

Microbial Database Building: What Have We Learned?

006475

Experience in developing predictive models for foodborne pathogens has resulted in improved design, efficiency of data collection, and precision

Richard C. Whiting

MICROBIAL MODELING PROGRAMS, PRINCIPALLY AT THE U.S. DEPT. of Agriculture and the United Kingdom's Ministry of Agriculture, Fisheries, and Food (MAFF), have created databases and calculated growth models for 10 pathogens, survival models for 4 pathogens, thermal-death-time models for 5 pathogens, and time-to-growth models for proteolytic and nonproteolytic types of *Clostridium botulinum*. We have learned that microbial behavior in a food is largely determined by 3–5 environmental factors, usually temperature, pH, NaCl level (water activity), and atmosphere.

Suitably accurate and precise predictions can be made with these models to rapidly and conveniently provide an initial estimate of microbial behavior for use in estimating microbial growth/survival in a food, guiding R&D efforts, developing Hazard Analysis Critical Control Point (HACCP) programs, assisting in planning laboratory experimentation, and educating nonmicrobiologists.

We have also learned the importance of the model developer's clearly stating the limitations of a particular model and the user's respecting those limits. Providing confidence limits about a prediction is necessary to convey the appropriate precision and reliance to the user. Factors controlling the lag phases need more research, and new growth models, such as the Baranyi model, that incorporate lag effects and fluctuating or variable conditions (e.g., temperature) are being developed.

In microbiology, distributions are critical. The mean value is frequently not as important as information about the fastest-growing strains, first spores to germinate, or the tailing of the resistant subpopulation of survivors during thermal inactivation. Modeling research has illuminated the need for knowledge of the microbial physiology that affects the observed behavior. Current modeling does not answer the question "Is this food/process is safe?" or "Will eating this food make me sick?" Modeling needs to be incorporated into a risk-assessment system that includes distribution of pathogens in raw materials, changes in pathogen populations during food manufacture, distribution, and preparation, and infectious dose models.

We have learned that modeling foodborne pathogens is laborious work. Multiple-factored models may need hundreds of treatment combinations and six months of lab work to become a single equation that runs on a personal computer in a fraction of a second.

Most important, we have learned that modeling works. The growth or survival of microorganisms in most foods was shown to be largely controlled by 3–5 environmental factors, and the microorganism's response in a food was similar to that in broth culture at corresponding levels of these factors (Ross and McMeekin, 1994). Although predictability is a fundamental assumption for all sciences, we have learned that microbial behavior is indeed predictable within error ranges which, in turn, are also quantifiable.

Using various mathematical relationships and statistical procedures, quantitative models have been developed that are useful to microbiologists in regulatory agencies, HACCP programs, and lab-

oratories. Usefulness reflects the compromises between a model's being complex enough to provide accurate predictions and being simple enough to have input parameters readily available to the user. A specific model may not be useful to everyone, it may not contain an important factor for the specific food of interest, the range of a factor may not include the value of interest, or the accuracy and precision may not be adequate.

Several reviews describe the development of this field and explain the various models that have been developed (McMeekin et al., 1993; Ross and McMeekin, 1994; Skinner et al., 1994; Whiting, 1995). Two modeling software programs currently exist, the Food MicroModel developed by MAFF and available from Leatherhead Food Research Association, Leatherhead, Surrey, U.K., and the Pathogen Modeling Program developed by USDA's Eastern Regional Research Center, Wyndmoor, Pa.

Conducting and Validating Modeling Research

Through experience, we have learned various ways to improve the design, efficiency of data collection, and precision of the resulting models. The present generation of models used a cocktail of bacterial strains, hoping to encompass within the model the characteristics of the strains most likely to be encountered in commercial products. This was a valid approach, but a better approach might be to first determine the growth rates of a number of outbreak strains at slow, medium, and fast growth conditions, then develop the model using one of the faster and more frequently occurring strains. The growth of other strains would then be related to the modeled strain. A risk-assessment model would factor the frequency of occurrence of each strain with its growth characteristics.

