

Dose-Response Relationships for Non-Typhoid *Salmonella*

March 18, 2009
DRAFT FINAL

Prepared by
Edmund Crouch, Ph. D.
Cambridge Environmental Inc.
58 Charles Street
Cambridge, MA 02141
www.CambridgeEnvironmental.com

Prepared for
Food Safety Inspection Service

under
Task order 4
Contract 43-3A94-3-5041
Subcontract 4400100198
and
Contract AG-3A94-C-08-0016

Table of contents

Table of contents.....	2
Acknowledgements.....	4
Glossary of symbols in the main text in approximate order of occurrence	4
1. Executive Summary	6
2. Introduction.....	8
2.1. Objectives of this document.....	8
2.2. End points	9
2.3. Previous analyses.....	10
2.4. Structure of this document.....	11
3. Available data	12
3.1. Human feeding trials used in FSIS (1998) and FAO/WHO (2002b).....	12
3.2. Outbreak studies used in FSIS (1998) and FAO/WHO (2002b)	14
3.3. Additional studies included in this analysis.....	17
3.4. Common analysis approaches.....	17
4. Doses and their uncertainties	18
4.1. Approach.....	18
4.2. Concentrations determined by MPN methods	19
4.3. Concentrations determined using plating methods.....	21
4.4. Generic uncertainty estimates for concentrations	22
4.5. Treatment of MPN non-detect concentrations.....	22
4.6. Both MPN and plating information available.....	22
4.7. Uncertainties in quantities ingested	22
4.8. Total uncertainties in doses.....	23
5. Uncertainties in attack rates	23
6. Sensitive sub-populations	23
7. Evaluation of individual reports.....	24
8. Dose response evaluation.....	32
8.1. General considerations for dose-response models.....	32
8.2. Primary dose-response model	34
8.3. Observed variation of virulence.....	38
8.4. Modeling the variation of virulence.....	45
8.5. The random isolate dose-response curve	49
8.6. Uncertainty in the virulence distribution parameters.....	53
8.7. Alternative dose-response model.....	56
8.8. Sensitivity analyses.....	62
9. A summary of the principal hypotheses assumed in this analysis.....	65
10. Practical application of the results	67
11. References.....	68
Appendix A. Evaluation of Individual Outbreaks	74
A.1. Outbreak data	74
A.1.1. Entries in Table 3	74
A.1.2. Entries in Table 4.....	84
A.1.3. Additional outbreak entries.....	89
A.2. Human feeding studies.....	94

Appendix B.	Analyses in the workbook “Epi Attack Data.xls.”	98
B.1.	Structure of the workbook “Epi Attack Data.xls”	98
B.2.	Attack rate summary pictures — worksheet “Attack-pict”	100
B.3.	Analyses of individual outbreaks — worksheet “Analyses”	100
B.4.	Analysis of Kasuga <i>et al.</i> (2004) — worksheet “Kasuga”	101
B.5.	Approximating loglikelihoods — worksheets “MultiDose” and “MultiDoseP”	101
B.6.	Adequacy of fit of dose-response models — worksheet “McCullough”	102
B.7.	Individual ML estimates of virulence — worksheet “All_data”	102
B.8.	Distribution of individual ML estimates of virulence — worksheet “Distrib”	103
B.9.	Estimates for the distribution of virulence — worksheet “All_data”	103
B.10.	The random isolate dose-response — worksheets “All_data” and “DR-pict”	104
B.11.	Uncertainties in the parameter estimates — worksheet “All_data”	104
B.12.	Alternative dose-response curve — worksheet “Probit” and “Distrib_p”	105
B.13.	Sensitivity analysis — worksheet “All_data”	105
B.14.	Calculation of approximations for an integral — worksheet “h_coeffs”	105
B.15.	Reference tables — worksheet “MPN tables”	106
B.16.	Reference specific to Appendix B	106
Appendix C.	Evaluation of the integrals.	107
C.1.	Binomial response, linear-exponential dose-response with lognormal parameters....	107
C.1.1.	Multiple doses, all doses known	107
C.1.2.	Upper bound on single dose.....	109
C.2.	Binomial response, probit-log-dose dose-response with lognormal parameters	111
C.2.1.	Multiple doses, all doses known	111
C.2.2.	Upper bound on single dose.....	113
C.3.	References specific to Appendix C.....	116
Appendix D.	The effects of a distribution of doses.....	117
Appendix E.	Conceptual difficulties with FSIS (1998) and FAO/WHO (2002a,b)	121

Acknowledgements

This work would not have been possible without the encouragement of the FSIS and the assistance in particular of Neal Golden. I would like to thank the five peer reviewers (Dr. A. Fazil, Dr. D. Crawford-Brown, Dr. J. Englehardt, Dr. N. Strachan, and Peter Tuenis) for their excellent suggestions that have substantially improved the document; and Sara Hendrix for organizing the reviewer comments. However, all errors and omissions remaining are my own.

Glossary of symbols in the main text in approximate order of occurrence

N	The total number of observational units (<i>e.g.</i> number of dose groups, number of dilutions in an MPN analysis) in some experiment
i	An integer index for observational units, generally ranging from 1 to N
r	An observed count
r_i	The number of results tagged as positive in the observational unit with index i
n_i	The total number of results in the observational unit with index i
z_i	Concentration of test organism in the dilution with index i in an MPN analysis
m_i	Mass or volume of original test material in each tube of an MPN analysis at the dilution with index i
z	Concentration (number per unit mass or per unit volume) of <i>Salmonella</i> in a food sample
m	Mass or volume of food containing <i>Salmonella</i>
z_0	Maximum likelihood estimate for concentration
z_{05}	Concentration at the approximate 5 th percentile of the uncertainty distribution
z_{95}	Concentration at the approximate 95 th percentile of the uncertainty distribution
z_{100x}	Concentration at the approximate 100x percentile of the uncertainty distribution
x, θ, t, w	Generic symbols used for intermediate quantities, with meanings depending on the local context
J	Loglikelihood (used generically, with application varying with context)
ξ	Uncertainty standard deviation of the logarithm ($\ln(\text{GSD})$) of concentration.
μ	An expected count
R_{100x}	The 100x percentile of the uncertainty distribution for the mean of a Poisson-distributed variable.
π	The ratio of the length of the circumference of a circle to the length of its diameter
Φ	The standard cumulative normal distribution function, defined by
$\Phi(\theta) = \left(1/\sqrt{2\pi}\right) \int_{-\infty}^{\theta} \exp(-x^2/2) dx$	
ζ	Uncertainty standard deviation of the logarithm ($\ln(\text{GSD})$) of food mass
p	Probability of illness
p	(in text) Approximate probability for some observation under a stated null hypothesis.
d	Dose (number of <i>Salmonella</i> ingested)
d_i	Dose (number of <i>Salmonella</i> ingested) in the observational unit with index i
χ	Generic virulence parameter(s)
λ	Virulence parameter for the linear-exponential dose-response model
σ	Uncertainty standard deviation of the logarithm ($\ln(\text{GSD})$) of dose d

σ_i	Uncertainty standard deviation of the logarithm ($\ln(\text{GSD})$) of dose d_i
m	Mean logarithm of virulence in a lognormal distribution of virulences
m_{se}	Mean logarithm of virulence for <i>S. enteritidis</i> (in a lognormal distribution)
m_{ot}	Mean logarithm of virulence for serotypes other than <i>S. enteritidis</i> (in a lognormal distribution)
Σ	Logarithmic standard deviation of a lognormal distribution of virulences
ω	Combined $\ln(\text{GSD})$ for doses and the virulence distribution
S	Relative sensitivity factor, with value 1, s_Y , s_M , or s_H
S_i	Sensitivity factor the observational unit with index i
s_Y	Sensitivity factor for the 0–5 year old sensitivity class
s_M	Sensitivity factor for the 6–12 year old sensitivity class
s_H	Sensitivity factor for the hospitalized sensitivity class
m'	Median parameter in normal curve approximating the random-isolate dose-response curve
Σ'	Standard deviation parameter in normal curve approximating the random-isolate dose-response curve
β	Slope parameter in a probit-log-dose dose-response curve
d_{50}	Dose causing 50% illness in a probit-log-dose dose-response model
D_{50}	Median of a lognormal distribution of d_{50} values.
Ω	Logarithmic standard deviation of a lognormal distribution of d_{50} values.

1. Executive Summary

Quantitative microbial risk assessments (QMRA) for food safety require information on dose-response relationships for the organisms evaluated. This document synthesizes the available information from experiments on and outbreaks of non-typhoid *Salmonella* to obtain illness dose-response relationships suitable for use in QMRA, specifically for use in probabilistic QMRA that are typically implemented by Monte Carlo simulations. The observed end points after human ingestion of *Salmonella* vary from nothing through infection, mild fevers, diarrheal illness, to death. The dose-response relationships evaluated here are for diarrheal illness. Previous analyses (FSIS, 1998; FAO/WHO, 2002a,b) have attempted to evaluate similar dose-response relationships using experimental human feeding studies combined with data obtained from outbreaks; however, those attempts failed to adequately fit the observed data. This document uses the same human experimental study data, combined with an updated and extended database of outbreak observations.

The crucial distinguishing feature of this analysis is the observation that a single dose-response curve cannot explain the observed attack rate versus dose information (Figure 1).

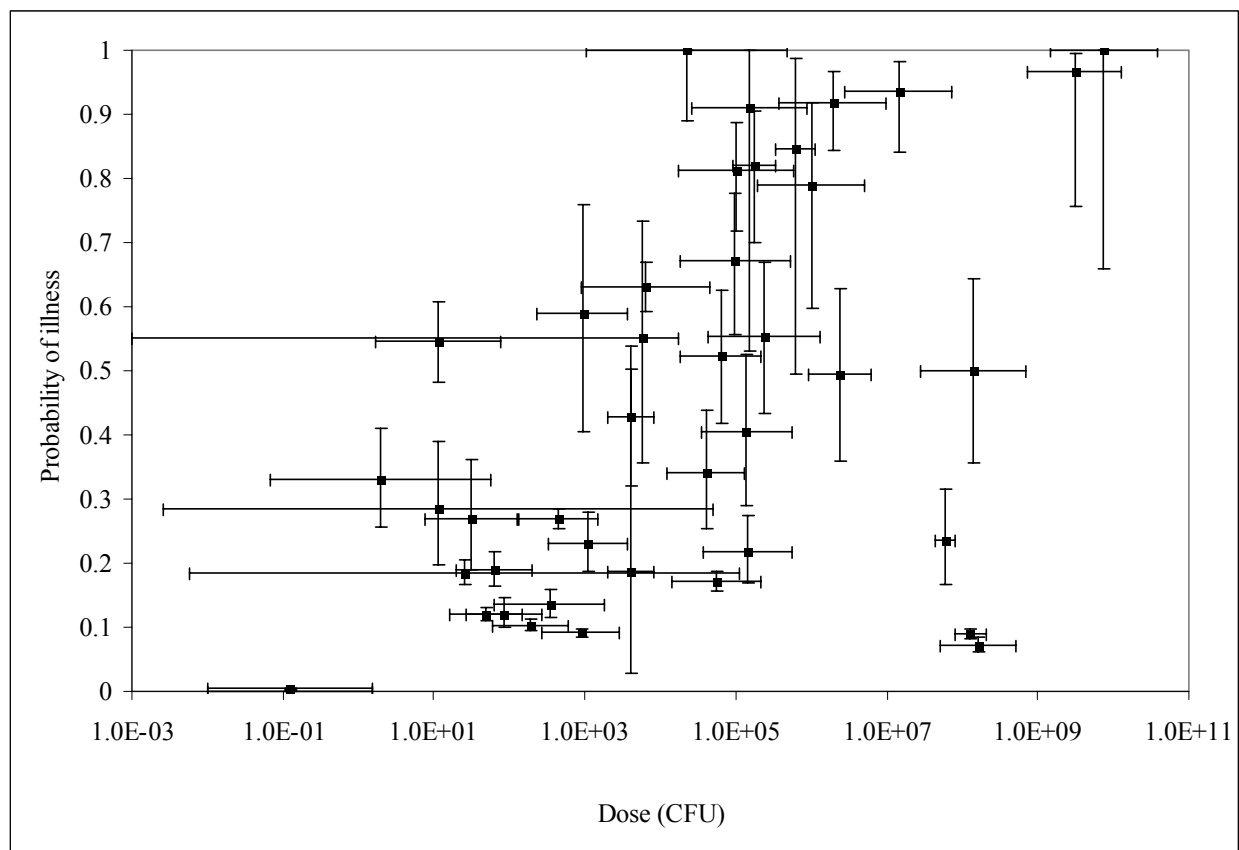


Figure 1 Summary of attack rate versus dose data; no single dose-response curve can be consistent with all the observations.

Instead it is plausible to postulate a distribution of virulence for different *Salmonella* isolates. Calculation of individual estimates for virulence for all available data supports the hypothesis that this distribution is lognormal (Figure 2).

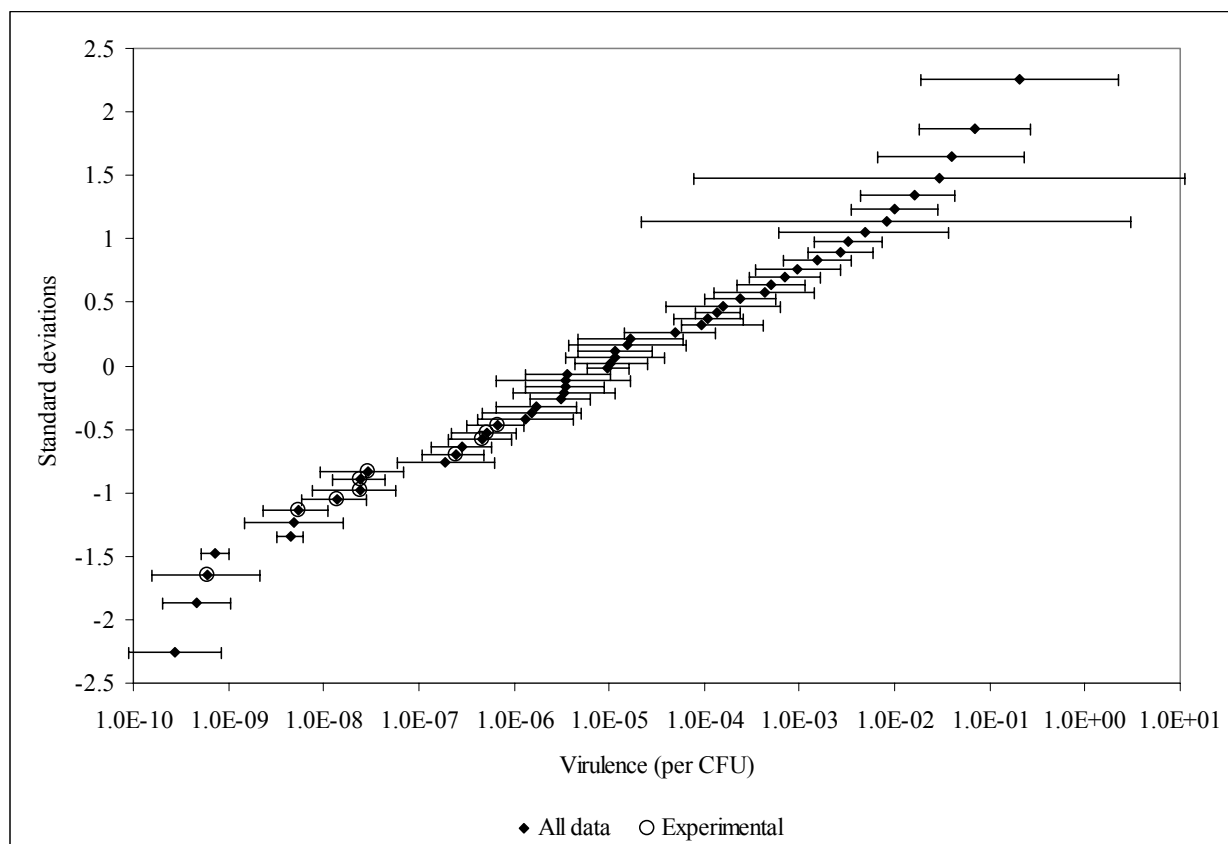


Figure 2 Individual virulence estimates; a straight line corresponds to a lognormal distribution.

Further refinement of the analysis then leads to evaluation (Table 1) of the parameters of lognormal distributions of virulence for *S. enteritidis* and for all other *Salmonella* serotypes combined (that *S. enteritidis* is singled out is probably because the majority of outbreak data are for this serotype; the data are too sparse to distinguish other serotypes individually). Moreover, different sensitivities can be evaluated for children in two age groups and adults, and the sensitivity of hospitalized patients is consistent with being equal to that of one of the child groups.

The virulence described here is the parameter λ in a linear-exponential dose-response model of the form

$$p = 1 - \exp(-S\lambda d) \quad (1)$$

where p is the probability for illness at a dose of d CFU in a population with sensitivity factor S (and S is unity by definition for non-hospitalized adults). This simple dose-response model describes what would happen for a single isolate of *Salmonella* given at different doses to the same population — a single value of λ would apply to that population. However, the implication of this analysis is that prediction of the particular value of λ that would apply in any particular situation is not currently possible; in general, all that can be predicted is that λ will come from the probability distributions defined by Table 1. However, for typical probabilistic QMRA that evaluates multiple independent exposures to different *Salmonella* isolates, this information is sufficient.

Table 1 Maximum likelihood estimates for virulence distribution parameters

Parameter	Parameter symbol	Value
Sensitivity factor, 0–5 yr old	s_Y	5.832
Sensitivity factor, 6–12 yr old or hospitalized	$s_M = s_H$	2.843
Mean logarithm of virulence, <i>S. enteritidis</i>	m_{se}	-10.426
Mean logarithm of virulence, all other serotypes	m_{ot}	-14.019
Standard deviation of the logarithm of virulence, all serotypes	Σ	4.853

Equation (1) may be considered simplistic as a dose-response model for *Salmonella* illness. However, it is here demonstrated to be adequate to model all available data, and a parallel analysis using a somewhat differently shaped dose-response model (a probit-log-dose model) gives similar results. Alternative dose-response models can thus be expected to produce similar results in probabilistic QMRA that evaluate multiple independent exposures to different *Salmonella* isolates; however, they may give substantially different predictions for extremely low doses (less than an average of about 1 CFU to the exposed population) if the distribution of doses in the population is uneven.

The analysis here is extended to provide uncertainty estimates for the parameters described in Table 1 and the method of using these estimates is described — such uncertainty analyses should be used in probabilistic QMRA. Sensitivity analyses show that larger uncertainties in the underlying data than evaluated here do not have much effect on the parameter estimates.

2. Introduction

2.1. Objectives of this document

Quantitative microbial risk assessments (QMRA) for food safety require information on dose-response relationships at the population level for the organisms evaluated. This document synthesizes the available information from experiments on and outbreaks of non-typhoid *Salmonella* to obtain population-level illness dose-response relationships suitable for use in QMRA, specifically for use in probabilistic QMRA that are typically implemented by Monte Carlo simulations. In such assessments the number of relevant contaminating organisms is tracked as a food passes through the various stages of preparation, so that the number of relevant organisms ingested is evaluated. What is therefore required is the probability for illness given the ingestion of some known number of organisms. For probabilistic assessments it is also required to estimate the uncertainty in the probability of illness; and indeed this uncertainty cannot be ignored in any evaluation of the available information on dose-response.

Here the need is for population-level results; the variability among individuals in the population is not addressed in this report, indeed the available data are entirely inadequate to evaluate such individual variability. The principal source of uncertainty turns out to be the variation in virulence (pathogenicity) of the organism; that is, the variation in probability of illness at a fixed dose for different strains or isolates of *Salmonella* under potentially different conditions of consumption. It is already known from analysis of human feeding experiments that virulence for illness varies between strains or isolates even within serotype under fixed consumption conditions (Coleman and Marks, 1998; Coleman *et al.*, 2004), and this observation is confirmed in the analysis that follows. For *Salmonella*, the determinants of the variation in virulence of the organism to cause illness are not known. Even if the determinants of virulence were known, it is unlikely that in a QMRA of a food the distributions of such determinants arising from the original contaminating events, the contaminated food, or the conditions of consumption would be known. This document therefore attempts to evaluate a probability for illness that incorporates the likely distribution of virulence for illness. The analysis is based primarily on information from outbreaks, with the assumption that the distribution of virulence observed in the outbreaks and human experimental studies corresponds to the distribution that would be seen generally for consumption of contaminated foods. The principle objective is thus not to obtain dose-response information on just one isolate of *Salmonella*, but to obtain an average dose-response taking account of the variation in virulence.

2.2. End points

Evaluation of dose-response relationships requires specification of the end point(s) to be considered. For *Salmonella*, the observed end points after human ingestion vary from nothing, to observation of infection, usually defined by detection of organisms in the feces, to mild fevers, to extensive diarrhea with cramping and fever, to death. The principal human feeding studies (McCullough and Eisele, 1951a,b,c,d) reported information on infection defined by detection in the feces, and on overt symptoms. The primary overt symptom in these human studies was diarrhea in 46/47 cases with the 47th case being borderline with soft stool but no frank diarrhea; the next most common symptom was abdominal cramps in 39/47 cases. The main source of information on dose-response, however, comes from studies of outbreaks in which a contaminated food item was identified and the *Salmonella* contamination quantified, the number of persons ingesting the food could be estimated, and the number becoming ill was also identified. For such outbreaks the identification of symptoms, sometimes recalled several days to weeks after the episode, requires that they have been readily apparent (*e.g.* diarrhea, cramping, fever with chills, or more severe). To evaluate the main concern of *Salmonella* ingestion, namely overt illness, and to maintain consistency between the symptoms identified during outbreaks and the human feeding trials, only such overt symptoms will be evaluated here. In particular, for the studies of McCullough and Eisele (1951a,b,c,d), evidence of infection (identification of *Salmonella* in feces) was not alone considered to constitute illness; the assignment of illness made by McCullough and Eisele was adopted. This required overt symptoms such as abdominal cramping, diarrhea, or vomiting. Asymptomatic shedding of the organism or mild fever were not considered sufficient to constitute illness. Outbreak reports are taken at face value, with reported illnesses accepted as constituting “illness” for this report; the definition undoubtedly varies from outbreak to outbreak, and is usually not explicitly specified in the reports. Some outbreak reports discuss “infection” rather than illness, but generally define this to mean the presence of overt symptoms. Where attempts were made to measure the

presence of *Salmonella* in feces, there are clear distinctions in the reports between this measure of infection and the measure of illness as defined by overt symptoms.

Practically all serotypes of *Salmonella enterica* subsp. *enterica* are considered capable of causing (human) illness in suitable circumstances, although it is possible that the nature of the illness varies depending on the serotype and/or the circumstances. It is also possible that some serotypes do not cause the overt symptoms in humans that are identified here; such serotypes will necessarily not figure in this analysis, since they will not be identified (except accidentally, if they happened to co-occur with illness-causing serotypes) in outbreaks or in human experiments that induced illness with the required end-points. If necessary, therefore, this analysis should be considered to be limited to serotypes of *Salmonella* that cause the end-points considered, but no information is evaluated to determine what fraction of *Salmonella* fall in this category.

2.3. Previous analyses

There have been multiple evaluations of the human feeding studies of McCullough and Eisele (1951a,b,c,d), usually examining infection as an initial end point (*e.g.* Teunis *et al.*, 1996;¹ Tuenis *et al.*, 1999; Fazil, 1999²) in a model framework that then considers the probability for illness given infection, although others have pointed out problems with this approach (Coleman and Marks, 1998; Coleman *et al.* 2004). The most complete recent evaluations of dose-response relationships for overt illness due to *Salmonella* spp. are FSIS (1998)³ and FAO/WHO (2002b),⁴ while others (*e.g.* Bollaerts *et al.*, 2008) have used the data summarized in these efforts as starting points for more complex analyses. FSIS (1998) was primarily concerned with evaluation of a dose-response curve for illness from *S. enteritidis*, but examined the infection endpoint (not the illness endpoint) for the human experimental data of McCullough and Eisele (1951a,b,c,d).

