11 The Art and Practice of Risk Assessment Modeling

11.1 INTRODUCTION

Quantitative risk assessment relies on models. A model is an abstraction of reality used to gain clarity about a problem or its solutions by reducing the variety and complexity of a situation to a level we can understand. This simplification enables us to gain insight into systems, processes, events, scenarios, problems, and their possible solutions. These insights can help us understand or control some aspect of a problem or opportunity. Models help us to address complex phenomena, force us to synthesize knowledge, and enable us to solve problems, test hypotheses, examine strategies, set priorities, and communicate all of this to others in a cost-effective and time-efficient manner.

It is not always easy to understand what people mean when they use the word "model." Sometimes there are models within models and the language can be confusing. A risk assessment model, for example, may actually comprise a consequence model, a likelihood model, and a risk characterization model along with numerous computational components. So, a model could be the entirety of all these parts or any one of the component parts. "Probability model," right or wrong, is a phrase often used to describe a probability distribution used in a Monte Carlo simulation. So, in this chapter, *model* refers to the overarching representation of a process, that is, the whole of the risk assessment model rather than its component parts. The parts that comprise a model are called modules or components. The content of this chapter applies to these as well.

Risk assessment models have many practical uses and yield a number of important benefits in addition to those mentioned previously. A valid model accurately represents the relevant characteristics of the decision problem. Valid models are a relatively inexpensive way to learn about the potential effects of uncertainty. Uncertainty often ensures that important data will not become available until some point in the future. Risk assessment is designed to address such uncertainty, and its models help bridge the gap between what we know and what we do not know in a way that aids timely decision making. These models enable us to explore more or less desirable futures from the present. They also help us explore the effectiveness of things that are currently impossible to do in reality. We use them to understand failure scenarios that have not occurred and improvement scenarios that have not been implemented. Assessing the risk of risk management option (RMO) failure modes (levee failures, nuclear accidents, financial collapses, oil spills, and other catastrophic incidents) affords us the opportunity to anticipate future risks. Models empower us to examine the likely efficacy of these measures before they are actually implemented.

A wide variety of model types are available for use by risk assessors, and this chapter begins by discussing some of them. The chapter then narrows its focus to mathematical simulation models built in a spreadsheet environment. This is one of the easiest model building environments in which one can work. A 13-step model-building process is offered to help guide the efforts of novice model builders. The remainder of the chapter is devoted to a discussion of model-building skills. These skills are broken into technical skills and craft skills. Technical skills might be likened to the science of modeling while craft skills are more like the art and practice of modeling.

11.2 TYPES OF MODELS

There are many ways to discuss and categorize models. The most common models are mental, visual, physical, mathematical, and spreadsheet models (Powell and Baker 2007). Mental models are conceptual notions and ideas about how things work in reality. They are internal representations of external realities, that is, mental images translated to verbal or written descriptions. Mental models are abstract models. The risk management and risk assessment models of Chapters 3 and 4 are good examples of mental models. Many of these models tend to be informal or conceptual frameworks, although well-worked-out formal theories can be mental models as well.

Visual models include maps, figures, graphics, and charts that show how things work in reality or how ideas are related to one another. When coaches draw plays on the sidelines, they are using visual models. When someone gets up at a meeting and goes to the whiteboard and begins to sketch the ideas being discussed, they are using visual models.

Physical models are usually analog or iconic models. Analog models look like the reality they represent. Cockpit simulators are used to train airline pilots, bridge simulators to train ship captains and pilots. Working in replicas of the planes and ships they will command, pilots can practice a wide variety of situations and circumstances. Analog simulators of all types are becoming more common. Model engines used in shop classes and models of human body parts used in doctors' offices are examples of other kinds of analog models.

Iconic models are scaled-down replicas of the object, system, or process under study. They have been used to design cars, buildings, and new pieces of equipment. Master planners rely on iconic models to communicate their visions. Beach cross sections can be designed and tested in a wave tank before beach nourishment projects are built. The plastic models we built as children are additional examples. In the not-too-distant past, iconic models of watersheds were quite common.

There is also a class of models intended to represent a critical or interesting physical dimension of reality or to give a physical dimension to an abstract concept without replicating it or scaling it down. Examples include models of DNA molecules and models of three-dimensional mathematical functions like the utility functions of economics.

There are many kinds of mathematical models, and they are used in every area of scientific endeavor. Mathematical models generally rely on functional relationships among dependent and independent variables. Every quantitative risk assessment model is a mathematical model. Their basic characteristic is that they use mathematics to describe a system, situation, object, or problem. Mathematical models can be categorized as prescriptive, predictive, or descriptive models (Ragsdale 2001).

A prescriptive model produces the best value for a dependent variable. Examples of prescriptive models include things like linear, integer, and nonlinear programming models, networks, and so on, that can be optimized. They prescribe the most unambiguous course of action for the risk manager when they are available.

Predictive models predict the value of a dependent variable based on the specific values of the independent variables. When the functional form of the relationship between the dependent and independent variables is known, we often have prescriptive models. When that functional form is unknown and has to be estimated, we have predictive models. Many risk assessment models are predictive models. Such models are used to route flood waters and to anticipate hurricane tracks, to estimate pathogen growth, as well as to predict the performance of markets and other complex systems. Examples of predictive models include regression and time-series analysis.

The third category of models, descriptive models, may be the most common in risk assessment. These models are characterized by their uncertainty. That uncertainty may include uncertainty about the functional form of a relationship or, when very precise functions are known, they may include uncertainty about the exact values of one or more independent variables. These kinds of models describe the range of outcomes or behaviors that are possible in a system that is plagued by uncertainty. Simulations, queuing, and inventory models are examples of descriptive models.

Mathematical models can also be described in accordance with the family tree of system models seen in Figure 11.1. A deterministic model has no stochastic, that is, random, components. Thus, the next state of the system is always determined by predictable actions built into the model. In a static model, time is not a significant variable, that is, the data do not "age." The passage of time is a significant consideration

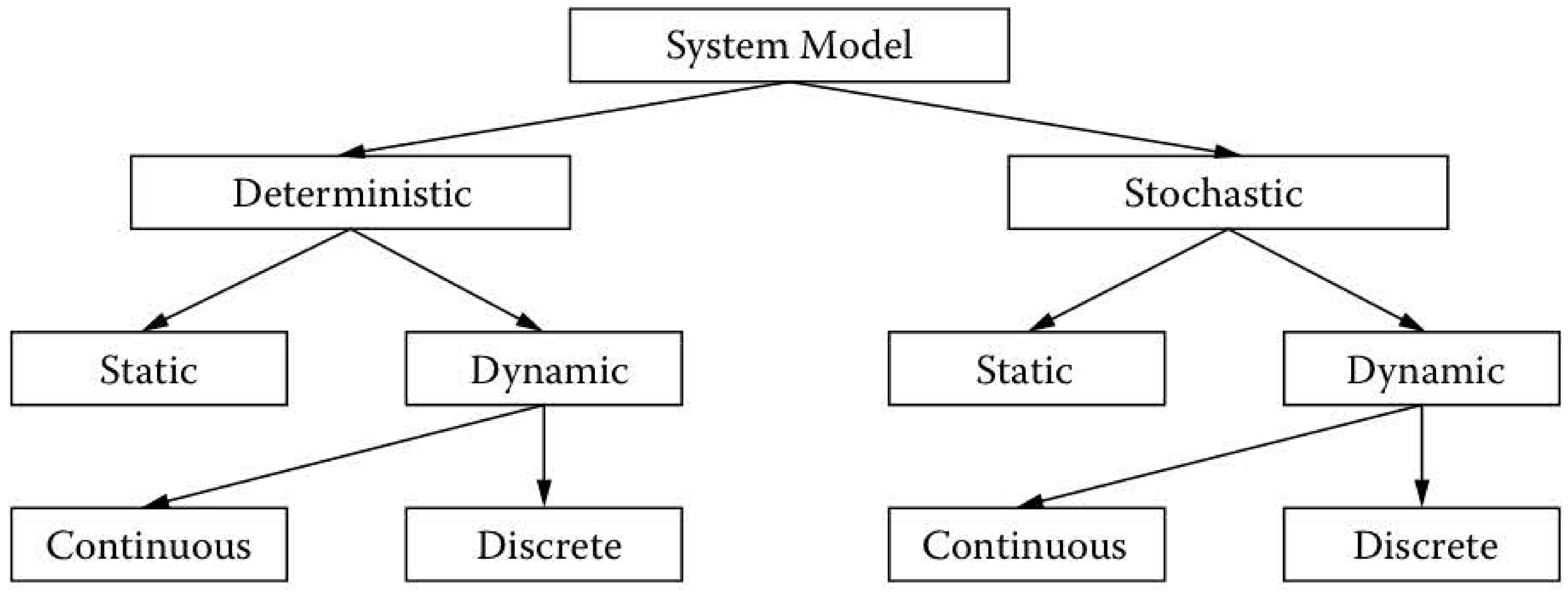


FIGURE 11.1 A taxonomy of model types.

in a dynamic model. Dynamic models often offer an explicit representation of the sequence in which events occur. Continuous dynamic models involve states that evolve continuously. They often rely on differential equations. In contrast, discrete dynamic models produce states that are a piecewise constant function of time. A discrete model may evolve a new state every 15 minutes, every day, or annually, whereas a continuous model is continuously evolving new states. The same kinds of models are found in the stochastic universe, where risk assessors deal with random components in the systems they model.

Simulation models, the focus of this chapter, are a subset of mathematical models that can exhibit any of the characteristics of the family tree of system models in Figure 11.1. Monte Carlo simulation models built in the spreadsheet environment enable us to explore the stochastic branch of models and will occupy most of our attention, but there are many more complex system simulation models in use that are worth a brief discussion. Evans and Olson (2002) identify four such simulation modeling approaches.

Many of these system models involve the flow of some entity or object through a system. The entity might be an attribute (purity, resistance, strength), a condition (new, fresh, complete), a piece of information (message, order, profile), or a physical object (an egg, a towboat, a human being, a pathogen). System simulation models reproduce the most important activities and dependencies that control the flow of one or more entities through a system over time. These sorts of models tend to be too complex for a spreadsheet.

Activity-scanning simulations describe the activities that occur during fixed time intervals. The model is incremented by some fixed time period and it describes the activities of interest that occurred during that time period. For example, the model will describe what happens in the first time period and then advance to the next time period and continue the process. If we model the movement of towboats on a waterway, this would include all the activities that occur within a given time increment such as distances traveled and positions, traffic maneuvers, and dockside operations. If our model follows eggs from laying to consumption, it might represent what activities occur in the first 24 hours, followed by the activities in the second 24 hours, and so on. If we are monitoring bacteria in an egg, it might describe all the physical activities that occur within the egg at fixed time intervals that offset bacterial growth in an egg.

Process-driven simulation models focus on the process an entity must move through in a system. So, for example, if we're concerned about how an egg gets from the farm to a store, the process could be described and modeled as follows:

> Egg forms→laid→transported to collection→collected→transported to processing→washed→graded→separated→transported to packaging→placed in carton→palletized→shipped to retail→received→stored

If we think in terms of discrete modules, this sort of model would use modules corresponding to the parts of the process rather than modules defined as increments of time. So, we could examine what happens in the washing process or the shipment to retail outlets.

Another type of simulation model is called event-driven. Events are occurrences in a system at which point changes in the system (or entity under consideration) occur. These events take place at a moment in time. Events are processed in chronological order, and time is advanced from one event to the next. These models focus on events that change the system, and so they are often computationally efficient. This kind of model describes changes in the system that occur at the instant an event occurs. For our egg example, contamination of the egg in utero, breakdown of the yolk membrane, and cooling on the pallet are events of potential interest that can change the egg entity from risk free to risky when assessing the risk of Salmonella in shell eggs. If we are modeling towboats, the risky events of interest might include traffic maneuvers when meeting, passing, or overtaking other vessels in the waterway or perhaps when navigating difficult turns.

When variables of interest change continuously over time, continuous simulation is appropriate. The amount of money moving through an economy, water in a stream, oil in a pipeline, or milk in a production run are all examples of continuous variables. These continuous variables are frequently called state variables. Continuous simulations rely on equations defining relationships among state variables that enable dynamic system behavior to be studied. System dynamics is a popular form of continuous simulation.

Simulation is a legitimate technique for solving problems when analytical solutions are not possible. Whereas analytical models represent reality, simulation models imitate it in a simplified manner. Simulation models allow us to conduct controlled experimentation that is otherwise impossible. Simulations can produce information that reveals new facts about the problems we work on. They are suitable for a broad range of applications and can be effective training tools.

Building models generates insight that enables us to make better decisions. Modeling improves our thinking about and understanding of a problem, and it enables us to experiment and learn about it as well. As this discussion suggests, a wide range of simulation modeling options is available. Spreadsheet models are an especially useful platform for building risk assessment models because they require no special programming language skills or abilities. These spreadsheet models are the focus of the remainder of this chapter.

11.3 A MODEL-BUILDING PROCESS

There is no one way to build a mathematical spreadsheet model for risk assessment. Modeling may be the most idiosyncratic part of the risk assessment process. Nonetheless, it can be useful to have a process in mind. The model-building process presented in this section has served me reasonably well over the years; feel free to adopt or adapt it. It assumes your data have already been collected or will be collected once step 4 is completed. Thus, data collection is not an explicit step in this process. The 13 steps in this process include:

1. Get the question right
2. Know the uses of your model
3. Build a conceptual model

4. Specify the model
5. Build a computational model
6. Verify the model
7. Validate the model
8. Design simulation experiments
9. Make production runs
10. Analyze simulation results
11. Organize and present results
12. Answer the question
13. Document the model and results

11.3.1 Get the Question Right

Know what information the model needs to produce. If the risk management questions are not clear or if they are not the right questions, then nothing that follows in the model-building process will make any difference at all.

Different questions can lead to very different models. Consider the following questions that relate to a food-safety concern with *Vibrio parahaemolyticus* (*Vp*) and oysters. How many people get sick from *Vp* in oysters? What is the probability of getting sick from eating a raw oyster? How effective would refrigeration of oysters within 60 minutes of harvest be in reducing illness from *Vp* in oysters? What is the risk of eating raw oysters to people with liver disease?

Each of these questions will require different data and a different model structure. Get the right question, then get the question right. The first step in developing any model is understanding what question(s) the model needs to be able to answer. Be sure to discuss and clarify the meaning and intent of each and every question with the risk managers before you begin to build your model.

11.3.2 Know the Uses of Your Model

Understand what the model is expected to do and how it will be used. Is the model identifying research needs, developing a baseline risk estimate, attributing risk to different hazards? Will it be used to evaluate risk management options? What kinds of options might they be? Is it a learning tool? Will it be shared with others or used again? Who will use it? Will it be added to over time, or is it to be completed once and for all? Is it the basis for a regulation?

Understand what the model cannot do. Risk managers, in particular, must be made aware of the limitations of the model and its outputs. The seemingly innocent request to evaluate an unanticipated RMO after a model is built may be impossible because the necessary data were not collected or the model structure does not support that analysis.

Anticipate as many potential uses of the model before you begin to build it as possible. This can only be done through close collaboration and clear communication between risk managers and risk assessors.

11.3.3 Build a Conceptual Model

This is where your model-building effort is most likely to succeed or fail. Models fail less often due to data and parameter value issues than they do for faulty conceptualization of the problem to be represented. This step proceeds from the abstraction of ideas and notions to the cold, hard reality of details. The best modelers will include the important processes and exclude unimportant ones. This step calls on both science and art.

You now have specific questions to answer. Work out your risk hypothesis for each question before you begin to build a model. Develop your risk narrative. That narrative should answer the four questions:

1. What can go wrong?
2. How can it happen?
3. What are the consequences?
4. How likely is it?

Use the qualitative generic process if it fits your problem better than the four questions do. Be practical, not bound to any one approach. This is the step where you work out the major ideas for your model. What are the stated variables? How are they related to one another? The conceptual model is a mental model. A sample for a pest outbreak associated with a fruit import is shown in Figure 11.2.

11.3.4 Specify the Model

Once the conceptual model has been developed it is time to convert it into a specification model. This means moving from concepts to develop the relationships, equations, and algorithms that will describe the ways the various components of your conceptual model will work together. This is the model construction step that usually stops short of actually building the final computer program. It may help to think of it as the paper and pencil exercise of figuring out the calculations that will be needed in your model. It is more than this, but that provides a good mental image of the essence of this task.