Models are usually developed in broths where compositions are easier to control than solids. Broths are more homogeneous and conveniently sampled. Repeated sampling of a single sample will generally result in more precise data points and better curve fits than will replicate packages that are sampled at different times. The number of data points necessary to fit a growth or inactivation model depends on the complexity of the model; however, 7–10 "quality" data points appropriately distributed are probably sufficient.

The design of our sampling plans is closer to a fractional factorial than a central composite design because the data are more variable toward the extremes of a factor's range than in the center. An adequate number of data points need to be collected at conditions where the microorganism's behavior is changing. After the data from the initial combinations of treatments are collected, additional combinations are frequently added at treatment levels of particular concern representing specific food products or at levels where the parameters are in transition. Replication of some treatment combinations is necessary to estimate the residual error, but additional combinations of the factor levels are better than extensive replications of single combinations. The range of the environmental factors must include values expected to be of interest to the model user. Extrapolating a prediction outside the range of the

data used to develop the model is very dangerous. Most of the polynomial-regression, secondary models show better compliance to the statistical requirement for normal variances when the logarithm or square root transformation of the primary parameter data is used.

Once a predictive model is calculated, the predictions must be carefully compared to the originating data set to determine if there are particular factor combinations which do not fit well or appear unreasonable. Additional data may need to be obtained and the model recalculated.

We have learned that it is critical that models be validated before confidence is placed in them. Validation means that the prediction is compared to inoculated-pack data. Comparisons of model predictions to inoculated-pack studies have shown that their predictive ability is good and usually somewhat conservative or "fail-safe" (McClure et al., 1994). However, it remains critical that the user be made aware of and respect the limitations of a specific model. Some foods will contain additional factors not included in the model that have a significant influence on the pathogen's behavior. Use of the model in this situation would be inappropriate, and the prediction would probably be inaccurate. A user should test a limited number of products or storage temperatures to ensure that the model is sufficiently accurate.

Uses of Models

Experience has shown that models can quickly provide information for making decisions in many situations:

- **Prediction of Risk.** Models can estimate the extent of growth or likelihood of survival of a pathogen after a period of normal or abuse storage, thereby highlighting problem products and processes. Growth models can assist in establishing "pull dates" by estimating the growth of likely pathogens. This information can be used in conjunction with information on the spoilage flora.

- **Quality Control.** Quantitative estimates at different levels of the environmental factors indicate the acceptable ranges and can aid in the development of HACCP programs. The setting of a critical control point (CCP) is currently based on subjective interpretations and experience. Modeling and risk assessment will increasingly be used to determine the CCP values. Models are also useful when an out-of-compliance event, such as an unexpected temperature rise, must be evaluated for microbial consequences. Decisions on whether to rework, utilize rapidly, or discard a food or ingredient can be made without waiting for microbial testing.

- **Product Development.** The microbial consequences from changes in the composition or processing can quickly be evaluated. New formulations can be compared to the new and old model predictions and to actual experiences with the old formulations.

- **Education.** Models can help explain microbiological behavior to nontechnical people. The generation of graphs or estimates of times to reach critical populations can dramatically illustrate the importance of CCPs or the importance of obtaining raw materials with low microbial counts. The models are useful in teaching food microbiology because of their ability to quickly illustrate the effects of environmental conditions on microbial behavior.

- **Data Analysis and Laboratory Planning.** Modeling techniques should become a routine tool for the description and analysis of microbial data even when the study is not intended to develop a model. Curve-fitting routines provide an unbiased quantified determination of growth rates. Standard deviations can be estimated and the parameter values compared and statistically tested. Laboratory efficiency is increased when models guide experimental design by suggesting treatment levels and appropriate sampling times.

Risk Assessment

The ultimate goal of modeling foodborne pathogens is to estimate the likelihood that consuming a particular food will make someone

sick (CAST, 1994). To model this risk, information is needed in four areas:

1. **Occurrence and Level of the Pathogen in the Starting Material.** Quantitative information on the initial numbers of pathogens in raw ingredients or at the beginning of the desired process to be modeled would typically be depicted as a histogram, representing the percentage of samples containing various concentrations of a microorganism. An example is the national survey for the presence of *Listeria monocytogenes* on the surface of beef carcasses, where 65% of the carcasses had less than 0.03 cfu/cm², 24% 0.03–0.3, 5% 0.301–3.0, 6% 3.001–30.0 (FSIS, 1994).