FSIS (1998) and FAO/WHO (2002b) obtained estimates of dose, numbers of persons exposed at each dose, and number of persons ill from that dose, from outbreak reports. In both analyses, a single dose-response curve for the probability of illness was assumed to apply to all the outbreak data simultaneously, so that implicitly it was assumed that all strains or isolates of *Salmonella* are of the same virulence. Appendix E discusses some of the conceptual difficulties of this assumption.

In FSIS (1998), point estimates of dose and attack rates were tabulated for the outbreaks (for all serotypes of *Salmonella*); some uncertainty in doses was taken into account by entering the same

¹ The tables on page 37 and 38 of Tuenis *et al.* (1996) summarizing the results of McCullough and Eisele (1951a) do not correspond exactly to the results reported (and used here) for *S. meleagridis* I (2nd dose, 4 infected not 3; top dose, 4 ill not 5), *S. meleagridis* II (3rd dose, 5 infected, not 6), and *S. anatum* II (6th dose, 5 infected not 6).

² The table on page 399 of Fazil (1999) summarizing the results of McCullough and Eisele (1951a,c) misplaces the entries in the “infected with” column for “*S. anatum* strain I” and “*S. malegriis* strain II” (the latter being a misprint for *S. meleagridis*), potentially misleading the reader about doses tested for various strains.

³ Dose-response information (“Evidence”) is discussed at pages 224-227 and Table E-3 (page 235) of FSIS (1998). FSIS (1998) is a draft of FSIS (2004), but the latter uses the FAO/WHO (2002a,b) dose-response material (see Section 4, page 117; Annex I of FSIS, 2004 is a draft of the Hazard Characterization section of FAO/WHO, 2002b).

⁴ Dose-response information is discussed in Section 3.3 (Human Feeding Trials, pages 27ff.) and Section 3.5 (Epidemiological Information, pages 46ff.) of FAO/WHO (2002b).

outbreak twice in the table with different estimates of dose. The dose-response curve for *S. enteritidis* specifically was assumed to correspond to that obtained by fitting data for a surrogate organism, *Shigella dysenteriae*, to a beta-Poisson model. Two variations of the *Shigella* dose-response curve were introduced to correspond to normal and susceptible individuals, the second variation having a virulence or pathogenicity parameter (in this case, the dose causing 50% incidence of illness) reduced by a factor of 10. No formal fitting procedure was used on the *Salmonella* outbreak data, only a graphical confirmation that the plotted dose-response curves were in the right range by comparison with the observed attack rates plotted against the dose estimates. Uncertainty was introduced by ascribing a truncated normal distribution to the virulence parameter just described, and matching the uncertainty in the *Shigella* dose-response fit at a particular dose. No justification was provided for the shape of the uncertainty distribution.

In FAO/WHO (2002b), uncertainty estimates were made of the dose(s) and attack rate(s) in the outbreak data. A beta-Poisson model was assumed adequate to fit all the outbreak data, and the uncertainty distributions for the parameters of the model obtained by re-sampling from the uncertainty distributions for the dose(s) and attack rate(s) in the outbreak data and re-fitting the beta-Poisson model using likelihood methods with a binomial incidence assumption. The combination of the use of a single beta-Poisson model for all the data, together with the binomial incidence assumption, is equivalent to assuming that all outbreaks were caused by *Salmonella* with the same virulence. No methodology is described that would allow an independent analyst to derive from the outbreak data the same minima and maxima in dose and attack rates given in FAO/WHO (2002b), nor is any methodology described that would allow derivation of the shapes of the uncertainty distributions used.

Both FSIS (1998) and FAO/WHO (2002b) suffer from substantial limitations. The first is not based on the observed *Salmonella* data, except insofar as the dose-response curve appears graphically to be in the right range. However, the outbreak data are of such a nature that such comparisons could be substantially misleading. For example, 10/18 of the attack rates were 100%, so provide little confirmation of predicted response rates less than 100%. The numbers of persons ill varied from 1 to 16,000, so the uncertainty in the observed attack rate varied substantially; but such variation was not taken into account in the graphical comparison. Some outbreaks (4/14) were entered twice with different dose estimates, up to 10 times different, so these outbreaks would be over-weighted graphically. There was no confirmation that the different outbreaks were consistent with the same dose-response curve, as assumed. Finally, the uncertainty estimate introduced had no relation to the observed *Salmonella* data (since it was based on matching *Shigella* data). FAO/WHO (2002b) substantially improved the approach of FSIS (1998) by accounting for uncertainty in doses and attack rates; but the methods used for this accounting, and the results obtained, were not reproducible from the available documentation. Moreover, there was again no confirmation that the different outbreaks were consistent with the same dose-response curve, as assumed (see Appendix E).

2.4. Structure of this document

This introductory section summarizes the objectives of the analysis, documents the end point (illness) evaluated, and points to previous analyses and some of their shortcomings. Section 3 summarizes the human feeding and outbreak data available in the literature, including what has been used in previous analyses, and introduces the problems of dose and response estimation.

Section 4 describes the approaches taken to estimate doses and their uncertainties. Section 5 describes the assumptions used to evaluate attack rates. Section 6 defines “sensitive populations” and how they are incorporated in the analysis. Section 7 lists all the individual reports that are used here, while Appendix A documents all the reports that have been identified and how they used here or why they are not used here. Section 8 performs and shows the results of dose-response analyses, and the uncertainties involved. Appendix B cross-references the analyses described in the text with their numerical implementation in an Excel workbook (“Epi Attack Data.xls”), and Appendix C sketches the mathematics involved in performing some of the calculations. Appendix D examines the effect of different distributions of doses in exposed populations, and Appendix E demonstrates some problems with the FSIS (1998) and FAO/WHO (2002a,b) approaches.

3. Available data

3.1. Human feeding trials used in FSIS (1998) and FAO/WHO (2002b).

The most complete recent summaries of the human data suitable for evaluation of dose-response relationships for illness due to *Salmonella* spp. are FSIS (1998) and FAO/WHO (2002b). These summaries together with evaluation of references cited in all documents examined and ad hoc searches of PubMed (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>), and Google (<http://www.google.com> and <http://scholar.google.com>), were used to obtain an updated list of potentially suitable human data. The literature in the reference list (Section 11), including that not directly cited, was included in this search. All available human data identified in this manner as potentially suitable for inclusion, including all data so identified by FSIS (1998) and FAO/WHO (2002b), are here re-evaluated for dose-response modeling.

FSIS (1998) and FAO/WHO (2002b) identified all known human feeding trials involving non-typhoid *Salmonella* spp. (Table 2).⁵

⁵ Inhalation studies, and ingestion studies involving *S. typhi*, are not discussed or evaluated here. The illustrative case report included in Mackenzie and Livingstone (1968) is also excluded since it involved introduction of *S. sofia* and *S. bovis-morbificans* into the sinus of one volunteer, not direct ingestion.

Table 2 Human feeding trials involving non-typhoid *Salmonella* spp.

Serotype and strain ^a	Reference
S. typhimurium	Hormaeche <i>et al.</i> , 1936
S. anatum	Varela and Olarte, 1942
S. meleagridis I, II, & III	McCullough and Eisele, 1951a,b
S. anatum I, II, & III	McCullough and Eisele, 1951a,b
S. newport	McCullough and Eisele, 1951c
S. derby	McCullough and Eisele, 1951c
S. bareilly	McCullough and Eisele, 1951c
<i>S. pullorum</i> I, II, III, and IV	McCullough and Eisele, 1951d
^a Hyperlinks lead to the discussions in Appendix A	

Hormaeche *et al.* (1936) administered 2×10^9 CFU⁶ *S. typhimurium* to two subjects, and 4×10^9 CFU to three others in liquid suspension (subjects were aged 24 to 51, sex not specified). One of the first two suffered only a mild elevation in temperature (38°C) the next day. All others suffered from diarrhea of varying severity on the second and/or third day following administration, with or without fever. For the purposes of this evaluation, the evidence of mild fever alone is not considered sufficient to constitute illness.

Varela and Olarte (1942) fed one 37 year old male 20 c.c. of an emulsion containing 5×10^8 CFU *S. anatum* per c.c., causing stomach pains and diarrhea on the day of administration and altered fecal consistency for some days thereafter.

McCullough and Eisele (1951a,b,c,d) fed multiple doses of three strains each of *S. typhimurium* and *S. anatum*, one strain each of *S. newport*, *S. derby* and *S. bareilly*, and four strains of *S. pullorum* to selected adult male volunteers (ages not specified) in a penal institution. Doses were administered in egg nog. Following FAO/WHO (2002b), the data on *S. pullorum* are here considered inappropriate for use, since this serotype is host-adapted for fowl, exhibits differences in motility from other serotypes (Holt and Chaubal, 1997), and human infections appear to be rare (see discussions in Blaser and Newman, 1982; and Coleman and Marks, 1998). Further, the nature of the illness caused in these feeding trials was not typical of that caused by *Salmonella* food poisoning (McCullough and Eisele, 1951d). As noted by the authors *S. pullorum* was never recovered from stool at doses lower than those required to produce illness, and even then it was recovered only on the first and second day following administration; and the clinical picture produced was distinct from that noted with the other serotypes (explosive onset, acute nature, high fever and prostration, with prompt recovery). In addition, although not noted by the authors, in some cases clinical illness was noted with no recovery of the organism from stool.

One further study involving deliberate dosing of humans with potentially contaminated material was located ([Greenberg, 1964](#)), but this is incorporated in the outbreak studies of Section 3.2

⁶ Colony forming unit; the smallest separable collection of bacteria, possibly a single bacterium, that can be obtained from a given sample and that will grow to a colony when suitably incubated.

since it corresponds more closely to such studies (e.g. it does not have experimenter-selected doses).

3.2. Outbreak studies used in FSIS (1998) and FAO/WHO (2002b)

FSIS (1998) identified the outbreak data shown in Table 3 as suitable for dose-response characterization. The number in the first column has been added to the material in FSIS (1998), and is used below for identification purposes. Several of the references identified in Table 3 are cited on more than one row — in these cases, FSIS (1998) used more than one estimate for dose and/or attack rate and entered both in the table.

Table 3 Outbreak dose-response data identified in FSIS (1998).

No. ^a	Serotype	Dose (CFU)	Attack rate	Number ill	Reference
FSIS #1	<i>S. typhimurium</i>	17	12%	16,000	Boring, <i>et al.</i> , 1971
FSIS #2	<i>S. schwarzengrund</i>	44	100%	1	Lipson, 1976
FSIS #3	<i>S. newport</i>	60	45%	48	Fontaine, <i>et al.</i> , 1978
FSIS #4	<i>S. eastbourne</i>	100	45%	95	D'Aoust, <i>et al.</i> , 1975
FSIS #5	<i>S. heidelberg</i>	100	28%	339	Fontaine, <i>et al.</i> , 1980
FSIS #6	<i>S. heidelberg</i>	200	100%	1	George, 1976
FSIS #7	<i>S. newport</i>	230	45%	46	Fontaine, <i>et al.</i> , 1978
FSIS #8	<i>S. heidelberg</i>	500	36%	339	Fontaine, <i>et al.</i> , 1980
FSIS #9	<i>S. typhimurium</i>	11,000	52%	1,790	Armstrong, <i>et al.</i> , 1970
FSIS #10	<i>S. cubana</i>	15,000	100%	28	Lang, <i>et al.</i> , 1967
FSIS #11	<i>S. cubana</i>	60,000	100%	28	Lang, <i>et al.</i> , 1967
FSIS #12	<i>S. zanzibar</i>	150,000	100%	6	Reitler, 1960
FSIS #13	<i>S. infantis</i>	1,000,000	100%	5	Angelotti, <i>et al.</i> , 1961
FSIS #14	<i>S. zanzibar</i>	1.00E+11	100%	8	Reitler, 1960
FSIS #15	<i>S. enteritidis</i>	6	6%	>1,000	Hennessy, <i>et al.</i> , 1996
FSIS #16	<i>S. enteritidis</i>	24	6%	>1,000	Vought and Tatini, 1998
FSIS #17	<i>S. enteritidis</i>	1,000	100%	39	Levy, <i>et al.</i> , 1996
FSIS #18	<i>S. enteritidis</i>	10,000	100%	39	Levy, <i>et al.</i> , 1996
^a Hyperlinks lead to the discussions in Appendix A.					

FAO/WHO (2002b) identified the outbreak studies listed in Table 4 as suitable for dose-response characterization. There is some overlap with Table 3, as noted in the more detailed evaluation in Appendix A. FAO/WHO (2002b) discussed 33 outbreaks that have been reported in the literature, but found that many could not be used for dose-response characterization, hence the numbering given in Table 4. The population (“Popn.”) column in Table 4 indicates whether the population involved is considered to be normal (N) or at high risk (S, for sensitive) due to age or

hospitalization. Outbreaks #18 through #33 were reported by personal communication(s) from the Japanese Ministry of Health and Welfare.

Table 4 Outbreak dose-response data used in FAO/WHO (2002a,b)

Case No. ^{a,b}	Serotype	Popn. ^c	Dose (log CFU)	Attack rate (%)	Reference
1a	<i>S. typhimurium</i>	N	2.31	10.63	Boring, <i>et al.</i> , 1971
1b	<i>S. typhimurium</i>	S	2.31	18.91	
2	<i>S. heidelberg</i>	N	2.22	32.76	Fontaine, <i>et al.</i> , 1980
3	<i>S. cubana</i>	S	4.57	70.93	Lang, <i>et al.</i> , 1967
4	<i>S. infantis</i>	N	6.46	100	Angelotti, <i>et al.</i> , 1961
5	<i>S. typhimurium</i>	N	3.79	55	Armstrong, <i>et al.</i> , 1970
7	<i>S. newport</i>	N	1.23	1.07	Fontaine <i>et al.</i> , 1978; Fazil, 1996
11	<i>S. enteritidis</i>	N	4.74	100	Levy <i>et al.</i> , 1996; FSIS, 1998
12	<i>S. enteritidis</i>	N	2.09	6.8	Vought and Tatini, 1998; Hennessy, <i>et al.</i> , 1996
13a	<i>S. typhimurium</i>	N	8.7	100	Taylor, <i>et al.</i> , 1984
13b	<i>S. typhimurium</i>	S	8	100	
18	<i>S. enteritidis</i>	N	5.41	60	Ministry of Health and Welfare, Japan, 1999, personal communications (see text)
19	<i>S. enteritidis</i>	N	6.31	93.93	
20	<i>S. enteritidis</i>	N	2.97	26.86	
22	<i>S. enteritidis</i>	N	6.3	56.01	
23	<i>S. enteritidis</i>	N	5.8	84.62	
24	<i>S. enteritidis</i>	N	1.72	16.41	
25a	<i>S. enteritidis</i>	N	3.63	18.75	
25b	<i>S. enteritidis</i>	S	3.63	42.74	
30	<i>S. enteritidis</i>	N	3.8	64.18	
31	<i>S. enteritidis</i>	N	2.65	27.33	
32	<i>S. enteritidis</i>	S	1.4	26.92	
33	<i>S. oranienburg</i>	N	9.9	100	

^a Discontinuity in numbering occurs because some numbered studies were considered unsuitable for analysis and were not used in the dose-response evaluation.

^b Hyperlinks lead to discussions in Appendix A.

^c N indicates a “non sensitive” population, S indicates a “sensitive” population; see text.

The literature cited in Table 3 and Table 4 was consulted in the current effort to define a consistent set of dose-response data. FSIS (1998) was used solely to identify literature, since minimal methodology is documented as to how the dose and response estimates were obtained (in many cases these are not obvious from the literature cited). Some material in FAO/WHO (2002a,b) was considered unreliable since it either misquoted or otherwise mischaracterized information provided in the original literature, and rejected for this assessment (see detailed evaluations in Appendix A). In two cases ([Greenwood and Hooper, 1983](#); and [Narain and Lofgren, 1989](#)), reports rejected by FAO/WHO (2002a,b) were retained, because it was found that the rejection was due to misunderstanding or failure to incorporate all available data provided in the literature. The series of outbreak reports ([FAO/WHO #18 through #33](#), Table 4) were replaced by the published report of Kasuga *et al.* 2004; this has the effect of omitting outbreaks FAO/WHO Case Numbers [21](#),⁷ [22](#), and [31](#) (Table 4) that are not reported by Kasuga *et al.* (2004), and incorporating 18 further outbreaks (that were not apparently reported earlier to FAO/WHO).

3.3. Additional studies included in this analysis

Additional searches of the literature using ad hoc searches of PubMed (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>), Google (<http://www.google.com> and <http://scholar.google.com>), and the literature in the reference list (Section 11), including that not directly cited, were used to obtain an updated list of potentially suitable data. This search located an additional six reports with sufficient data for inclusion in this analysis. The following material was therefore also incorporated, in addition to that already discussed.

[Greenwood and Hooper, 1983](#); this is FAO/WHO Case Number 17; it was missed by FAO/WHO that 3 people, not just 1, were reported as exposed.

[Narain and Lofgren, 1989](#); this is FAO/WHO Case number 8; it was missed by FAO/WHO that there was sufficient information provided to estimate the attack rate.

[Matsui *et al.*, 2004](#); a further report from Japan not incorporated in Kasuga *et al.* 2004.

[Abe *et al.*, 2004](#); three outbreaks, numbered 5, 6, and 8 in Table 1 of this reference, were not reported by Kasuga *et al.*, 2004 or Matsui *et al.*, 2004 (all others were so reported).

[Urfur *et al.*, 2000](#); a single outbreak is reported in sufficient detail to allow inclusion.

[Greenberg, 1964](#); this is effectively a human feeding study, but the dose was not experimenter-selected. Although no illnesses resulted from the infections, sufficient dose information is provided to allow inclusion in this analysis.

3.4. Common analysis approaches

All the human feeding studies and outbreak reports were examined individually in order to estimate doses of *Salmonella* organisms, numbers of people exposed to those doses, and the

⁷ FAO/WHO Case Number 21 was included in the text but excluded from analysis in FAO/WHO (2002b), so does not occur in Table 4; it is mentioned here for completeness.

number of individuals becoming ill due to exposures. In addition, the uncertainties in these estimates were also evaluated using reproducible methodologies fully explained below (by contrast, neither FSIS, 1998 nor FAO/WHO, 2002b, described reproducible methodologies to explain the ranges of values used).

The uncertainties included in the analysis are those that are known. There are also uncertainties that are not included quantitatively, because they are known in nature but unknown in magnitude (e.g. the extent to which the *Salmonella* density in the foods examined deviates from homogeneity, or the food or the population examined is non-representative) or unknown in both nature and magnitude. The sensitivity analysis of Section 8.8 evaluates the potential effect of such unknown uncertainties.

Some common methods were used for evaluating dose uncertainty. Doses are generally estimated as the product of a quantity of material ingested, and a concentration⁸ in that material. In general, both these quantities and their uncertainties were characterized by parameters of lognormal distributions, since the uncertainties in both generally are right skewed in a manner similar to lognormal, and their product is likely to be closer to lognormal than either one individually (a consequence of the central limit theorem and the independence of the estimates for concentration and amount ingested).

4. Doses and their uncertainties

4.1. Approach

The methodology used here relies on the assumption that the food analyzed in the original report is representative in every way of the food that was ingested by the population examined. Even more strongly, it is assumed that the contamination is sufficiently homogeneous that the sample analyzed adequately represents the mean concentration in food eaten by the population, since the distribution of contamination in the food is not known (except perhaps in the human feeding studies, where homogeneity should have been ensured by the preparation methods). The entire analysis is therefore conditional on the representativeness of the concentration measurements and intake estimates given in the original reports (or as modified here).

In this document, “dose” is defined to be the number of *Salmonella* organisms ingested by an individual. As described further in Section 8.2, the measure of dose required in the following analyses is the mean to the population exposed and in which the attack rate is known. In the available human feeding studies, the original investigators measured the doses administered; these measurements are used here. In the outbreak studies, dose estimates were generally obtained (by the original investigators) as the product of an amount of food ingested, and a concentration in that food. The same approach is used here, using data from the original literature with additional analysis where necessary, primarily to evaluate uncertainty. When there were multiple ingestions of contaminated food by individuals, the total amount ingested

⁸ The term “concentration” is used throughout this report to denote the number of CFU per unit volume or the number of CFU per unit mass of *Salmonella* in food, food samples, or other media. The precise definition, where needed, should be clear from the context.

was used, since this correctly meshes with the primary dose-response approach used (see Section 8.2).

In cases where average doses of the same *Salmonella* isolate are known for different population subgroups, and the attack rate in those subgroups are also known, the subgroups can be used to provide information on the dose-response shape, and are so analyzed here. This requires that the circumstances of exposure of the subgroups (other than dose) are similar, and that the only potential difference (other than dose) between the subgroups is their susceptibility, which is taken into account in Section 8.2 by a relative sensitivity factor.

Uncertainties in dose estimates are usually quite large, and are generally skewed to high values in a manner approximated here by lognormal distributions; so their uncertainties are represented here by the standard deviation of their logarithms (the logarithm of the geometric standard deviation, abbreviated throughout as $\ln(\text{GSD})$). The total uncertainties in doses have in most cases been obtained by adding the uncertainty estimates in concentrations and quantities ingested, since these estimates are independent in all cases examined. Such independent addition of uncertainties corresponds, mathematically, to adding in quadrature⁹ the $\ln(\text{GSD})$ estimates of concentration and quantity ingested.

In the feeding studies (Serial #s [42–50](#), and [59](#)), and in several outbreaks (Serial #s [1](#), [7](#), [8](#), [51](#), and [52](#)) dose and attack rate information is available for multiple subgroups the members of which were exposed to the same isolate. Such subgroups have to be analyzed simultaneously, so the relative uncertainties of the doses to different subgroups might be important as well as the absolute uncertainties in doses. In all the feeding studies the doses to individual subgroups were measured independently, so that independence was taken into account in the analysis (see [Human 02](#) in Appendix A for the method used). For the outbreaks, the uncertainties in dose estimates were generally obtained as a quadrature sum of the uncertainties in an amount of food eaten and a measured concentration in the food. For the purposes of the main analysis it was assumed that the relative doses were exactly known, and the quadrature sum applied to the overall uncertainty in dose. The effect of this, if the uncertainties in food amounts are independent, is to slightly overestimate the uncertainty in absolute dose (since the uncertainties in food amounts would average out somewhat). However, in view of the low sensitivity of the results to the uncertainty estimates (Section 8.8), this approximation was judged acceptable.¹⁰

4.2. Concentrations determined by MPN methods

In the standard MPN (most probable number) approach, known aliquots of a series of well-mixed dilutions (usually dilutions by factors of 10) of the test material are added to a number of sample tubes (usually 3 or 5 at each dilution) containing a medium that promotes growth of the test organism. After sufficient incubation that even one original test organism (or at least one

⁹ That is, taking the square root of the sum of the squares; see, *e.g.* Slob (1994).

¹⁰ The same type of approximation as used for the human feeding studies could have been applied to the food amount uncertainties, and the resultant estimate of absolute dose uncertainty added to the concentration uncertainty in quadrature. The improvement in accuracy would be trivial compared with unknown uncertainties.

CFU) would be detected, each tube is tested for the presence or absence of the test organism.¹¹ Let N be the number of dilutions,¹² and m_i the mass of (the original) test material¹³ added to each tube at dilution i ($1 \leq i \leq N$), and suppose that r_i is the number of tubes with a positive result at dilution i , and n_i is the number of tubes tested at dilution i . Then if z is the concentration of the test organism in the test material, the expected number of test organisms deposited in each tube at dilution i is zm_i , but the actual number will vary randomly according to the Poisson distribution. The probability for one or more of the test organisms to be deposited in a given tube at dilution i , and hence the probability for a positive result from that tube, is then:

$$1 - \exp(-zm_i) \quad (2)$$

and the observation of r_i out of n_i tubes positive at dilution i is then a binomial observation of this probability. The loglikelihood (J) for a given set of results is then given as a function of concentration z by

$$J(z) = \sum_{i=1}^N [r_i \ln(1 - \exp(-zm_i)) - (n_i - r_i)zm_i] \quad (3)$$

(this is just the binomial loglikelihood with probabilities given by Equation (2)). The usual reported MPN, and that used here, is z_0 , the maximum likelihood estimate, so that

$$J(z_0) \geq J(z) \quad \text{for all } z > 0 \quad (4)$$

and the approximate 5th and 95th percentile confidence limits z_{05} and z_{95} are obtained as the two solutions z to

$$2(J(z_0) - J(z)) = (\Phi^{-1}(0.95))^2 = 2.7055 \quad (5)$$

These two estimates define the 90% confidence interval.