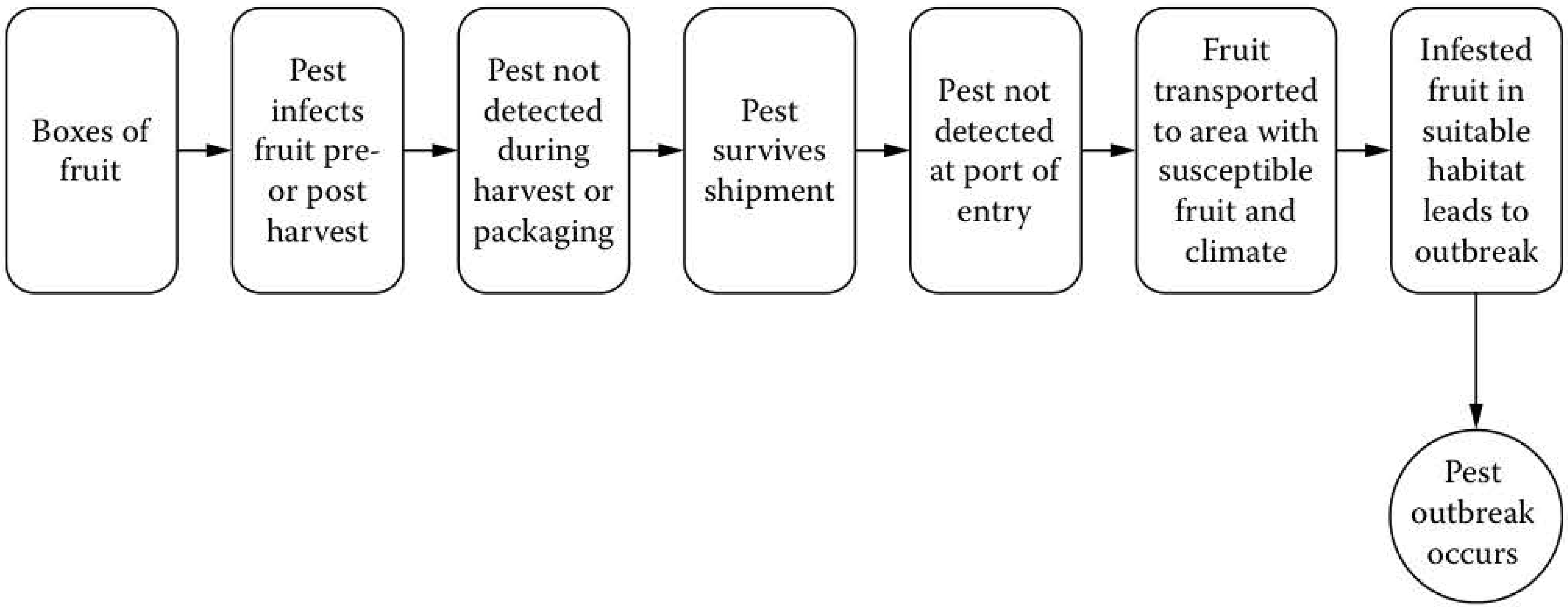


FIGURE 11.2 Conceptual model for fruit import pest infestation risk.

Parameter specification from minimal to maximal precision:

Guesstimates ("hand waving")
Expert opinion
Bounded estimate
Survey-based data
Data-based parameters (averages, etc.)
Curve-fit parameters (trade-offs)
Parameter sensitivity testing
Entire parameter sets by optimization

Source: Risk Analysis 101 USDA, APHIS, PPQ.

The functional form of all relationships as well as the model logic are built in this step. Placeholders and dummy values can be used in lieu of data. The specification model defines the calculations, logic, and other inner workings that will make the model run. It often includes a detailed sketch of the spreadsheet model. In the current example of Figure 11.2, this will concern such things as the mathematical process that determines how many boxes of fruit are imported, how we identify which are infected with pests, how we calculate the percentage of the pests that avoid detection and survive transit, and so on.

11.3.5 Build a Computational Model

If we think of the specification model as the skeleton, then the computational model puts flesh on the bones. In this step you complete the computer program needed to run the model you have specified and analyzed and organized the data you will need to make the model operational. The output of this step is a fully functional spreadsheet model.

Placeholders are replaced with real data. Dummy values may be replaced by probability distributions. At this point you have a working risk assessment model. You will learn how to choose probability distributions to represent knowledge uncertainty and natural variability in Chapter 13. For now, we focus on the steps required to build a model rather than on the details.

11.3.6 Verify the Model

Verification is an extremely important separate and formal step in the model-building process. You want to get the equations, calculations, logic, cell references, and all the details just right. This is when you ensure that your computational model is consistent with the specification model and that the specification model is correct. You want to make certain that you have built your model correctly. So, you debug, review, and test your model to ensure that the conceptual model and its solution are implemented correctly and that the model works as it was intended to work.

Many model builders verify their own work. The obvious advantage of this approach is their intimate knowledge of the model's structure. The disadvantage is that they have often been working with the model for so long they can no longer see potential problems. Having a model verified by someone other than its builder(s) can be time consuming and more costly, but it is often more effective in finding flaws. It is certainly more effective in increasing the credibility of the model and confidence in its outputs.

11.3.7 Validate the Model

Once your model is verified to ensure that it does what you want it to do in the way you want it done, it is time to validate your model. Is it a good representation of reality? Have you built a model that represents reality closely enough to provide valid information to support decision making? The weakest form of model validation is a model that accurately describes the system being modeled. Next, we ask if the model produces plausible results by considering whether the model can reproduce independent results like statistics or real data.

There are basically three ways to validate your model. You can validate model outputs, model inputs, or your modeling process. Historical data validity uses historical inputs to see if your model reproduces historical outputs. If weather conditions that always produced rain in reality produce sunny days in your model, you have a problem. There are at least two practical issues with this form of validation in risk assessment.

The first, and most pressing for risk assessment, is the lack of historical data. You must hold back some of your data and not use it in the model-building process to use it to validate your model. A common technique is to divide the available data. Part of the data, say half, is used to build the model and develop all the requisite relationships in the model. The other half is used to test the model. If actual data are input into the model, we expect to get predictions that closely match the actual output data from a valid model. Risk analysis is a decision-making paradigm designed for conditions of uncertainty. Consequently, data gaps and insufficient information frequently make historical data validation impossible. Frequently there are not sufficient data for confidently building a model, much less enough to hold some back.

A second problem is how closely a model must replicate historical conditions. If you are fortunate enough to have the data, how many rainy days must your model replicate for us to declare it valid? The answer to that question is most often a pragmatic one rather than a statistically rigorous one, a matter of art more than of science.

A second validation approach is data validity. When the outputs cannot be validated, it may help to validate the inputs. Data validity means you have clean, correct, and useful data. Furthermore, it ensures that all input data and probability distributions are truly representative of the system modeled. When simulating situations that have never existed, it is often impossible to assure data validity.

When neither the inputs nor the outputs can be validated we have few options but to try to validate the reasonableness of the process. Face validity, the third option, asks experts to examine the model or its results to determine if they are reasonable. A model has face validity when it looks like it will do what it is supposed to do in a way that accurately represents reality. One face validity test is to present knowledgeable experts with real-world data and model output data. If they cannot distinguish the two or find no significant divergence between the two, the model that produced the data may be considered valid on the face of its output data. However, a model that looks like it should work is quite different from a model that has been shown to work.

Validation can be checked at several levels. For instance, each module in a model should be validated. In a microbiological risk assessment, for example, it may be possible to validate a pathogen growth model or a dose-response curve even though the entire model cannot be validated. It may be advisable to validate components of some of the modules. A probability distribution chosen to represent a variable value within a module might be validated, for example. Of course, the whole-system level, that is, the entire model, should be validated as should any benchmark cases. Graphical comparisons, confidence intervals, statistical tests, and hypothesis tests can be useful validation tools to help determine whether model predictions are within an acceptable range of precision.

Verification and validation, together, provide the basis for the assessors' degree of confidence in the model. Risk managers should always inquire after the specific steps taken to verify and validate the model. Risk assessors should always inform risk managers about the verification and validation effort and its results. Some organizations use a model certification process to address questions of model verification and validity for models to be adopted for use throughout the organization.

11.3.8 Design Simulation Experiments

Characterizing the range of outcomes in an uncertain world is not a simple thing to do. There are often many scenarios to investigate. Assessors may be asked to explore historical conditions, existing conditions, a benchmarked baseline condition, future conditions, improved conditions, worst case, best case, different geographic locales, different seasons, and so on. Consequently, the assessor must carefully plan the simulation experiments that must be run to answer the risk manager's questions and to fulfill the intended uses of the model. This means deciding what parameters and inputs will vary, what assumptions will be used for which experiments, and so on.

If you sit down with your model and just start making runs, it is likely that, sooner or later, you will get what you need. It is also likely you'll waste a lot of time making runs you did not need. Carefully identifying the various conditions for the simulations you want to run is essential to an efficient risk assessment process. Arrange your series of experiments efficiently. It may make sense to do them in a specific order if significant adjustments must be made to your model for different conditions. Write down the runs you will do in the order you will do them. Take care to verify all alterations to a model and save significantly different versions of your model as separate files.

11.3.9 Make Production Runs

This could easily be an extension of the preceding step. Once you have identified the range of experiments you want to run, the next step is to run them. It is very easy to fall prey to the excitement of running your model as soon as it is built. Many assessors skip over the verification, validation, and experiment design steps. You have likely spent a good bit of time and effort to get to the point of a computational model, and you want to use it. Unfortunately, few if any models are error free. If you make a slew of runs with your model and then find a decimal point error in the formula, everything must be repeated. Resist the temptation to jump directly to production runs.

One of the most frustrating, common, and avoidable problems is failure to document a production run. Every time you run your model you should carefully record the nature and purpose of the run (e.g., existing risk estimate to establish a baseline measure of the risk) and make note of your model's initial conditions, input parameters, outputs, date, the analysts, and so on. Enter this information into a log you keep on a separate worksheet in your model. Keep it up to date.

It can also help to be systematic in making your runs. For instance, using our current example, if we are investigating pest establishment risks for several fruits, several countries of origin, and several pests, it makes a lot of common sense to work smart. If setting up the model for a specific commodity takes time, it may make sense to do all the runs for avocados before moving to the peaches. On the other hand, if it takes more effort to set up the pest or country parameters, it may make sense to sequence the runs in another fashion. This is why you design your experiments so that you can run them efficiently.

It is especially important to take great care when making runs that reflect the presence of a new risk management option. These runs need to be carefully documented. It may also be important to record any fixed seeds* that were used to initiate your simulation process in case there is a desire to reproduce the simulation at a future date.

Take special care to save all outputs from a production run and to carefully identify them. Unless you are absolutely sure about the outputs you will and will not need to complete your risk assessment, save all of your simulation outputs if possible. It is far better to save outputs you will never need than to need outputs you never saved.

11.3.10 Analyze Simulation Results

This is the all-important process of getting useful information from data. As the runs are completed, the assessors need to analyze the results and learn what the simulations have taught us about the problem. This analysis is statistical in nature, and it needs to account for the uncertainty in the inputs and to carefully convey that and the resulting uncertainty in the outputs to the risk manager. It is through the analysis of your simulation results that you will craft answers to the risk manager's questions, always taking care to characterize the remaining uncertainty in useful and informative ways.

11.3.11 Organize and Present Results

The information assessors glean from the simulations needs to be organized and presented to decision makers in a form that is useful. At this stage of the modeling

* Seeds are explained in Chapter 15 and Appendix A.

process, useful information is information that supports decision making and carefully addresses the significant uncertainties encountered. Presenting useful information for decision making is the topic of Chapter 18.

11.3.12 Answer the Question(s)

The entire reason for doing a risk assessment as well as for building an assessment model is to provide risk managers with the specific information they need to make a decision. Be sure you answer the questions. Do it in a question-and-answer format. Do not assume that your well-organized and well-presented report narrative and results will accomplish this purpose. Answer each question specifically. To the extent that lingering uncertainty affects those answers, this must be carefully portrayed as well. Once you have adequately answered the questions, feel free to summarize the insights you have gained, offer specific observations, and even to conjecture in a responsible way.

11.3.13 Document Model and Results

Many risk assessments are going to be documented by some kind of report. Printed risk assessments are as common as crows. Careful documentation of your results is assumed to be an essential part of your model-building process. Few would disagree with that statement. However, there is more that needs to be done.

You need to document your model. That means explaining the structure of the model, including relevant descriptions of the preceding steps, your conceptual and specification models, the source and quality of your data, the results of the verification and validation efforts, as well as the history of production runs. And that is not all of it! You also need to provide the equivalent of a user's manual in your documentation.

You have likely spent so much time on this risk assessment model that your inbox is piled high with tasks that all had to be done yesterday. There may be no time to document your model, but you intend to do so the first chance you get. Alternatively, you are sick of living with the model; you know it inside and out; and you can't stand to spend another minute dealing with it. Both of these scenarios pretty much ensure that your model will never get documented. That could be a problem.

As familiar as you are with the model today, in six months' time there is a good chance it will look like someone else's work to you. Inevitably there will be questions you are asked to answer, and it will take 10 times longer than it would have had you documented your model.

Then consider what happens if the model builder gets hit by a bus or changes jobs? With the builder goes all of the organization's model expertise unless someone has carefully documented the model and how it works. As much of a pain that it is to document your model, you cannot afford not to.

11.4 SIMULATION MODELS

Simulation is the process of building a model of a system or a decision problem and experimenting with the model to obtain insight in to the system's behavior or to assist in solving the decision problem. This can be done with physical models. Do you recall the coin-operated mechanical horse of your youth that sat outside the mall? That was a physical simulation model. Physical simulation models have grown increasingly sophisticated and are now used for all sorts of training and testing.

Simulation models are often used in training exercises. Food companies hold simulated recall exercises; first responders use simulated emergency exercises for training. Everyone has participated at one point in time in the fire drill, another simulation model exercise. These simulations are valuable training tools in situations where it is prohibitively expensive or too dangerous to allow trainees to use the real equipment in real situations. Tabletop exercises provide another form of simulation training that enables decision makers to learn valuable lessons in a safe virtual environment.

While risk analysis can make effective use of these kinds of simulation models, the most common simulation models and the focus of this chapter are computer simulation models. Computer simulations are used to model actual or hypothetical situations on a computer so that they can be studied to see how the system works. By changing variables and other model inputs, predictions may be made about how the system will behave under a wide variety of circumstances/scenarios.

Simulation models have been used for modeling the function of natural systems (flood, drought, hurricane tracking, ecosystems) and human systems (traffic flow, transportation systems, engineering, economics, the social sciences). They are used in all the natural sciences and have been used to model processes of all kinds. Clearly, analytical solutions to problems are preferred and should be sought when they exist. Computer simulations arose as an adjunct to, or substitute for, modeling systems for which closed-form analytic solutions do not exist. Simulation models are able to generate a number of representative scenarios describing the possible behavior of a system when it is either impossible or impractical to consider the entire range of possible scenarios.

Simulation models are most helpful when problems exhibit significant uncertainty, a situation commonly characterizing most risk assessment problems. Monte Carlo simulation tends to be used in static models, while systems simulation (described previously) are used more often in dynamic models. The Monte Carlo process, a sampling experiment designed to estimate the distributions of outcome variables that depend on one or more probabilistic input variables, is described in detail in Chapter 15.

Simulation is a legitimate technique for solving problems when analytical solutions are not possible. They have a broad range of applications and, in risk assessment, are often constructed in a patchwork style, that is, a hazard module may be built first, followed by an exposure/likelihood module and a hazard characterization/consequence module. These may all be pulled together subsequently in a risk characterization module. Readily available commercial software makes simulation modeling easy to do.

Simulations are not without their downside. The larger models can be costly or time consuming to build and run. Powerful computers may be needed for the largest simulations. Model results are very sensitive to model formulation, and there is no guarantee of an optimal solution being identified from the results.

Perhaps the greatest disadvantage of simulation modeling is that it has become so easy to do. It is tempting, when faced with a problem, to sit before the computer and start to build a simulation model. Too often this is easier than thinking the problem through and possibly finding an analytical solution to a problem. The ease of simulation modeling can encourage overlooking other techniques.

11.5 REQUIRED SKILL SETS

Building a good risk assessment model requires a special skill set. Powell and Baker (2007) have proposed breaking this skill set into technical skills and craft skills. Their distinction is a useful one and one I will build on in the remainder of this chapter. Technical skills are necessary for getting the right answer. Getting the right answer is important, in case you had not guessed that! Craft skills are extremely useful for simplifying complex problems and for modeling in new and poorly structured situations. Craft skills represent the creative side of modeling. Both skill sets are invaluable.

The discussion that follows is generally applicable, but there is a wide range in the formality with which risk analysis is conducted. In a regulatory setting, the risk analysis process and the risk assessment in particular may be quite formal. In a large organization where risk analysis has penetrated the culture, it may be practiced less formally in branches, sections, teams, and even by individuals. Some readers will find some of the ideas that follow useful; others may find them less so. For example, if you address the knowledge uncertainty and natural variability in a model used only by you because it makes sense to do so, you may not feel it necessary to design your model to communicate effectively with others. Feel free to take what is useful from the sections that follow and leave the rest. Modeling is an art, and ultimately each artist must develop their own style.

11.5.1 Technical Skills

Technical modeling skills comprise the knowledge and proficiencies required to build a specific risk assessment model. These are the skills that get you to a right answer. No one person may possess all the necessary technical skills for a complex risk assessment, but all those skills must be present on the assessment team.

