

Considering uncertainty in comparing the burden of illness due to foodborne microbial pathogens[☆]

Mark Powell^{a,*}, Eric Ebel^b, Wayne Schlosser^c

^a Office of Risk Assessment and Cost Benefit Analysis, U.S. Department of Agriculture, SW 5248 S. Ag. Bldg., 1400 Independence Ave., Washington, DC 20250, USA

^b USDA Food Safety and Inspection Service, Ft. Collins, CO, USA

^c USDA Food Safety and Inspection Service, College Station, TX, USA

Received 7 October 2000; received in revised form 2 February 2001; accepted 20 February 2001

Abstract

The uncertainty attendant to burden-of-illness estimates should be taken into account in comparing the public health impact of different foodborne pathogens. In this paper, decision analysis concepts are applied to the comparisons of pathogen-specific burden-of-illness estimates. In situations wherein the magnitude of uncertainty varies, the rank order of pathogen-specific burden-of-illness estimates is sensitive to the decisional criteria applied. To illustrate the magnitude of attendant uncertainty in pathogen-specific foodborne-illness estimates, probabilistic risk assessment methods are used to characterize the uncertainty regarding the burden of illness due to *Escherichia coli* O157:H7. The magnitude of uncertainty about the burden of food-related illness due to *E. coli* O157:H7 is substantial, ranging from less than 50,000 to more than 120,000 cases/year. This example underscores the importance of considering the uncertainty attendant to burden-of-illness estimates in comparing the public health impacts of different pathogens. Although some would argue that the expected value of the number of illnesses provides the “best estimate” for decision-making, this merely reflects a decision-making rule of convention and not a scientific truism. Published by Elsevier Science B.V.

Keywords: Comparative risk assessment; Uncertainty; *E. coli* O157:H7

1. Introduction

Comparative risk is defined as any comparison among the risks of two or more hazards with respect to a common scale (Society for Risk Analysis, 2001). Comparative risk assessment is an analytic-deliberative process used to systematically measure, characterize, compare, and rank different hazards to inform value-laden risk management decision-making under scientific uncertainty (National Research Council, 1996).

[☆] Disclaimer: The opinions expressed herein are the views of the authors and do not necessarily reflect the official policy or position of the U.S. Department of Agriculture. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply the endorsement, recommendation, or favorance by the United States government.

* Corresponding author. Tel.: +1-202-720-9786; fax: +1-202-720-4240.

E-mail address: mpowell@oce.usda.gov (M. Powell).

Recently, Mead et al. (1999) undertook the enormous and commendable task of estimating the overall burden of foodborne disease in the United States. The burden of illness due to foodborne pathogens was quantified in terms of the annual number of illnesses, hospitalizations, and deaths. From the viewpoint of risk analysts concerned with food safety regulatory decision-making, such data and analysis have manifold uses. For example, the U.S. Department of Agriculture Food Safety and Inspection Service is currently employing much of the data underlying the estimated burden of illness due to *Escherichia coli* O157:H7, as well as other epidemiologic data, in developing a farm-to-table process risk model for *E. coli* O157:H7 in ground beef.

The pathogen-specific estimates of foodborne disease are also particularly useful to consider in making resource allocation decisions and setting the risk assessment agenda. However, the priority assigned to a pathogen depends upon a number of factors, including the metric used to estimate public health impact. Table 1 was derived from Mead et al. (1999) and lists the top three-ranked pathogens for the annual number of foodborne illnesses, hospitalizations, and deaths. Note that the pathogen rankings in Table 1 vary depending on health outcome measure. Consequently, using a common metric such as Quality Adjusted Life Years (QALY's) or Disability Adjusted Life Years (DALY's) is conceptually appealing for the purpose of comparing food safety risks. (QALY's and DALY's combine duration of life (quantitative dimension) and health-related quality of life (qualitative dimension) into a single public health measure.) There are certainly other relevant factors, however, that need to be taken into account in setting public health priorities and regulatory agendas in-

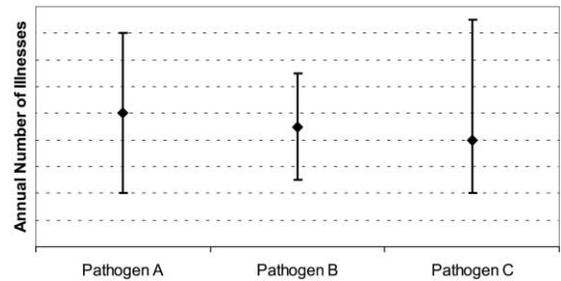


Fig. 1. A comparison of uncertainties of three pathogens.

cluding the feasibility and cost-effectiveness of preventive measures, legal authorities under the operative statutes, and the economic costs associated with the burden of illness (Buzby and Roberts, 1997).