For concentrations estimated using MPN methods, the pattern of detects could usually be determined from the available information (e.g. if a 3- or 5-tube MPN method is specified, knowledge of the resulting MPN estimate and an assumption that the usual 10-fold dilution approach was used gives the pattern of results with high probability). In most cases, the 90% confidence interval around the maximum likelihood MPN estimate was evaluated, and the corresponding logarithmic standard deviation ($\ln(\text{GSD})$) estimated as

$$\xi = \frac{\ln(z_{95}/z_{05})}{2 \times \Phi^{-1}(0.95)} \quad (6)$$

where ξ is the logarithm of the GSD estimate, and $\Phi^{-1}(0.95) = 1.6449$ is the 95th percentile of a standard normal distribution. In some cases where the distribution is very wide, ± 1 standard deviation estimates around the maximum likelihood MPN were computed in place of the 90% confidence interval, giving the estimate

¹¹ The procedure described can be modified to account for a probability for individual organisms to fail to grow during the incubation, if that probability is known. The standard approach is to assume that even one organism of interest will grow sufficiently to be detected, and that approach is adopted here.

¹² Usually 3 dilutions are included in the analysis — the highest dilutions with no observable organisms, and the lowest dilutions with all tubes positive, may be omitted.

¹³ The original test material is often diluted and homogenized in suitable medium before the dilution procedure; it is necessary to take account of any such initial dilution in calculating the masses of original test material placed in each sample tube.

$$\xi = \frac{\ln(z_{100x}/z_{100(1-x)})}{2} \quad \text{where } \Phi^{-1}(x) = 1 \quad \text{so } x \approx 0.8413 \quad (7)$$

The lognormal approximation used here compensates quite well for the skew nature of the MPN distribution obtained.

Roussanov *et al.* (1996) have evaluated a Bayesian approach to the likelihood distribution (treating the likelihood itself as a distribution, rather than treating the variation from the maximum likelihood as half the square of a normal variate). That approach gives very similar results to those obtained here. In a few cases examined, the reported MPN was calculated by the Roussanov *et al.* (1996) procedure, so tables of these values were used to infer the pattern of detects in the MPN procedure.

4.3. Concentrations determined using plating methods

Where concentrations were estimated using plating techniques, sufficient information was sometimes available to evaluate the count of organisms obtained. In such cases, the uncertainty in concentration was assumed to be lognormal with a median given by the count, and an uncertainty derived from the Poisson nature of counts. That is, the logarithm of the GSD (ξ) was estimated from

$$\xi = \frac{\ln(R_{95}/R_{05})}{2 \times \Phi^{-1}(0.95)} \quad (8)$$

where R_{95} and R_{05} are 95th and 5th percentile estimates for the expected count μ using a conventional likelihood approach with a Poisson loglikelihood

$$J(\mu) = r \ln(\mu) - \mu \quad (9)$$

where r is the observed count.

For a single plating experiment, the maximum likelihood is

$$J_m = r \ln(r) - r \quad (10)$$

in which case 100x percentile estimate R_{100x} is obtained from

$$2(J_m - J(\mu)) = 2r \left(\ln \frac{r}{R_{100x}} - 1 + \frac{R_{100x}}{r} \right) = (\Phi^{-1}(x))^2 \quad (11)$$

where Φ^{-1} is the inverse of the standard normal distribution function, so that

$$\begin{aligned} \text{if } t &= \Phi^{-1}(x)/\sqrt{r} \\ \text{and } e^\theta - \theta - 1 &= t^2/2 \quad \text{with } \theta \sim t \quad \text{for small } t \\ \text{then } R_{100x} &= r \exp(\theta) \end{aligned} \quad (12)$$

(the condition on θ for small t forces the sign of θ to be the same as that of t).¹⁴

¹⁴ The middle equation of (12) is solved for θ by the VBA function Subtr_exp in the workbook “Epi Attack Data.xls”, so that θ may be obtained using the syntax =Subtr_exp(T) where T is the value t or a cell reference that evaluates to t .

4.4. Generic uncertainty estimates for concentrations

Where no details are provided for the methods used to estimate concentrations, it is assumed that typical MPN or plating methodology was used. The former is assumed to be 3-tube or 5-tube, 3-dilution (with 10-fold dilutions); the latter to be some form of plating (*e.g.* spiral plating, or 10-fold dilutions and standard plates) that results in a count of order 40 being used to define the concentration. Such MPN methods typically have estimates of about 0.5 for ξ ($\ln(\text{GSD})$), while for plating the typical value is about 0.16 (obtained by using a count of 40 and assuming Poisson statistics, see Section 4.2). Where MPN results or plate counts are known, they are used to estimate the uncertainties; otherwise a default value for ξ of 0.5 for MPN, or 0.2 for plate counts is used. Where no method is specified, MPN methodology is assumed.

4.5. Treatment of MPN non-detect concentrations

In some cases, illness was observed after consumption of material known to be contaminated (*e.g.* by plating from a large volume of enriched specimen), but MPN measurements on smaller aliquots of the material were negative. In all such cases included here, the methodology providing the positive result is stated (or was obtained by private communication), so it was possible to incorporate that positive result in the generic MPN approach (Section 4.2). For example, if a single test on a large but specified mass (*e.g.* 25 g) of test material was positive, but a standard MPN test (*e.g.* 3-dilution 5-tube with 1 g, 0.1 g, and 0.01 g of test material per tube at the 3 dilutions) was entirely negative, the positive result corresponds to a single tube test at a different “dilution” (with the stated test material mass) and this extra result was incorporated in the generic MPN approach described (in the example, the augmented MPN results are 1/1, 0/5, 0/5, 0/5 at 25 g, 1 g, 0.1 g, and 0.01 g respectively).

4.6. Both MPN and plating information available

Where both plating and MPN results for the same material are specified, both are treated as equally valid and the analyses given in Sections 4.2 through 4.5 performed. The two results are then combined by averaging on a logarithmic scale with weighting given by the inverse of their logarithmic variances (the square of the standard deviation estimates). Using obvious subscript notation

$$z = \exp \left(\frac{\frac{\ln z_{MPN}}{\xi_{MPN}^2} + \frac{\ln z_{PL}}{\xi_{PL}^2}}{\frac{1}{\xi_{MPN}^2} + \frac{1}{\xi_{PL}^2}} \right) \quad (13)$$

$$\xi = \sqrt{\frac{1}{\frac{1}{\xi_{MPN}^2} + \frac{1}{\xi_{PL}^2}}}$$

where z is the combined concentration estimate, and ξ its estimated $\ln(\text{GSD})$.

4.7. Uncertainties in quantities ingested

Uncertainties in the quantities ingested were estimated in an ad hoc fashion that depended on the description provided in the published reports. Where the quantities were accurately determined, (*e.g.* [Human 02](#) in Appendix A), no uncertainty in quantity was incorporated. In other cases, the estimated uncertainty in quantity dominates the dose uncertainty (*e.g.* [Outbreak 17](#) in Appendix

A). The method used for each report is described in Appendix A. Undoubtedly the uncertainties were not estimated entirely consistently, because of the wide variety of information available in the outbreak reports. However, the results obtained are extremely insensitive to the uncertainties estimated (Section 8.8).

4.8. Total uncertainties in doses

Doses were generally obtained as the product of quantity and concentration, and the uncertainty estimates for the two were generally obtained independently. The uncertainty distributions are approximately lognormal, so the maximum likelihood estimates (MLE) are the medians. The median estimate for total dose and its ln(GSD) are then given by

$$\begin{aligned} d &= mz \\ \sigma &= \sqrt{\xi^2 + \zeta^2} \end{aligned} \quad (14)$$

where the symbols indicate:

z	Concentration of <i>Salmonella</i> in food
m	Mass or volume of food ingested
d	Dose (number of <i>Salmonella</i> ingested)
ξ	Uncertainty standard deviation of the logarithm (ln(GSD)) of concentration
ζ	Uncertainty standard deviation of the logarithm (ln(GSD)) of food mass
σ	Uncertainty standard deviation of the logarithm (ln(GSD)) of dose

For some cases where distinguishable groups were exposed to the same *Salmonella* isolate, the procedures used are more complex (see Section 4.1) and are documented in Appendix A.

5. Uncertainties in attack rates

The uncertainty estimates for attack rates are generally obtained through the assumption that the observed numbers ill are binomial observations of the probability of illness (attack rate) given the numbers exposed. In two outbreaks where additional uncertainties are apparent from the literature, hypothetical “effective” numbers ill and at risk have been obtained from the observed numbers ill and at risk in order to effectively quantify the uncertainties. The methodologies used are included in the descriptions of these particular outbreaks ([Outbreak 15](#) and [Outbreak 26](#)) in Appendix A

6. Sensitive sub-populations

Sensitive subpopulations are assumed to be children and hospital patients. In an outbreak due to contaminated drinking water (Collaborative, 1971; [Outbreak 1](#) in Appendix A, [Serial #1](#)¹⁵) a significantly different attack rate was observed in children aged less than 5 years, compared with ages 5–9 and older age groups. The last two groups had attack rates that were not significantly different; although because of different drinking water rates in the different age groups the estimated virulence differed monotonically with age between all three age groups. There was no difference in attack rates by single years of age below 5 years. Matsui *et al.* (2004; [Outbreak 38](#) in Appendix A, [Serial #51](#)) reported a significantly higher attack rate in elementary school

¹⁵ Serial numbers refer to Table 5, and are hyperlinked to that table.

pupils, age range 6–12 years, compared with equally exposed junior high pupils (age >12 years), while Kasuga *et al.* (2004; [Outbreak 37](#) #6, [Serial #14](#)) reported different attack rates in young children (<5) versus their equally exposed adult teachers. In view of these reports, observations are classified by age ranges of 0–5, 6–12, and >12 years where such information is available, so that there are three classes of sensitive population (hospitalized, ages 0–5, ages 6–12). The population with ages >12 and not hospitalized is considered non-sensitive or normal.¹⁶ Where exact classification by these age ranges is not possible, individuals and groups have been classified into the nearest age range (details for individual outbreaks are given in Appendix A and in the workbook “Epi Attack Data.xls”). Where the outbreak report does not specifically mention the age or hospitalization status of the population exposed, or does not provide sufficient detail to obtain attack rates separately for the subpopulations here considered normal and sensitive, the entire exposed population is treated as normal.

Sensitive populations are assumed to be some fixed factor more sensitive than normal (non-sensitive) populations, in the sense that their response corresponds to the response of the non-sensitive population at a dose higher by that fixed factor. For the two “outbreaks” where the subpopulation (in these cases single individuals) can be classified as both sensitive by reason of age and hospitalization, only one sensitivity factor has been applied (see the discussions for [Outbreak 2](#) and [Outbreak 6](#) in Appendix A). It is a hypothesis of this analysis that, on average, the hospitalized patient class and one age class are equally sensitive.

Only the three outbreaks mentioned in the first paragraph of this section ([Outbreak 1](#), [Serial #1](#); [Outbreak 37](#) #6, [Serial #14](#); and [Outbreak 38](#), [Serial #51](#)) providing strong direct evidence of different sensitivities between the chosen sensitivity classes exposed to the same *Salmonella* isolate under similar conditions. The remaining outbreaks that involve more than one sensitivity class similarly exposed provide only weak comparisons because of the small numbers involved and/or the 0 or 100% attack rates in one or more classes (see Table 5). The outbreaks in only one sensitivity class provide no direct evidence of differences between classes; however, examining virulence estimates for all such outbreaks provides indirect evidence of differences through differences in the shapes or locations of the distributions of those virulence estimates for different sensitivity classes. The analysis method used here (Section 8) incorporates both sorts of comparison, and accounts for the relative strengths and weaknesses of the evidence.

It is quite possible that other sensitive subpopulations exist, such as the immuno-compromised and the elderly. However, introducing any other such subpopulations would not alter the analysis presented here, since there are very few (elderly) or no (immuno-compromised) observational data that allow separation by such identifiable subpopulations.

7. Evaluation of individual reports

Table 5 summarizes the dose-response information used in this document, with MLE for doses, the number ill and exposed at each dose, and the sensitivity class for each group exposed. Each entry in Table 5 is hyperlinked to the individual discussions in Appendix A, and to the corresponding entries in Section 3. A more complete table (together with all the analyses

¹⁶ Non-sensitive is used here as the antithesis of sensitive; non-sensitive populations are the norm.

discussed in Section 8) is included in the accompanying workbook “Epi Attack Data.xls”. Included in the workbook table are the estimates of uncertainty for the doses.

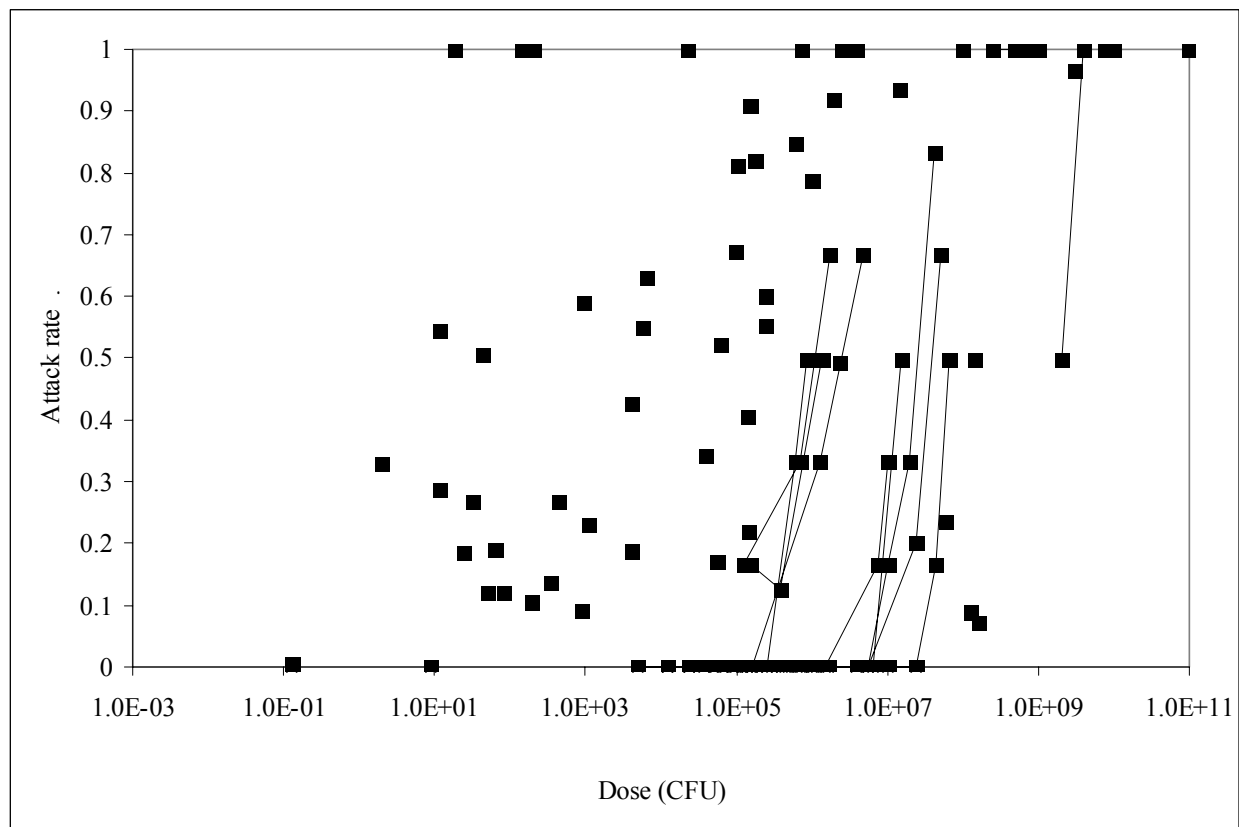


Figure 3 Summary of available data; all distinct point estimates, with dose-response data shown by lines connecting the points.

The complete set of data are summarized in Figure 3, which shows attack rate versus average dose for all the data in Table 5, with lines joining points in the same dose-response series (*i.e.* where exposure is known to be to the same isolate of *Salmonella*). Many of the data points shown, however, are based on very few illnesses in very few exposed individuals. More informative is Figure 4, showing estimates of probability of illness versus average dose in all cases where there were more than 10 persons exposed, together with 98% confidence intervals on both dose and probability of illness.¹⁷ The confidence intervals on probability of illness assume binomial statistics for the number ill and exposed, while those for the dose are based on the dose uncertainty discussed in Section 4 and in the individual analyses of Appendix A.

¹⁷ In these figures, the two entries in Table 5 where only an upper bound on dose is available (Serial #s [3](#) and [41](#)) have been arbitrarily plotted at 1/2 that upper bound and with a dose uncertainty extending to the left side of the graph. Serial #[41](#) is censored from Figure 4. See the worksheet Attack-pict in workbook “Epi Attack Data.xls” for the derivation of these figures.

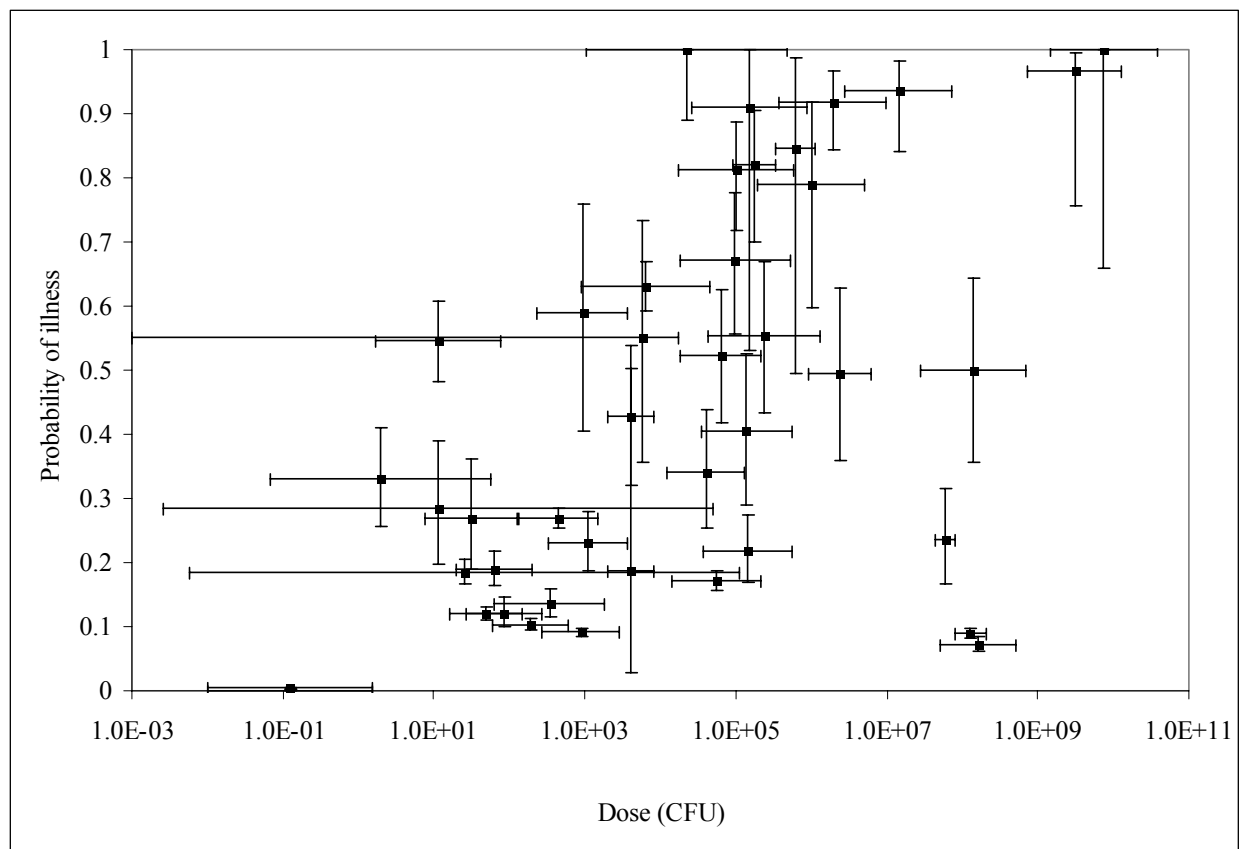


Figure 4 Summary of available data, censoring all exposure groups smaller than 11, with 98% confidence intervals on dose and probability of illness.

Table 5 Summary of dose-response information used in this evaluation. Doses are MLE.