First, technical skills require the modeler to know or to be able to understand the relevant science represented in the model. This does not mean the modeler must be a microbiologist to build a food-safety model or a civil engineer to build an engineering-reliability model. It does mean that discipline-specific knowledge must be present and available and that the modeler must be capable of translating that knowledge into a workable and reasonable model structure.

Second, there is the technology knowledge that is required. R, Unix, Linux, Java, C+++, Perl, MySQL, Microsoft C, and other language skills may be required for some models. In this book, we will focus on spreadsheet models, so it is sufficient to have knowledge of Microsoft Excel or a similar spreadsheet program. If additional software tools, for example, Palisade's @RISK, used in this text, are used, the modeler must be proficient in their use as well. (An introduction to the use of several modules of Palisade's DecisionTools® Suite is found in Appendix A.)

Third, there is a wide variety of narrow, well-defined tasks the modeler must be able to carry out. There are techniques and methodologies for all sorts of calculations that may not belong to any one discipline. These include basic math and language skills, for example, as well as more specialized skills. For economic and financial models, this could include handling present-value calculations. For engineering models, it might include working with a reliability beta index model, a hazard function, or a fragility curve. For food-safety models, this task might be constructing or programming a dose-response curve. These skills also tend to be discipline based.

A good risk assessment team will have all these technical skills available. It is not uncommon for a risk assessment model builder to come from a quantitative disciplinary background that may be quite different from that of the problem's context. The technical skills of risk assessment model building remain one of the scarcer resources in risk analysis. When the requisite skills are spread across many people, frequent team interaction and effective communication are essential.

11.5.2 Craft Skills

This discussion of craft skills has some general applicability, but it is intended primarily for those who work in the spreadsheet environment. Craft skills are separated into two topical areas for presentation here: the art of modeling and the practice of modeling, as this chapter title suggests. The discussion that follows attempts to flesh out the 13-step process presented previously with some art and practice tips.

11.5.2.1 The Art of Modeling

Powell and Baker (2007) offer eight modeling heuristics that are simply too good for anyone to ignore. They form the basis for this discussion of the art of modeling as supplemented by my own experiences. To see these heuristics in their original form, the work of Powell and Baker is not to be overlooked. The box introduces the example followed in the remainder of this section.

THE EXAMPLE

To illustrate the heuristics described here, we will use the U.S. FDA *Vibrio parahaemolyticus* Risk Assessment, July 19, 2005, Quantitative Risk Assessment on the Public Health Impact of Pathogenic *Vibrio parahaemolyticus* in Raw Oysters found at https://www.fda.gov/downloads/Food/FoodScienceResearch/UCM196915.pdf (accessed July 16, 2018).

The risk assessment model, Appendix 3, can be obtained from the following link https://www.fda.gov/downloads/Food/FoodScienceResearch/UCM196915.pdf (accessed July 16, 2018).

Vibrio in raw oysters represent a serious health risk. This risk assessment considered different geographic sources of oysters and different seasons of the year. To simplify the exposition, we'll use only the Gulf Coast of Louisiana in the summer as the single geographic region and season. In addition, we will consider only one of several questions posed to risk assessors, "What reductions in risk can be anticipated with different potential intervention strategies?"

11.5.2.1.1 Identify the Question(s) and Simplify the Problem

We will assume that the modeler begins with a well-defined decision context and a list of questions the risk managers would like to have answered. Let us also assume that the risk assessment team has considered these questions, answered those that could be answered with the available information and knowledge, and now seeks to build a model to predict answers to the remaining questions. This is where we would like the art of modeling to begin. The truth is that few risk assessment modelers ever find themselves in such a situation. So, let us back up and begin a bit more realistically with the risk manager who has never read Chapter 3.

You, the modeler, simply must know what question(s) you are trying to answer with your risk assessment model. There is no way around this if you are to build a useful model. Ideally, you will be the beneficiary of a good risk management approach, such as the one described in Chapter 3 of this book. If you do not know what the risk manager's questions are, sit down with your risk managers and ask them, "What do you need to know to solve the problem you are facing?" Write down their responses. Ask if there is more they need to know. Get them to understand what they have asked for and what you will provide and not provide.

Albert Einstein has been quoted as saying, "Any intelligent fool can make things bigger, more complex, and more violent. It takes a touch of genius—and a lot of courage—to move in the opposite direction." This brings us to the starting point for modeling: it is time to begin to simplify.

Build the simplest model you can. Power it with test-case data. Can it answer the question(s) risk managers have asked you? If yes, stop model building and go get real data. If not, revise the model and repeat the process. Make your model as simple as you can, so long as it meets the needs of your risk managers. It is always easier to add to a simple model than it is to simplify a complex one.

Sanchez (2007) formalizes this advice in four very concise suggestions for building a model:

1. Start small. Begin with the simplest possible model you can. It should cut away all the complexity that is not essential while still capturing the essence of the system. This can sometimes best be done by focusing on the questions you are trying to answer and the outcomes you'll need to provide those answers.

   Simplifying anything requires us to make assumptions. Document your assumptions as you make them. Remember that you are most dangerous as a model builder when you begin making assumptions without recognizing that you are doing so. Peer review is one of the best ways to guard against this danger.
2. Improve your model incrementally. It is easy to add features to a basic working model. You can often improve a model by relaxing the assumptions with a more complete representation of reality. Even this should be done simply. Prioritize the changes you could make. Make one small change and see if it is adequate.
3. Test your model frequently. You want your model to conform to reality, not to duplicate it. How do you know when the model is adequate? If it answers

the questions you have been tasked with answering, it is quite likely to be adequate.
4. Backtrack if you must. In every model there comes a point when additional improvements are not worth it. When the latest change produces no measurable benefit to the answers to your questions, do not be afraid to go back to the earlier, simpler model when all you have added is complexity.

The question for this example (see preceding box) is, "What reductions in risk can be anticipated with different potential intervention strategies?" This question requires risk assessors to first estimate the annual number of illnesses before it can address the risk reductions. The simplest model that does this is:

$$\text{Eat oysters} \rightarrow \text{get sick}$$

Clearly, this is not going to be sufficient, but it is a start. Perhaps there are things that happen before the oysters are eaten that are important or maybe there are alternative outcomes. Most importantly, we have begun to build our model.

11.5.2.1.2 Break Problem into Modules/Sections

One of the most effective ways to simplify a problem is to decompose it, that is, to break it down into simpler components or modules. This decomposition is the basis for the conceptual risk assessment model discussed in the 13-step procedure. We can always break any risk assessment into a hazard/opportunity identification, a likelihood assessment, a consequence assessment, and a unifying risk characterization. Likewise, each of these components can be further decomposed.

When working on a specific decision problem, it is usually simpler to decompose the problem into components. The reason for doing this is that it is easier to think about and work with the components of the problem than it is to work with the whole problem. The components, when well chosen, provide a natural structure for the assessment model.

The question then becomes: how many components do you need? Where do we draw the line? Powell and Baker (2007) suggest that a problem should be divided into components that are as independent of one another as possible. This number remains a subjective judgment. Each component of the model would then be specified and built separately, with the builder remaining cognizant of where each fits into the overall conceptual model. A model module would have its own worksheet(s) and would be physically separated from other modules.

In our example, as shown in Figure 11.3, we might look at that simplistic model and decide it does not capture the key components of this decision problem. It might be useful to break that eating-oysters problem down into several components such as harvesting, handling, storage, consumption, and health outcome.

It is time for a brief caveat to the reader. The risk assessment chosen to illustrate the art of modeling is a real one that was actually used for public decision making. It is presented as a real example rather than an ideal one. Thus, you will find that it does not always conform to the practices described in this chapter.

FIGURE 11.3 Key model components of *Vibrio* risk assessment model.

11.5.2.1.3 Rapid-Iteration Prototyping

Word-processing software did not spring from the womb the way it looks today. It began with someone figuring out how to get software that would type words neatly arranged on a page. It began with a prototype, and then features were added to that model until it evolved into the word-processing software we use today. Echoing the "keep it simple" message discussed previously, the way to approach modeling is to build and refine it in the same way. Build a prototype of your model; use it; examine the gaps between what it gives you and what you want; and then refine it.

> **PROTOTYPE**
>
> A prototype is a working model suitable for testing. It takes inputs from the user and produces outputs.

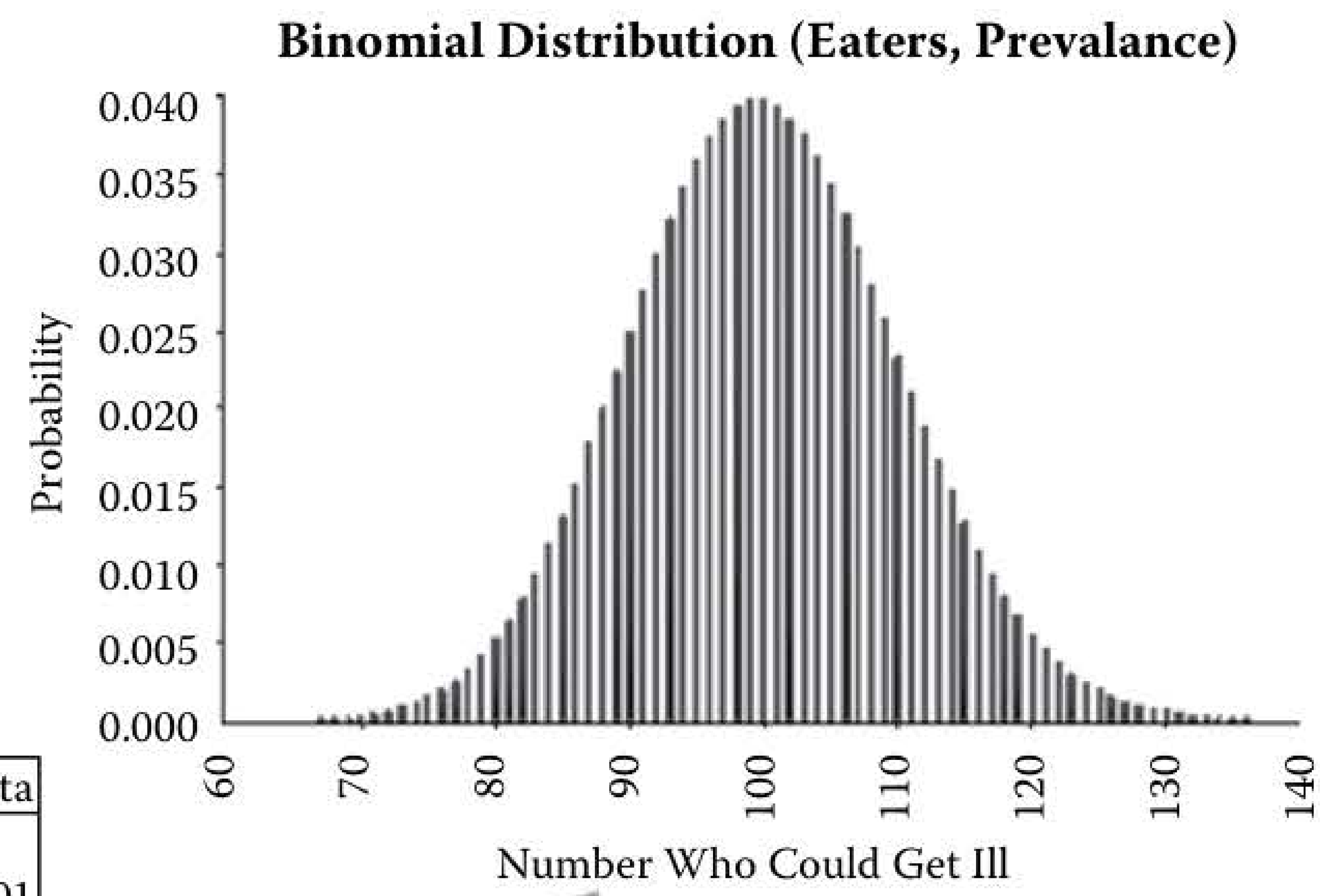
In practice, that means building a prototype of each module and testing it. Rapid-iteration prototyping means building a prototype as quickly as possible, testing it to see if it works as intended, and if it does not, moving on to the next prototype as quickly as possible. This is all done without worrying about finishing the model. The emphasis is on improving the model.

This is an especially useful approach for new model builders who struggle with wrapping their heads around a big problem. Simplify, break it into modules, and work quickly. Expect errors, insufficiencies, and problems, and this will free you to be more creative and crafty in your model building.

Risk assessment is often needed for ill-structured problems with lots of uncertainty. There are few things more satisfying in this situation than having a working model, no matter how crude. Beginning to bring order from chaos is psychologically very satisfying, even if the final model is many iterations off into the future. The most important thing is to begin. Add detail, refine your thinking, improve calculation efficiencies, and simplify the problem further as you go. Iterate, iterate, iterate your model.

Figure 11.4 shows a simple model that could estimate the number of people who get ill annually. Note that it is using test-case data. With a probability of illness, the number of eaters, and the assumption that this is a binomial process, we actually have a working prototype of the simple "eat oysters → get sick" model described previously. In the random iteration shown, 94 people become ill.

Does it give us everything we need? Probabilities could, conceptually, be estimated for different regions and seasons, but we'd have to be able to get the



	Dummy Data
Probability of illness from an oyster meal	0.0001
Number of different eaters	1000000
Number who get ill	94

FIGURE 11.4 A simple *Vibrio* risk assessment model prototype.

data and estimate them. Keep in mind that our question requires us to be able to investigate the effectiveness of risk reductions associated with different RMOs. If they could be expressed as changed probabilities of illness, we might be onto something here.

As it turns out, all the data inputs in the binomial model are uncertain. To estimate the probability of illness from an oyster meal we have to decompose it into more pieces, ultimately, the pieces shown in Figure 11.3. In addition, there are other questions asked by the risk managers that this early iteration model does not answer. There is a desire to model watermen's behavior and the other components identified in Figure 11.3. Nevertheless, the very first primitive model might look like this. We know we are not done because the model does not answer all the questions!

11.5.2.1.4 Graph Key Relationships in X–Y Plane

How do you begin to "complexify" a model? Express your intuition! Modeling is an abstract activity. You often begin with nothing but a weakly defined problem with no definitive structure and a crude conceptual model. From this you are to bring order and answers! It can be difficult to clarify your thoughts and to get started.

Powell and Baker (2007) point out that many people have a good intuition about the modeling challenge they face but lack the skills to represent those intuitions in a useful way. Experienced modelers often have a variety of ways to consider a problem. They may try to identify the parts of a risk assessment model. If working on a food safety problem, they may look for the hazard identification, hazard characterization, exposure assessment, and risk characterization pieces. Some may identify different components of these four steps. Still others will use analogies, draw influence diagrams, arrange objects on a table, perform experiments, develop risk hypotheses, or drink a beer. Figure 11.5 illustrates how the actual risk assessment represents the components of Figure 11.3.

Gulf Coast Louisianna
Summer

	No mitigation		Model Module
Water parameters			
mean m	28.650722		
mean s	1.3282188		Harvest
Water temperature	28.650722	degrees C	
	2.4080866		
Log Vp level in environment	2.408087	log counts/gram	
min time on water	5		
likely time on water	9		
max time on water	11		
Time on the water	8.6666667	hours	
Time unrefrigerated	4.8333333	hours	
Air temperature parameters			
m	-1.66		Handling
s	1.33		
Ambient air temp	26.990722	degree C	
sqrt(max growth rate)	0.1901383		
Estimate growth rate in oysters	0.2007189	log counts/hr	
outgrowth1	0.9701415		
Predicted counts at 1st refrigeration	3.378228	log counts/gram	
Duration of cooldown	5	hours	
outgrowth2	0.6021568		
Predicted counts after cooldown	3.980385		Storage
Length of refrigeration time	7.7	days	
Predicted level after die off	3.483585	log counts/gram	
Grams oysters consumed	383.227	grams	
			Consumption
Total Vp exposure in one meal	1166919.8	log counts	
Pathogenic Vp consumed	2929		
	3.4667194	log counts	
probability of illness	0.0001033		Health Outcome
Log(risk)	-3.9857685		
ill? 1=yes 0=no	0		

FIGURE 11.5 Key model components of actual FDA *Vibrio* risk assessment model.

Some will draw a picture of the problem as they understand it. Mind maps, introduced in Chapter 7, are great for fleshing out model components. Visual representations of problems can make the abstract tangible, and a problem we can look at is often easier to deal with than a mathematical or conceptual one that is stuck in our minds. Few of us are blessed with the artistic skills or confidence to think we can sketch a problem or a model. But we can all graph a freehand relationship between two variables, and that is a good place to begin.

So, assuming that we have identified the components of a conceptual model, as shown in Figure 11.3, and we have identified some critical variables in each component, we begin to ask some simple questions. What is the relationship between water temperature and the number of *Vibrio* in the water? How about the number of *Vibrio* and time on the water or the length of refrigeration? What about the probability of getting sick and the number of *Vibrio*? Even if you have no background in microbiology you may have some intuition about the nature of these relationships, and visualizing them is a good way to start.