2. **Rates of Growth, Thermal Death, and Survival.** The next step is to estimate the changes in pathogen numbers during food processing, distribution, and final preparation. This involves linking a series of growth, survival, and cooking steps of various durations. Most previous efforts in modeling have estimated changes in microbial populations under constant conditions.

3. **Amount of Food Consumed.** The serving size, frequency of consumption, and other consumer practices affect the total number of viable pathogens that an individual is actually challenged with.

4. **Infectious Dose.** The infectious dose depends on the virulence of the specific strain when consumed, the food matrix, and the susceptibility of the individual. The overall process involves infection, morbidity, and mortality. These models are discussed by Haas et al. (1997).

Information is needed on the consequence of ingesting a specific number of pathogens. Currently, it is believed that there is no threshold dose, that one viable pathogen has a definable probability of causing illness. Models developed for waterborne pathogens show a sigmoidal relationship between number of *Shigella* and the probability of infection. However, probabilities of only 1% ($P = 0.01$) are unacceptably high. Current estimates of 25 million cases of illness in the United States from foodborne microorganisms per year would calculate to be a probability of 10^{-1} per year or about 10^{-4} per meal (CAST, 1994).

Prototype Risk-Assessment Model

A simple model was written to estimate the probability of an infection from *Salmonella* in a cooked poultry patty and demonstrates the kinds of information needed in a risk-assessment model. The model's steps are the initial populations, storage, cooking, consumption, and infectious dose (Fig. 1). The initial distribution is adapted from data (inset) presented by Surkiewicz et al. (1969), where 3.5% of the samples have high populations (> 0.44 cfu/g). The growth model uses the exponential growth rate calculated by the *Salmonella* model of Gibson et al. (1988). Only the temperature and storage time are input; in this example, 21°C and 5 hr. The thermal death model was determined from published D values for *Salmonella* in eggs (Anellis et al., 1954). Input is for temperature (60°C) and time (6 min). The amount of food consumed (100 g) represents a typical serving. The infectious dose model for *Salmonella* is an exponential function with a parameter value of 0.00752 (Rose, 1994). For this model, one *Salmonella* has the probability of 0.007 of being an infectious dose.

Most steps have an uncertainty about the calculated mean value. Some uncertainties may have a definable distribution such as a normal distribution with its standard deviation. Others will not, such as the histogram describing the initial populations. Because of the complexities of the complete model and the discrete distributions, simulation modeling techniques are used to determine the occurrence pattern of various probabilities of infection. In this method, the entire model is recalculated (iterated) many times. As the calculation advances through the steps and it encounters a variation or distribution, a value is randomly selected according to that distribution. This means that most iterations will cluster near the mean, but a few iterations will start with higher numbers and

progress through faster growth and slower death rates to give a higher probability of infection. Conversely, some iterations will result in a lower probability. After 1,000 iterations, a frequency pattern results for the probabilities of infection for the process being modeled.

For this demonstration model, the standard deviation was estimated to be 0.2 log units for the growth period and 0.1D for the cooking. The calculated median value for each step is given in Fig. 1, and the mean probability that this process is an infectious dose is $10^{-4.6}$. This appears to be a relatively low-risk process, however, when this process is iterated and the distribution of infectious doses is observed, a tailing toward higher risk is apparent. About 3% of the iterations exceeded a risk of 10^{-3} , clearly an unacceptable situation.

Examination of the individual iterations showed this tailing to originate from the relatively small number of initial samples with high *Salmonella* contamination. Rerunning the simulations without these samples reduced the median risk only to $10^{-4.7}$. However, the high-probability tailing is gone, and none of the iterations now exceed 10^{-3} . If this situation were still judged to be unacceptable, or the samples with high initial numbers could not be avoided, the cooking temperature could be raised, e.g., to 61°C , and the model calculates that the increased thermal death decreases the median probability to $10^{-7.4}$. This probably would be considered a "safe" process, although there is no public consensus on what constitutes an acceptable risk from foodborne pathogens.