Serial # ^a	ID ^b	FAO/WHO # ^c	Reference	Serotype and Phage type ^d	S ^e	Food	Dose (CFU)	Ill	Exposed
<u>1</u>	FSIS #1	<u>1</u>	Boring <i>et al.</i> , 1971; Collaborative Report, 1971	<i>typhimurium</i> PT 2		Water	1.92E+02	666	6,423
					M		8.30E+01	139	1,149
					Y		6.37E+01	230	1,216
<u>2</u>	FSIS #5 FSIS #8	<u>2</u>	Fontaine <i>et al.</i> 1980	<i>heidelberg</i>		Cheese	1.93E+00	71	215
<u>3</u>	FSIS #9	<u>5</u>	Armstrong <i>et al.</i> 1970	<i>typhimurium</i> PT 2a		Chiffonade	<11,300	22	40
<u>4</u>	FSIS #13	<u>4</u>	Angelotti <i>et al.</i> , 1961	<i>infantis</i>		Ham	2.44E+06	8	8
<u>5</u>	FSIS #15 FSIS #16	<u>12</u>	Hennessy <i>et al.</i> 1996;	<i>enteritidis</i> PT 8		Ice cream	4.34E+01	2.67 ^f	5.28 ^f

			Vought & Tatini, 1998						
<u>6</u>	<u>FSIS #17</u> <u>FSIS #18</u>	<u>11</u>	Levy <i>et al.</i> , 1996	<i>enteritidis</i> PT 8		Hollandaise sauce	2.26E+04	39	39
<u>7</u>	<u>FAO/WHO #13</u>	<u>13</u>	Taylor <i>et al.</i> , 1984	<i>typhimurium</i>	Y	Ice cream	1.00E+08	1	1
					M		2.50E+08	1	1
					M		5.00E+08	2	2
							5.00E+08	1	1
							7.50E+08	2	2
							1.00E+09	1	1
<u>8</u>	<u>FAO/WHO #17</u>	<u>17</u>	Greenwood & Hooper, 1983	<i>napoli</i>	M	Chocolate	1.71E+01	1	1
							8.84E+00	0	2
<u>9</u>	<u>Addl. 1 #1</u>	<u>26</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 1	M	Peanut dressing	2.50E+01	418	2,267
<u>10</u>	<u>Addl. 1 #2</u>		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>	M	Peanut dressing	3.42E+02	179	1,320
<u>11</u>	<u>Addl. 1 #3</u>	<u>20</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 22	M	Beef and bean sprouts with sesame dressing	8.80E+02	967	10,552
<u>12</u>	<u>Addl. 1 #4</u>	<u>24</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 1	M	Spinach with peanut dressing	4.84E+01	644	5,320
<u>13</u>	<u>Addl. 1 #5</u>		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>	Y	Macaroni salad	4.00E+04	52	152
<u>14</u>	<u>Addl. 1 #6a</u>	<u>25</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Chicken and eggs on rice	4.05E+03	3	16
	<u>Addl. 1 #6b</u>				Y		4.05E+03	50	117
<u>15</u>	<u>Addl. 1 #7</u>	<u>32</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i>	Y	Egg salad	3.11E+01	42	156
<u>16</u>	<u>Addl. 1 #8</u>	<u>29</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 1	H	Tartar sauce	1.13E+01	36	126
<u>17</u>	<u>Addl. 1 #9</u>	<u>27</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i>	H	Natto with raw eggs	6.00E+07	45	191
<u>18</u>	<u>Addl. 1 #10</u>	<u>28</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 4	H	Grated yam diluted with soup	1.44E+05	75	343
<u>19</u>	<u>Addl. 1</u>		Kasuga <i>et al.</i>	<i>typhimurium</i>	H	Grated yam	1.39E+05	40	99

	#11		2004			diluted with soup			
20	Addl. 1 #12		Kasuga <i>et al.</i> 2004	<i>typhimurium</i>	H	Grated yam diluted with soup (with quail eggs)	2.40E+06	39	79
21	Addl. 1 #13	18	Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 4		Seared beef	2.40E+05	3	5
22	Addl. 1 #14	30	Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 1		Thin omelet	6.35E+03	558	885
23	Addl. 1 #15		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Omelet	1.50E+05	10	11
24	Addl. 1 #16		Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 1		Broiled tiger prawn with egg yolk	9.60E+04	70	104
25	Addl. 1 #17		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Plain rolled egg	1.14E+01	198	363
26	Addl. 1 #18		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Scallop cream sauce	1.00E+06	30	38
27	Addl. 1 #19		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Natto with raw eggs	7.20E+05	9	9
28	Addl. 1 #20	19	Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Grated yam diluted with soup	1.92E+06	113	123
29	Addl. 1 #21		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Spaghetti salad	1.44E+07	73	78
30	Addl. 1 #22		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Omelet	2.40E+05	57	103
31	Addl. 1 #23		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Bavarois	1.00E+05	100	123
32	Addl. 1 #24		Kasuga <i>et al.</i> 2004	<i>bareilly o7</i>		Sauce for octopus pancake	1.40E+08	34	68
33	Addl. 1 #25	33	Kasuga <i>et al.</i> 2004	<i>oranienburg o7</i>		Grated yam diluted with soup	7.50E+09	11	11
34	Addl. 1 #26		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Green-tea ice cream (cheese cake ignored)	3.80E+06	5	5
35	Addl. 1 #27		Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 4		Three-layer cake	4.39E+02	1,371	5,103

<u>36</u>	<u>Addl. 1 #28</u>		Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Tiramisu	1.28E+08	697	7,873
<u>37</u>	<u>Addl. 1 #29</u>	<u>23</u>	Kasuga <i>et al.</i> 2004	<i>enteritidis</i>		Cake	6.00E+05	11	13
<u>38</u>	<u>Addl. 1 #30</u>		Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 1		Sherbet	1.76E+05	68	83
<u>39</u>	<u>Addl. 1 #31</u>		Kasuga <i>et al.</i> 2004	<i>enteritidis</i> PT 4		Homemade mayonnaise in bread	5.47E+04	498	2,907
<u>40</u>	<u>FSIS # 6</u> <u>FAO/WHO #15</u>		George 1976	<i>schwarzengrund</i>	Y	Pancreatic extract	2.00E+02	1	1
<u>41</u>	<u>FSIS #2</u>	<u>16</u>	Lipson 1976	<i>schwarzengrund</i>	Y	Pancreatin (porcine, powdered)	<291	1	1
<u>42</u>	<u>M&E A_I</u>		McCullough and Eisele, 1951a	<i>anatum</i>		Eggnog	1.20E+04	0	5
							2.40E+04	0	6
							6.60E+04	0	6
							9.30E+04	0	6
							1.41E+05	0	6
							2.56E+05	0	6
							5.87E+05	2	6
							8.60E+05	3	6
<u>43</u>	<u>M&E A_II</u>		McCullough and Eisele, 1951a	<i>anatum</i>		Eggnog	8.90E+04	0	6
							4.48E+05	0	6
							1.04E+06	0	6
							3.90E+06	0	6
							1.00E+07	0	6
							2.39E+07	0	6
							4.45E+07	1	6
							6.73E+07	4	8
<u>44</u>	<u>M&E A_III</u>		McCullough and Eisele, 1951a	<i>anatum</i>		Eggnog	1.59E+05	0	6
							1.26E+06	2	6
							4.68E+06	4	6
<u>45</u>	<u>M&E M_I</u>		McCullough and Eisele,	<i>meleagridis</i>		Eggnog	1.20E+04	0	6
							2.40E+04	0	6

			1951a				5.20E+04	0	6
							9.60E+04	0	6
							1.55E+05	0	6
							3.00E+05	0	6
							7.20E+05	0	5
							1.15E+06	0	6
							5.50E+06	0	6
							2.40E+07	1	5
							5.00E+07	4	6
46	M&E M II		McCullough and Eisele, 1951a	<i>meleagridis</i>		Eggnog	1.00E+06	0	6
							5.50E+06	0	6
							1.00E+07	1	6
							2.00E+07	2	6
							4.10E+07	5	6
47	M&E M III		McCullough and Eisele, 1951a	<i>meleagridis</i>		Eggnog	1.58E+05	0	6
							1.50E+06	0	6
							7.68E+06	1	6
							1.00E+07	2	6
48	M&E N		McCullough and Eisele, 1951c	<i>newport</i>		Eggnog	1.52E+05	1	6
							3.85E+05	1	8
							1.35E+06	3	6
49	M&E B		McCullough and Eisele, 1951c	<i>bareilly</i>		Eggnog	1.25E+05	1	6
							6.95E+05	2	6
							1.70E+06	4	6
50	M&E D		McCullough and Eisele, 1951c	<i>derby</i>		Eggnog	1.39E+05	0	6
							7.05E+05	0	6
							1.66E+06	0	6
							6.40E+06	0	6
							1.50E+07	3	6
51	Addl. 2		Matsui <i>et al.</i> 2004	<i>enteritidis</i> PT 1	M	Dessert bun	1.26E-01	92	18,571
							1.26E-01	3	4,265
52	Addl. 3		Hormaeche <i>et al.</i> 1936	<i>typhimurium</i>		Physiological solution	2.00E+09	1	2
							4.00E+09	3	3

<u>53</u>	<u>FSIS #12</u> <u>FSIS #14</u>		Reitler <i>et al.</i> , 1960	<i>zanzibar</i>	M	Goat cheese	1.00E+11	2	2
							1.00E+11	3	3
<u>54</u>	<u>Addl. 4</u>		Varela and Olarte, 1942	<i>anatum</i>		Emulsion	1.00E+10	1	1
<u>55</u>	<u>Addl. 5 #5</u>		Abe <i>et al.</i> 2004	<i>enteritidis</i>	M	Boiled vegetables with sauce dressing	1.09E+03	107	465
<u>56</u>	<u>Addl. 5 #6</u>		Abe <i>et al.</i> 2004	<i>enteritidis</i>	Y	Three foods with sauce dressing	9.22E+02	26	44
<u>57</u>	<u>Addl. 5 #8</u>		Abe <i>et al.</i> 2004	<i>enteritidis</i>	M	Cabbage, seachicken, and harusame, dressed with sauce	6.31E+04	69	132
<u>58</u>	<u>Addl. 6</u>		Urfur <i>et al.</i> 2000	<i>braenderup</i>		Meat pies	1.61E+08	215	3,000
<u>59</u>	<u>Addl. 7</u>		Greenberg, 1964	<i>typhimurium</i>		Atole	2.50E+05	0	1
							5.00E+05	0	1
							1.20E+05	0	1
							5.00E+04	0	1
							4.00E+05	0	1
							6.40E+05	0	1
							5.00E+03	0	1
							3.80E+04	0	1
<u>60</u>	<u>FAO/WHO</u> <u>#8</u>	<u>8</u>	Narain and Lofgren, 1989	<i>newport</i>		Meat sandwiches	3.06E+09	29.48 ^f	30.53 ^f

^a Hyperlinking is to the discussions in Appendix A.

^b Hyperlinking is to the first mention of this outbreak in Section 3.

^c Hyperlinking is to the entry in Table 4, where appropriate.

^d Phage type is listed where it is reported in the publication(s). Where more than one is reported, the majority phage type is listed.

^e Sensitive population. Y indicates young age (generally less than 5 years old), M indicates medium age (generally 6 to 12 years old), H represents a hospitalized population. Blank indicates a non-sensitive (normal) population.

^f See the discussion of these studies in Appendix A for an explanation of the non-integer values given here.

8. Dose response evaluation

8.1. General considerations for dose-response models

For each individual isolate of *Salmonella*, we assume a heuristic population dose-response model of the form

$$p = f(\chi, d) \quad (15)$$

where the symbols mean:

- d the (same) dose, an integer number of *Salmonella* organisms, to every member of that population,
- χ a virulence parameter (more generally a list of parameters one or more of which may be assigned as a virulence),
- p the probability for illness in that population in the sense that the expected number of illnesses in finite samples of the population, with expectation taken over all possible samples (without replacement), is given as the product of p and the size of the finite population sample exposed, and
- f the mathematical form of the dose-response relationship. f is expected to be a monotonically increasing function of d and to be monotonic in χ .

This formulation is chosen as the most general that applies to the situations examined. The dose-response model is heuristic in that no inquiry is here made about mechanistic implications, if any, for individuals in the population or for populations as a whole. It is assumed that individuals in the population are indistinguishable at the level of detail examined, although that does not mean that individuals are identical.

There are multiple ways in which such a (population) dose-response curve could arise from individual dose-response curves, assuming such individual dose-response curves can be defined. For example:

A. Entirely probabilistic and equal: Every member of the population has the same independent probability to become ill at a given dose, that probability is given by an equation of the same form as Equation (15), and there are no interactions between individuals.

B. Entirely probabilistic and unequal: Every member of the population has a different independent probability to become ill at a given dose, and the average value of these probabilities over the population is given by Equation (15), again with no interactions between individuals.

C. Entirely deterministic: Every member of the population will become ill if fed more than a certain threshold number of organisms that may differ for every member of the population, and will remain well if fed any less than that number of organisms. If the cumulative distribution P of threshold doses D among the population is given by

$$P = f(\chi, D) \quad (16)$$

then Equation (15) will be the result for the population dose-response.

D. Combinations of these, and other possibilities.

Given this wide variety of possibilities, the assumption of a particular form for the population dose-response of Equation (15) cannot have strong mechanistic implications, since mechanisms act primarily at the individual level for *Salmonella*-induced illnesses and no evaluation of secondary cases or other potential interactions between individuals is performed here. In the same way that mechanistic implications are masked, variations between individuals are also incorporated within the dose-response function used. Such variations include those intrinsic to the individual, such as any inherent variation between individuals in susceptibility, as well as those that depend on exposure circumstances, such as differences in the foods contaminated with *Salmonella* and in other foods eaten around the same time as the contaminated food. Again, no inquiry is made here into such variations.

In any finite size randomly selected population sample, Equation (15) gives the illness probability, and the observed number of illnesses is a binomial observation of that illness probability. The binomial distribution arises from the randomness of the selection and the indistinguishability of the individuals, not from any putative probabilistic nature of individual dose-response relationships. For the analysis of the outbreaks and experiments examined here, it is assumed that the populations exposed are randomly selected from the relevant populations. The relevant populations are the non-sensitives (all persons >12 years of age and not hospitalized), and the three sensitive populations (0–5 years of age, 6–12 years of age, and hospitalized).

The dose-response model given by Equation (15) is for a single *Salmonella* isolate at a fixed dose to every person exposed, so matches well the experimental situation (*e.g.* the experiments of McCullough and Eisele, 1951a,c) provided different virulence parameters χ are allowed for different isolates. It is not assumed *a priori* that a single dose-response model with identical parameters applies to all isolates of *Salmonella*. Such an assumption is untenable for the experimental observations (Coleman and Marks, 1998; Coleman *et al.*, 2004), and is shown below to be untenable for outbreak observations also. By contrast, the analysis of FSIS (1998) and FAO/WHO (2002b) assumed that an identical dose-response model applied to all *Salmonella* isolates, without inquiring into the truth or falsity of the proposition.

Equation (15) defines a model for the probability of illness in a population in which every individual is fed exactly the same dose. Even in experimental settings, unless every *Salmonella* organism is counted individually, such a circumstance is unlikely. In practice experimentally, and certainly in outbreak situations, there will be different doses to different individuals. Such differences may be modeled by a distribution $g(d)$ of doses in the population examined. It is assumed here that for the cases examined, and to an adequate approximation, the probability of illness is independent of the distribution of doses and may be modeled adequately by using Equation (15) with the mean dose D . That is, it is assumed that the approximation

$$\sum_{d=1}^{\infty} g(d) f(\chi, d) \approx f(\chi, D) \quad \text{where} \quad D = \sum_{d=1}^{\infty} d g(d) \quad (17)$$

is adequate. Appendix D examines this approximation for its practical effect on the main analysis herein.

8.2. Primary dose-response model

There is wide variation in illness-producing doses in different outbreaks, even those involving the same *Salmonella* serotype (in particular for *S. enteritidis*), indicating that a single dose-response curve can only be expected for individual isolates (or strains), not serotypes.

Theoretically it is possible that doses as low as 1 CFU might cause illness, and observations in outbreaks (particularly Matsui *et al.* (2004) and D'Aoust (1985) confirm that very low doses¹⁸ are capable of causing illness with reasonably high probability.

It appears (see Section 6) from two particular outbreak observations in which both young children and adult (or older children) populations were involved, so that both were infected with the same isolate of *Salmonella*, that children are more likely to become ill. The same is postulated for hospitalized patients, and possibly the immuno-compromised.

There are no data that adequately distinguish between possible different shapes of dose-response curve for illness due to single isolates of *Salmonella*. Theoretically, it may be expected that individual infecting cells might act independently, leading to the possibility of linearity at low doses.

In view of the foregoing observations, and as a practical matter, the dose-response curve assumed for exposure to a single isolate is assumed to be of the form:

$$p = 1 - \exp(-S\lambda d) \quad (18)$$

where the terms are

- p probability for a random member of a population to become ill after each member of that population is dosed with d CFU *Salmonella*,
- S a factor to account for the sensitivity class of the population, unity for non-sensitive population,
- λ a virulence for the *Salmonella* isolate to cause illness in a non-sensitive population, and
- d the dose of the *Salmonella* isolate (CFU).

The dose-response curve given by Equation (18) is here called linear-exponential, abbreviated LE, or just linear. Others have called this dose-response curve “exponential.”

Strictly speaking, Equation (18) applies only for integer values of dose d in situations where every member of a population gets the same dose. In practice (see the discussion near Equation (17)) it is used for exposed populations using the population mean dose (see also Appendix D).

This is the dose-response curve that would be expected if every *Salmonella* cell ingested had the same probability ($S\lambda$) to cause illness, although that hypothesis is not necessary for such a dose-response relationship to hold (see the discussion in Section 8.1). The practical considerations mentioned include the need for a one-parameter model, since only a single parameter can be

¹⁸ At least as inferred from available measurements of CFU; it is possible that the measurements are biased low (*e.g.* by failure to recover organisms, or by clumping of the organisms), but in that case if all measurement methods are similarly biased the analysis of this work applies to the measured concentrations.

estimated from most of the available observations (both outbreaks and feeding studies).¹⁹ The linear-exponential form conveniently summarizes (in the virulence parameter λ) the propensity of a *Salmonella* isolate to cause illness. This dose-response curve is called, elsewhere in this work, the “primary” dose-response curve.

The shape assumed for the dose-response curve for individual isolates of *Salmonella* is of minor importance in the analysis performed in this work. This is confirmed in Section 8.7, where an alternative, very different shape is assumed, but with a similar overall outcome. For this analysis, the linear-exponential dose-response curve has the following advantages:

- a. there is at least one biologically plausible mechanistic model that gives this dose-response shape at least for individuals,
- b. it adequately fits the available human experimental dose-response data (see the following paragraphs),
- c. it contains only one parameter, the virulence, that conveniently summarizes the potential for a particular *Salmonella* isolate to cause illness — only one parameter can be estimated from many of the outbreaks, since only a single attack rate and single dose estimate can be obtained from published information,
- d. for those outbreaks where only the average dose to a potentially exposed population can be estimated, not individual doses, the virulence parameter estimated using this average dose is a very good approximation to the parameter that would be obtained using individual doses²⁰ (see Appendix D).

The only data suitable for evaluating the shape of dose-response curves are those in which multiple, known doses (or at least, known relative doses) of exactly the same isolate of *Salmonella* are administered to populations under identical circumstances. Only the experiments of McCullough and Eisele (1951a,c) provide suitable information. Table 6 and Table 7 demonstrate that the illness observations in these experiments are adequately explained by dose-response curves of linear-exponential form. Table 6 shows standard statistics for evaluation of fits with the additional assumption that the relative doses are known exactly, while Table 7 shows that the uncertainties in dose estimates do not affect this evaluation.²¹

Table 6 shows the maximum loglikelihood for the observations, assuming linear-exponential dose-response curves and the loglikelihood for binomial observations, that is

$$J = \sum_{i=1}^N \left(r_i \ln \left(\frac{p_i n_i}{r_i} \right) + (n_i - r_i) \ln \left(\frac{(1 - p_i) n_i}{n_i - r_i} \right) \right) \quad (19)$$

¹⁹ The linear-exponential dose-response curve contains only one parameter. That parameter is here separated into two components, S and λ , that cannot be separately estimated for any single experiment or outbreak on a uniform population. In evaluating a single experiment or outbreak on a uniform population S should be regarded as known — it is unity by definition for non-sensitive populations. S can be evaluated for sensitive populations by evaluating those situations where both sensitive and non-sensitive populations are simultaneously exposed to the same *Salmonella* isolate; or (with much less precision) by comparing the distribution of values obtained in sensitive and non-sensitive populations exposed to multiple *Salmonella* isolates; or both. The approach adopted here does both.

²⁰ This is a consequence of the near linearity of the primary dose-response curve. The same is not true for highly non-linear dose-response curves (see Section 8.7).

²¹ The calculations for Table 6 and Table 7 are in the sheet McCullough in “Epi Attack Data.xls”.

where for $i = 1$ to N dose groups there are observations of r_i illnesses in a population of n_i fed an exact dose d_i , and the probability of illness p_i is given by the linear-exponential dose-response Equation (18) with $S = 1$ and $d = d_i$ (terms in the sum with $r_i = 0$ or $r_i = n_i$ are interpreted as zero). The only free parameter (for each experiment on the same *Salmonella* isolate) is the virulence λ , and the loglikelihood is maximized with respect to this parameter. The normalization adopted here for the likelihood makes the loglikelihood vanish if the observations fit the data exactly, so that twice the absolute value of the maximum loglikelihood may be roughly compared with a chi-squared variate with $N - 1$ degrees of freedom. The p-values shown are obtained from an X^2 statistic defined at the maximum likelihood parameter values by

$$X^2 = \sum_{i=1}^n \frac{(r_i - p_i n_i)^2}{n_i p_i (1 - p_i)} \quad (20)$$

which should behave approximately like a chi-squared variate with approximately $N - 1$ degrees of freedom, although number of degrees of freedom may be closer to one less than the number of dose groups with non-zero response (both estimates are shown). None of the measures of goodness of fit indicate any lack of fit.

Table 6 Adequacy of fit of the linear-exponential model to the human experimental data (McCullough and Eisele, 1951a,c; Serial #s [42](#) through [50](#)) assuming exactly known doses.

	Dose groups	Dose groups with non-zero response	Maximum loglikelihood	X^2 statistic	X^2 -based p-value using # of dose groups	X^2 -based p-value using # of non-zero doses
S. anatum I	8	2	-2.147	2.68	0.91	0.10
S. anatum II	8	2	-1.973	2.74	0.91	0.10
S. anatum III	3	2	-0.305	0.38	0.83	0.54
S. meleagridis I	11	2	-1.080	1.46	1.00	0.23
S. meleagridis II	5	3	-1.620	2.19	0.70	0.33
S. meleagridis III	4	2	-0.409	0.54	0.91	0.46
S. newport	3	3	-0.363	0.87	0.65	0.65
S. bareilly	3	3	-0.258	0.64	0.73	0.73
S. derby	5	1	-1.788	2.39	0.66	0.53
All combined ^a	50	20	-9.944	13.90	1.00	0.24
^a Summing all the individual loglikelihood and X^2 values, and accounting for the 9 parameters for the p-values.						

Table 7 takes account of the uncertainties of the doses, approximated here by lognormal uncertainty distributions and characterized by their $\ln(\text{GSD})$ (see Section 4.8). The loglikelihood in this case is given by

$$J = \sum_{i=1}^N \ln \left\{ \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left(-\frac{(t - \ln d_i)^2}{2\sigma_i^2} \right) \left(\frac{p_i n_i}{r_i} \right)^{r_i} \left(\frac{(1-p_i)n_i}{n_i - r_i} \right)^{n_i - r_i} dt \right\} \quad (21)$$

using the same nomenclature as for the exactly known dose case (and terms in the integrands with $r_i = 0$ or $r_i = n_i$ are interpreted as unity), except that d_i is now considered the median of a lognormal uncertainty distribution with $\ln(\text{GSD})$ σ_i , and within each integral,

$$p_i = 1 - \exp(-\lambda e^t) \quad (22)$$

This loglikelihood may also be written as²²

$$J = \sum_{i=1}^N \ln \left\{ \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left(-\frac{(x - \ln \lambda)^2}{2\sigma_i^2} \right) \left(\frac{p_i n_i}{r_i} \right)^{r_i} \left(\frac{(1-p_i)n_i}{n_i - r_i} \right)^{n_i - r_i} dx \right\} \quad (23)$$

with now

$$p_i = 1 - \exp(-e^x d_i) \quad (24)$$

For these McCullough and Eisele (1951a,c) experiments, $\sigma_i = 0.2$ (see the discussion of these experiments at [Human 2](#) in Appendix A). This likelihood is normalized in the same way as for Equation (19). Again, Table 7 gives no indication of any lack of fit.

While Table 6 and Table 7 do not show any lack of fit of the linear-exponential dose-response curve shape, the available data do not discriminate much against different shapes. It is shown in Section 8.7 (Table 16 and Table 17) that an alternative, very different, dose-response shape fits equally well.

²² In each integral in Equation (21) make the change of variable $\lambda e^t = d_i e^x$. Each term of Equation (23) is a special case of the integral in Equation (25) and is evaluated in the same way (see Appendix C).

Table 7 Adequacy of fit of the linear-exponential model to the human experimental data (McCullough and Eisele, 1951a,c; Serial #s [42](#) through [50](#)), accounting for the uncertainties in doses.

	Dose groups	Dose groups with non-zero response	Maximum loglikelihood	X^2	X^2 -based p-value using # of dose groups	X^2 -based p-value using # of non-zero doses
S. anatum I	8	2	-2.195	2.76	0.91	0.10
S. anatum II	8	2	-1.994	2.82	0.90	0.09
S. anatum III	3	2	-0.404	0.39	0.82	0.53
S. meleagridis I	11	2	-1.133	1.49	1.00	0.22
S. meleagridis II	5	3	-1.683	2.21	0.70	0.33
S. meleagridis III	4	2	-0.459	0.55	0.91	0.46
S. newport	3	3	-0.445	0.88	0.64	0.64
S. bareilly	3	3	-0.373	0.64	0.73	0.73
S. derby	5	1	-1.793	2.45	0.65	0.52
All combined ^a	50	20	-10.477	14.19	1.00	0.22
^a Summing all the individual loglikelihood and X^2 values, and accounting for the 9 parameters for the p-values.						