Even simple visualizations are powerful. Consider the simple model that follows:

RMOs → Risk assessment model → Public health outcomes

It focuses our attention on three ideas. There are risk management options to consider. Ultimately, they have public health outcomes, being more or less effective in lowering *Vibrio* illnesses due to oyster consumption. The model in the middle step suggests we must identify relationships that connect risk management options to public health outcomes.

Research suggests these visualization techniques help us by externalizing the analysis. They move ideas from inside our minds to the external universe, where we and others can more readily consider them. This is especially important for risk assessment because it is usually a team effort, and externalizing ideas is essential for group work. Debate, progress, and revision are not possible until ideas are externalized and made more concrete.

Sketching the relationships between pairs of variables is also appealing because there are a limited number of relationships that are possible. A few examples are shown in Figure 11.6. As a practical matter, the linear relationship is most commonly considered. In the prototyping spirit, it makes a lot of sense to begin by ascertaining whether a relationship between two variables is positive or negative. A linear relationship assumption can always be amended as theory or increasing intuition suggests a more complex relationship.

Some relationships cannot be expressed as a simple positive or negative one. The relationship between health and a specific nutrient, for example, may look like the relationship shown in Figure 11.7. At very low levels of the nutrient, health may decline due to insufficient levels of the nutrient for peak bodily function. Across a range of levels of the same nutrient, health may increase. Once excessive amounts of

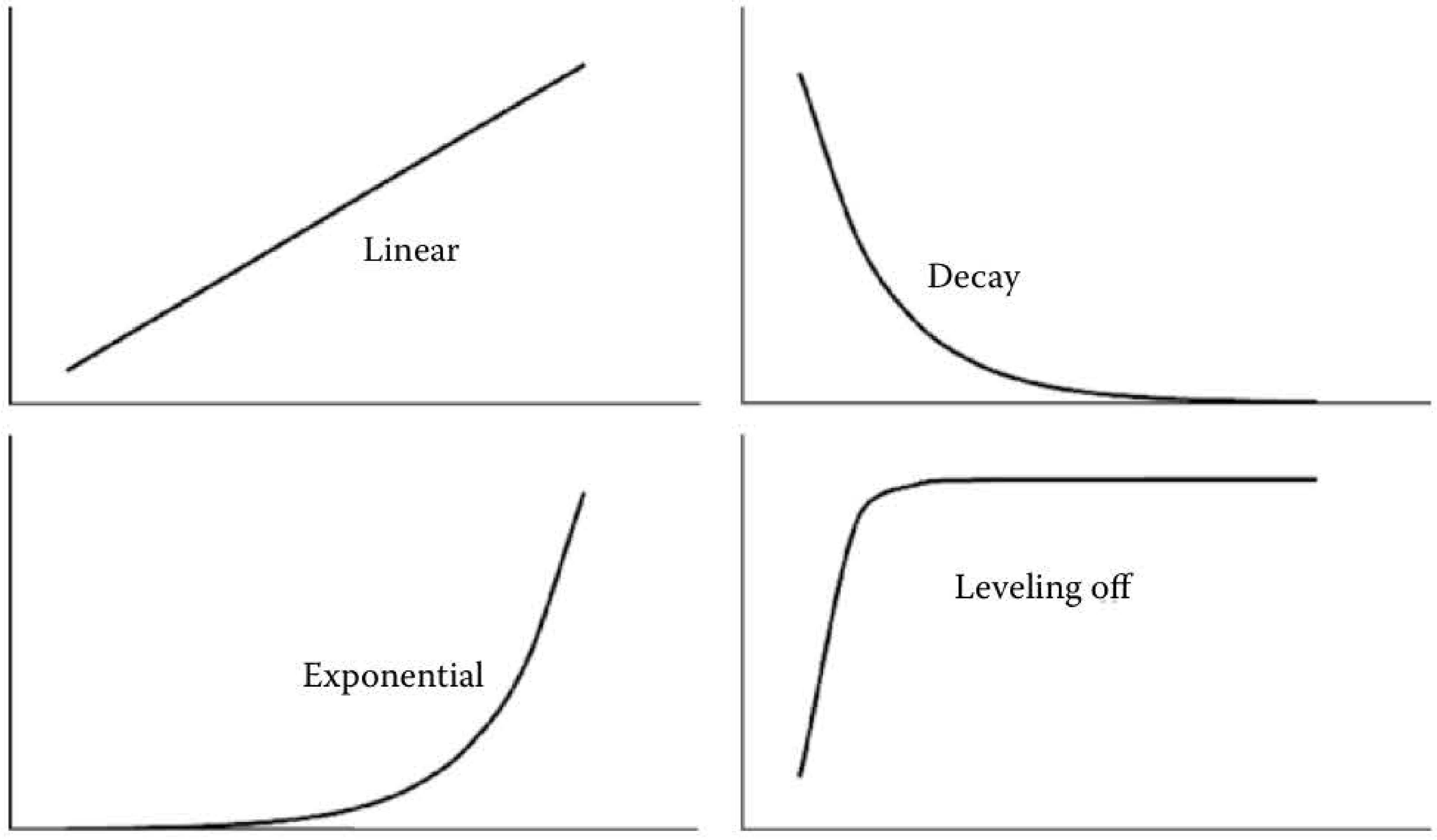


FIGURE 11.6 Sample relationships between pairs of variables.

the nutrient are consumed, health may once again be adversely affected, resulting in the odd negative/positive/negative slopes seen in Figure 11.7.

The point, not to be lost, is that once these ideas are reduced to paper they become clearer. This makes them easier to understand or to challenge should others disagree. Visually representing key relationships can be an effective heuristic for beginning to build a risk assessment model. Choose the most important pairs of variables in your model and sketch the relationship you think is possible. You need not be correct in the beginning. You only need to begin. So, pick up a pencil and sketch.

The *Vibrio*-in-oysters model offers many possible relationships of interest. A few were mentioned previously. The rough hand-drawn sketches of Figure 11.8 could

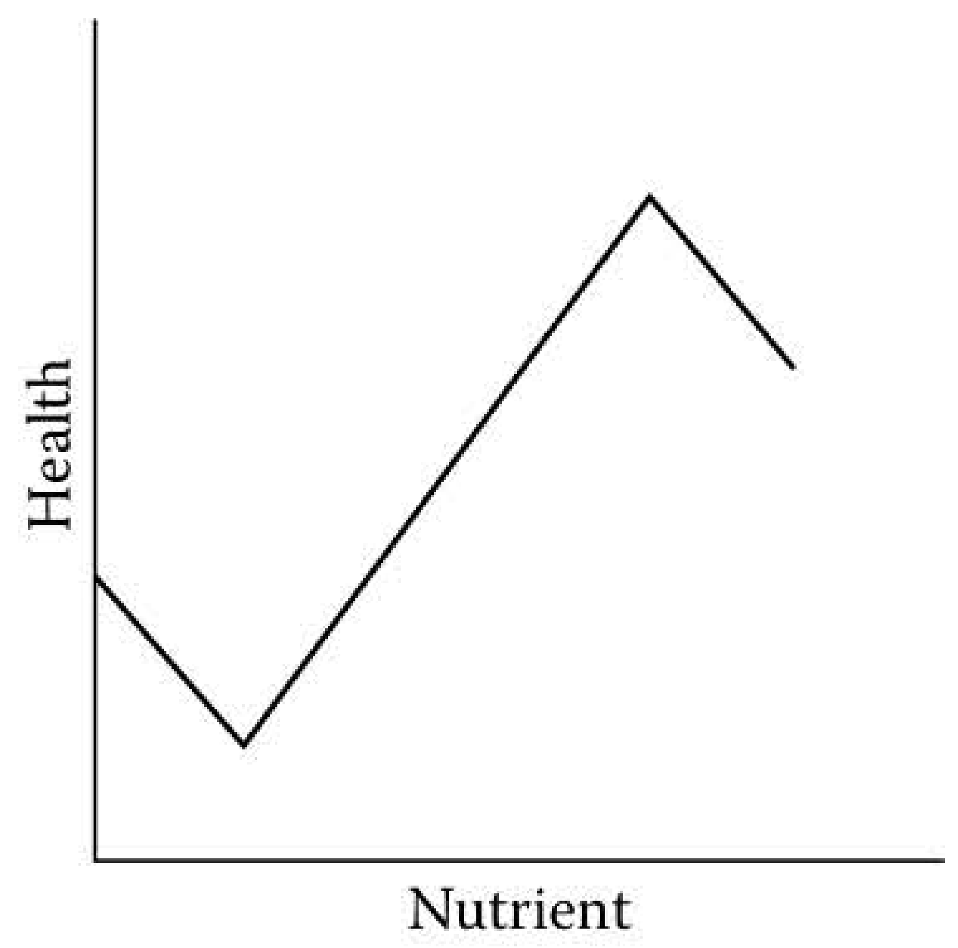


FIGURE 11.7 A nonmonotonic relationship between variables.

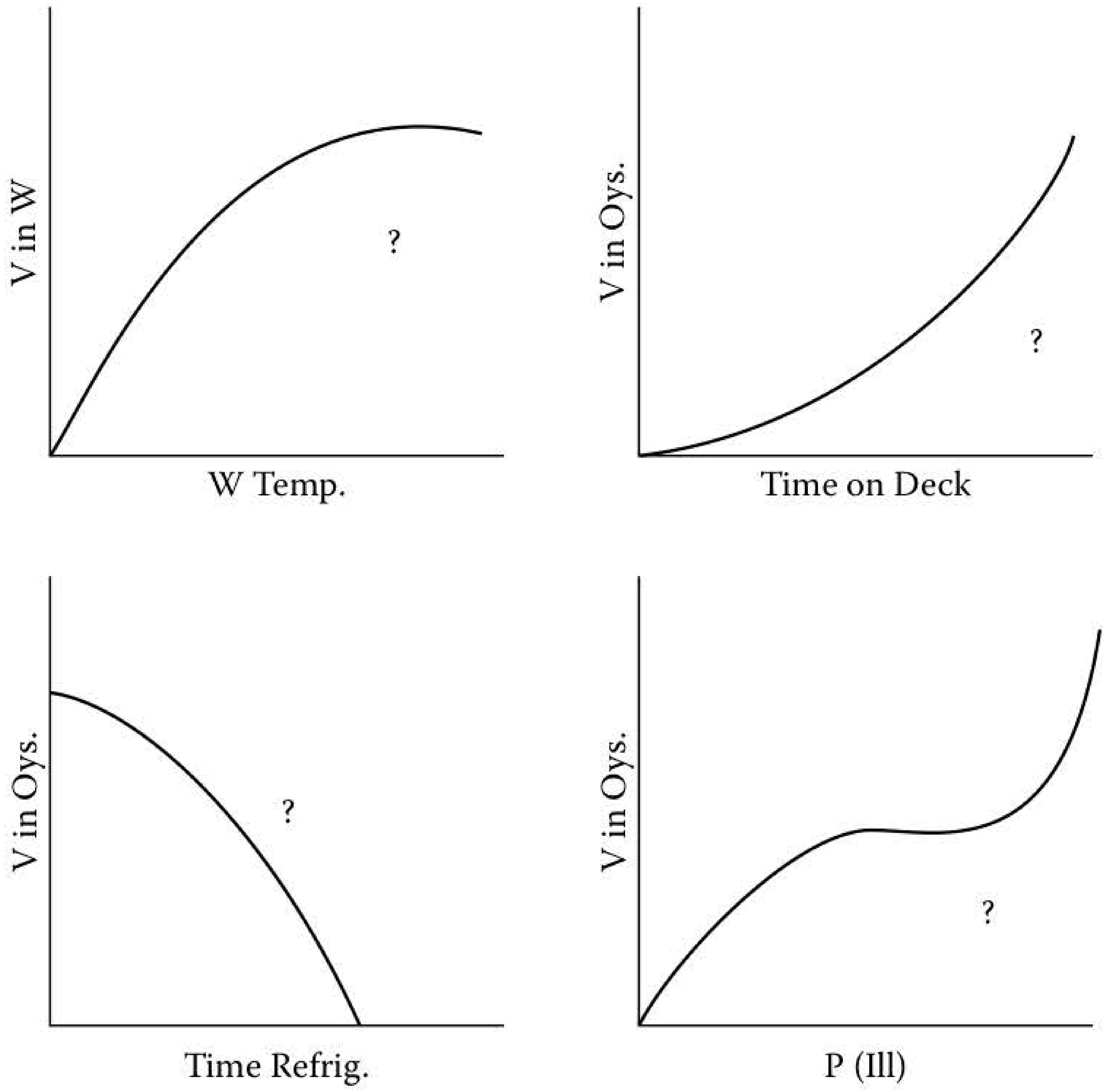


FIGURE 11.8 Hand-sketched relationship hypotheses for key assessment model variables.

represent initial ideas about the potential natures of those relationships. V stands for the number of *Vibrio*. The first thing to capture in a relationship is the nature or direction of the relationship. Is it positive or negative? Then it may be useful to represent ideas you may have about that relationship, whether they are right or wrong at this point. Presenting them visually gives others a chance to review and critique your ideas.

The upper-left sketch of Figure 11.8 shows the number of *Vibrio* present in a given volume of water over a range of temperatures. This sketch suggests the number of *Vibrio* increases up to some maximum and then levels off. As the modeler is an economist, it is useful to commit these ideas to paper so that those trained in predictive microbiology might either point out the shortcomings or confirm the wisdom of the modeler's intuition. The other sketches suggest hypotheses about how other variables are related to the number of *Vibrio*. They can be vetted with others before relationships are specified.

FAMILIES OF CURVES

It is useful to know the functional forms of different families of curves in order to produce sketches of them. The function, its basic behavior, and its mathematical form are shown here:

Linear functions—constant returns, $y = ax + b$
Power functions—increasing returns, $y = ax^b$ for $b > 1$, for diminishing returns $b < 1$
Exponential functions—decay, $y = ae^{-bx}$; leveling off at asymptote $y = a(1-e^{-bx})$
S-shaped curve—rapid then slowing growth, $y = b + (a - b)(x^2/(d + x^c))$

Source: Powell, S. G., and K. R. Baker. 2007. Management Science: The Art of Modeling with Spreadsheets. 2nd ed. Hoboken, NJ: John Wiley and Sons.

11.5.2.1.5 Identify Parameters

At some point we must move from intuition to functions and logical formulas. This is a necessary step to get from a visual representation to a mathematical formulation. In the model-building process presented earlier, this is the specification step on the way to building a computational model.

Several components or modules have been identified for the *Vibrio*-in-oysters model. Example relationships of interest were introduced in the previous section. Parameterization of a model is the step where we get specific about the model's structure and form. This heuristic may be most effectively developed and illustrated with our example.

Consider the health outcome component, specifically the relationship between the number of pathogenic *Vibrio* ingested and the probability of illness. This relationship between exposure to a hazard and probability of illness is called a dose-response relationship. A sample schematic is shown in Figure 11.9.

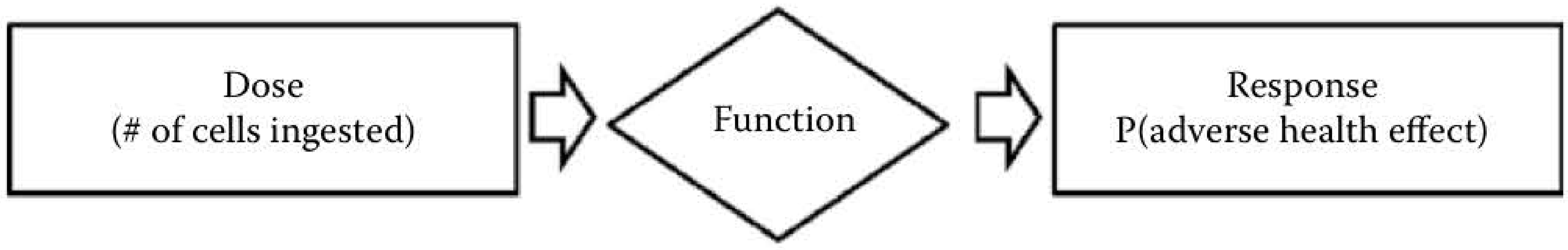


FIGURE 11.9 Dose-response links input (dose) to output (response).

Dose-response data tend to have a sigmoid shape. When available data are fit to a function, it is common practice to use any one of a number of mathematical functions that exhibit this shape. One such function is the Beta-Poisson, which is of the form:

$$P_{ill} = 1 - \left(1 - \frac{\text{Dose}}{\beta}\right)^{-\alpha} \tag{11.1}$$

where P_{ill} is the probability of illness, Dose is the number of pathogenic cells ingested, and α and β are parameters to be determined based on the available data. Such a function specifies the nature of the relationship between the dose and the health response. The function in this general form represents a family of curves. Sometimes, as in the case of the current example, it is the modeler's job to select one from among this family by choosing specific values for α and β. Identifying these parameters takes the conceptual relationship to the specified level.

FDA risk assessors determined the values for the parameters α and β to be 18,912,766 and 536,058,368, respectively. These values were determined fitting the Beta-Poisson function to the available data. The resulting dose-response curve is shown in Figure 11.10.