Another factor that needs to be considered in comparing risks from different sources is the uncertainty attendant to the risk estimates. Comparing risks on the basis of point estimates may omit information relevant to the decision. Consider the hypothetical scenario shown in Fig. 1, wherein the minimum, most likely, and maximum estimates of the annual number of illnesses are presented for three pathogens. The pathogen assigned the highest priority depends on the decisional criterion employed. Pathogen A maximizes the most likely estimate of the number of illnesses. Pathogen B maximizes the lower-bound estimate. Pathogen C maximizes the upper-bound estimate. Although some may argue that the expected value of the number of illnesses provides the “best estimate” for decision-making, this merely represents a conventional rule and not a scientific truism.

A specific example illustrates how the application of probabilistic methods to available epidemiologic

Table 1

Top-ranked foodborne pathogens in terms of illness, hospitalization, and death

Note: percentage of total foodborne illnesses attributed to the pathogen is presented in parentheses.

Pathogen	Illness	Hospitalization	Death
<i>Salmonella</i> (non-typhoidal)	3 (10%)	2 (26%)	1 (31%)
<i>Listeria monocytogenes</i>			2 (28%)
<i>Campylobacter</i>	2 (14%)	3 (17%)	
Norwalk-like viruses	1 (67%)	1 (33%)	
<i>Toxoplasma gondi</i>			3 (21%)

data can readily generate a more robust estimate of the burden of illness due to a particular pathogen with attendant uncertainty. In the following example, FoodNet surveillance and other epidemiologic data are used to characterize the uncertainty about the annual number of cases of *E. coli* O157:H7 in the US.

2. Methods

The analysis begins with the reported population-based rate in the five original FoodNet catchment areas during 1996–1998. The crude average rate across the five sites during 1996–1998 ranged from 2.3 to 2.8 cases/100,000 person-years (Centers for Disease Control and Prevention, 1999a). A cluster-weighted average was used to extrapolate to the national level (i.e., the rate reported in each of the sites was weighted by the population of the state in which the site is located). Consequently, the rate reported from the catchment area in a large state like California is given greater weight than the rate in a state like Minnesota, which has a smaller population. It is appropriate to use a cluster-weighted average to extrapolate to the national level, but doing so does not necessarily imply a correction for a hypothesized “northern-tier” effect (i.e., that the rate of illness from *E. coli* O157:H7 is correlated with northern latitudes). It seems equally plausible, for example, to hypothesize a state-level effect in the active surveillance system. In comparison to the crude rates, the population-based rates of reported *E. coli* O157:H7 illness for the five original FoodNet sentinel sites for 1996–1998 weighted by cluster (state) size are estimated to be 2.04, 1.25, and 1.51/100,000 person-years (Hedberg et al., 1997; U.S. Census Bureau, 2000a; U.S. Department of Agriculture, 1997, 1998, 1999). (Preliminary FoodNet data for 1999 have been reported indicating that the crude average rate across the five original sites during 1999 was 2.1 cases/100,000 person-years (Centers for Disease Control and Prevention, 2000), but the summary data presented are insufficiently disaggregated to update the present analysis.)

To represent the annual variability in the number of reported cases, the cluster-weighted rates from the 1996–1998 surveillance (2.04, 1.25, and 1.51

cases/100,000 person-years, respectively) are placed into a discrete uniform distribution:

Population – Weighted Rate of O157 per 100,000 person-years (unadjusted for underreporting)
 \sim Discrete Uniform (2.04, 1.25, 1.51)

As statistical simulation is performed, with each iteration of the model, the rates are drawn at random with equal probability from this distribution (Vose, 1996). To extrapolate to the national level, we multiply this distribution by the estimated US population in 1998, 269.4 million (U.S. Census Bureau, 2000b).