8.3. Observed variation of virulence

The virulence to cause illness (λ) is expected to vary between *Salmonella* isolates; and that variation may contain various components. To obtain an initial characterization of the variation, maximum likelihood and 95% confidence limits were obtained for the virulence for each observation in Table 5, using the likelihood of Equation (23).²³ Where populations from more than one sensitivity class (Section 6) were exposed, virulence estimates were obtained independently for each such sensitivity class. Table 8 lists the resulting maximum likelihood virulence estimates and their 95% confidence limits. There are 68 estimates, 44 for non-sensitive populations, 8 for children 0–5 yrs, 11 for children 6–12 yrs, and 5 for hospitalized individuals; although only 52 of the maximum likelihood estimates are both non-zero and finite.²⁴

²³ Table 5 does not contain the estimates of dose uncertainty σ_i . These may be found in worksheet All_data in the accompanying workbook “Epi Attack Data.xls”.

²⁴ Appendix B.7 summarizes the mechanical details of obtaining these estimates.

It was found that the 52 non-zero and finite maximum likelihood estimates were consistent with being random samples from a lognormal distribution ($p = 0.33$, Shapiro-Wilk test²⁵). Figure 5 shows these 52 estimates together with their estimated confidence limits on a lognormal plot (Cunnane, 1978).²⁶ Subsequent analyses evaluate sub-sets of these data, and generally such sub-sets were found also to be consistent with random samples from lognormal distributions, insofar as testing for agreement was possible. In view of these observations that a simple distribution shape (lognormal) was adequate to describe the variation in virulence estimates, no further search for other potentially suitable distribution shapes was undertaken.

Figure 6 shows the same data as Figure 5, but separated by sensitivity class and with the non-sensitive class separated by serotype into *S. enteritidis* and all others. *S. enteritidis* is specifically separated because most of the data are for this serotype (18/34 plotted non-sensitive entries in Table 8, 10/13 plotted entries in the 6–12 yr and hospitalized class, and 4/5 plotted entries in the 0–5 yr class). In both figures, the experimental data (as opposed to outbreak data), serial numbers [42](#) through [50](#), [52](#), [54](#), and [59](#), are distinguished by circles round the individual symbols.

²⁵ This p-value, and others below giving Shapiro-Wilk test statistics, should be interpreted cautiously: the values being tested are themselves uncertain, so the assumptions of the Shapiro-Wilk test are not valid. The p-values are given as a convenient and objective indication of the deviations of the set of values plotted from a lognormal distribution, but they probably underestimate the true p-value (hence may tend to indicate deviations from lognormality where none exist).

²⁶ Random samples from lognormal distributions will plot roughly as a straight line on such plots, with random deviations. The Shapiro-Wilk test measures conformity to the expected shape of these plots.

Table 8 Maximum likelihood estimates and 95% confidence limits for virulence

Serial # ^a	Serotype	Phage Type ^b	S ^c	MLE Virulence (per CFU)	LCL95 ^d	UCL95 ^e
1	<i>S. typhimurium</i>	PT 2		5.1E-04	2.2E-04	1.1E-03
1	<i>S. typhimurium</i>	PT 2	M	1.5E-03	6.7E-04	3.6E-03
1	<i>S. typhimurium</i>	PT 2	Y	3.3E-03	1.4E-03	7.5E-03
2	<i>S. heidelberg</i>			2.1E-01	1.9E-02	2.2E+00
3	<i>S. typhimurium</i>	PT 2a		9.5E-05	5.7E-05	4.1E-04
4	<i>S. infantis</i>			infinity	7.2E-07	infinity
5	<i>S. enteritidis</i>	PT 8		1.6E-02	4.5E-03	4.3E-02
6	<i>S. enteritidis</i>	PT 8		infinity	7.6E-05	infinity
7	<i>S. typhimurium</i>		Y	infinity	2.9E-09	infinity
7	<i>S. typhimurium</i>		M	infinity	2.5E-09	infinity
7	<i>S. typhimurium</i>			infinity	1.6E-09	infinity
8	<i>S. napoli</i>		M	infinity	1.2E-02	infinity
8	<i>S. napoli</i>			zero	zero	9.8E-02
9	<i>S. enteritidis</i>	PT 1	M	8.1E-03	2.2E-05	3.1E+00
10	<i>S. enteritidis</i>		M	4.3E-04	1.3E-04	1.4E-03
11	<i>S. enteritidis</i>	PT 22	M	1.1E-04	4.7E-05	2.5E-04
12	<i>S. enteritidis</i>	PT 1	M	2.7E-03	1.2E-03	5.8E-03
13	<i>S. enteritidis</i>		Y	1.0E-05	4.3E-06	2.5E-05
14	<i>S. enteritidis</i>			5.0E-05	1.5E-05	1.3E-04
14	<i>S. enteritidis</i>		Y	1.4E-04	7.9E-05	2.3E-04
15	<i>S. enteritidis</i>		Y	1.0E-02	3.6E-03	2.8E-02
16	<i>S. enteritidis</i>	PT 1	H	2.9E-02	7.8E-05	1.1E+01
17	<i>S. enteritidis</i>		H	4.5E-09	3.2E-09	6.2E-09
18	<i>S. enteritidis</i>	PT 4	H	1.7E-06	6.5E-07	4.5E-06
19	<i>S. typhimurium</i>		H	3.7E-06	1.3E-06	1.0E-05
20	<i>S. typhimurium</i>		H	2.8E-07	1.4E-07	5.8E-07
21	<i>S. enteritidis</i>	PT 4		3.5E-06	6.4E-07	1.7E-05
22	<i>S. enteritidis</i>	PT 1		1.6E-04	4.0E-05	6.2E-04
23	<i>S. enteritidis</i>			1.6E-05	3.8E-06	6.5E-05
24	<i>S. enteritidis</i>	PT 1		1.2E-05	3.5E-06	3.9E-05

25	<i>S. enteritidis</i>			6.9E-02	1.8E-02	2.6E-01
26	<i>S. enteritidis</i>			1.5E-06	4.7E-07	5.1E-06
27	<i>S. enteritidis</i>			infinity	2.1E-06	infinity
28	<i>S. enteritidis</i>			1.3E-06	4.1E-07	4.2E-06
29	<i>S. enteritidis</i>			1.9E-07	5.9E-08	6.2E-07
30	<i>S. enteritidis</i>			3.3E-06	9.7E-07	1.1E-05
31	<i>S. enteritidis</i>			1.7E-05	4.7E-06	5.9E-05
32	<i>S. bareilly o7</i>			4.9E-09	1.5E-09	1.6E-08
33	<i>S. oranienburg o7</i>			infinity	2.2E-10	infinity
34	<i>S. enteritidis</i>			infinity	3.7E-07	infinity
35	<i>S. enteritidis</i>	PT 4		7.1E-04	3.0E-04	1.7E-03
36	<i>S. enteritidis</i>			7.2E-10	5.2E-10	1.0E-09
37	<i>S. enteritidis</i>			3.1E-06	1.5E-06	6.2E-06
38	<i>S. enteritidis</i>	PT 1		9.7E-06	5.8E-06	1.6E-05
39	<i>S. enteritidis</i>	PT 4		3.4E-06	1.3E-06	9.0E-06
40	<i>S. schwarzengrund</i>		Y	infinity	1.4E-03	infinity
41	<i>S. schwarzengrund</i>		Y	infinity	2.2E-03	infinity
42	<i>S. anatum</i>			4.7E-07	2.0E-07	9.3E-07
43	<i>S. anatum</i>			5.5E-09	2.3E-09	1.1E-08
44	<i>S. anatum</i>			2.4E-07	1.1E-07	4.8E-07
45	<i>S. meleagridis</i>			1.4E-08	5.9E-09	2.9E-08
46	<i>S. meleagridis</i>			2.4E-08	1.2E-08	4.4E-08
47	<i>S. meleagridis</i>			2.9E-08	9.2E-09	6.8E-08
48	<i>S. newport</i>			5.2E-07	2.2E-07	1.0E-06
49	<i>S. bareilly</i>			6.8E-07	3.2E-07	1.3E-06
50	<i>S. derby</i>			2.4E-08	7.7E-09	5.7E-08
51	<i>S. enteritidis</i>	PT 1	M	3.9E-02	6.6E-03	2.3E-01
51	<i>S. enteritidis</i>	PT 1		4.9E-03	6.1E-04	3.6E-02
52	<i>S. typhimurium</i>			5.9E-10	1.5E-10	2.1E-09
53	<i>S. zanzibar</i>		M	infinity	5.4E-12	infinity
53	<i>S. zanzibar</i>			infinity	7.5E-12	infinity
54	<i>S. anatum</i>			infinity	2.8E-11	infinity
55	<i>S. enteritidis</i>		M	2.4E-04	1.0E-04	5.7E-04

56	<i>S. enteritidis</i>		Y	9.5E-04	3.4E-04	2.7E-03
57	<i>S. enteritidis</i>		M	1.2E-05	4.8E-06	2.8E-05
58	<i>S. braenderup</i>			4.6E-10	2.0E-10	1.1E-09
59	<i>S. typhimurium</i>			zero	zero	6.8E-07
60	<i>S. newport</i>			2.7E-10	9.1E-11	8.4E-10

^a Hyperlinking is to the entries in Table 5

^b Where known; the majority phage type where multiple phages types were given.

^c Y = Children 0–5 yrs, M = Children 6–12 yrs, H = Hospitalized

^d LCL95 = lower 95th percentile (profile likelihood estimate)

^e UCL95 = upper 95th percentile (profile likelihood estimate)

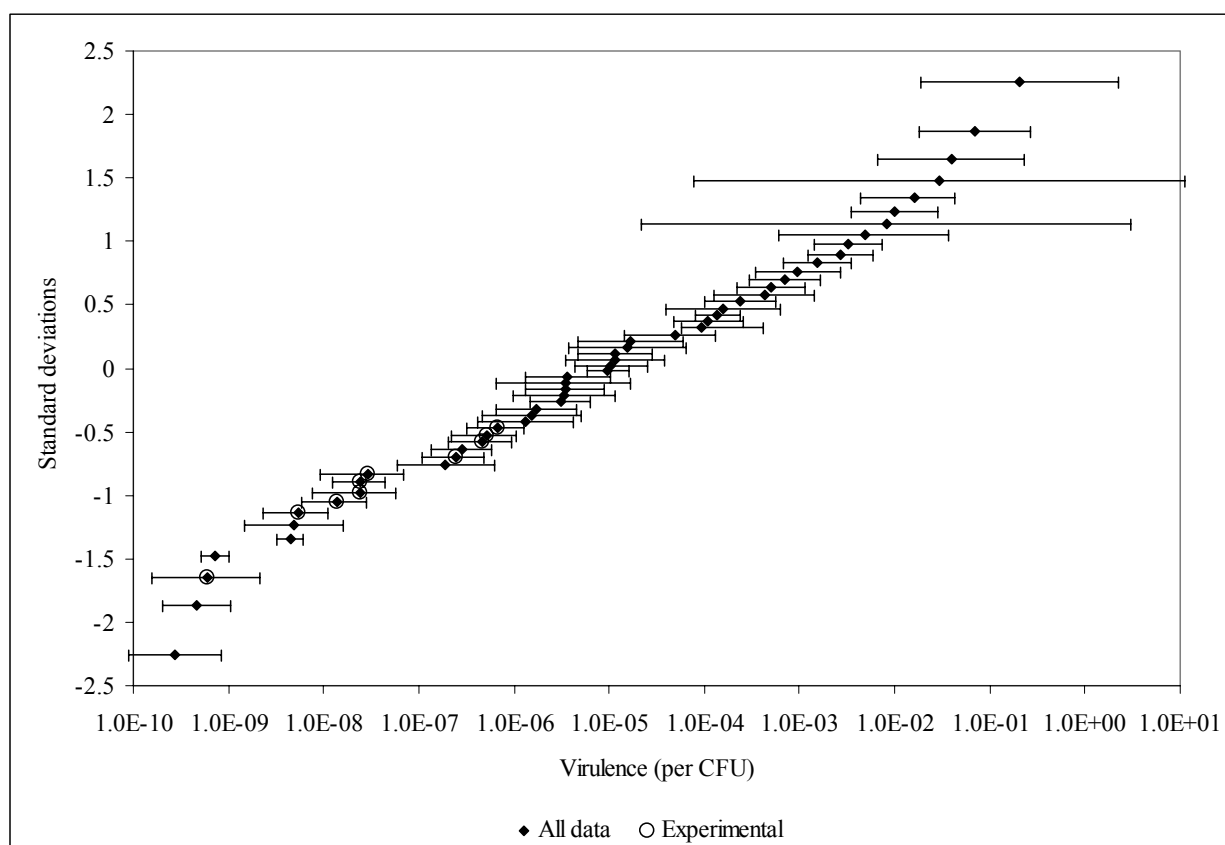


Figure 5 Individual ML virulence estimates and their confidence limits.

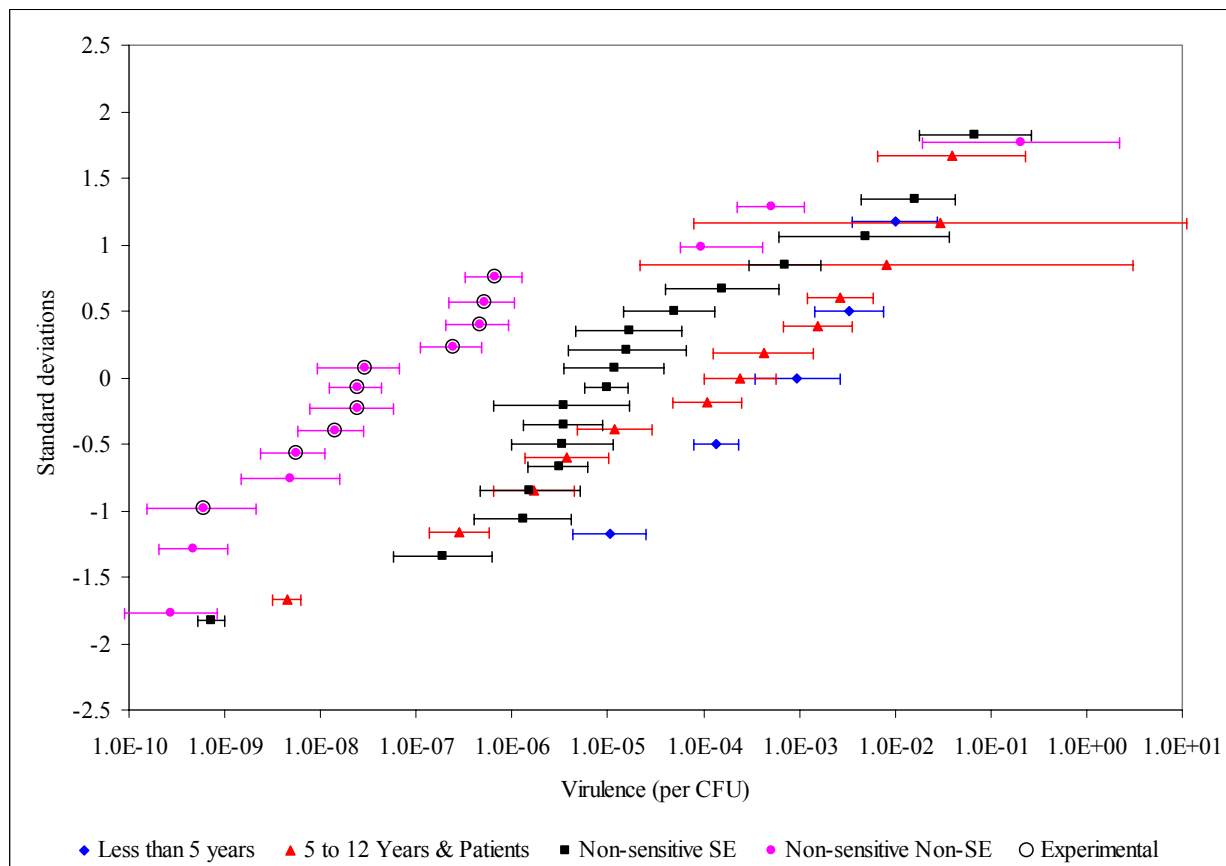


Figure 6 As for Figure 5, showing distributions for sensitivity classes and by serotype (SE = *S. enteritidis*).

Figure 5 demonstrates that there is a very large variation (nearly nine orders of magnitude) in the virulence as defined here (the coefficient of dose in the linear-exponential dose-response curve). The variation seen in Figure 5 is necessarily composed of multiple components, including variation due to:

- inherent differences between strains
- circumstances of exposure, including:
 - food or food type
 - physiological state of the organism
 - previous history of the food
 - co-eaten foods
- residual variation in the human groups exposed

For outbreaks, the primary source of the data used here, the only data generally available are a very general description of the population exposed, the food involved, and the serotype (with occasionally the phage type). The selection of the data is from rather special, investigated, outbreak-causing circumstances (hence from some combination of food, circumstances of exposure, the exposure of two or more people, from circumstances sufficiently distinguished to have been investigated and in which dose estimates were made). It is an open question whether the selection is in any sense random, and there are clearly insufficient data to separately evaluate

all the potential components of the variation seen in Figure 5. For the three particular components for which there is information (population characteristics, food involved, and serotype), the following approach was adopted.

Population characteristics: there may be variation between individuals' susceptibilities to illness due to exposure under standard conditions, and an individual's susceptibility may vary depending on the circumstances of exposure and the previous history of the individual. However, individual variation beyond that captured by the classification into susceptibility classes (Section 6) is here supposed subsumed into the dose-response curve (see Section 8.1), since that is defined as an average over a population. Nevertheless, because of the finite size of the populations exposed, there will be some residual variation due to any such individual variations.

Food type: among the 60 outbreaks and experimental data sets examined, there are 45 distinctly named foods. No attempt was made here to evaluate variation with food type. Bollaerts *et al.*, (2008) attempted such an analysis using the summary information collected by FAO/WHO (2002b), and found that the heterogeneity due to food matrix and serotype could not be separated.

Serotype: from Figure 6 it is apparent that even within a single serotype that there is a large variability in virulence to cause illness — for *S. enteritidis*, the variability covers nearly eight orders of magnitude. Among serotypes, Table 9 shows there are sufficient data to evaluate *S. enteritidis* and perhaps *S. typhimurium* separately; or at least to test whether they differ significantly from one another and from all other serotypes combined.

The approach is thus to evaluate those components of variation that may practically be separated given the available data, and to assume that the observations are otherwise random samples from all other variability distributions. The variabilities obtained for the evaluated components will thus incorporate whatever further variabilities exist and have not been separately evaluated. To apply the results obtained, it is further necessary to assume that the observations are representative of the specific application (so that the same un-evaluated variability components apply). The variabilities that are separately evaluated here are therefore the variation among the susceptibility classes defined in Section 6, and between *S. enteritidis*, *S. typhimurium*, and all other serotypes combined.

Figure 5 and Figure 6 suggest that to a good approximation the isolates used in the human dosing and occurring in investigated outbreaks do indeed represent a random selection of all *Salmonella* isolates in human illness-causing circumstances, and that the distribution of virulences for such *Salmonella* isolates can be adequately represented by a lognormal distribution²⁷ (the Shapiro-Wilk p-values for departure from lognormal for the sub-groups shown

²⁷ These two inferences are interdependent. The wide range of values observed suggests reasonably random selection, since most mechanisms for non-random selection would presumably select for high virulence. The data are from the combination of human dosing experiments and from investigated outbreaks (indeed, from the subset of investigated outbreaks where dose estimates were made). It is a hypothesis that these observations can be extrapolated to all illness-causing circumstances (combinations of food and situation). The lognormality comes

in Figure 6 are: 0–5 yr: 0.81; 6–12 yr & hospitalized: 0.63; Non-sensitive SE: 0.30; Non-sensitive, non-SE: 0.045). These hypotheses will be assumed for the rest of this work.

Table 9 Counts of serotypes listed in Table 8; total number of maximum likelihood virulence estimates, and number with finite, non-zero values.

Serotype	Total	Finite, non-zero
<i>S. anatum</i>	4	3
<i>S. bareilly</i>	1	1
<i>S. bareilly</i> o7	1	1
<i>S. braenderup</i>	1	1
<i>S. derby</i>	1	1
<i>S. enteritidis</i>	35	32
<i>S. heidelberg</i>	1	1
<i>S. infantis</i>	1	0
<i>S. meleagridis</i>	3	3
<i>S. napoli</i>	2	0
<i>S. newport</i>	2	2
<i>S. oranienburg</i> o7	1	0
<i>S. schwarzengrund</i>	2	0
<i>S. typhimurium</i>	11	7
<i>S. zanzibar</i>	2	0

The experimental data (serial numbers [42](#) through [50](#), [52](#), [54](#), and [59](#)) are distinguished in Figure 5 and Figure 6 by circles round the corresponding symbols. It is apparent from those figures (and statistical tests confirm) that the experimental data alone cannot be considered random samples from the distributions of which they are part — the experimental data lie in the lower ends of the distributions. It is a hypothesis of this analysis that the combination of experimental data and outbreak data form a random sample. The sensitivity analysis performed Section 8.8, in which the experimental data are omitted from the analysis, shows that this hypothesis is critical for evaluation of the non-*S. enteritidis* serotypes as a group.

8.4. Modeling the variation of virulence

Based on the empirical evidence, primarily the distributional shapes shown in Figure 5 and Figure 6, it is here assumed that variability of virulence may be represented by lognormal

directly from the shape of the plots, together with a bias for lognormality in natural phenomena since this is the distribution expected where values depend multiplicatively on multiple independent characteristics. However, inference from the shape of the plots requires random selection of entries to plot.

distributions in each of the sub-groups separately analyzed. To evaluate the parameters of the lognormal distribution of virulence, to test if there are any observable differences between serotypes, and to evaluate quantitatively how sensitive are “sensitive” populations, the data given in Table 5 are here evaluated with the assumption of random selection from lognormal distributions.²⁸

A lognormal distribution of virulences may be parameterized by two values — its mean (m) and standard deviation (Σ) on a logarithmic scale. Estimates of these parameters and the uncertainties of those estimates in the form of a variance-covariance matrix were obtained using likelihood methods applied to all the relevant data in Table 5. With the assumptions stated the loglikelihood for the observations in a particular outbreak or feeding experiment in which the relative doses are known exactly is (up to an additive constant):

$$J_B(m, \Sigma, \{r_j, n_j, d_j; j = 1 \dots N\}) = \ln \left[\frac{1}{\omega \sqrt{2\pi}} \int_{-\infty}^{\infty} \left\{ \exp \left(-\frac{(x-m)^2}{2\omega^2} \right) \prod_{i=1}^N \left(\frac{p_i n_i}{r_i} \right)^{r_i} \left(\frac{(1-p_i) n_i}{n_i - r_i} \right)^{n_i - r_i} \right\} dx \right] \quad (25)$$

where the terms are:

- N number of independent dose groups,
- n_i number of people exposed in dose group i ,
- r_i number of people responding in dose group i ,
- d_i the dose in dose group i ,
- ω is given by

$$\omega = \sqrt{\Sigma^2 + \sigma^2} \quad (26)$$

- where σ is the ln(GSD) for the doses (Section 4.8) for this set of observations, and
- p_i the probability for illness at dose d_i for a virulence of e^x , given by the dose-response function:

$$p_i = 1 - \exp(-S_i e^x d_i) \quad (27)$$

where S_i is a factor that accounts for higher sensitivity in sensitive subpopulations, with value unity (by definition) for non-sensitive subpopulations and s_Y , s_M , or s_H for sensitive subpopulations in the sensitivity classes (Section 6) 0–5 yrs, 6–12 yrs, and hospitalized respectively. The normalization adopted here for the likelihood is such that the loglikelihood vanishes if the model fits the data exactly (terms in the product in the integrand with $r_i = 0$ or $r_i = n_i$ are interpreted as unity).