This risk model can be expanded by incorporating distributions for storage times, cooking temperatures, or amounts of meat consumed. Additional processing steps and storage periods could be added in a modular manner. This type of model can quickly show which factors have a major influence on the final probability of infection and where process control must be maintained. In this example, a change of only a degree or two in the cooking temperature of a process designed to achieve a 4-log reduction can greatly influence the probability of infection.

Limitations of Current Modeling

The variation about both the individual steps as well as the final risk assessment is of paramount importance in evaluating the risk. We are mostly interested in evaluating the marginal situations, i.e., processes that are not overtly dangerous but exhibit potential to cause illness after one in 1,000 or one in 100,000 consumptions. Knowing the average value for a growth rate or any of the other steps is not sufficient to make this evaluation. The dangerous situations will arise principally from the outliers. The few samples with high contamination, the first pathogenic cells to adjust to a new environment and begin to grow, or the subpopulation with the greatest thermal resistance is most likely where the outbreaks will arise.

Even for our current one-step models, an estimation of the variance would greatly assist in evaluating the prediction. We frequently want the models to estimate behavior at conditions of slow growth or extended survival times. This is where the variation is greatest and the model user needs to be aware of its magnitude. An

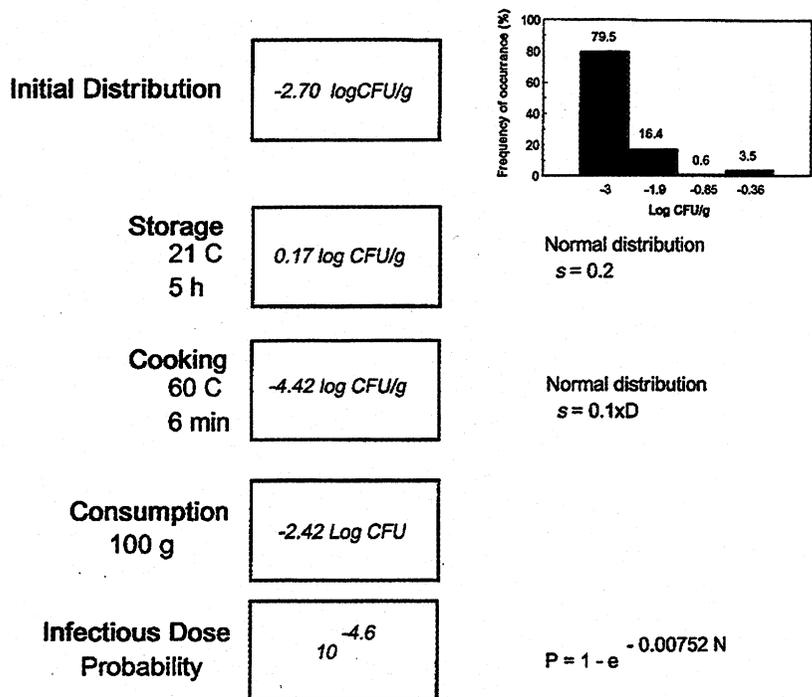


Fig. 1—Risk-assessment model for *Salmonella* in a cooked poultry patty

example is the model for *Staphylococcus aureus* survival that ranges from less than one day to 5 months, depending on the environmental conditions (Whiting et al., 1996). In broth at 25°C and pH 5.0, for example, the time for one log decline is 62 days, with lower and upper confidence intervals at 29 and 133 days, respectively. At pH 4.0, the times are 11, 7, and 19 days, respectively.

Many studies of the occurrences of pathogens in raw materials or final products in the marketplace report only the presence or absence of the pathogen, as the percentage of positive samples. However, for risk analysis when pathogens are present, we need to know how many of them are present in the various samples. This can be illustrated as an unstructured frequency histogram, or it may closely resemble a normal curve with a mean of $4.0 \log \text{cfu/cm}^2$ and a standard deviation of $1.9 \log \text{cfu/cm}^2$ that was found for *Pseudomonas* on retail pork carcasses (Coates et al., 1995).