Cases of O157 per year in the US
(unadjusted for underreporting)

$\sim 2694 * \text{Discrete Uniform} (2.04, 1.25, 1.51)$

Next, these estimates need to be adjusted for recognized sources of underreporting because some ill persons do not seek medical care, physicians do not obtain stool specimens from all patients, laboratories do not culture all stool samples for *E. coli* O157:H7, and some proportion of the lab results are false negatives. With the exception of test sensitivity, each of these proportions is treated as dependent on whether the infected person presents with (exhibits symptoms of) bloody or non-bloody diarrhea. Therefore, the proportion of bloody and non-bloody diarrheal cases must be estimated first. The uncertainty about the proportions of cases at each node in the pathway leading to a reported case is then characterized. These proportions feed into a sequence of negative binomial distributions used to estimate the number of cases missed at each step. The two resultant uncertainty distributions about the number of diarrheal cases (both bloody and non-bloody) are summed to estimate the total annual number of cases in the US. For the severe cases—defined as bloody diarrhea where the ill person seeks medical care—the progression of the illness to more severe health outcomes, such as hospitalization, hemolytic uremic syndrome and thrombotic thrombocytopenic purpura (HUS/TTP) or death, are estimated subsequently.

We proceed to determine the prevalence of reported bloody and non-bloody diarrheal cases by observing that 409 out of 480 reported cases presented with bloody diarrhea (Hedberg et al., 1997; Ostroff et al., 1989). The beta distribution is the conjugate prior to the binomial probability distribu-

tion and is therefore often used to describe uncertainty about proportions (Vose, 1996). Therefore, these data provide the parameters for beta distributions that characterize the uncertainty about these proportions, with corresponding point estimates of 85% and 15%, respectively:

$$P(\text{Bloody Case Reported}) \\ \sim \text{Beta}(409 + 1, 480 - 409 + 1)$$

$$P(\text{Non-Bloody Case Reported}) \\ \sim 1 - \text{Beta}(409 + 1, 480 - 409 + 1)$$

In conjunction with the active surveillance system, FoodNet has conducted a number of companion surveys to estimate the degree of underreporting in the sentinel sites. The FoodNet Laboratory survey found that 182 of 230 (79%) labs reported testing bloody stool for *E. coli* O157:H7, but only 108 of 230 (47%) labs reported testing all stool samples for *E. coli* O157:H7 (Hedberg et al., 1997). These data feed into beta distributions characterizing the uncertainty about the proportion of labs that culture for *E. coli* O157:H7 in bloody and non-bloody stool specimens, respectively:

$$P(\text{Laboratory Cultures Bloody Stool Sample for O157}) \\ \sim \text{Beta}(182 + 1, 230 - 182 + 1)$$

$$P(\text{Laboratory Cultures Non-Bloody Stool Sample for O157}) \\ \sim \text{Beta}(108 + 1, 230 - 108 + 1)$$

The sensitivity of the sorbitol MacConkey agar (or SMAC) test used by the labs to identify *E. coli* O157:H7 in stool samples is assumed to be 75% (Hedberg et al., 1997):

$$P(+ve \text{ SMAC test} | \text{sample} +ve) = 0.75$$

In a survey conducted in the FoodNet catchment area, 1515 of 1943 responding physicians (78%) reported that they obtained stool specimens from patients presenting with bloody diarrhea and 699 of 1943 physicians (36%) reported obtaining specimens from patients with non-bloody diarrhea (Hedberg et al., 1997). These data feed into a beta distribution characterizing the uncertainty about the proportion of physicians that obtain specimens from patients presenting with bloody and non-bloody stool symptoms, respectively:

$$P(\text{Physician Obtains Culture from Patient with Bloody Diarrhea}) \\ \sim \text{Beta}(1515 + 1, 1943 - 1515 + 1)$$

$$P(\text{Physician Obtains Culture from Patient with Non-Bloody Diarrhea}) \\ \sim \text{Beta}(699 + 1, 1943 - 699 + 1)$$

Regarding the proportion of ill-seeking medical care, Cieslak et al. (1997) found that 32 of 58 (55%) bloody diarrhea cases in an *E. coli* O157:H7 outbreak in Las Vegas reported seeking medical care. These data are used to characterize the uncertainty about the proportion of bloody diarrheal cases seeking medical care. For the non-bloody diarrheal cases, we use the results of a FoodNet population survey in which 88 of 1100 respondents (8%), who had a bout of diarrhea, reported seeking medical attention (Hedberg et al., 1997). These data feed into beta distributions characterizing the uncertainty about the proportion of patients with and without bloody diarrhea that seek medical care:

$$P(\text{Patient with Bloody Diarrhea Seeks Medical Care}) \\ \sim \text{Beta}(32 + 1, 58 - 32 + 1)$$

$$P(\text{Patient with Non-Bloody Diarrhea Seeks Medical Care}) \\ \sim \text{Beta}(88 + 1, 1100 - 88 + 1)$$

