of these methods to environmental health. It suggests that the complexity of the mathematical structure must be carefully balanced with the nature and extent of application-specific data, which exist to meaningfully evaluate and build confidence in its behavior. Modeling is of great value in organizing diverse knowledge and data of problem-specific importance, but unless used with skill and insight it is no panacea for reducing the variance of quantitative predictions of the future.

Robert Spear
School of Public Health
University of California, Berkeley
Berkeley, California
E-mail: spear@uclink4.berkeley.edu

Robert Spear, an engineer, is a long-time faculty member in the Environmental Health Sciences Division. Much of his work concerns applications of mathematical and statistical techniques in exposure assessment and control. He is Founding Director of the University's Center for Occupational and Environmental Health.

REFERENCES AND NOTES

1. Spear RC. Large simulation models: calibration, uniqueness and goodness of fit. Environ Modeling and Software 12:219–228 (1997).
2. Beck MB, Ravetz JR, Mulkey LA, Barnwell TO. On the problem of model validation for predictive exposure assessments. Stochastic Hydrology and Hydraulics 11:229–254 (1997).
3. Oreskes N. Evaluation (not validation) of quantitative models. Environ Health Perspect 106(suppl 6):1453–1460 (1998).
4. Levin SA, Grenfell B, Hastings A, Perelson AS. Mathematical and computational challenges in population biology and ecosystems science. Science 275:334–342 (1997).
5. Neutel AM, Heesterbeek JAP, de Ruiter PC. Stability in real food webs: weak links in long loops. Science 296:1120–1124 (2002).