For an outbreak in which only an upper bound on the dose is available, the likelihood is assumed to be proportional to the integral over doses of that in Equation (25), so the loglikelihood is (up to an additive constant)

²⁸ All the results in Table 5 and Table 8 are used. 16/68 results in Table 8 could not be plotted, because they provide only bounds (so cannot be placed in order, as required for such figures). The analysis performed here also takes correct account of those outbreaks where distinct populations were exposed to the same isolate — the different populations are modeled as being exposed to the same isolate, but may have different sensitivities.

$$J_{BI} = \ln \left[\int_0^d \left\{ \frac{1}{\omega\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left(-\frac{(x-m)^2}{2\omega^2} \right) \left(\frac{p(x,t)n}{r} \right)^r \left(\frac{(1-p(x,t))n}{n-r} \right)^{n-r} dx \right\} dt \right] \quad (28)$$

using the same nomenclature (for such outbreaks, there is only one dose group) with

$$p(x,t) = 1 - \exp(-Se^x t) \quad (29)$$

This loglikelihood incorporates an assumption of equal likelihood for any dose below the upper bound (a uniform prior distribution), which for attack rates not equal to 0% or 100% is equivalent to an implicit bias against high virulence. For outbreaks in which the attack rate is not 0% or 100%, the loglikelihood given by Equation (28) thus has a maximum, giving a point estimate for virulence even in the case where only an upper bound on dose is available. This is correct in the situation where the doses to every member of the population are identical, since that dose could not be less than 1 CFU; but may be misleading with a distribution of doses that can have arbitrarily low mean value. However, the upper uncertainty estimate on the virulence estimated using Equation (28) is sufficiently large that the potential error introduced by using this equation are sufficiently small to be ignored, as they are here.

The loglikelihood for all outbreaks and the experiments combined is the sum of the loglikelihoods for each outbreak or experiment separately. Each serotype was modeled with its own mean and standard deviation estimate for the lognormal distribution of virulence, but these individual estimates were then combined into groups consisting of *S. enteritidis*, *S. typhimurium*, and all other serotypes combined (there are too few estimates for other serotypes to evaluate them separately, see Table 9). Estimates for means and standards deviation of the lognormal distributions, and for s_Y , s_M , or s_H , the sensitivity factors, are obtained by maximizing the combined likelihood for all observations in Table 5 with respect to these parameters. Uncertainty estimates are obtained by standard profile likelihood methods.

To test for differences²⁹ between s_Y , s_M , and s_H , and for different means or standard deviations in the lognormal virulence distributions for different serotypes, the likelihood for all experiments in Table 5 was first maximized while allowing for different s_Y , s_M , s_H , and different means and standard deviations for *S. typhimurium*, *S. enteritidis*, and all other serotypes combined. Comparing non-zero Σ with $\Sigma = 0$ for all three distributions showed that the differences in virulence required this additional variance ($p < 10^{-300}$), as is also evident by inspection of Figure 5 and Figure 6. It was found that s_Y , s_M , and s_H are not homogeneous ($p = 9.6 \times 10^{-12}$), and that s_Y and s_M are significantly different from unity ($p = 6.2 \times 10^{-87}$ and 1.6×10^{-26} respectively). With the available data, s_H cannot be distinguished from unity (95% confidence limits 0.076 to 3.38), but s_M and s_H are not significantly different ($p = 0.12$), allowing acceptance of the hypothesis of this analysis (Section 6) that hospitalized patients should be considered as sensitive as one of the selected age classes. With $s_M = s_H$, there is no significant difference between serotypes in standard deviation ($p = 0.20$). With $s_M = s_H$ and equal standard deviations, the means are not homogeneous ($p = 0.04$), but the means for *S. typhimurium* and all other serotypes except *S. enteritidis* are not significantly different ($p = 0.98$). Accepting all these constraints

²⁹ All the p-values shown in this paragraph are from likelihood ratio tests. These results are all shown in sheet All_data of "Epi Attack Data.xls".

simultaneously ($s_M = s_H$, equal standard deviations for all serotypes, equal means for all serotypes except *S. enteritidis*) gives no significant difference in loglikelihood ($p = 0.23$) compared with the unconstrained case, so these constraints are considered to acceptably define the available information. The resulting (constrained) maximum likelihood parameter estimates are given in Table 10:

Table 10 Maximum likelihood estimates for virulence distribution parameters³⁰

Parameter	Parameter symbol	Value
Sensitivity factor, 0–5 yr old	s_Y	5.832
Sensitivity factor, 6–12 yr old or hospitalized	$s_M = s_H$	2.843
Mean logarithm of virulence, <i>S. enteritidis</i>	m_{se}	-10.426
Mean logarithm of virulence, all other serotypes	m_{ot}	-14.019
Standard deviation of the logarithm of virulence, all serotypes	Σ	4.853

Figure 5 and Figure 6, the functional form of Equation (25), individual likelihood profiles for the various parameters (Section 8.6), and numerical experimentation suggest that the loglikelihood is concave with respect to all parameters, which would ensure that the maximum located is the unique global maximum, although this has not been mathematically proved. To illustrate, Figure 7 shows the loglikelihood surface above the (m_{se}, Σ) plane with the other parameters fixed at the values shown in Table 10.

³⁰ The accuracy of these values is far lower than the precision given; for their uncertainties, see Table 13.

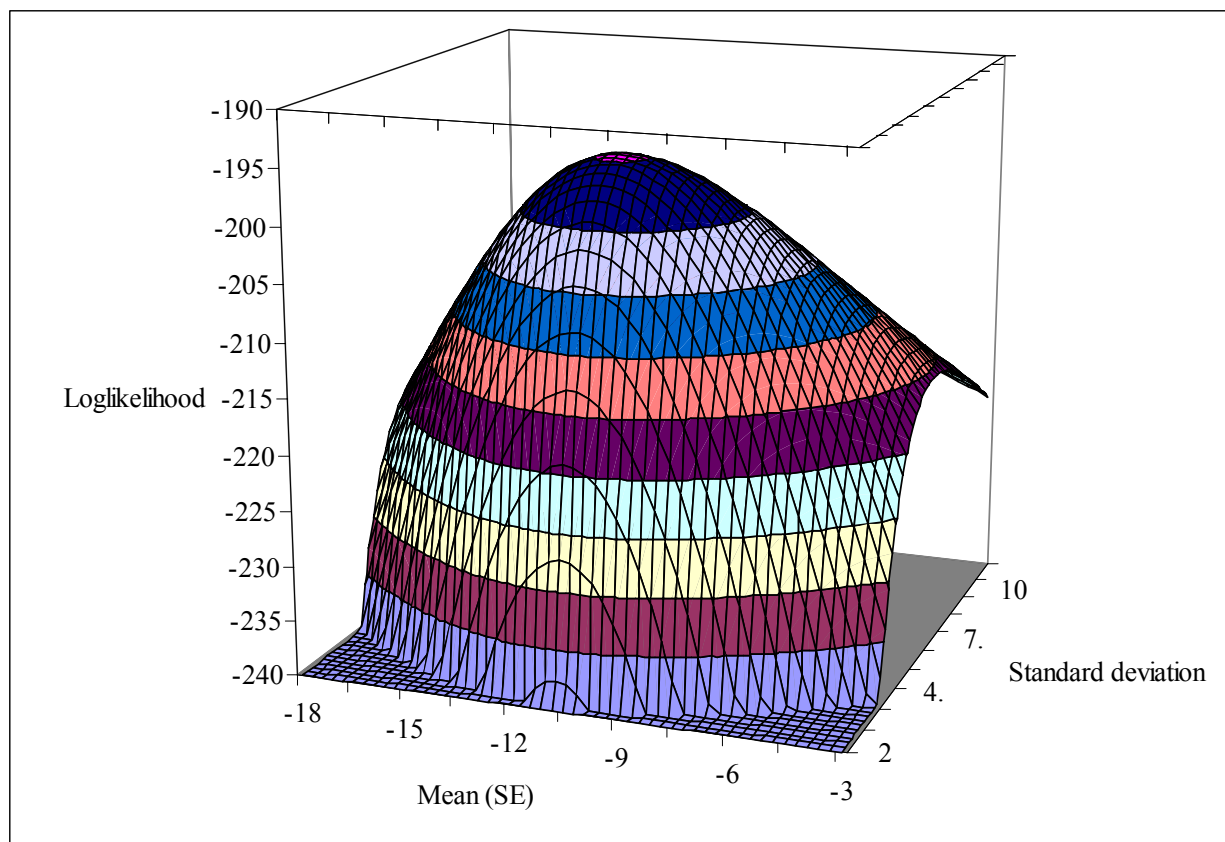


Figure 7 Loglikelihood surface above the (m_{se}, Σ) plane with the other parameters fixed at the values shown in Table 10 ($SE = S. enteritidis$).

8.5. The random isolate dose-response curve

The dose-response curve for any particular *Salmonella* isolate in this analysis is linear-exponential. In typical Monte Carlo simulation applications each iteration of the simulation will select a random situation, and the isolate will be different in each such iteration. Typically only a dose and serotype will be simulated, with no further information about the *Salmonella* isolate. In such circumstances, what is required is then to select at random a virulence to use in the linear-exponential dose-response function, and that virulence should be selected from the virulence distribution for that serotype. It may be required to select at random for particular exposure circumstances, perhaps for a specific type of food. The analysis performed here has not attempted to be specific about exposure circumstances, in the belief that there are too few data to adequately evaluate how exposure circumstances may affect illness probabilities. The virulence distributions obtained here are for distributions of exposure circumstances that presumptively correspond to typical food consumption and exposure circumstance patterns.

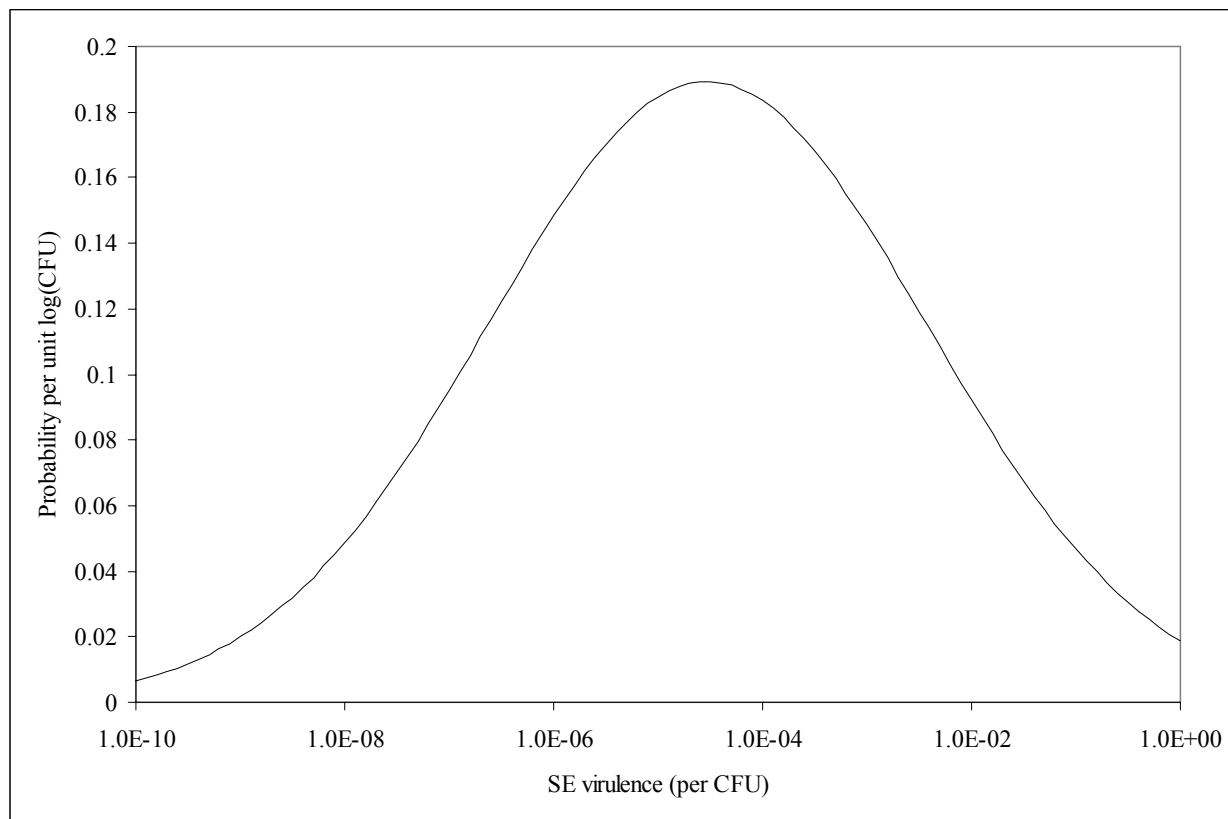


Figure 8 The maximum likelihood estimate of the virulence distribution for *S. enteritidis* for non-sensitive populations.

Given this, the virulence for (say) an *S. enteritidis* isolate will be selected from a lognormal distribution with parameters given in Table 10 (in practice, the uncertainties evaluated in Section 8.6 should be taken into account, as detailed in Section 10). This virulence distribution is shown in Figure 8. The selected virulence (together with the appropriate sensitivity factor) may then be applied in a linear-exponential dose-response curve, Equation (18), to obtain the probability of illness. Figure 9 shows examples of such dose-response curves for virulences selected at particular percentiles of the *S. enteritidis* virulence distribution.

In these circumstances, which average over different isolates, the probability for a given dose to cause illness can also be computed for a particular serotype in a population with a given sensitivity. That probability is the just average that will be obtained for a fixed dose using the preceding procedure, averaging the differences in virulence between isolates. Treating the average probability as a function of that fixed dose gives what is called here the “random isolate” dose-response curve; it has the functional form³¹

$$p(d) = \frac{1}{\Sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-m)^2}{2\Sigma^2}\right) (1 - \exp(-e^x Sd)) dx \quad (30)$$

³¹ The integral here is a special case ($N = 1$, $r = n = 1$) of Equation (25), and may be computed in the same way.

where $S = 1$, s_Y , s_M , or s_H for non-sensitive populations, those 0–5 years, 6–12 year-olds, and hospitalized populations, respectively. It is just the average of the linear-exponential dose-response curve over the virulence distribution.

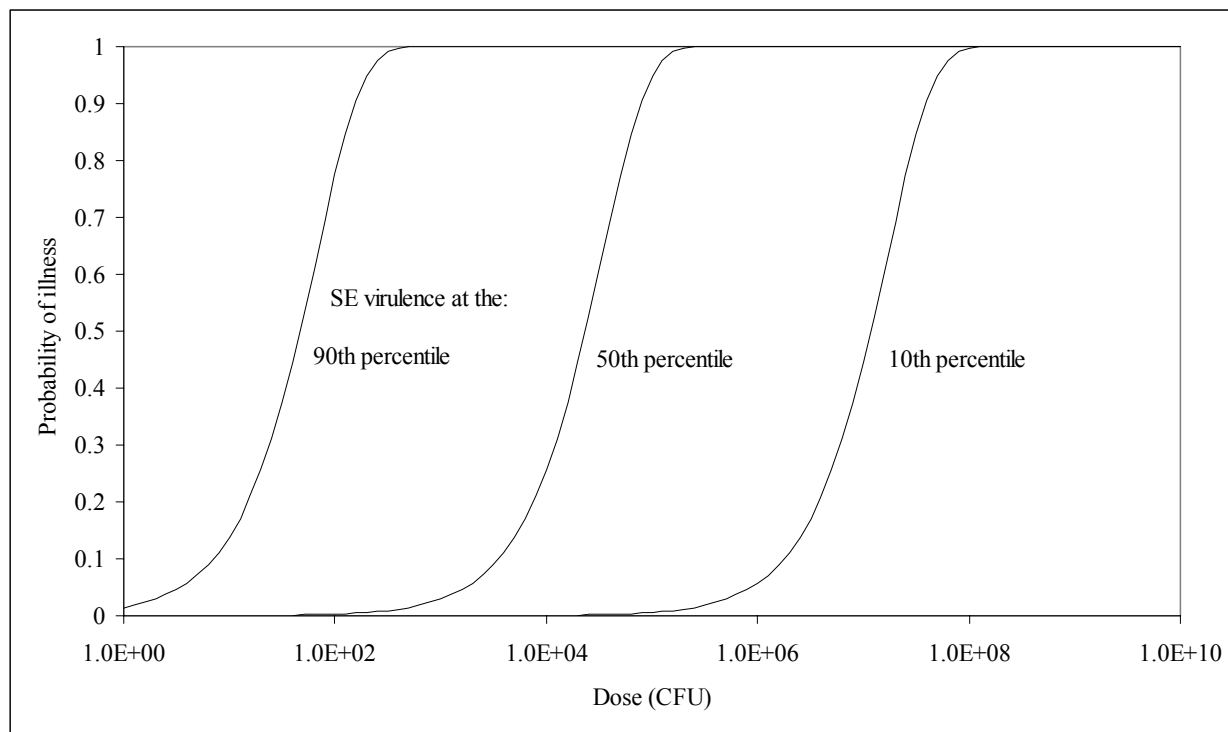


Figure 9 Dose-response curves with virulence selected from the *S. enteritidis* virulence distribution

The random isolate dose-response curves for *S. enteritidis* and for other serotypes, for normal (non-sensitive) populations, are plotted in Figure 10, together with envelopes of the 90% profile likelihood confidence intervals around each point on the curves (labeled as UCL95 and LCL95 in Figure 10). Figure 10 should strictly be interpreted as being the random isolate dose-response curve for the same fixed dose to every member of the population exposed, as for the linear-exponential dose-response curve on which it is based (Section 8.2). As discussed in Section 8.1 and Appendix D, using the average dose in a population should provide an adequate approximation for arbitrary distributions of dose provided the dose is above about 1 CFU and the distribution is reasonably narrow. However, for average doses substantially less than 1 CFU, the distribution of doses in the exposed population may have a substantial effect on this average probability (see Appendix D). Also plotted for comparison is the “dose-response” curve identified as the central estimate from outbreak data in FAO/WHO (2002b; see Appendix E) — the random isolate dose-response curve corresponds most closely with the intent of the FAO/WHO (2002b) analysis.

Numerically, the random-isolate dose-response curve of Equation (30) may be very well approximated by a probit-log-dose curve over a very large range of doses. That is, Equation (30) is well approximated by a curve of the form

$$p(d) = \Phi\left(\frac{\ln(Sd) - m'}{\Sigma'}\right) \quad (31)$$

although there appears to be no simple relationship between m' , Σ' and m , Σ . Plotting the inverse normal function of the probability p from Equation (30) versus the logarithm of dose therefore leads to approximately straight lines, as shown in Figure 11.

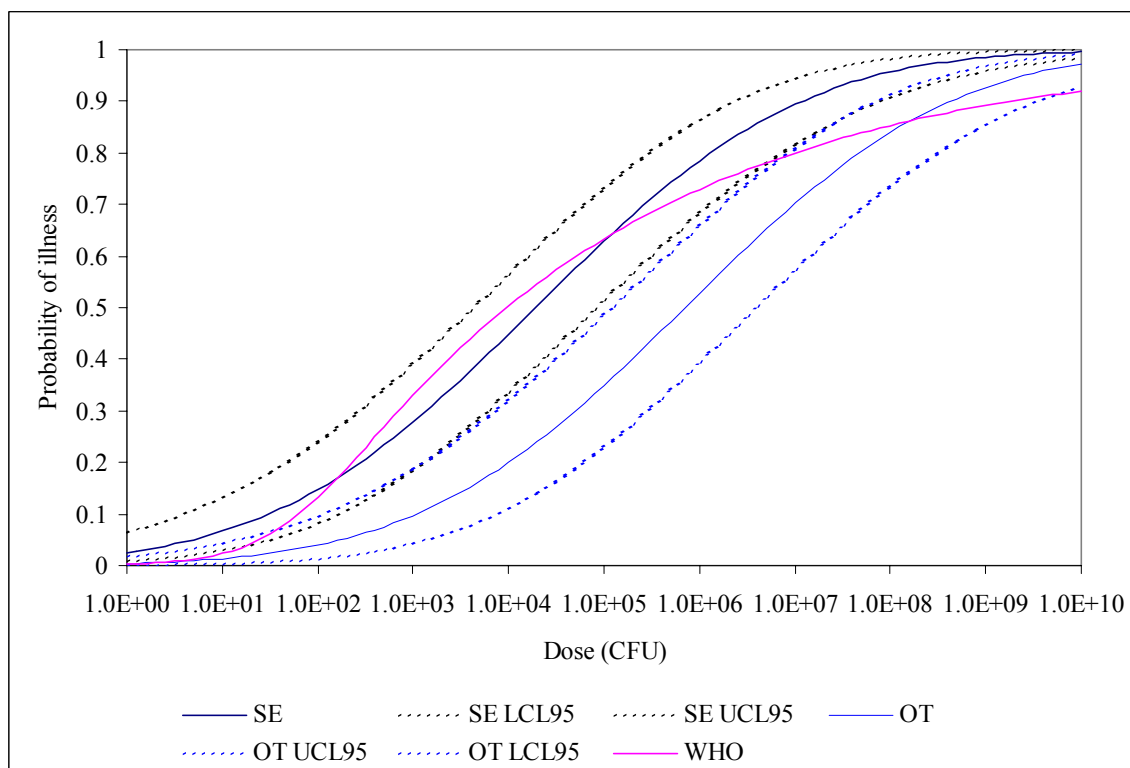


Figure 10 Random isolate dose-response curves for *S. enteritidis* (SE) and other serotypes (OT) (non-sensitive population), with 95% confidence limits and FAO/WHO (2002b) dose-response curve.

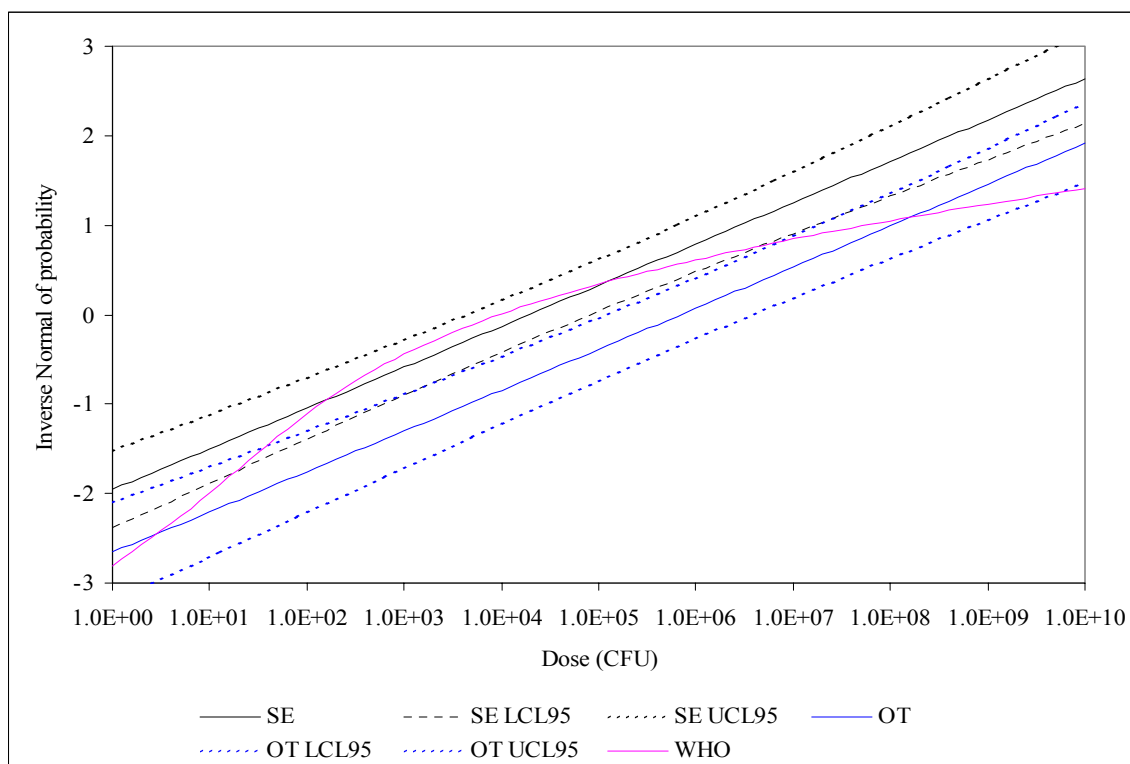


Figure 11 As for Figure 10, but applying inverse normal function to the probability of illness

8.6. Uncertainty in the virulence distribution parameters

The uncertainties in the parameters of Table 10 were approximated using the profile likelihood method.³² Profile loglikelihoods were evaluated for each parameter over a range of values such that at the extremes of the range the profile loglikelihood changed by approximately 5 (within the range 5.0 to 5.5),³³ then simple monotonic transformations of the parameters selected so that the profile likelihoods of the transformed parameters were almost perfectly normal in shape over this range. This allows approximating the joint uncertainty distributions for the transformed parameters by a multinormal distribution,³⁴ using the variance-covariance matrix (obtained from the information matrix of the likelihood function) to obtain its parameters.³⁵ The transformations

³² The profile likelihood of a parameter at a particular value is the maximum likelihood (with respect to all other parameters) evaluated at that value of the parameter.