From this point, the negative binomial distribution (Vose, 1996) is employed in stepwise fashion to add the number of cases that are missed by the surveillance system due to test insensitivity, laboratories not culturing stool samples for *E. coli* O157:H7, physicians not obtaining stool specimens from patients, and ill persons not seeking medical care:

$$\text{Total No. cases } (N) = \text{No. cases reported } (s) \\ + \text{Negative Binomial } (s, p)$$

where p = probability of positive test, lab culturing stool for O157, physician obtaining specimen, or ill person seeking care.

This procedure presumes independence among cases, an assumption that may be violated, for example, if behaviors change in response to heightened awareness during publicized outbreaks. The survey data used to characterize the uncertainty about the conditional probabilities, however, reflect a combination of sporadic and outbreak conditions.

Proceeding from the estimated number of severe cases (a subset of bloody diarrheal cases for which the patient seeks medical care), we can characterize the uncertainty regarding the proportion of such cases that progress to more severe health outcomes—hospitalization, HUS/TTP, and death. For 203 outbreaks occurring during 1982–1998, for which preliminary disaggregated health outcome data are currently available (i.e., outcomes are broken down into four categories: affected, hospitalized,

HUS/TTP, and death), 968 out of 4478 cases (21.6%) resulted in hospitalization, 228 out of 4478 cases (5.1%) progressed to HUS/TTP, and 28 out of 4478 cases (0.6%) resulted in death (Centers for Disease Control and Prevention, 1999b). These data feed into beta distributions characterizing the uncertainty about the attack rates:

Proportion of severe cases resulting in hospitalization

$$\sim \text{Beta}(968 + 1, 4478 - 968 + 1)$$

Proportion of severe cases progressing to HUS/TTP

$$\sim \text{Beta}(228 + 1, 4478 - 228 + 1)$$

Proportion of severe cases resulting in death

$$\sim \text{Beta}(28 + 1, 4478 - 28 + 1)$$

Applying these attack rates to all cases involves an assumption that the severity of outbreak and sporadic *E. coli* O157 strains are similar.

Using the data described above, the number of cases in each category is estimated using Monte Carlo simulation methods (Vose, 1996). Monte Carlo simulation procedures were performed with Latin Hypercube sampling (10,000 iterations) using Palisades®@Risk™ (Ver. 3.5.2), an add-on to Microsoft® Excel™ (*97).

3. Results

Table 2 summarizes the results of Monte Carlo simulation characterizing the uncertainty about the

Table 2

Results of epidemiologic Monte Carlo simulation—*E. coli* O157:H7 from all sources

Note: To avoid a misleading impression of precision, the values in this table were rounded off to two significant figures, with the exception of the estimated number of deaths, which were rounded off to the nearest increments of 5.

	Median	95% Credible interval
<i>Total results</i>		
Total annual cases O157 US	75,000	50,000–120,000
Total bloody	14,000	9600–22,000
Total non-bloody	60,000	38,000–100,000
<i>Severe results</i>		
Severe (bloody, seeks care)	7500	5800–11,000
Hospitalized	1600	1200–2400
HUS/TTP	380	280–590
Deaths	50	25–90

number of cases of *E. coli* O157:H7 annually in the US.

4. Discussion

Note that there is considerable uncertainty regarding the total number of cases of *E. coli* O157:H7. Also, the upper- and lower bounds of the credible intervals for the estimated number of severe cases, hospitalizations, and HUS/TTP differ by about a factor of 2. The 95% credible interval for the estimated annual number of deaths due to *E. coli* O157:H7 from all sources is approximately 25–90. (The 95% credible interval is the span between the 2.5th and 97.5th percentiles of the Monte Carlo simulation output distribution.)