Spoilage organisms have not received much attention for development of comprehensive models, although much information exists in the literature and in the food industry's experience for spoilage of specific products. Often, the safety of semipreserved foods becomes a race between spoilage and pathogenic microbes. Whenever possible, we want the spoilage flora to produce off-odors and appearances before the pathogens can grow or produce enough toxin to cause illness. It would be desirable to model both spoilage and pathogenic microorganisms through the processing and storage periods to the consumer's table. This approach would not be appropriate for pathogens, such as *Escherichia coli* O157:H7 or *Campylobacter jejuni*, which are infectious in low numbers; the concern is about their presence and survival, rather than growth.

Growth, death, and survival models are the most developed aspect of the various components of the risk-assessment process. However, major shortcomings exist in two areas. The first is the previously discussed determination of variation. We are in the process of recalculating our existing models and adding confidence limits to the predictions. This will give users a much better idea of how to evaluate the models' estimates. However, this only esti-

mates the variation in the data set used to create the specific model. Additional variation exists between strains that we have not incorporated into existing models. Model developers usually use a cocktail of 3–5 “typical” strains. In effect, the models made from a cocktail represent the fastest-growing or most-resistant strain in the cocktail. This has the advantage of making the models conservative or “fail-safe.” But further information is needed about the natural diversity in growth rates, D values, or survival times of strains of all the pathogens.

Information such as that for 39 strains of *Listeria monocytogenes* will increase the confidence in the predictions (Barbosa et al., 1994). The average lag phase duration at 4°C was 150.9 hr with a standard deviation of 29.5 hr, and the generation time was 43.1 ± 10.7 hr. A more-sophisticated risk model could link the parameter values of a strain with the frequency of occurrence of that strain. Otherwise, it is difficult to evaluate the importance of a strain such as *Salmonella senftenberg* 775W, which has a D value approximately five times longer than other *Salmonella* strains, but has not caused a foodborne outbreak. Basing a model on *S. senftenberg* 775W, which would be the effect if it were part of a cocktail, would be excessively conservative.

The other source of variation not incorporated into current models is the physiological state and prior history of the cell. Most of the current models use inocula grown for 18–24 hr at favorable temperatures (37°C) in nutritious broths. Hudson (1993) has shown that the lag phase duration of *Aeromonas hydrophila* depends on the temperature the cells came from as well as the new temperature. Growth rates after the lag phase ended were not affected by the previous temperatures. A pathogen adapted to a refrigerated temperature in a meat processing plant at 10°C would have a much shorter lag time and earlier beginning of the exponential growth phase after accidentally contaminating a piece of refrigerated meat than would be predicted by the current model—a “fail-dangerous” situation. The model by Baranyi and Roberts (1994) considers the apparent lag phase a consequence of two processes, the first reflecting the prior state of the cell (q_0) and the other the rate of the adjustment from one environment to the next (v).

Heat shock and other stress proteins, lag vs exponential growth phases, and starved, injured, or biofilm cells are all biochemical and physiological phenomena that affect the length of the lag phase and need to be incorporated into models. A current objective of modeling is to combine growth, death, and survival models through a series of process steps or fluctuating storage temperatures to simulate a food product from manufacture to consumer. It is critical to know the actual extent of the lag phase, if present at all, when individual steps and different models are sequentially linked. In many commercial situations where the food is in bulk and packaged, the microorganisms undoubtedly adapt their growth rates as the temperature changes, and no lag period reoccurs.

Other microbial situations that need modeling are growth in heterogeneous foods, on surfaces or boundaries, in microenvironments, and in biofilms. Most models use percentage of NaCl as a factor, and generalization to water activity and other salts and nonionic humectants is needed. The suppression of growth by sodium lactate and other buffered organic acids show that the concentrations of acid anion (or undissociated acid) can be a significant factor for growth or survival.

Current models do not factor competition from other microorganisms. In most situations, this probably doesn't become a significant factor until relatively high populations are achieved, above the numbers of importance for pathogens. However, high numbers of spoilage flora could suppress growth of low numbers of pathogens during the period just before organoleptic unacceptability is reached. *Staphylococcus aureus* is a pathogen traditionally consid-

ered to be inhibited by competition. Two other situations where other microorganisms affect the pathogens would be the lowering of pH from formation of lactic acid and production of bacteriocins in fermented foods.