³³ Using the conventional interpretations of the loglikelihood, this range encompasses approximately 99.8% of the uncertainty distribution.

³⁴ A distribution that is multinormal in its variables is here defined to be distribution with a differential distribution function that is proportional to the exponential of a non-positive quadratic in the variables.

³⁵ The usual approach is to omit the transformations, and locally approximate the joint uncertainty distribution using a multinormal distribution. By including the transformations, the approximation to the likelihood function is improved over a relatively large region around the maximum likelihood estimates. Asymptotically the approximations with and without transformations are the same, but the transformation approach takes better account of the asymmetries of the likelihood function (which is much more symmetrical with respect to the transformed parameters). Choice of suitable transformations also allows enforcement of natural boundaries; for example, with the transformations chosen here s_Y , s_M , s_H , and Σ can never be negative (as required) in simulations based on the

used here are shown in Table 11, together with the correlation coefficient over the chosen range between the signed square-root of the change in profile loglikelihood from its maximum and the transformed parameter (a correlation coefficient of 1.0 corresponds to a perfectly normal shape of profile likelihood).³⁶

Table 11 Transformations of the parameters to give nearly normal profile likelihoods, and the correlation coefficients obtained.

Parameter	Transformed variable	Correlation coefficient
s_Y	$\ln(s_Y)$	0.99995
$s_M = s_H$	$\ln(s_M)$	0.99990
m_{se}	m_{se}	0.99983
m_{ot}	m_{ot}	0.99992
Σ	$\ln(\Sigma - 1.69)$	0.999990

Table 12 shows numerical estimates for the variance-covariance matrix with respect to the transformed variables, Table 13 the corresponding standard deviations and correlation coefficients, and Table 14 the Cholesky decomposition of the variance-covariance matrix in Table 12 (see Section 10 for the use of Table 14). Table 15 lists the corresponding estimates for standard deviation of the untransformed parameters, obtained by transforming back from the estimates given on the main diagonal of Table 13. All these estimates were obtained using a numerical step-size for the transformed parameters that is approximately equal to their standard deviations. The estimated variance-covariance matrix thus takes account of correlations that may occur over considerable ranges of the parameters. The transformations applied to the parameters ensure that the choice of numerical step size is relatively unimportant. There are only small relative differences in the standard deviation estimates (<2.2%), and in the absolute differences in the correlation coefficient estimates (<0.09), for a numerical step size 100 times smaller.

multinormal approximation. This approach thus better represents the uncertainties than use of the untransformed variables.

³⁶ Graphs, indistinguishable by eye from straight lines, are included in sheet All_data of workbook "Epi Attack Data.xls", where all the calculations of this section are performed.

Table 12 Variance-covariance matrix representing uncertainties in dose-response parameters

	m_{se}	m_{ot}	$\ln(\Sigma-1.69)$	$\ln(s_M)$	$\ln(s_Y)$
m_{se}	0.7962	0.0258	0.0229	-0.00268	-0.00133
m_{ot}	0.0258	1.1066	0.0308	-0.00121	-0.00084
$\ln(\Sigma-1.69)$	0.0229	0.0308	0.0278	-0.00017	-0.00019
$\ln(s_M)$	-0.00268	-0.00121	-0.00017	0.00787	0.00146
$\ln(s_Y)$	-0.00133	-0.00084	-0.00019	0.00146	0.00574

Table 13 Standard deviations (main diagonal) and correlation coefficients (lower left) corresponding to Table 12

	m_{se}	m_{ot}	$\ln(\Sigma-1.69)$	$\ln(s_M)$	$\ln(s_Y)$
m_{se}	0.892				
m_{ot}	0.027	1.052			
$\ln(\Sigma-1.69)$	0.154	0.175	0.167		
$\ln(s_M)$	-0.034	-0.013	-0.011	0.0887	
$\ln(s_Y)$	-0.020	-0.010	-0.015	0.216	0.0758

Table 14 Cholesky decomposition of the variance-covariance matrix of Table 12

	m_{se}	m_{ot}	$\ln(\Sigma-1.69)$	$\ln(s_M)$	$\ln(s_Y)$
m_{se}	8.923E-01	0	0	0	0
m_{ot}	2.886E-02	1.052E+00	0	0	0
$\ln(\Sigma-1.69)$	2.566E-02	2.854E-02	1.622E-01	0	0
$\ln(s_M)$	-3.008E-03	-1.071E-03	-3.550E-04	8.868E-02	0
$\ln(s_Y)$	-1.496E-03	-7.534E-04	-7.951E-04	1.635E-02	7.399E-02

Table 15 Approximate standard deviations for the untransformed parameters

Parameter	Symbol	MLE	Estimated standard deviation
Mean logarithm of virulence, <i>S. enteritidis</i>	m_{se}	-10.426	0.89
Mean logarithm of virulence, other serotypes	m_{ot}	-14.019	1.05
Standard deviation of virulence distribution, all serotypes	Σ	4.853	0.81
Sensitivity factor, 6–16 year age group and hospitalized patients	$s_M = s_H$	2.843	0.25
Sensitivity factor, 0–5 year age group	s_Y	5.832	0.44

8.7. Alternative dose-response model

To confirm that the results shown in Sections 8.3 through 8.5 are relatively insensitive to the shape of the dose-response model chosen for individual isolates, the same analysis was performed using a threshold-like dose-response model of the form

$$p = \Phi(\beta \ln(Sd/d_{50})) \quad (32)$$

where Φ is the cumulative normal function (this dose-response model is sometimes called the probit-log-dose model). The model has two parameters, the slope β and the dose d_{50} for 50% response (S is the sensitivity factor for the population, as in Section 8.2, and should be regarded as a known quantity — see footnote 19). Evaluation of both parameters for individual isolates would require multiple dose groups, and moreover would require that the doses to all individuals within each dose group be identical.³⁷ Only the McCullough and Eisele (1951a,c) human feeding experiments (for all of which $S = 1$) adequately satisfy both requirements, so the value of β ($= 0.967194$) was obtained as the maximum likelihood estimate from those experiments and fixed for the rest of the analysis.³⁸

Table 16 and Table 17 show the adequacy of fit of this alternative dose-response model to the illness data of McCullough and Eisele (1951a,c), in exactly the same way as for Table 6 and Table 7 respectively for the primary dose-response model of Section 8.2.³⁹ Again, there is no evidence of lack of fit. The loglikelihood for Table 16 is identical to Equation (19) except for the substitution of the dose-response model

$$p_i = \Phi(\beta \ln(d_i/d_{50})) \quad (33)$$

³⁷ For non-linear dose-response curves, the mean dose to the group strictly should not be used for evaluation of the dose-response curve, although here there is no alternative. Attempting to estimate β using all the data results in underestimates. See Appendix D for further discussion.

³⁸ See the individual discussion of the McCullough and Eisele (1951a,c) experiments ([Human #2](#) in Appendix A) for the detailed methodology, and sheet MultiDoseP in the workbook “Epi Attack Data.xls” for the calculations.

³⁹ The calculations for Table 16 and Table 17 are in the sheet McCullough of the workbook “Epi Attack Data.xls”. The “All combined” entries account for 10 parameters here (versus 9 in Table 6 and Table 7) because the slope parameter is also estimated.

while the loglikelihood for Table 17 (after a substitution of the same nature as that described in Footnote 22) takes a form comparable to Equation (23):

$$J = \sum_{i=1}^N \ln \left\{ \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left(-\frac{(x - \ln d_{50})^2}{2\sigma_i^2} \right) \left(\frac{p_i n_i}{r_i} \right)^{r_i} \left(\frac{(1-p_i)n_i}{n_i - r_i} \right)^{n_i - r_i} dx \right\} \quad (34)$$

with

$$p_i = \Phi \left(\beta \ln(d_i / e^x) \right) \quad (35)$$

Table 16 Adequacy of fit of the probit-log-dose model to the human experimental data (McCullough and Eisele, 1951a,c; Serial #s [42](#) through [50](#)), assuming exactly known doses.

	Dose groups	Dose groups with non-zero response	Maximum loglikelihood	X^2 statistic	X^2 -based p-value using # of dose groups	X^2 -based p-value using # of non-zero doses
S. anatum I	8	2	-0.798	0.96	1.00	0.33
S. anatum II	8	2	-0.958	1.37	0.99	0.24
S. anatum III	3	2	-0.203	0.41	0.81	0.52
S. meleagridis I	11	2	-0.342	0.55	1.00	0.46
S. meleagridis II	5	3	-0.798	1.15	0.89	0.56
S. meleagridis III	4	2	-0.098	0.16	0.98	0.69
S. newport	3	3	-1.211	4.32	0.12	0.12
S. bareilly	3	3	-1.261	4.65	0.10	0.10
S. derby	5	1	-0.991	1.31	0.86	0.79
All combined ^a	50	20	-6.661	14.86	1.00	0.14
^a Summing all the individual loglikelihood and X^2 values, and accounting for the 10 parameters for the p-values.						

Table 17 Adequacy of fit of the probit-log-dose model to the human experimental data (McCullough and Eisele, 1951a,c; Serial #s [42](#) through [50](#)), accounting for the uncertainties in doses.

	Dose groups	Dose groups with non-zero response	Maximum loglikelihood	X^2 statistic	X^2 -based p-value using # of dose groups	X^2 -based p-value using # of non-zero doses
S. anatum I	8	2	-1.008	1.07	0.99	0.30
S. anatum II	8	2	-1.118	1.49	0.98	0.22
S. anatum III	3	2	-0.279	0.41	0.81	0.52
S. meleagridis I	11	2	-0.486	0.56	1.00	0.45
S. meleagridis II	5	3	-1.018	1.16	0.88	0.56
S. meleagridis III	4	2	-0.221	0.18	0.98	0.67
S. newport	3	3	-1.160	4.43	0.11	0.11
S. bareilly	3	3	-1.171	4.58	0.10	0.10
S. derby	5	1	-1.081	1.41	0.84	0.75
All combined ^a	50	20	-7.542	15.29	1.00	0.12
^a Summing all the individual loglikelihood and X^2 values, and accounting for the 10 parameters for the p-values.						

Evaluation of d_{50} for those outbreaks where such an evaluation is possible shows a distribution of values that is not distinguishable from lognormal ($p = 0.31$; Shapiro-Wilk test for all estimates combined, 0.81 for the <5 year age class, 0.96 for the combined 5–12 age class and the hospitalized class, 0.20 for the non-sensitive *S. enteritidis* class, and 0.02 for the non-sensitive other serotype class; but see footnote 25).⁴⁰ Figure 12 shows the distribution for the individual estimates with estimated 95% confidence limits; this figure corresponds to Figure 5 for the primary dose-response model. Figure 13 shows the distributions by sensitivity class and separated between *S. enteritidis* and other serotypes for the non-sensitives; this figures is analogous to Figure 6.

⁴⁰ The calculation of these values is performed in sheet Probit of workbook “Epi Attack Data.xls”, while figures are plotted in sheet Distrip_p.

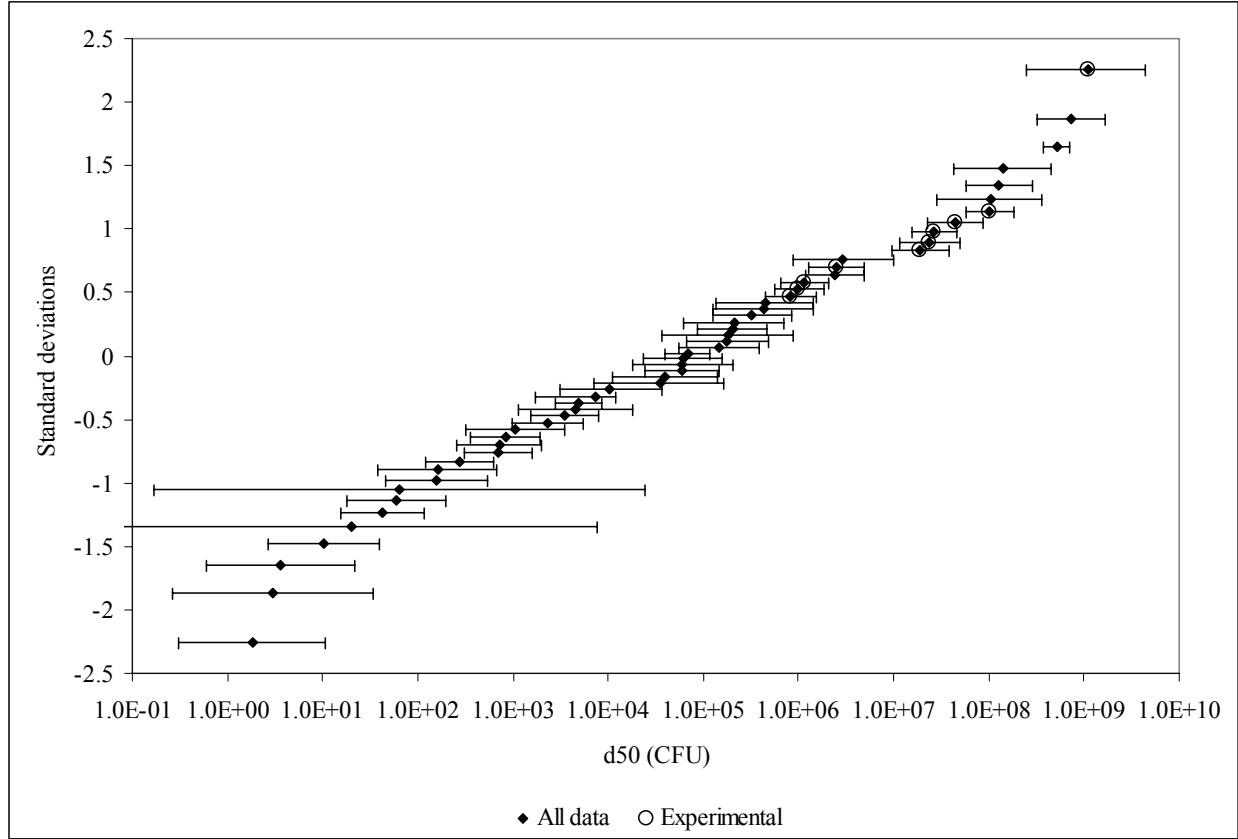


Figure 12 Distribution of individual ML estimates for the logarithm of d_{50} .

Hypothesizing that the distribution of d_{50} values is indeed lognormal and performing the same analyses as in Sections 8.3 and 8.4 leads to corresponding results. The likelihood function is identical to Equation (25), except that Equation (27) is replaced by

$$p_i = \Phi\left(\beta \ln(S_i d_i / e^x)\right) \quad (36)$$

The random isolate dose-response curve (Section 8.5) becomes

$$\begin{aligned} p(d) &= \frac{1}{\Omega\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \exp\left(-\frac{(x - \ln(D_{50}))^2}{2\Omega^2}\right) \Phi\left(\beta \ln(Sd/e^x)\right) \\ &= \Phi\left(\frac{\beta \ln(Sd/D_{50})}{\sqrt{1 + \beta^2 \Omega^2}}\right) \end{aligned} \quad (37)$$

where D_{50} and Ω characterize the lognormal distribution of values of d_{50} (they are the median and standard deviation of the distribution).⁴¹

⁴¹ In this case the required integral is analytically expressible in the form of a standard function, as shown.

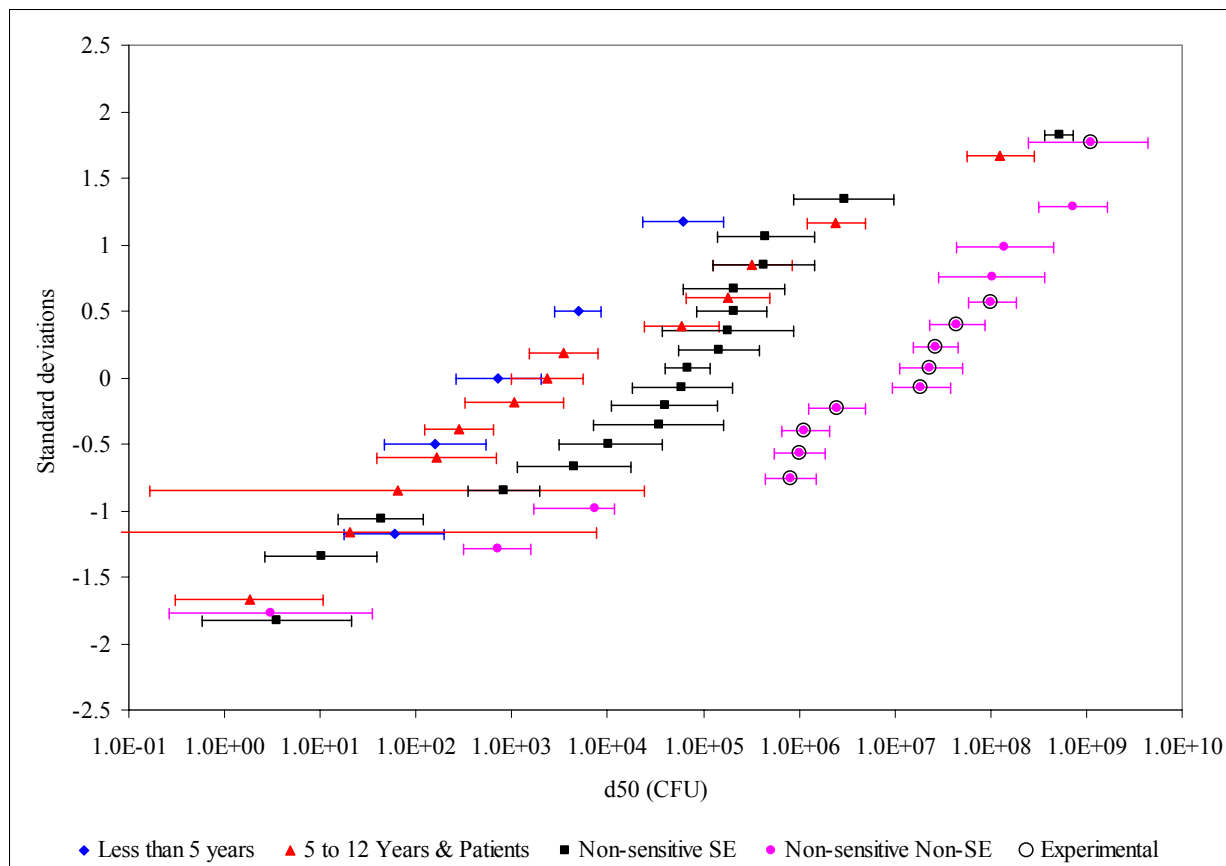


Figure 13 As for Figure 12, showing distributions for sensitivity classes and by serotype (SE = *S. enteritidis*).

The results obtained for the alternative dose-response curve parallel those for the primary dose-response function as discussed in Section 8.4; final numerical results are shown in Table 18, corresponding to Table 10 for the primary dose-response curve. These are derived in the sheet Probit in the workbook “Epi Attack Data.xls.”

Table 18 Maximum likelihood estimates for d_{50} distribution parameters⁴²

Parameter	Parameter symbol	Value
Sensitivity factor, 0–5 yr old	s_Y	4.422
Sensitivity factor, 6–12 yr old or hospitalized	$s_M = s_H$	2.496
Mean logarithm of D_{50} <i>S. enteritidis</i>	D_{50se}	9.719
Mean logarithm of D_{50} , all other serotypes	D_{50ot}	13.313
Standard deviation of the logarithm of d_{50} , all serotypes	Ω	4.941

Figure 14 shows a comparison of the random isolate dose-response obtained with the ML estimates for D_{50} and Ω (“Probit”) compared with the similar dose-response obtained with the primary dose-response curve (“Linear-exponential”) in Section 8.5 (see that section for interpretation of these curves). The difference between the two is negligible compared with the uncertainties involved.

Uncertainty estimates for the parameters shown in Table 18 could be obtained in the same manner as described in Section 8.6; but this was considered superfluous in view of the practical identity of results obtained using the two dose-response curves.

⁴² Again, the accuracy of these values is far lower than the precision given.

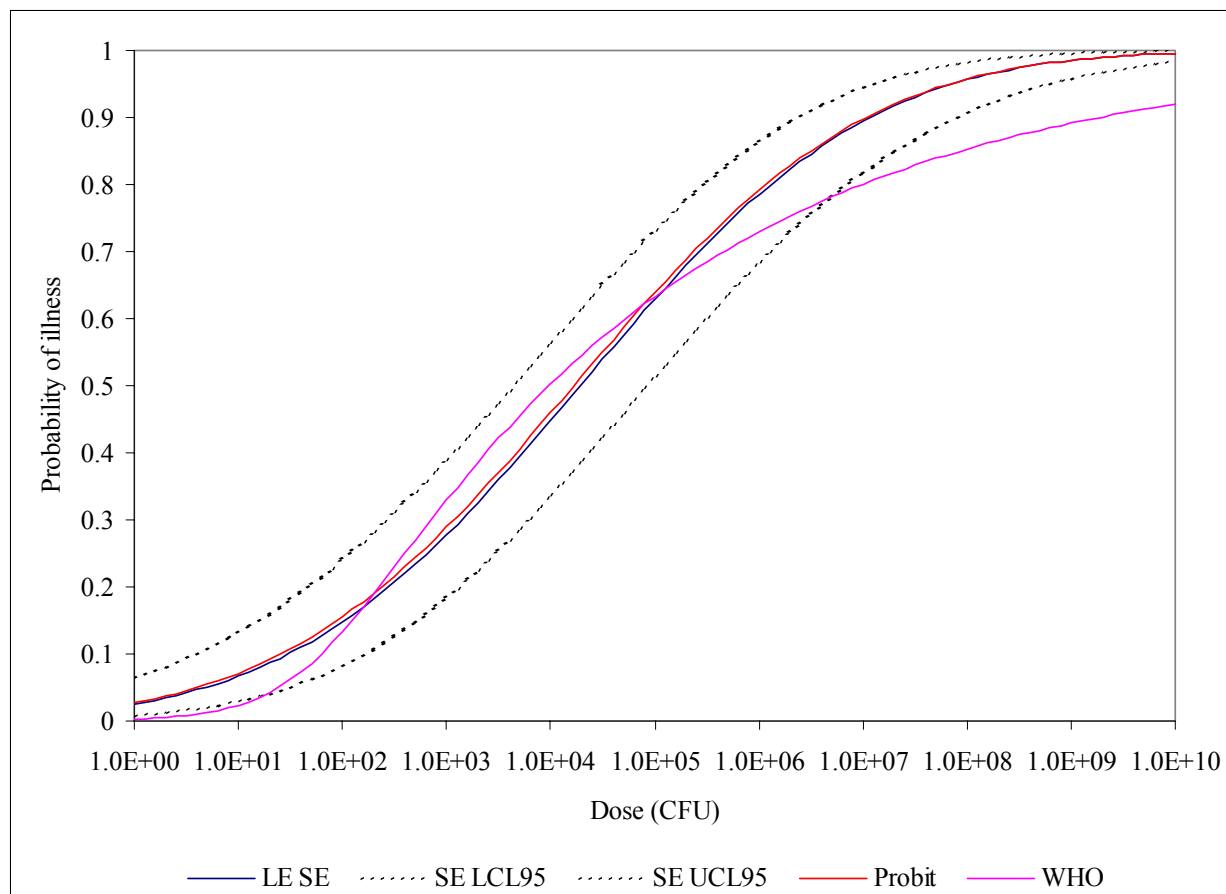


Figure 14 Random isolate dose-response for *S. enteritidis* (SE) for the alternative model (“Probit”) compared with the primary model (LE SE, with envelopes of 90% confidence intervals “SE LCL95” to “SE UCL95”) and with the FAO/WHO (2002b) dose-response (“WHO”).