In addition to allowing us to avoid misleadingly precise point estimates for the annual incidence of diseases such as *E. coli* O157:H7, the methods of probabilistic risk assessment can also help us to explore the data to better understand which of the many factors contribute most to the overall uncertainty. Sensitivity analysis can help guide our thinking about where to invest in additional epidemiological data acquisition to most effectively improve the precision in our overall estimates. Fig. 2 presents, in descending order, the rank correlations of various factors in the model with the total number of cases of *E. coli* O157:H7. The pattern that emerges is that if we seek to reduce our uncertainty in the overall number of cases, we should focus on enhancing the data on the non-bloody cases, beginning with those that do not seek medical care. Toward the other end of the range, it seems that if we seek to have a more precise estimate of the overall number of cases, then we may gain relatively little from decreasing our uncertainty about the proportion of *E. coli* O157:H7 cases that are bloody.

As is typically the case, however, the results of the sensitivity analysis depend upon what question we are trying to answer. If rather than trying to reduce the uncertainty in the overall number of cases, we seek instead to improve the estimate of the number of deaths due to *E. coli* O157:H7, then the pattern that emerges from the sensitivity analysis results, presented in Fig. 3, is that we are keenly interested in improving our knowledge about the

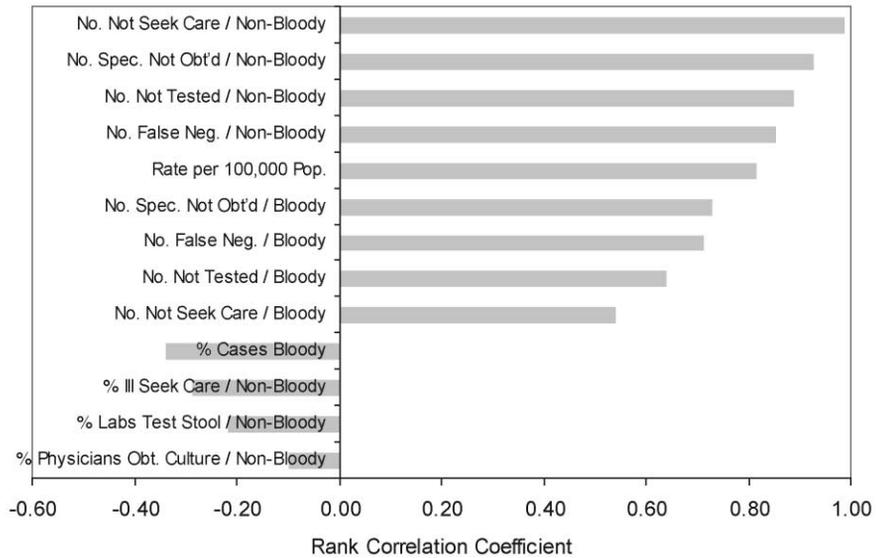


Fig. 2. Rank correlations for the total number of cases.

disposition of the bloody cases and about the overall rate of *E. coli* O157:H7 in the population.

In summary, this paper attempted to motivate the importance in a decision-theoretic framework of taking uncertainty into account in comparing the public health impact of different foodborne pathogens. The case of *E. coli* O157:H7 is developed to provide an illustrative example of how to

characterize and analyze the uncertainty attendant to pathogen-specific estimates of foodborne illness. As part of a comprehensive strategic planning and resource allocation process, the US President's Food Safety Council has requested the Interagency Food Safety Risk Assessment Consortium, which consists of the Departments of Health and Human Services and Agriculture and the Environmental Protection