Probabilities of illness reach about 10% when the eater has ingested about 5.5 log of pathogenic *Vibrio* cells (about 350,000 cells). The probability of illness rises rather steadily to a 100% probability of illness at log 8.8 cells (about 630 million cells). This

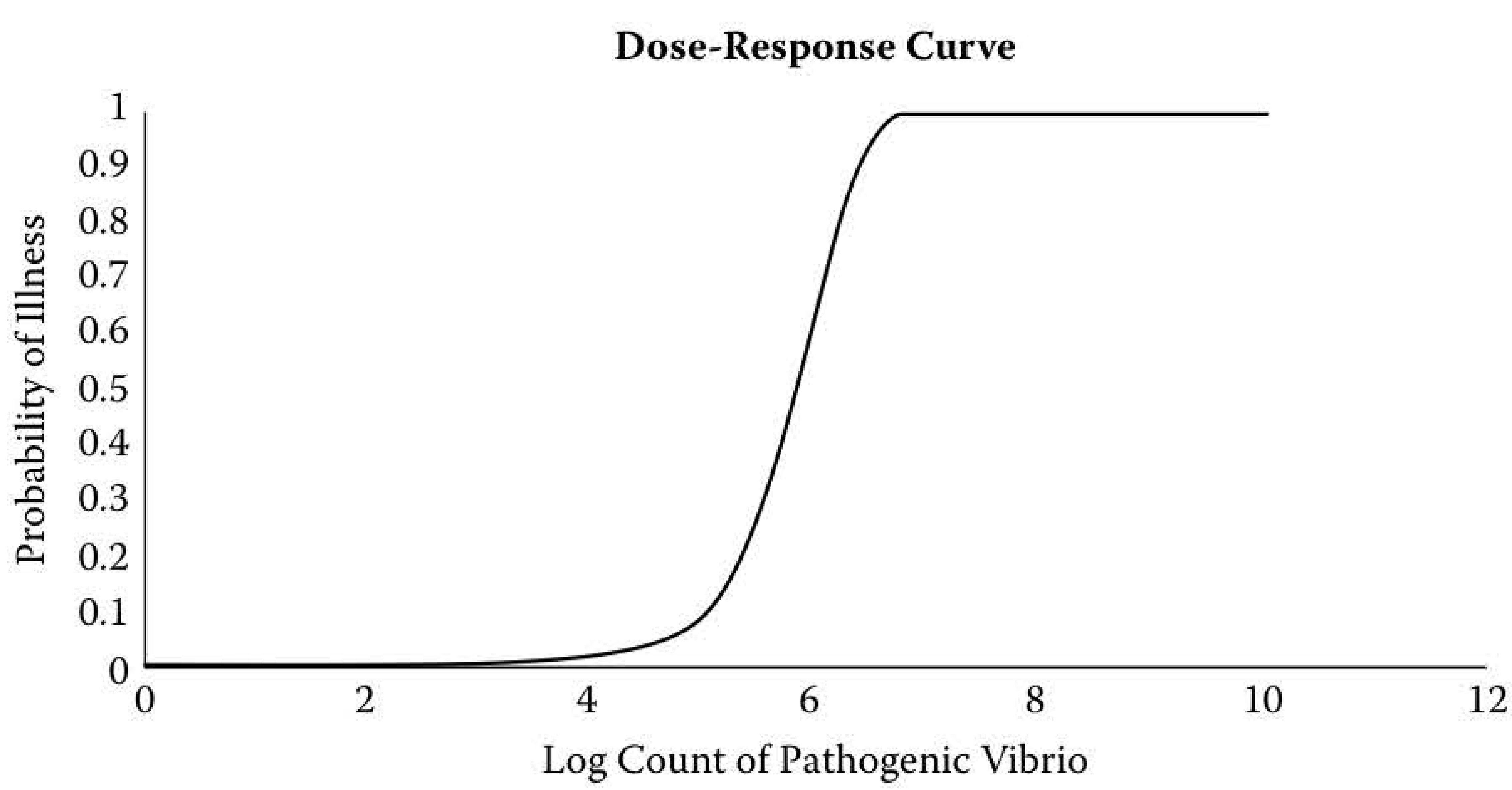


FIGURE 11.10 Dose-response curve.

is just one of an infinite number of possible relationships, however. Each pair of α and β values yields a different curve. Identifying parameters, with or without uncertainty, is a critical step in the art of model building that is often well served by statistical techniques and judgment. When point estimates of parameters are not possible, they can be estimated by distributions.

Once you have parameterized your model it is time to consider protecting those parts of your model that should not be changed. This is especially important if users will include people not familiar with the structure and details of the model. It is a simple matter to format selected cells with protection. This can avoid accidental changes to parameters that should not change. The protection can be removed if changes become necessary at some point in the future.

11.5.2.1.6 Generate Ideas and Scenarios but Do Not Evaluate Them Too Soon

Best-practice modeling is a creative activity. It is not easy to be creative. The rewards for creativity are rarely obvious, while the sanctions for being wrong are often quite evident. We are often biased by our education, training, and experience to be logical and to seek the right answer. But life is often ambiguous and even maddeningly illogical. The wicked problems of risk analysis rarely have right answers; there are only answers that are better or worse than others.

Following the rules and being practical are lessons we learn at home, at play, and at work. While these are wonderful attributes, they do little to aid creativity. One of the more persistent blocks to creativity is our tendency to judge an idea prematurely. Many of our education systems do a better job teaching us how to criticize ideas than they do teaching us how to create ideas. Consequently, many of us have a strong preference for judging ideas rather than generating them, and we are quick to turn it loose against ourselves.

Model building, like much of life, requires a period of divergent thinking, followed by some intensive convergent thinking. It is terribly easy to rely on our biases, mindsets, and belief systems when we approach tasks that cry out for creativity. When trying to model a problem, begin with divergent thinking. Generate as many alternative ideas as possible. Don't evaluate. Be creative.

Brainstorming, the topic of Chapter 8, is a creative thinking technique that is both effective and easy. The only real difficulty is getting a group to commit to use the technique and to have fun with it. One of the tenets of the best brainstorming techniques is a "no evaluation/no criticism" rule. A second valuable rule is "the wilder the better." Use them both early in the model-building process.

Do not fall in love with your first idea. When starting to build a model, it is important to consider as many different approaches to the problem and to answering the risk manager's questions as possible. That cannot happen in an environment where ideas are criticized and defeated before they have had a chance to fully develop. The harshest critic of most ideas is the self. We are very quick to censor our own thoughts and only slightly less quick to criticize the ideas of others.

When the critical self is in charge, the open-ended job of developing a model can seem overwhelming. The downside of every approach, the considerable problems we can anticipate, the hurdles that seem insurmountable are all very evident from the start. Modelers must learn how to quiet the critic in themselves and in others. When building a model you will make mistakes. There will be blind alleys. You may get halfway through a model and realize there is no way out. It is okay to begin again.

Begin by generating as many ideas and approaches as possible. Think of as many ways to model *Vibrio parahaemolyticus* and oysters as you can. Let them ferment for a while. Refine them, explore them, flirt with them, and get to know them. Only then is it time to systematically evaluate them.

11.5.2.1.7 Work Backwards from the Desired Form of the Answer

If you have done a good job at divergent thinking, you may have many possible approaches to take with your model. Your problem now is to choose one. To do so it is often helpful to begin at the end of your model.

What do you want your answer to the risk manager's question(s) to look like? Imagine the specific form the best answer will take. Pay attention to the units of measurement your answer(s) may require. Then work backwards from that point to the best approach or the most logical starting point for your model.

Powell and Baker (2007) propose the "PowerPoint heuristic." Imagine that your answer must be summarized in a single slide. What is the essential message for that slide? Is it a figure, a graph, a table, a procedure, a number, a distribution, a recommendation? Working through this exercise helps you recognize what your critical outputs are and what the essential message of your model must be.

In the model for *Vibrio* in oysters, we want to know how many people get sick annually the way things are done now. We also want to know how those numbers of people might be reduced by a heat treatment, a freeze treatment, a rapid cooling treatment, and regulation of the harvest.* Let us stay with the estimate of the existing risk for now. What do we need to get the annual number of people who get sick? First, a rate of illness or probability of illness per eating occasion is needed along with the number of eaters. It is a simple matter to understand how these two outputs get to the essential message for us.

What do we need to get the rate of illness? We need to know how many eaters out of, say, 100,000 get sick. What do we need to get this? We are going to need a dose-response relationship and a dose. This approach suggests that we might want a model that provides the dose for a given eater at a given meal. If that is the case, we will need a way to estimate the eater's exposure to pathogenic *Vibrio*. This is an input to the dose-response relationship.

Continuing to work backwards, to get the number of pathogenic *Vibrio* in a meal, we need to know the size of the meal and the *Vibrio* load per gram of oyster flesh. Then the *Vibrio* per gram times the number of grams will give us a log count of the pathogenic *Vibrio* the eater ingests.

Preliminary risk assessment work may have provided insight into the nature of the available data. For example, if data on pathogen loads at the time of consumption are available, our model may have reached its beginning point. Get the data on pathogen loads and meal size and work through to the number of annual cases of illness.

---

* These details on the nature of the risk interventions under consideration are found in the risk assessment.

In the *Vibrio* example, these data were not available, so it was necessary to use predictive microbiology to estimate it. Using the available science, the model estimates the amount of *Vibrio* present in a gram of oyster flesh at the time of harvest. That number is then allowed to grow as the oysters remain on the deck of the workboat and during the period when the oysters are being cooled down. This produces a maximum number of *Vibrio* per gram of oyster flesh. The model estimates the amount of die-off during refrigeration and uses this value to estimate the total *Vibrio* load in a meal. Along the way, the modeler learned it was easier to work with all *Vibrio* and to estimate the percentage of them that were pathogenic later in the model. Working backward from the form of the answer you want often provides the focus you need to begin an open-ended process like model building.

11.5.2.1.8 Focus on Model Structure, Not on Available Data or Data Collection Issues

"Models before data" would be a great bumper sticker for modelers. A good modeler focuses on the structure of the model and getting it to represent reality in a way that meets the information and decision-making needs of risk managers. One of the most common, most appealing, and most limiting approaches to model building is building to the data. Consequently, many modelers begin by searching for, collecting, and analyzing the data. They then build a model that uses the data they have instead of the model that should have been built.

It is tempting for a modeler to build a model that makes use of the existing data. Likewise, it is tempting to avoid including model components for which data do not yet exist. This approach is flawed from the outset. There is no automatic reason to assume the available data are accurate for your purposes. At the other extreme, if we avoid model structures that require new data, we may be failing in an opportunity to "grow the science." Research can be as integral a part of risk assessment as is addressing uncertainty in our analysis.

Build the model that will solve the problem and answer the risk manager's question(s). Do not worry about the data. A good model improves the risk assessment process and often expands the body of knowledge because of new data that must be collected and organized.

The existing data are often flawed and need to be carefully screened. All our data describe the past, and all our risk management decisions are about the future. There is always good reason to suspect the temporal relevance of our data. Some of the more common sources of bias and errors in the data are sampling error, data collected for different purposes, masking, inappropriateness, definitional differences, geographic or cultural irrelevance, flawed experimental design, and so on. See others in the accompanying box.

The best-practice method is to build the model you need. Let it tell you what data you need rather than letting the data tell you what model to build. Once the appropriate specification model is built, you can then use the available data to refine the model. This sequence may also avoid unnecessary collection of additional data. If necessary, your better data-free model can be revised to accommodate the available data, but you will know what you are sacrificing in the process. Getting new data and getting better data are not always obligatory activities.

TYPES OF BIAS

Selection bias
Volunteer or referral bias
Nonresponse bias
Measurement biases
Instrument bias
Insensitive measure bias
Expectation bias
Recall or memory bias
Attention bias
Verification or work-up bias
Intervention (exposure) biases
Contamination bias
Co-intervention bias
Timing bias(es)
Compliance bias
Withdrawal bias
Proficiency bias

Source: Hartman, J. M. et al. 2002. Tutorials in clinical research: Part IV: Recognizing and controlling bias. Laryngoscope, 112: 23–31.

Good data do matter. But good data in a flawed model are not necessarily better than imperfect data in a good model. Good decisions are frequently driven more by good models than the data that are in them. A well-structured model can often reveal valuable insights with a rough estimate of model parameters. Frequently, insights about the problem gained by building the model are sufficient for informing the decision-making process. Data are not always necessary for informed decision making.

11.5.2.2 The Practice of Modeling

Modelers have developed many handy and helpful ideas about how to build a model in a spreadsheet environment. These are the focus of the remainder of the chapter. You will notice a bit of overlap with the art of modeling heuristics at times. The difference is that here we assume you have tackled and solved the conceptual problems and now look to practice the art of modeling.

11.5.2.2.1 Sketch the Model—Document the Logic Flow

Once you understand the model, after your initial visualization efforts and before you begin to build your spreadsheet, sketch the logic flow or the components of the model. The visual representation will help you decide how best to organize your efforts. Figure 11.11 is taken from FDA's *Vibrio* risk assessment. It provides a very effective visual summary of the model's construction. This diagram makes it easier to identify component modules at a finer level of detail and more intuitively than does Figure 11.3.

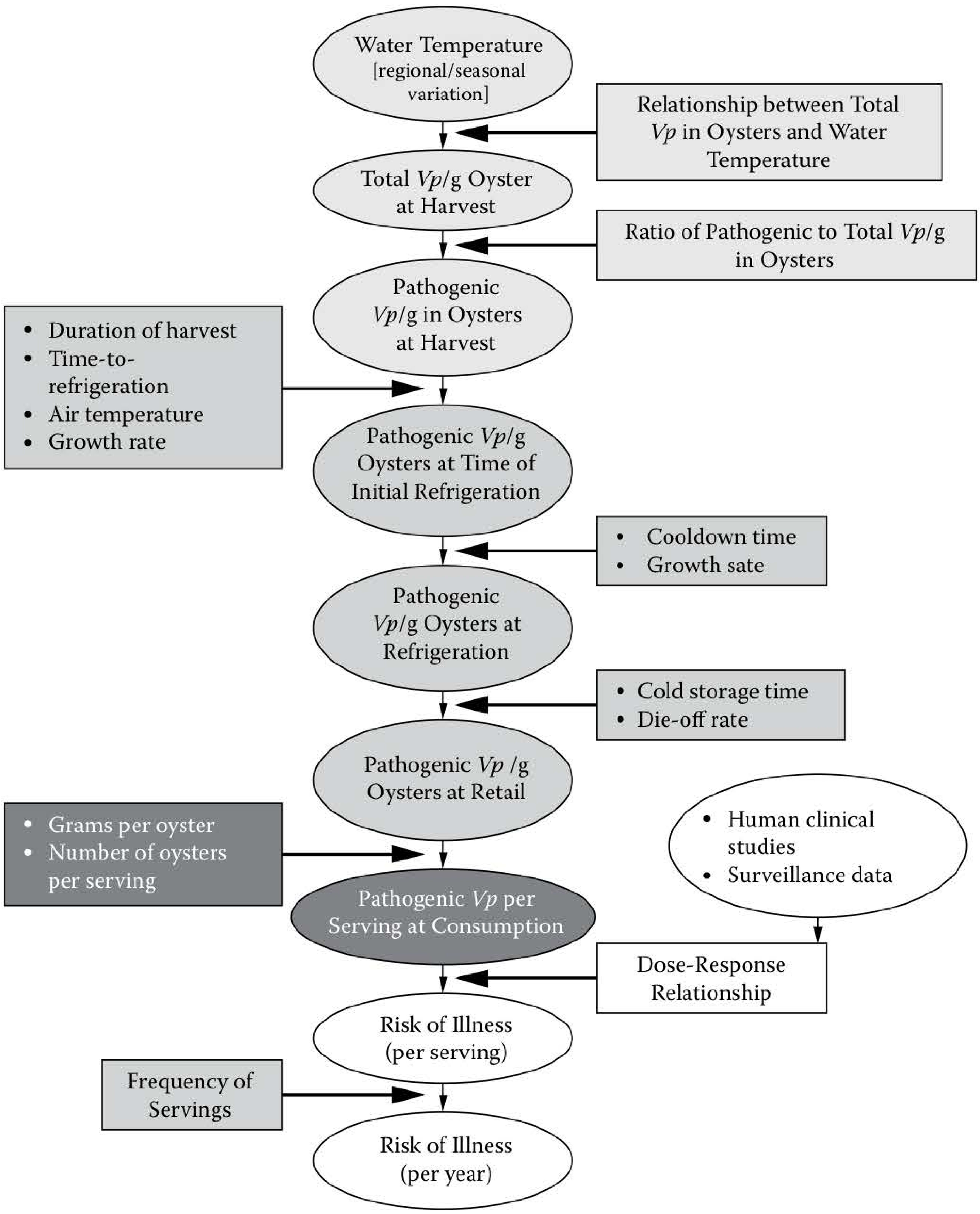


FIGURE 11.11 Schematic representation of the *Vibrio parahaemolyticus* risk assessment model. (From Food and Drug Administration. 2005. Center for Food Safety and Applied Nutrition. Quantitative risk assessment on the public health impact of pathogenic *Vibrio parahaemolyticus* in raw oysters. http://www.fda.gov/Food/ScienceResearch/ResearchAreas/RiskAssessmentSafetyAssessment/ucm050421.htm.)