Powerful, But Not a Substitute

Essentially, modeling is a technique to quantitatively analyze microbial behavior. It provides a more accurate prediction in complex situations than previous subjective methods. However, the usefulness and accuracy of models are critically dependent on the quality of the data that go into each step. The expectations for a model must be within the limitations of the data that went into its creation. Great progress has been made in modeling growth and inactivation, but much still remains to be done. There is a lack of appropriate quantitative data for most of the steps in the risk-assessment process at present.

Models can be a powerful tool for microbiologists, quickly providing an initial estimate of a microorganism's behavior. However, they are only one of several sources of information. They do not replace the experience and judgment of a trained microbiologist.

References

- Anellis, A., Lubas, J., and Rayman, M.M. 1954. Heat resistance in liquid eggs of some strains of the genus *Salmonella*. *Food Res.* 19: 377-395.
- Baranyi, J. and Roberts, T.A. 1994. A dynamic approach to predicting bacterial growth in food. *Intl. J. Food Microbiol.* 23: 277-294.
- Barbosa, W.B., Cabedo, L., Wederquist, H.J., Sofos, J.N., and Schmidt, G.R. 1994. Growth variation among species and strains of *Listeria* in culture broth. *J. Food Protect.* 57: 765-769, 775.
- CAST. 1994. Foodborne pathogens: Risks and consequences. Task Force Rept. 122. Council for Agricultural Science and Technology, Ames, Iowa.
- Coates, K.J., Beattie, J.C., Morgan, I.R., and Widders, P.R. 1995. The contribution of carcass contamination and the boning process to microbial spoilage of aerobically stored pork. *Food Microbiol.* 12: 49-54.
- FSIS. 1994. Nationwide Beef Microbiological Baseline Data Collection Program: Steers and Heifers, 1992-1993. Food Safety and Inspection Service, U.S. Dept. of Agriculture, Washington, D.C.
- Gibson, A.M., Bratchell, N., and Roberts, T.A. 1988. Predicting microbial growth: Growth responses of salmonellae in a laboratory medium as affected by pH, sodium chloride, and storage temperature. *Intl. J. Food Microbiol.* 6: 155-178.
- Haas, C.N. 1983. Estimation of risk due to low doses of microorganisms: A comparison of alternative methodologies. *Am. J. Epidemiol.* 118: 573-582.
- Haas, C.N., Rose, J.B., Gerba, C.P., and Crockett, C.S. 1997. Risk assessment: What food microbiology can learn from water microbiology. *Food Technol.* 51(4): 91.
- Hudson, J.A. 1993. Effect of pre-incubation temperature on the lag time of *Aeromonas hydrophila*. *Lett. Appl. Microbiol.* 16: 274-276.
- McClure, P.J., Blackburn, C. de W., Cole, M.B., Curtis, P.S., Jones, J.E., Legan, J.D., Ogden, I.D., Peck, M.W., Roberts, T.A., Sutherland, J.P., and Walker, S.J. 1994. Modelling the growth, survival and death of microorganisms in foods: The UK Food Micromodel approach. *Intl. J. Food Microbiol.* 23: 265-275.
- McMeekin, T.A., Olley, J., Ross, T., and Ratkowsky, D.A. 1993. “Predictive Microbiology: Theory and Application.” John Wiley, New York.
- Rose, J.B. 1994. Personal communication, Dept. of Marine Sciences, Univ. of Southern Florida, St. Petersburg.
- Ross, T. and McMeekin, T.A. 1994. Predictive microbiology. *Intl. J. Food Microbiol.* 23: 241-264.
- Skinner, G.E., Larkin, J.W., and Rhodehamel, E.J. 1994. Mathematical modeling of microbial growth: A review. *J. Food Safety* 14: 175-217.
- Surkiewicz, B.F., Johnston, R.W., Moran, A.E., and Krumm, G.W. 1969. A bacteriological survey of chicken eviscerating plants. *Food Technol.* 23: 1066-1069.
- Whiting, R.C. 1995. Microbial modeling in foods. *CRC Crit. Rev. Food Sci. Nutr.* 35: 467-494.
- Whiting, R.C., Sackitey, S., Calderone, S., Morley, C., and Phillips, J.G. 1996. Model for the survival of *Staphylococcus aureus* in nongrowth environments. *Intl. J. Food Microbiol.* 31: 231-243.