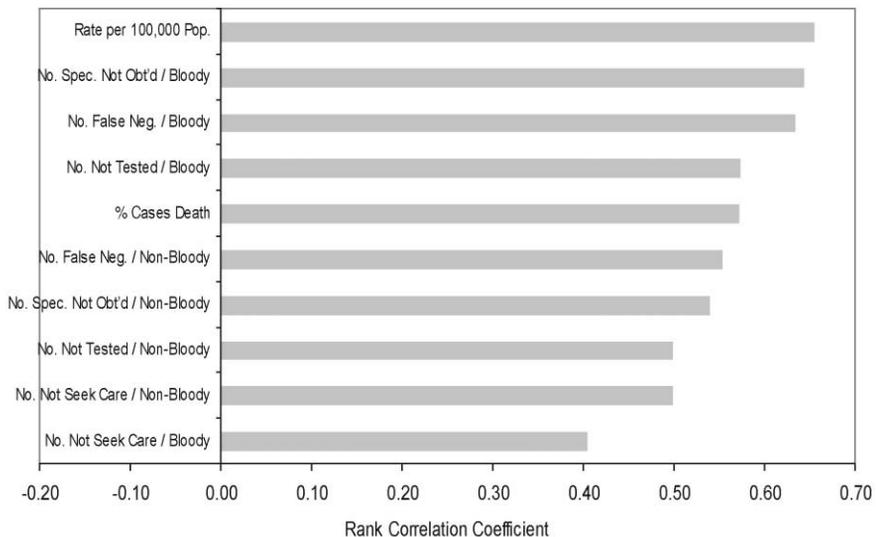


Fig. 3. Rank correlations for the number of deaths.

Agency, to consider how to develop a comparative risk analysis for food safety. An important element of this effort is exploring the available methods to assess the comparative health risks associated with foodborne pathogens (President's Council on Food Safety, 1999).

References

- Buzby, C., Roberts, T., 1997. Guillain–Barre syndrome increases foodborne disease costs. *FoodReview: Mag. Food Econ.* 20, 36–39.
- Centers for Disease Control and Prevention, 1999a. FoodNet 1998. *Morb. Mortal. Wkly. Rep.* 48, 189–194.
- Centers for Disease Control and Prevention, 1999b. Unpublished data.
- Centers for Disease Control and Prevention, 2000. Preliminary FoodNet data on the incidence of foodborne illnesses—selected sites, United States, 1999. *Morb. Mortal. Wkly. Rep.* 49, 201–205.
- Cieslak, P.R., Noble, S.J., Maxson, D.J. et al., 1997. Hamburger-associated *Escherichia coli* O157:H7 infection in Las Vegas: a hidden epidemic. *Am. J. Public Health* 87, 176–180.
- Hedberg, C., Angulo, F., Townes et al., 1997. Differences in *Escherichia coli* O157:H7 annual incidence among FoodNet active surveillance sites. Poster Presented at FoodNet Conference. Baltimore, MD, 22–26 June.
- Mead, P.S., Slutsker, L., Dietz, V. et al., 1999. Food-related illness and death in the United States. *Emerging Infect. Dis.* 5, 607–625.
- National Research Council, 1996. *Understanding Risk: Informing Decisions in a Democratic Society*. National Academy Press, Washington, DC.
- Ostroff, S.M., Kobayashi, J.M., Lewis, J.H., 1989. Infections with *Escherichia coli* O157:H7 in Washington state: the first year of statewide disease surveillance. *J. Am. Med. Assoc.* 262, 355–359.
- President's Council on Food Safety, 1999. Assessment of the NAS Report, Ensuring Safe Food from Production to Consumption. March. Available at: <http://www.foodsafety.gov/~fsg/creport.html>.
- Society for Risk Analysis, 2001. Glossary of Risk Analysis Terms. Available at: <http://www.sra.org/glossary.htm>.
- U.S. Census Bureau, 2000a. State Population Estimates: Annual Time Series, July 1, 1990 to July 1, 1999. Available at: <http://www.census.gov/population/estimates/state/st-99-3.txt>.
- U.S. Census Bureau, 2000b. Historical National Population Estimates: July 1, 1900 to July 1, 1999. Available at: <http://www.census.gov/population/estimates/nation/popclockest.txt>.
- U.S. Department of Agriculture, F.S.I.S., 1997. Report to Congress: FSIS/CDC/FDA SENTINEL SITE STUDY: The Establishment and Implementation of an Active Surveillance System for Bacterial Foodborne Diseases in the United States. February. <http://www.fsis.usda.gov/OPHS/fsisrep1.htm>.
- U.S. Department of Agriculture, F.S.I.S., 1998. Report to Congress: FoodNet: An Active Surveillance System for Bacterial Foodborne Diseases in the United States. April. Available at: <http://www.fsis.usda.gov/OPHS/rpcong97/text.htm>.
- U.S. Department of Agriculture, F.S.I.S., 1999. Report to Congress: FoodNet: An Active Surveillance System for Bacterial Foodborne Diseases in the United States. April. Available at: <http://www.fsis.usda.gov/OPHS/rpcong98/rpcong98.htm>.
- Vose, D., 1996. *Quantitative Risk Analysis: A Guide to Monte Carlo Simulation Modelling*. Wiley, New York.