Make the shapes in your sketch be meaningful. Influence diagrams are a common model-sketching tool that rely on the common convention of showing a chance event as a circle, a decision as a square, a calculation as a rounded square, and a payoff as a diamond. As long as the shape and color schemes in your model sketch have meaning, their information content is enhanced.

11.5.2.2.2 Think Communication as You Design the Model

Before you begin to build your computational model, think about how you can organize the model so that it aids the risk communication process. Spreadsheet models consist of tabbed worksheets. If you use these worksheets to give your model a logical structure, it can save you and others who may use it a great deal of time in the future. Following are some suggestions for organizing a spreadsheet model.

Overview sheet: The first sheet should provide an overview to the risk assessment model. Why was it prepared and what does it do? When was it developed and by whom? Using what software? The overview sheet should be brief, and it should at some point include the list of risk management questions the model was built to answer.

Instructions sheet: Documenting your model is a critically important part of the modeler's job. Spreadsheet models are not likely to be so complex as to require an actual user's manual. However, they should provide instructions sufficient to enable a future user of the model to understand how to update data and other model components as well as how to run the model and access the desired results. The second sheet should include these instructions. If the model has been verified and/or validated, the methods by which this was accomplished and the time frame should be documented here as well.

Log of use sheet: This feature is rarely included, but it can be extremely useful, especially when it is being used by a government agency for regulatory, stewardship, or other public purposes. The log of use sheet identifies who used or modified the model for what purposes and when. It can also be used to record the planned sequence of simulation experiments and the dates they were completed.

Guide to model sheets: A modular risk assessment is likely to consist of many separate worksheets. The organization of a worksheet can be difficult to discern from the short phrase on the tab. In a large model, all the tabs may not even be visible at the same time. This guide identifies the various sheets in the model and briefly describes the purpose or function of each sheet. The guide functions as a table of contents and is an indispensable aid for a large model.

Influence diagram/conceptual model: The sketch of the model needs to be presented somewhere in the model itself. It may be helpful to add a narrative to accompany the influence diagram or model layout. This may be combined with the guide to your model sheets.

Assumptions: Assumptions are critical to the construction of any risk assessment model. All explicit assumptions and as many implicit assumptions as possible should be identified and listed in an assumptions sheet. It may be helpful to organize the assumptions by model module.

Inputs: One of the most practical things you can do is create a sheet where all input variables and parameters can be entered. Separate the variables from the constants. This will make it easier for you to correct, change, revise, and update data throughout the model's lifetime.

Relationships: Adequately identifying and modeling dependencies among variables in your model is one of the modeler's greatest challenges. It is critically important to document and to convey to others the nature of these dependencies. Each interrelationship between or among variables should be described here in a narrative form. Include the formulas used to capture these dependencies when possible and indicate where in the model they can be found. In the *Vibrio* model this would include the dose-response relationship and the growth and die-off equations, for example.

Performance measures: Many models will have performance measures. These are cell values that are of interest to the risk assessor or the risk manager. They tend to be intermediate calculations rather than outputs. The number of *Vibrio* cells per gram of oyster flesh is a performance measure in our example. A model that calculates a benefit-cost ratio would have costs and benefits as performance measures. A transportation model might have travel time as a performance measure. Identify all the performance measures and include the formula used to calculate them when possible. An example is presented in Table 11.1. In some instances, it may be useful to include a cell reference for the performance measure (not shown in the table) so it can be easily located in the risk assessment model and observed.

Outputs: Collecting all model output cell values in a single place is extremely helpful with complex models. Models constructed in a modular fashion may have outputs of interest spread over several worksheets. In an event tree, model outputs may be scattered over many cells in a worksheet. Consolidating all output cell values on a single page is not only convenient for observing outputs, but it can also be an effective editing and debugging aid. Note that these outputs are not detailed statistical and graphical outputs of simulations. They are only images of the output cells, similar to the model values shown for performance measures in Table 11.1.

Modules: Each model component should have its own module, and each module should have its own worksheet(s). The worksheets should be labeled descriptively, and each should include a brief narrative explaining the function of the module. If a model sketch is used, it is helpful to indicate what part of the model is being represented by the module. The module sheets are the guts of the model.

Historical data used to generate model inputs: Often, model inputs are derived from historical data. When this is the case, it is desirable to include these data in the risk assessment model itself if it is practical to do so. The size or proprietary nature of the data files may make it impractical. When the data cannot be included, its source and location should be noted.

Graphics page: Some models make use of dynamic graphics so that model effects and relationships can be more easily observed. If your model makes use of dynamic graphics, gather and present them all in a single worksheet when practical to do so.

TABLE 11.1
Performance Measure Examples for a Risk Assessment Model

Performance Measure	Model Value	Formula
Log *Vp* level in environment	2.0963456	−1.03 + 0.12 * E10 + RiskNormal(0,0.886)
Predicted counts at first refrigeration	2.8192952	IF((E12 + E27) > 6,6,E12 + E27)
Predicted level after die-off	2.6691434	E31 − 0.0027 * 24 * E32

11.5.2.2.3 Organize Your Model into Modules

My best advice to new and aspiring modelers is: do not attempt to work out the entire model in your head all at once. Put things on paper. Sketch the model's major parts. Practice the art of modeling. Once you have identified the components of your model, you have the framework for a modular model design. Build it simply first, then add to it.

Repeat the model-building process for each module. Treat the module like a mini-model. Build and work on one module at a time. Sketch it unless it is simple and obvious. Identify its subcomponents. Consider making subcomponents separate modules. Don't go overboard with the decomposition but make an effort to define modules that make sense to yourself and others.

If you are modeling a process, the modules are often easy to define. The egg production process in Figure 11.12 provides an easy identification of modules. If there is no clear process, allocate specific tasks in the model (e.g., assessment and hazard characterization) to different sheets.

Use common sense. Some models are compact even though they have several discrete components. Though it is always wise to think in terms of modules and to build in terms of modules, it is not always necessary to separate the model physically along those lines. The FDA model we have been looking at, for example, does not use a modular approach.

11.5.2.2.4 Handling Inputs and Outputs

There are reasons beyond good communication for segregating your input and output values. Efficiency and accuracy are at the top of that list. A spreadsheet model should provide a single point of entry for the values of all model inputs, that is, variables and parameters. Isolate the data inputs to a single sheet of the model and it will be easier to input data, check for errors in inputs, avoid errors in formulas, and revise and update your inputs.

Formulas in your model should not include constants of numerical values nor distribution formulas for variables. Enter all such values into the cell of a single input

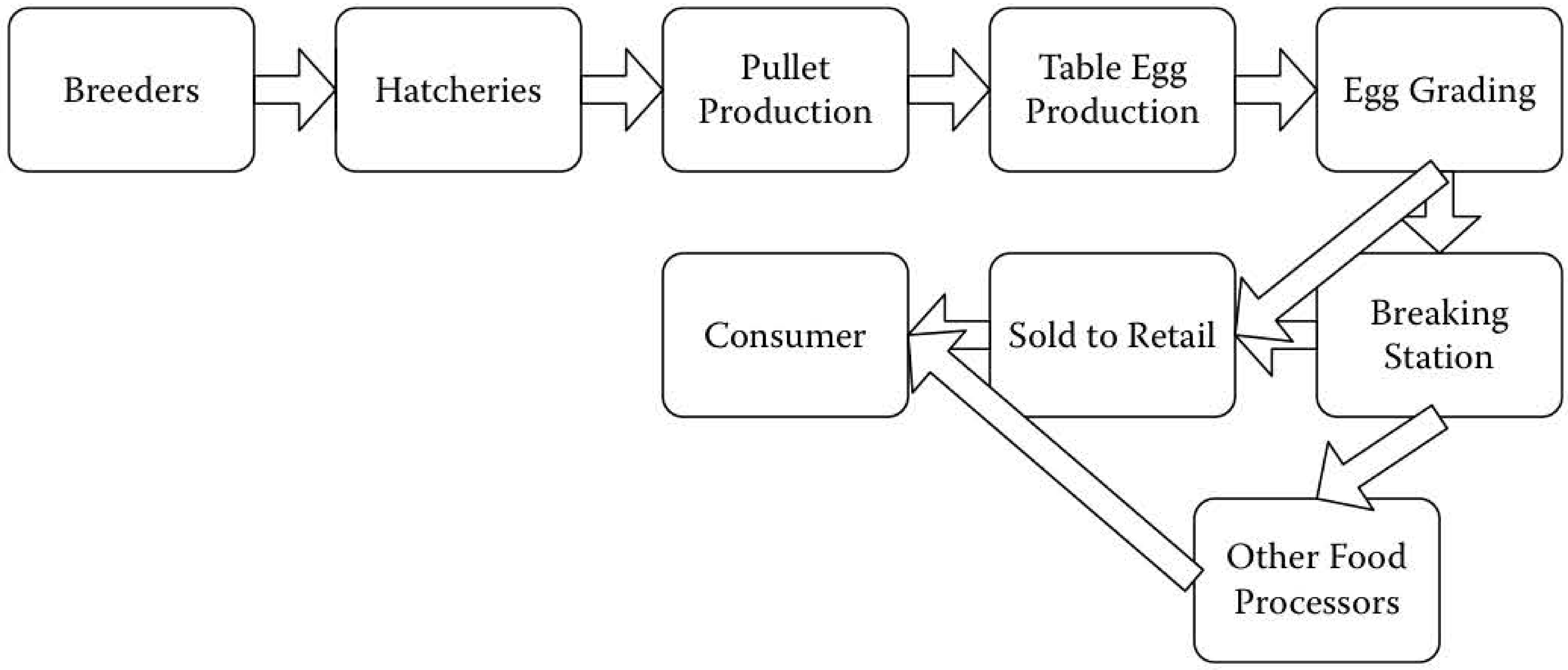


FIGURE 11.12 Process flow for shell eggs and egg products.

	A	B	C
1		Direct Entry Model	
2		Weight in pounds	Formula Syntax
3	Man 1	178	RiskNormal(182,16)
4	Man 2	194	RiskNormal(182,16)
5	Sum	371	
6			
7		Cell Reference Model	
8		Weight in pounds	Formula Syntax
9	Mean	182	182
10	SD	16	16
11	Man 1	188	RiskNormal(C9,C10)
12	Man 2	158	RiskNormal(C9,C10)
13	Sum	346	

FIGURE 11.13 Comparison of direct parameter and cell reference parameter entry.

sheet and then make sure all your formulas use the cell references of that cell in the input sheet in place of the actual values.

Consider a simple example of a model that calculates the combined weight of two men. Figure 11.13 shows two ways to enter the data. Notice that column and row identifiers are found in the figure's margins. The direct entry model shows the formula using the actual numerical values as arguments in the distribution. The cell reference model shows arguments for the distributions using cell references.

The advantage to using cell references and a single point of entry for input data is that if new information shows the standard deviation to be 17 pounds instead of 16, then a single change in the cell reference model at cell B10 updates the model. Using direct entry, the modeler must be sure to make the changes in every instance where the standard deviation appears in the model.

In a more complex model, all input values like the mean and standard deviation would be entered on a separate input worksheet and then referenced there.

On that input sheet, separate your inputs into constants and variables. Constants are numerical values that do not change for a model run. Variables are numerical values that do change in a model run. There are two kinds of variables: deterministic and stochastic. Deterministic variables vary in a known pattern, as the hours of operation might reflect a systematic pattern of change from opened to closed over a week. These values may well change within a model run but they do so in a determined way. Decision variables and value variables are also deterministic variables, although we may be uncertain about what the most appropriate values for these variables are. In

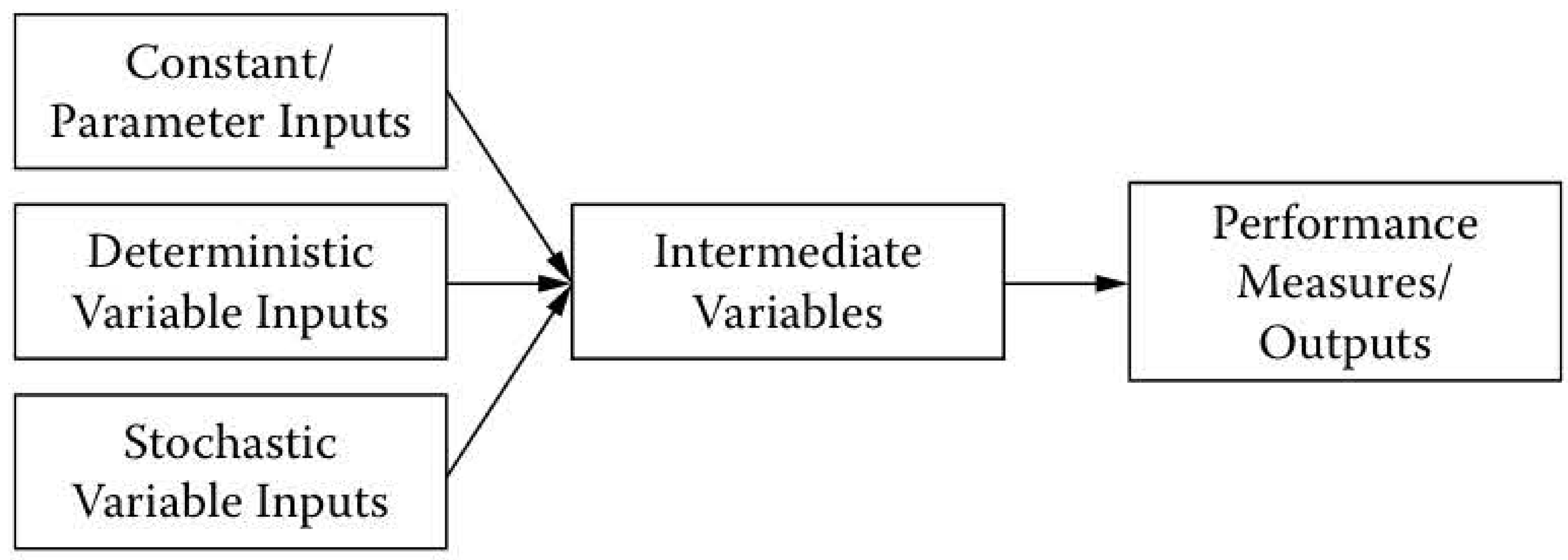


FIGURE 11.14 The role of intermediate variables in model input/output relationships.

that case, different values may be chosen for multiple simulations of an otherwise identical model, or we may simply want to do some sensitivity analysis of specific variables that have no true value. Stochastic variables have an element of randomness to them. They are usually represented by probability distributions. Segregate and identify constants, deterministic variables (including decision and value variables), and stochastic variables in your input sheet.

You may want to segregate the intermediate variable calculations of your model in their own worksheet as well, as represented in Figure 11.14. These include the performance measures mentioned earlier. Intermediate variables link inputs and outputs through intermediate calculations. The ability to monitor intermediate values can be a valuable debugging feature of a model. When intermediate variables are performance measures, there may be intrinsic value in being able to observe them as one steps through a model. Consider gathering an image of all significant intermediate variables in a single place in your model.

Gathering an image of all model outputs in a single sheet provides a convenient way to find all the values of interest for answering the risk manager's question(s) in a single place. Consider developing modeling conventions for your organization or at least for yourself. One of the simplest of these is to use color-coded fonts. Choose one color for any cell that includes an input value, another color for intermediate calculations, a third for outputs, and perhaps a fourth for data. This way the color patterns of your spreadsheet cells immediately convey useful information to anyone who knows the color code.

11.5.2.2.5 Output Reports

Each planned model run will produce outputs and performance measures of interest. These outputs should be documented in data arrays, tables, and graphics, that is, in output reports. There should be an output report for each run of the model. A large number of model runs may necessitate storing them in a separate spreadsheet file.

11.5.2.2.6 Design for Use

Design your model to be used. We have already said, "Build the simplest model possible but no simpler." The corollary to that rule of thumb is to build your model as simply as possible. To make your model easier to use: calculate first, simulate second. Although there may be no closed-form analytical solution to your overall model, there may be analytical parts of the model. Whenever you have a choice between analyzing and simulating, choose analyzing. Keep your cell formulas simple. When you have a complex formula, break it down into several intermediate calculations. This will help avoid difficult-to-find errors, and it may speed up your model calculations.

As you determine your model's structure and begin to quantify the variables and their interrelationships, one of the issues you must resolve is the level of detail or the resolution of your model. One way to frame this issue is as a trade-off between the realism of the model and its utility.

As the terms are used here, realism means the extent to which the model faithfully represents the details of the real-world process being modeled. Utility means the extent to which the model provides useful answers to the risk manager's questions. A simple model may be sufficient for answering the risk manager's questions, but it may not be a very realistic model. On the other hand, a realistic model is not guaranteed to produce useful answers.

Realistic models tend to be larger if only because they include more details. Greater detail is rarely worthless. However, greater detail also requires significantly more data to enliven those details, and data availability is always a real concern for realistic models. Realistic models do have the advantage of getting to a tangible level of understanding, a level of resolution that is easy to envision. In a large, detailed model, no one part of the model is likely to be "too" critical. Indeed, aggregating the results of many smaller components to get an output can be more accurate than directly estimating an output. If these are important concerns, you will want a more detailed, more realistic model. Expect such a model to require a lot of data.

On the other hand, a more abstract model with less detail can be a lot smaller. Sometimes a more abstract model is a necessity and not a choice, as it may not be possible or practical to include the additional detail in a model—although you know better than to build to your data! There may be too much knowledge uncertainty about the system being modeled, or it could simply be cost prohibitive to develop a realistic model. Oftentimes, capturing the key elements of a system in your model is sufficient to support decision making.

In general, when a model provides useful answers to the risk manager's questions, it is ready to use. Design your model for use. Then, if you decide that incremental additions to the model's realism are justified by the value of the information these enhancements produce, you can add as many as you like. But first, build a model you can use.

11.5.2.2.7 Minimize Stupid Mistakes

Take extra care with modeling dependencies and relationships among variables and parts of your model. We live in a complex universe, and many phenomena are treated as independent of one another, often because it is too difficult to deal with the more complex reality. Independence is not to be an assumption of convenience for the modeler. Things must truly be at least functionally independent to be so represented in your model.

Your model's structure should prevent the generation of any unrealistic scenarios. To ensure this, you must recognize and model any interdependencies among your model's variables, calculations, and components. Things that always happen together should happen together in your model. Things that can never happen together, things that move in the same direction, things that move in opposite directions—all must be accurately represented in the structure of your model. Dependencies can be built

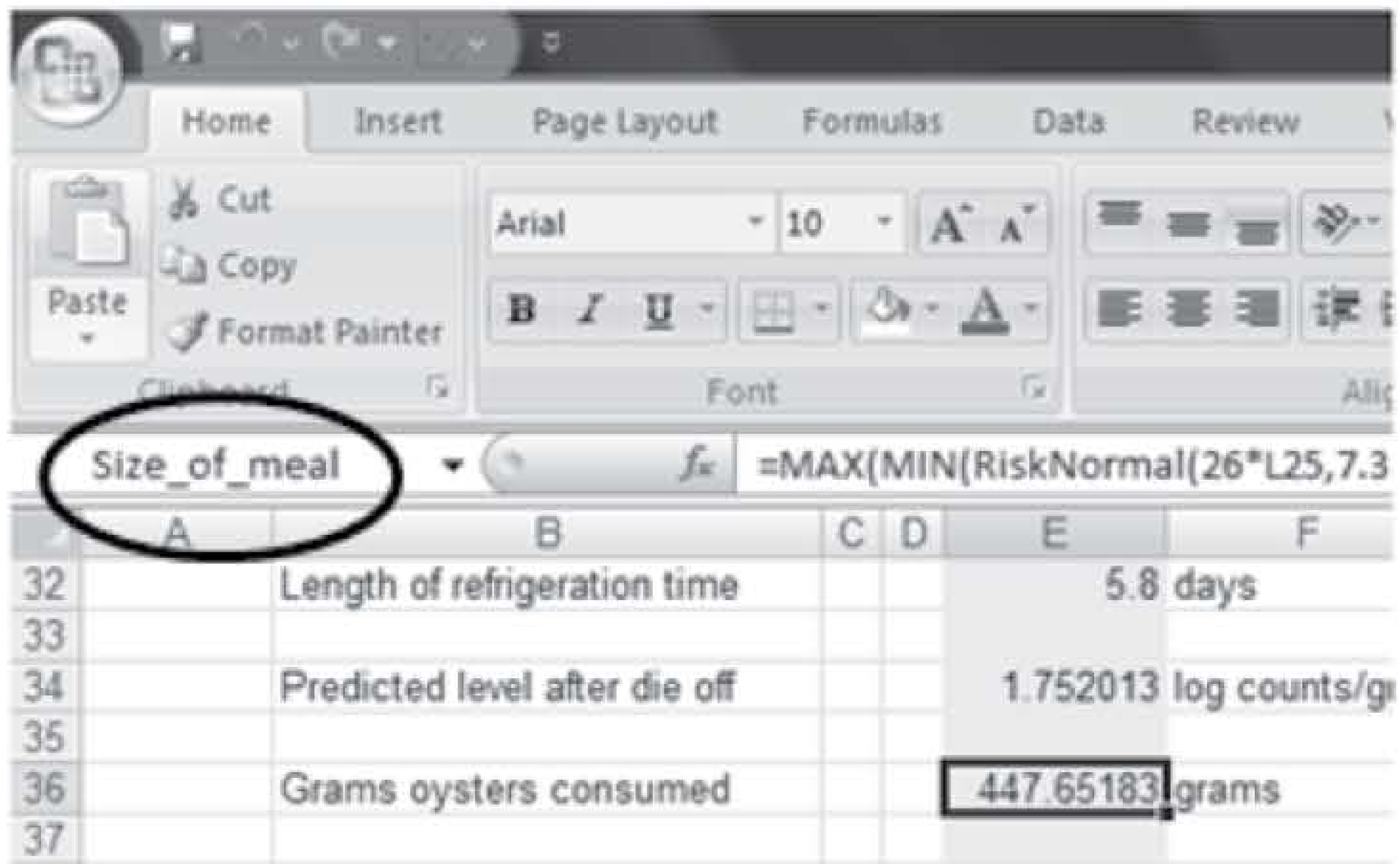


FIGURE 11.15 Use range names for key variables.

into a model using complex logic arguments, functions, correlations, linkages, and embedded arguments, among other means. Use whatever means necessary to ensure that you have carefully tended to important real-world dependencies and relationships, so they are reflected accurately in the workings of your model.

Copy and paste carefully. Remember that when you paste a cell formula your software is adjusting cell references for relative changes in their new location. This can result in unanticipated errors. Always make sure cell references are as desired after pasting copied material. Remember to use dollar signs ($) when you want to preserve the absolute location of a cell when you copy and paste.

Use range names for key inputs and intermediate values. Using simple descriptive names for a cell or range of cells makes formulas that reference these cells much easier to read and understand. In Figure 11.15, cell E36 has been named "Size_of_meal" in the example shown here. This is far more informative than using "E36."

Compare the difference between the mathematically identical formulas in Table 11.2. To understand the first formula, you would have to locate the cells referenced and see what the referenced values are before it is understood. The second formula uses range names and is far easier to understand. The more descriptive quality of the range names limits the likelihood of a stupid mistake like a cell reference typo where you type E35 instead of E36.

Use comments generously in your model to document assumptions and explain things. Figure 11.16 presents an example to demonstrate the concept. This comment is not found in the original FDA model; it was created to illustrate the point made here about the value of comments. Comments can help eliminate stupid mistakes by

TABLE 11.2
Comparison of Cell Address and Range Name Formulas

=10^E34*E36
=10^Log_vibrio_per gram*Size of meal

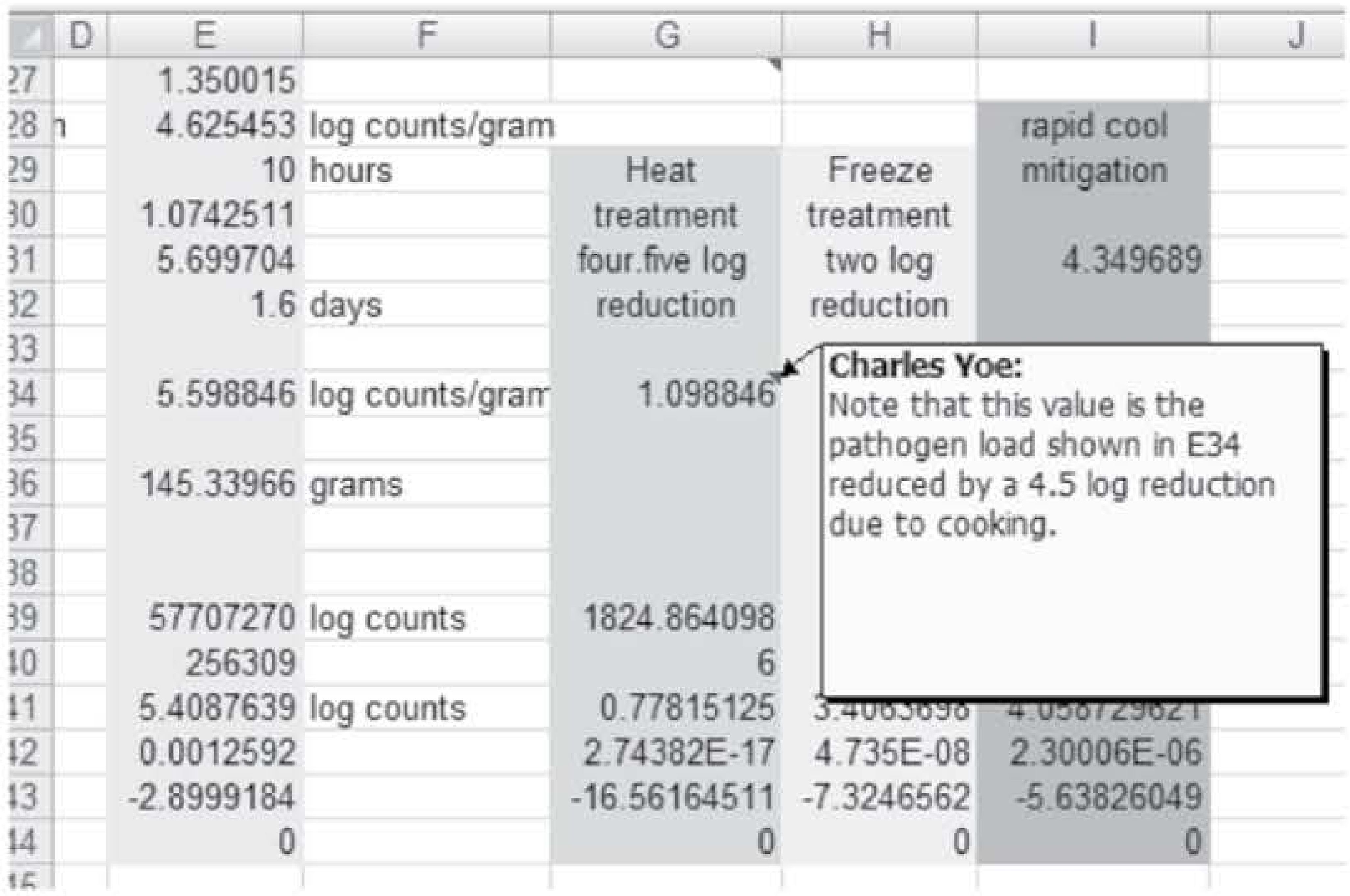


FIGURE 11.16 Use comments to document and explain the model.

reminding you and the naïve user of your model of things you might dangerously forget without the reminder. They are also excellent documentation aids.

Create generations of backup models. If you have ever saved changes in a file and then realized your changes were incorrect and mourned the loss of the correct file, you may already be doing this! When developing a model, each time you begin to revise it, save the revision as a new version, for example, MyModel01, MyModel02, etc. Then, when your work in progress veers off in a bad direction, you can simply return to an earlier generation of your model.

Be careful with external links. Minimize the number of links you have to other workbook files. Limit yourself to links with external databases if you must link at all. Once you link to another workbook, it may be linked to yet another workbook, and you may be entering a web of interconnections that you do not understand. While speaking of data, make sure your data are in a legitimate number format. Data can be formatted as text when imported carelessly. Text will not perform numerically!

Do not use cell formatting (see Figure 11.17) to round numbers. They are not truly rounded, they are simply displayed without trailing decimals. Figure 11.18 shows a

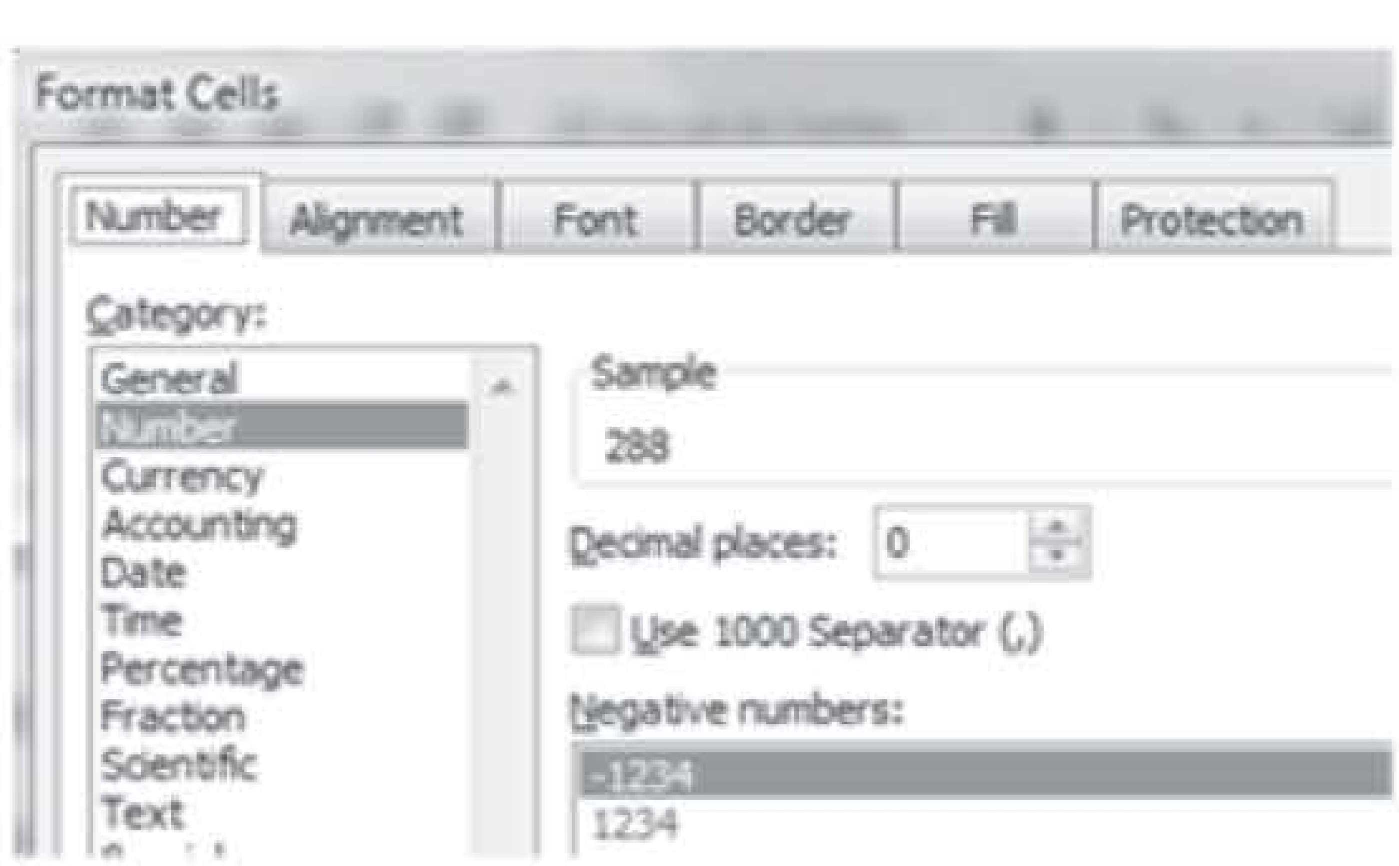


FIGURE 11.17 Truncating a value using cell format to mask decimals.

	K	L	M
15		Cell Format Formula	
16	Parameter	Value	Cell Formula in L
17	n	18	17.931
18	p	0.2	
19	Binomial	#NAME?	=riskbinomial(L16,L17)
20		Rounding the Value	
21	n	18	=ROUND(17.931,0)
22	p	0.2	
23	Binomial	4	=RiskBinomial(L21,L22)

FIGURE 11.18 Comparing the effect of cell format and rounding in a binomial distribution.

hypothetical binomial distribution. Values for n and p are given in column L. The exact syntax of the column L entry is given in column M. Thus, if you wanted to change the number 17.931 to the rounded-off integer, 18, use the ROUND function of your software. Cell formatting will display an 18 in the cell, but it will treat the cell value as 17.931. To illustrate this problem, imagine you want to use a number like 17.931 as an argument for the sample size, n, in a binomial distribution as shown in Figure 11.18. The binomial requires n to be an integer. In the top example, cell formatting is used for faux rounding. Notice it returns an error message (#NAME in cell L19) when entered into the binomial distribution. The second example uses the round function of Excel and enables the binomial function to work as expected.

Use the function wizard to minimize syntax errors. It can be found by clicking the f_x on the standard Excel toolbar. Figure 11.19 shows the function wizard listing all the @RISK probability distributions. Beneath the list is the Excel template that can be used to enter @RISK functions, as well as any of the Excel functions. Using the wizard will help minimize typos and inadvertent errors in formulas.

Be careful with parentheses in long formulas. It is common practice to use parentheses to establish the order of operations. It is easy to make mistakes with the placement of parentheses.

When you are building a model, it can be useful to "break your model" at times. To do this, you enter intentional error values into a cell. An example would be to enter a probability outside the acceptable range or to attempt to divide by zero. Once in a while error messages can be comforting. Your model should fail when it is supposed to. In a related fashion, it is useful to use simple values like 0.1, 1, 10, and other multiples of 10 to test your calculations. They need not be realistic values when you are building your model. Numbers like this make it easy to spot-check the accuracy of calculations.

11.5.2.2.8 Test and Audit the Model

Expect your spreadsheet model to contain errors. If you do, you'll make a more serious effort to find and fix those errors. Few things are worse for a modeler than to discover an error well after you've begun to use your model for decision analysis.

This step was called verification earlier in the chapter. Its purpose is to verify and ensure the model is free from logical errors. If your spreadsheet model is a simulation model, your goal is to ensure that every iteration of the simulation is both accurately

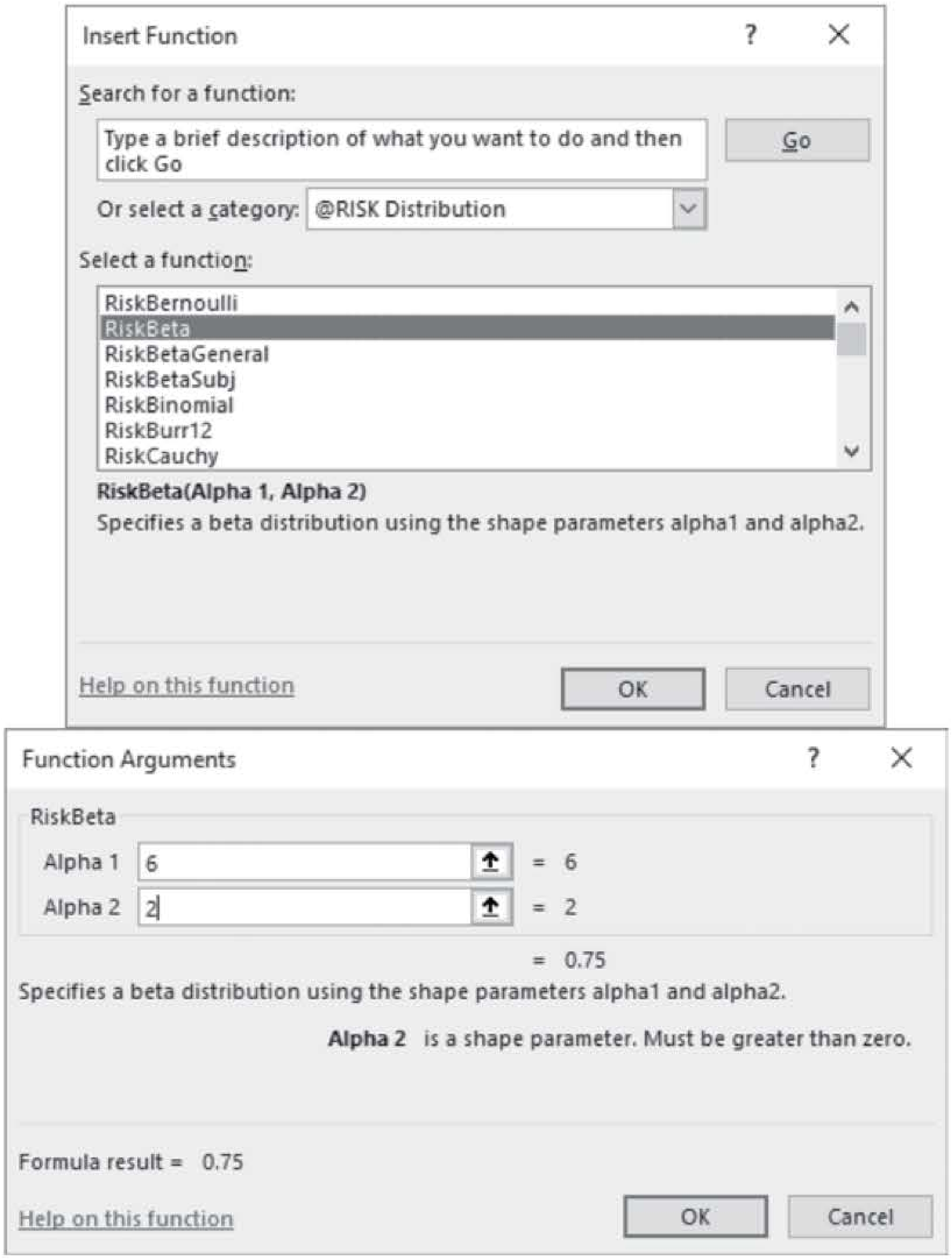


FIGURE 11.19 Excel's function wizard helps minimize typographical and syntax errors.

calculated and possible. This is when the care taken in designing your model for use and communication will first pay off.

Testing and verifying your model can be laborious and monotonous. It is all the more difficult to catch errors when your focus wanders. One of the best ways to test your model is to have experts review the model. This is a luxury few risk management activities can afford. It is also difficult to find these experts. They usually need software, modeling, and subject-matter expertise. This is hard to find in a single person. So, assuming you will have to verify your model, how can you best do that? Approach your model with a skeptical attitude. Powell and Baker (2007), once again, offer the single best description of a plan for testing your spreadsheet model. In the following discussion, I have supplemented their suggestions with those of the Fuqua School of Business at Duke University (2006), Mather (1999), and a few of my own.

11.5.2.2.9 Debugging

Debugging is a methodical process for finding and correcting errors in your model's logic and data. There is a rich literature on debugging, but the truth is the method of that process is highly individualized and often idiosyncratic. There is no simple or foolproof way to debug a model. The simplest advice is to build your model in small modules and debug each module carefully before you link them together. The most basic form of debugging is to methodically check and verify every line of code in your model, or in the case of a spreadsheet model, the contents of every cell.

As you link your models, check carefully to ensure that interactions between the modules do not create new bugs. Debugging, as used here, is usually an ongoing process rather than a discrete event in the life of a model. You are always debugging your model right up to the time that it is verified.

11.5.2.2.10 Make Sure Numerical Results Are Reasonable

As your model takes shape and numerical results (either intermediate calculations or outputs) begin to appear, or when it is initially completed and you begin verification, check the reasonableness of your numbers. Prepare rough estimates of what the values should be before you complete the calculation. Probabilities must lie between zero and one. We usually have or can estimate a range of plausible outcomes for a calculation. If that reasonable range is exceeded, verify the formula.

Check your calculations with a calculator. You may have a formula that functions exactly the way you programmed it, but if you programmed it incorrectly you may never catch the errors. Calculators can catch errors in formulas.

Sound logic in your model should provide logical results for realistic sets of values. One way to test this logic is to use the zero test, where zero is regarded as an extreme value. Set one or more variables or groups of variables to zero and make sure outputs are as they should be if this value is zero. In our *Vibrio* model, zero *Vibrio* at the time of harvest should produce zero *Vibrio* at the time of consumption given the structure of the model, which does not permit cross-contamination. Likewise, zero *Vibrio* should never result in a positive probability of illness. Thus, the zero test often provides a quick check of the model logic at this extreme value. When zero is not a reasonable extreme value, set the variables to known or reasonable extremes and make sure the intermediate calculations and outputs are as they should be.

Logic tests can be used to debug your model. Increase and then decrease the values of each input value one at a time to ensure that the changes to the outputs are in the logical direction. For example, as water temperature or air temperature increase in the *Vibrio* model, the number of bacteria should increase as well. They should not only increase but be a reasonable magnitude as well.

Excel has some useful auditing capabilities. Two of them are the Trace Precedents and Trace Dependents commands. They can be used to graphically display or trace the relationships between precedent and dependent cells and formulas with tracer arrows. This enables you to ensure that the proper relationships are in place. An example of Trace Precedents is shown for the *Vibrio* model in Figure 11.20. It shows all the cells that influence the probability of illness and the route of their influence. You need not understand the nature of those complex relationships from the figure,

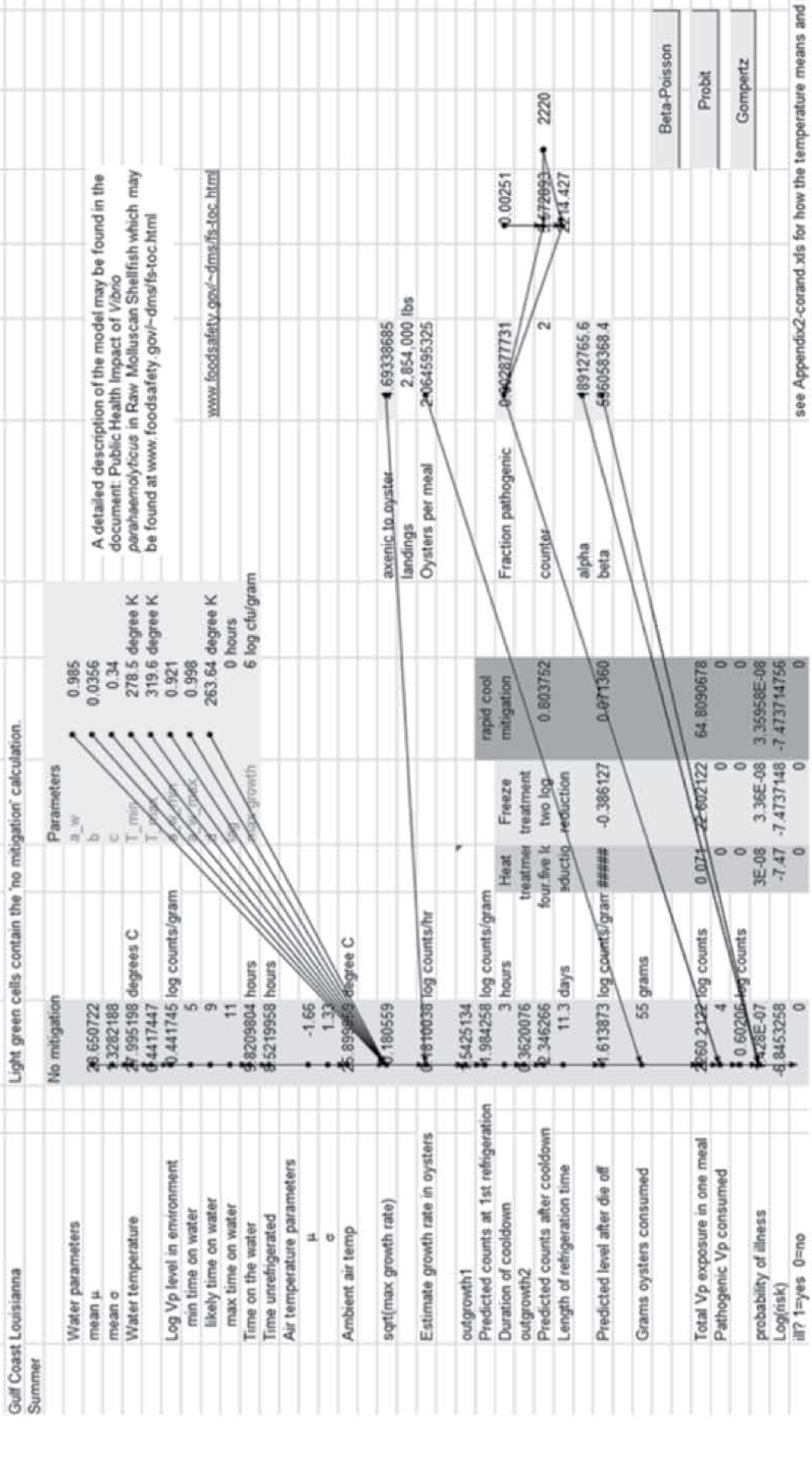


FIGURE 11.20 Trace precedents shows all the variables required to calculate the value in a single cell.

just appreciate the potential utility of the feature. Trace Dependents, as you might imagine, shows all the values that depend on a given cell. With a little practice, these tools can be very helpful in the verification process.

11.5.2.2.11 Check the Formulas

Perhaps this section should have been titled, "No Duh!" Formulas can be checked visually. This is the tedious and monotonous part of the verification. It is also the dangerous part, because if you built the model you may not be able to see your errors. Nonetheless, visual inspection of each cell formula is a minimal expectation for verification.

Excel has several nice features to aid this task. Putting your cursor in a model cell and pressing F2 or double-clicking in that cell will reveal the cell formula, and it will use a color-coded system to highlight every cell in the model that is referenced in that formula.

To reveal all the cell formulas at once, press Control ~. This is especially useful when cells in adjacent rows or columns have similar formulas. In that case, a break in the formula pattern may indicate an error.

Excel also offers an error-checking feature. If errors are found, each potential error is flagged by the appearance of a small triangle in the upper left corner of the cell. Excel's error checking can find eight different errors. Excel has an error-tracing feature as well. This feature most often points to the source of the error.

The Simulation Settings of Palisade @RISK offer a Pause on Output Error option that is very useful. If a logical error is found in any cell, the simulation will stop on that iteration, enabling you to find the cell with the error message. Excel's error-checking capabilities or your own trace back can then reveal the logical source of the error.

Excel offers a helpful Evaluate Formula option in the Formula Auditing group of the Formulas tab. To use it, position the cursor in the cell of interest and select Evaluate Formula. Evaluate, Step In, and Step Out options provide an excellent audit trail for checking formulas across a chain of cells. Using name ranges is a decided advantage for getting maximum value from this feature. If these built-in auditing features are not sufficient, there is a wide array of spreadsheet auditing software available on the Internet.

11.5.2.2.12 Test the Model Performance

If you cannot wait to begin to use your model before verifying your model—an ill-advised approach—examine your results as you would intend to do when using your model to answer the risk manager's question(s). Simply using the model may reveal some logic flaws, and unusual results may help you find bugs in the calculations. If your model is logical and accurate, your results should be also. Doing some sensitivity analysis may help uncover problems in the model. Note that this is not considered an adequate method for verifying your model. It is, however, sometimes helpful for finding gross errors in a model.

If you build models, you will spend a lifetime developing and refining your craft skills. Like model building itself, these can tend to be rather idiosyncratic. The ideas offered here are but the tip of the craft skills iceberg. But, if you actually follow them, they will advance you up the spreadsheet modeler's learning curve more quickly than your own trial-and-error methods will.

11.5.2.2.13 Documentation

Verification and validation are important steps in model development. They are often overlooked in the more casual application of risk analysis principles and in the less formal risk assessments. When a model has been verified and/or validated, the methods and time frame for accomplishing these tasks should be documented and appended to the model.

11.6 SUMMARY AND LOOK FORWARD

Not many risk assessors are trained as model builders. This is generally a skill that is picked up in practice by people with a quantitative orientation. As a result, the model builder is usually not the subject-matter expert. Model building is, consequently, often a group effort.

A 13-step modeling process is offered in this chapter to help those struggling with the notion of what it means to build a risk assessment model. Model building is iterative. It begins with a conceptual model, evolves to a specification model, and ends up as a computational model. Most model builders manage to muddle through those steps somehow, although it is rare to see the iterations as discrete and separate tasks. Verification, validation, and documentation are, in my experience, the most often overlooked tasks in model building.

The craft skills for model building are what experience teaches you. This chapter attempts to assist your development of craft skills by building on the eight heuristics offered by Powell and Baker (2007). Chapter 12 addresses one of the technical skills quantitative risk assessors must have: a basic understanding and command of the essentials of probability. That is followed by two more chapters on probability topics that will expand your technical skills while, hopefully, helping you to develop your craft skills. These are "Choosing Probability Distribution" (Chapter 13) and "Characterizing Uncertainty Through Expert Elicitation" (Chapter 14).

REFERENCES

Evans, J. R., and D. L. Olson. 2002. *Introduction to Simulation and Risk Analysis*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Food and Drug Administration. 2005. Center for Food Safety and Applied Nutrition. Quantitative risk assessment on the public health impact of pathogenic *Vibrio parahaemolyticus* in raw oysters. http://www.fda.gov/Food/ScienceResearch/ResearchAreas/RiskAssessmentSafetyAssessment/ucm050421.htm.

Fuqua School of Business. 2006. *Excel Design & Audit Tips*. Durham, NC: Duke University.

Hartman, J. M., J. W. Forsen, M. S. Wallace, and J. G. Neely. 2002. Tutorials in clinical research: Part IV: Recognizing and controlling bias. *Laryngoscope*, 112: 23–31.

Mather, D. 1999. A framework for building spreadsheet based decision models. *Journal of the Operational Research Society* 50: 70–74.

Powell, S. G., and K. R. Baker. 2007. *Management Science: The Art of Modeling with Spreadsheets*. 2nd ed. Hoboken, NJ: John Wiley and Sons.