8.8. Sensitivity analyses

Dose uncertainty estimates

The estimates of virulence obtained from individual outbreaks have considerable uncertainty. The average $\ln(\text{GSD})$ for the individual estimates is about 0.71 (a factor of two), and it ranges up to 5.4 (a factor of 230).⁴³ These estimates of uncertainty are primarily due to the uncertainty in doses. Using the procedures adopted here, the uncertainty estimates come primarily (in most cases) from the dose measurement uncertainties. There are substantial additional uncertainties that are not quantified by this procedure, for example in most outbreaks there is considerable uncertainty about the representativeness of the material used for the dose measurements. To evaluate the effect of additional uncertainty in doses on the estimates obtained for the distribution of virulence, all the $\ln(\text{GSD})$ estimates of dose uncertainty were doubled, or unity was added to those $\ln(\text{GSD})$ estimates, and the changes in maximum likelihood parameter

⁴³ See sheet Distrib in workbook “Epi Attack Data.xls.”

estimates evaluated.⁴⁴ Table 19 shows that the effect on the maximum likelihood estimates is minimal, substantially less than the standard deviations of the estimates even for such substantial changes in uncertainty estimates.

Table 19 The effect on MLE parameters of doubling ln(GSD) or adding unity to it for all individual outbreaks and experiments.

Parameter	Parameter symbol	MLE (Table 10)	Std. Dev. of MLE (Table 15)	Change in parameter on	
				doubling all ln(GSD) estimates	adding 1 to all ln(GSD) estimates
Sensitivity factor, 0–5 yr old	s_Y	5.832	0.44	0.002	0.001
Sensitivity factor, 6–12 yr old or hospitalized	$s_M = s_H$	2.843	0.25	-0.002	-0.001
Mean logarithm of virulence, <i>S. enteritidis</i>	m_{se}	-10.426	0.89	-0.188	-0.068
Mean logarithm of virulence, all other serotypes	m_{ot}	-14.019	1.05	-0.164	-0.065
Standard deviation of the logarithm of virulence, all serotypes	Σ	4.853	0.81	-0.212	-0.245

Since the uncertainties in doses and those in incidence add independently, this sensitivity analysis also shows that uncertainties in incidence beyond those due to binomial observations would also have relatively small effect on the MLE estimates for the distribution of virulence.

This insensitivity of the distribution parameters to the uncertainty estimates for the individual outbreak and experimental observations is not surprising. In evaluating the likelihood, the individual observational uncertainties (median ln(GSD) of 0.71) add in quadrature to the variability of the virulence distribution, which is much higher (ln(GSD) equal to 4.85). The net effect (4.90 for the median observational ln(GSD)) is hardly affected even by substantial changes in the observational uncertainty.

Hospitalized patient subgroup

One of the hypotheses of this investigation is that hospitalized patients form a sensitive sub-population, with sensitivity equal to one of the younger age groups. The analysis of Section 8.4 shows that the observations are consistent with this hypothesis, in that the sensitivity of hospitalized patients is not significantly different from that of the 6–12 year-old age group (Table 10). Evaluating the hospitalized patient group separately, unconstrained estimates of s_H are less than unity, but with a very wide confidence range (90% confidence interval 0.002 to 3.4 with no

⁴⁴ The calculations are performed in sheet All_data of workbook “Epi Attack Data.xls.” Although the same change is made to each uncertainty estimate, the effect is the addition of independent uncertainties to each outbreak or experiment analyzed.

constraints on any parameters). However, the available data do not distinguish this group from any of the others (normal, 0–5 year age group, or 6–12 year age group); and removing the constraint that the hospitalized group is equivalent to the 6–12 year-old age group changes the parameter estimates for the other groups only slightly (less than 0.25 of their standard deviations; calculations not shown).

Experimental observations versus outbreaks

A further hypothesis of this analysis is that the combination of experimental and outbreak observations constitute a random sample from the virulence distribution for *Salmonella* isolates affecting food, including the variation in exposure conditions. Examination of Figure 5 and particularly Figure 6 indicate that the experimental observations by themselves may be from a different distribution of *Salmonella* isolates than the outbreak data. The same analysis as in Section 8.4 may be performed omitting experimental observations (Serial #s [42–52](#), [54](#), [59](#)). In that case, however, there are too few outbreaks involving *S. typhimurium* to evaluate separately, so *S. typhimurium* has in this sensitivity analysis been combined with other serotypes. Once again, the sensitivity of the hospitalized patient subgroup is consistent with being equal to the sensitivity of the 6–12 year age group ($p = 0.051$). Now, however, the means of the distributions for *S. enteritidis* and other serotypes are consistent with being equal, but the standard deviations of those distributions are different. In place of Table 10, the results shown in Table 20 are obtained. Comparison of these tables and Table 15 shows that the sensitivity factors and the distribution for *S. enteritidis* are hardly affected (well within their standard deviations). The distribution for other serotypes, however, is now estimated to have substantially higher mean and wider distribution than if the experimental observations are included.

Table 20 Maximum likelihood estimates for virulence distribution parameters, omitting experimental observations.

Parameter	Parameter symbol	Value
Sensitivity factor, 0–5 yr old	s_Y	5.817
Sensitivity factor, 6–12 yr old or hospitalized	$s_M = s_H$	2.836
Mean logarithm of virulence, all serotypes	$m_{se} = m_{ot}$	-10.478
Standard deviation of the logarithm of virulence, <i>S. enteritidis</i>	Σ_{se}	4.114
Standard deviation of the logarithm of virulence, other serotypes	Σ_{ot}	7.227

There is no way using the data examined here to select one or other of the sets of results shown in Table 10 or Table 20 as superior; it remains a hypothesis as to which better represents the virulence distribution in food for the conditions of exposure to that food.

9. A summary of the principal hypotheses assumed in this analysis

There are various hypotheses that have been used explicitly or implicitly in the analyses. The principal ones are:

- a. The reported outbreak data combined with the human dosing experiments constitute a random sample from the distributions of virulence for *Salmonella* occurring in human foods in circumstances that may lead to illness. This hypothesis is a *sine qua non* for any analysis that attempts to use these data as predictive for *Salmonella* virulence. Even if the hypothesis is true, application of the results may be incorrect in specific circumstances (since the distribution may then differ from the general case). There is no way to test the hypothesis, and many potential reasons that it may fail; but there is currently no alternative approach available. The distribution of virulence may depend on location; the majority of outbreaks examined here were reported from Japan so may not represent conditions in the U.S. The distribution may also depend strongly on conditions of exposure, which differ with food type, location, and perhaps with calendar time. The observed data are from investigated outbreaks in which dose estimates were made, so represent a very small subset of all *Salmonella*-caused illnesses and may be a biased sample. Section 8.8 suggests that the non-*enteritidis* outbreaks and the human dosing experiments may represent different conditions (there have been no reported experimental dosing experiments with *S. enteritidis*), but again there is no way to test various potential alternative hypotheses.
- b. The distribution of virulence within a serotype is lognormal. If hypothesis a. is true, the observed data are consistent with lognormal distributions. Other distributions were not examined in detail,⁴⁵ in view of the agreement with the hypothesis and the bias to selection of lognormal distributions in natural attributes. This bias is based on the observation of lognormal distributions for many biological phenomena, together with the observation (based on the central limit theorem) that the lognormal is the expected limiting distribution for any phenomenon that depends on multiple multiplicative factors.
- c. The large predominance of *S. enteritidis* outbreaks does not bias the analysis. *S. enteritidis* appears to have a distribution distinct from all other serotypes combined; but there are many fewer observations on non-*enteritidis* serotypes. Possibly each serotype has a different distribution, but there are too few observations to currently distinguish them. Once again, the approach taken is based on necessity; there is no practical alternative available.
- d. The dose-response for individual isolates is linear-exponential. For average doses in typical outbreak-sized populations within the range 1 CFU to 10¹⁰ CFU, the assumed shape of the individual isolate dose-response is unimportant in deriving the distribution of virulence (or equivalent parameters for other dose-response shapes), as shown by the comparison of Section 8.7. However, extrapolations to lower average doses may be problematic. In practice, it is impossible to derive more than one virulence parameter for each outbreak observation, so there is no practical way to distinguish dose-response shapes for illness with available data. For population-level

⁴⁵ Referee suggestions of alternatives were shown to be inconsistent with the data (Crouch and Hendrix, 2009).

- estimates of illness rates due to population average doses in the range 1 CFU to 10^{10} CFU averaged over *Salmonella* isolates (*i.e.* averaged over multiple independent outbreaks), the exact shape of the dose-response curve should not make much difference.
- e. The reported outbreak data on doses, the population sizes, and illness incidence are unbiased. For example, there is the possibility of uniformly overestimating the population size for large populations. If reported doses are uniformly biased, for example by failure to account for lack of recovery, or clumping of organisms, the analysis applies to the biased doses. Lack of bias is more important than accuracy; the analysis of Section 8.8 shows that the results for the distribution of virulence would be hardly affected even with considerably larger uncertainties in dose (and implicitly in illness incidence or population size also).
 - f. The food items selected for dose estimation in the outbreaks are representative, and the organisms are homogeneously distributed on the scale of the test samples from the selected food items. The analysis depends on having unbiased estimates of dose for each outbreak, and a lognormal uncertainty distribution for the dose. Considerable further uncertainties in dose would only marginally affect the estimates of virulence distribution (Section 8.8), provided they were adequately represented by lognormal uncertainty distributions. Non-representative food samples or substantial inhomogeneities at the scale of the test samples might introduce uncertainties that could be considerably larger than any considered here, or that might not be adequately represented by lognormal uncertainties. Rejection of the hypothesis would require substantial further analysis of the potential effects of large uncertainties and different uncertainty distributions.
 - g. Sensitive populations are uniformly more sensitive. The analysis performed here uses a crude selection of age ranges (0–5, 6–12, >12 years) and the hospitalized/non-hospitalized dichotomy, and assumes that all subpopulations with the relevant characterization will be equally sensitive. In the only large outbreak with a reported single-year age distribution ([Outbreak 1](#)), this age characterizations seem adequate; but the approach has not been adequately tested. For example, it is plausible that different age ranges might respond differentially to different serotypes or isolates. The initial draft of this analysis used only a single sensitive age range, but subsequent examination of the individual reports showed that two were necessary. The hospital/non-hospital dichotomy is based on a few case study reports, and the theoretical assumption that the population of immuno-compromised should be higher in hospitals and that the immuno-compromised should be more sensitive. The current analysis shows that there is no strong evidence for any higher sensitivity in general for hospitalized patients, but cannot rule it out. Practically, there is little alternative to the current approach using these data, because characteristics of the populations exposed are either not adequately known or not reported in sufficient detail (*e.g.* only crude age ranges are reported). Better evaluation of the hospital/non-hospital dichotomy would require data on outbreaks in which members of both classes are simultaneously exposed to the same *Salmonella* isolates.
 - h. Where subgroups are evaluated, the various subgroups are exposed in similar conditions. Where dose differences (inferred from differences in food quantities, or through measurements of different food items) between subgroups are reported, they

- are taken into account; and the differences in sensitivity between the subgroups are incorporated in the analysis. However, conditions of exposure other than dose are assumed in the analysis to be constant between the subgroups. Any biases in these differences might affect the estimates of relative sensitivity between groups. There is no practical alternative to this hypothesis, however, because differences between subgroups (except with respect to food quantities) are generally not reported.
- i. Insofar as there are *Salmonella* serotypes that do not cause illness of the type evaluated, this report omits any consideration of them. No data have been evaluated that would indicate what fraction of *Salmonella* serotypes or isolates are so included or excluded, although it is expected that only a small fraction would be excluded. Using the results obtained here as described in Section 10 either requires explicit selection of only those *Salmonella* serotypes that can cause illness, or implicitly assumes that all *Salmonella* serotypes can cause illness.

10. Practical application of the results

The results obtained are designed for use in population-level probabilistic assessments that examine multiple independent exposures. As described in Section 2.1, the prototypical such assessment would be a Monte Carlo QMRA that tracks the number of contaminating *Salmonella* CFU as a food passes through the various stages of preparation, so that the number of CFU ingested is evaluated. As described in Section 8.5 each iteration of the Monte Carlo simulation describes a random situation. To fully use the results obtained in Sections 8.4 through 8.6 each such iteration will simulate:

1. the dose to which some random member of a population is exposed,
2. the sensitivity class of the population from which that member is selected, and
3. the serotype of the *Salmonella* (*S. enteritidis* or other serotype).

The randomly selected member of the population is considered representative of the population; the results obtained here do not attempt to model variability within the population, since that variability is here subsumed within the dose-response relationship (see Section 8.1). Given this information, the results of Section 8.6 allow selection of a set of parameters for the linear-exponential dose-response curve, which then allows an estimate for the probability of illness for the individual modeled in the QMRA. Specifically:

1. generate a vector of 5 independent standard normal (mean 0, standard deviation 1) random numbers,
2. pre-multiply by the matrix given by Table 14 to obtain a vector of values for $(m_{se}, m_{ot}, \ln(\Sigma-1.69), \ln(s_M), \ln(s_Y))$,
3. apply the inverse transforms to obtain a vector of values $(m_{se}, m_{ot}, \Sigma, s_M, s_Y)$,
4. sample randomly from a lognormal distribution with median m_{se} or m_{ot} , depending on the modeled serotype, and with $\ln(\text{GSD}) = \Sigma$, to obtain a virulence λ ,
5. select the applicable sensitivity factor S ($1, s_M = s_H$, or s_Y) depending on the sensitivity class of the population modeled,
6. apply the linear-exponential dose-response model of Equation (18) with the virulence λ , the sensitivity factor S , and the modeled dose to obtain the probability of illness.

Steps 1 and 2 generate samples from a multinormal distribution with the variance-covariance matrix of Table 12, corresponding to the uncertainty distributions for the transformed parameters. Step 3 recovers the sample values for the parameters, Step 4 obtains a sample from the variability distribution for virulence, Step 5 obtains the correct sensitivity factor, and Step 6 obtains the probability of illness.

As described in Section 8.1, the dose-response models and this analysis are designed for population-level assessment that average over multiple independent exposures, not for individual assessments either of people or isolates. While the procedure just described selects individuals in a QMRA, those individuals are treated as average representations of population classes. This analysis does not attempt to evaluate inter-individual variations; there is insufficient information in the outbreak or other literature to attempt this on the basis of illness data alone. Application of the dose-response models evaluated in this assessment to individuals will therefore fail to capture potential inter-individual variations.

11. References

- Abe, K., Saito, N., Kasuga, F., and Yamamoto, S. 2004. Prolonged incubation period of Salmonellosis associated with low bacterial doses. *J. Food Prot.* 67(12):2735–2740.
- Angelotti, R., Bailey, G.C., Foter, M.J., and Lewis, K.H. 1961. *Salmonella infantis* isolated from ham in food poisoning incident. *Public Health Reports*, 76:771–776.
- Armstrong, R.W., Fodor, T., Curlin, G.T., Cohen, A.B., Morris, G.K., Martin, W.T., and Feldman, J. 1970. Epidemic *Salmonella* gastroenteritis due to contaminated imitation ice cream. *Am. J. Epidem.* 91(3):300–307.
- Bemrah, N., Bergis, H., Colmin, C., Beaufort, A., Millemann, Y., Dufour, B., Benet, J.J., Cerf, O., and Sanaa, M. 2003. Quantitative risk assessment of human salmonellosis from the consumption of a turkey product in collective catering establishments. *Int. J. Food Microbiol.* 80:17–30.
- Bille, B., Mellbin, T., Nordbring, F. 1964. An Extensive Outbreak Of Gastroenteritis Caused By *Salmonella* Newport. I. Some Observations of 745 Known Cases. *Acta Med. Scand.* 175:557–567.
- Blaser, M.J., and Newman, L.S. 1982. A review of human salmonellosis: I. Infective dose. *Rev. Infect. Dis.* 4(6):1096–1106.
- Boring, J.R., III, Martin, W.T., and Elliott, L.M. 1971. Isolation of *Salmonella typhimurium* from municipal water, Riverside, California, 1965. *Am. J. Epidemiol.* 93: 49–54.
- Buchanan, R.L., Smith, J.L., and Long, W. 2000. Microbial risk assessment: dose-response relations and risk characterization. *Int. J. Food Microbiol.* 58:159–172.
- Coleman, M., and Marks, H. 1998. Topics in Dose-Response Modeling. *J. Food Protect.* 61(11):1550–1559.
- Coleman, M.E., and Marks, H. 1999. Qualitative and quantitative risk assessment. *Food Control* 10:289–297.

- Coleman, M.E., and Marks, H.M. 2000. Mechanistic modeling of salmonellosis. *Quant. Microbiol.* 2:227–247.
- Coleman, M.E., Marks, H.M., Golden, N.J., and Latimer, H.K. 2004. Discerning strain effects in microbial dose-response data. *J. Toxicol. Environ. Health* 67A(8–10):667–85.
- Collaborative Report. 1971. A waterborne epidemic of Salmonellosis in Riverside, California, 1965; epidemiologic aspects. *Am. J. Epidemiol.* 93(1):33–48.
- Craven, P.C., Mackel, D.C., Baine, W.B., Barker, W.H., and Gangarosa, E.J. 1975. International outbreak of *Salmonella* Eastbourne infection traced to contaminated chocolate. *Lancet* Apr 5 1(7910):788–192.
- Crouch, E., and Hendrix, S. 2009. *Peer Review Comments and Responses for The Dose-Response Relationship for Non-Typhoid Salmonella*. March, 2009. Cambridge Environmental Inc. Prepared for Food Safety Inspection Service under Contract AG-3A94-C-08-0016.
- Cunnane, C. (1978). Unbiased plotting positions — a review. *J. Hydrology* 37:205–222.
- Curry, J.C. 1976. Infecting dose of *Salmonella*. Letter to the editor. *Lancet* Jun. 12 1(7972):1296.
- D'Aoust, J.Y. 1985. Infective dose of *Salmonella typhimurium* in cheddar cheese. *Am J. Epidemiol.* 122:717–720.
- D'Aoust, J.Y. 1991. Pathogenicity of foodborne *Salmonella*. *Int. J. Food Microbiol.* 12(1):17–40.
- D'Aoust, J.Y., and Pivnick, H. 1976. Small infectious doses of *Salmonella*. Letter to the editor. *Lancet* Apr. 17 1(7964):866.
- D'Aoust, J.Y., Aris, B.J., Thisdele, P., Durante, A., Brisson, N., Dragon, D., Lachapelle, G., Johnston, M., and Laidely, R. 1975. *S. eastbourne* outbreak associated with chocolate. *Journal Institut Canadien de Technologie Alimentaire* 8(4):181–184.
- D'Aoust, J.Y., Warburton, D.W., and Sewell, A.M. 1985. *S. typhimurium* phage type 10 from cheddar cheese implicated in a major Canadian foodborne outbreak. *J. Food Protect.* 48:1062–1066. [Cited by FAO/WHO, 2002b, but not examined.]
- EPA (Environmental Protection Agency). 1997. *Exposure Factors Handbook*. 3 vols. EPA/600/P-95/002Fa-c, August 1997. (Available: <http://www.epa.gov/ncea/pdfs/efh/>; accessed 9/8/05).
- FAO/WHO. 2002a *Risk assessments of Salmonella in eggs and broiler chickens - Interpretative Summary (Microbiological Risk Assessment Series 1)*. ISBN 92-5-104873-8. (Available: http://www.fao.org/es/esn/jemra/Salmonella_en.stm; accessed 6/8/05).
- FAO/WHO. 2002b *Risk assessments of Salmonella in eggs and broiler chickens. (Microbiological Risk Assessment Series 2)*. ISBN 92-5-104873-8. (Available: http://www.fao.org/es/esn/jemra/Salmonella_en.stm; accessed 6/8/05).
- Fazil, A.M. 1996. *A quantitative risk assessment model for Salmonella*. Drexel University, Philadelphia PA. [Dissertation]. [This dissertation is cited by FAO/WHO (2002a,b) but was not consulted for this work.]

- Fazil, A.M. 1999. *Salmonella*. Pp. 389–411 in: Haas, C.N., Rose, J.B., and Gerba, C.P., *Quantitative Microbial Risk Assessment*, John Wiley & Sons, Inc., New York, 1999, ISBN 0-471-18397-0.
- Fontaine, R.E., Arnon, S., Martin, W.T., Vernon, T.M. Jr, Gangarosa, E.J., Farmer, J.J. III, Moran, A.B., Silliker, J.H., and Decker, D.L. 1978. Raw hamburger: an interstate common source of human salmonellosis. *Am. J. Epidemiol.* 107(1):36–45.
- Fontaine, R.E., Cohen, M.L., Martin, W.T., and Vernon, T.M. 1980. Epidemic salmonellosis from cheddar cheese: Surveillance and Prevention. *Am. J. Epidemiol.* 111: 247–253.
- Fowinkle, E.W., Armes, W.H., Barrick, J.H., Cutter, E., Powell, J., Edgett, J.A., Curlette, W., Zellner, S.R., and Dunn, C.G. 1970. A *Salmonella thompson* outbreak traced to barbecued pork — Tennessee. Washington, DC: US Department of Health, Education, and Welfare, Center for Disease Control. *Salmonella Surveillance, Report No. 99*, 1–3.
- FSIS. 1998. *Salmonella Enteritidis Risk Assessment. Shell Eggs and Egg Products*. Final Report. Submitted June 12, 1998, Revised with editorial corrections August 10, 1998. Prepared for The Food Safety and Inspection Service by the *Salmonella* Enteritidis Risk Assessment Team. (Available: <http://www.fsis.usda.gov/OPHS/risk/index.htm>; accessed 9/6/05).
- FSIS. 2004. *Draft Risk Assessments of Salmonella Enteritidis Shell Eggs and Salmonella spp. in Egg Products*. (Available: http://www.fsis.usda.gov/Science/Risk_Assessment_SE_Eggs/index.asp; accessed 6/8/05).
- George, R.H. 1976. Small infectious dose of *Salmonella*. Letter to the editor. *Lancet* May 22 1(7969):1130.
- Glynn, J.R., and Bradley, D.J. 1992. The relationship between infecting dose and severity of disease in reported outbreaks of *Salmonella* infections. *Epidemiol. Infect.* 109:371–388.
- Glynn, J.R., and Palmer, S.R. 1992. Incubation period, severity of disease, and infecting dose: evidence from a *Salmonella* outbreak. *Am. J. Epidemiol.* 136(11):1369–1377.
- Glynn, J.R., Hornick, R.B., Levine, M.M., and Bradley, D.J. 1995. Infecting dose and severity of typhoid: analysis of volunteer data and examination of the influence of the definition of illness used. *Epidemiol. Infect.* 115:23–30.
- Greenberg, B. 1964. Experimental transmission of *Salmonella Typhimurium* by houseflies to man. *Am. J. Hygiene* 80(2):149–156.
- Greenwood, M.H., and Hooper, W.L. 1983. Chocolate bars contaminated with *Salmonella* napoli: an infectivity study. *Br. Med. J. (Clin. Res. Ed.)* 286(6375):1394.
- Havelaar, A., Garssen, J., Takumi, K., Koedam, M., Ritmeester, W., de la Fonteyne, L., Bousema, T., and Vos, J. 2004. Intraspecies variability in the dose-response relationship for *Salmonella* enteritidis associated with genetic differences in cellular immune response. *J. Food. Prot.* 67(9):2008–2015.
- Havelaar, A.H., Garssen, J., Takumi, K., Koedam, M., Dufrenne, J.B., van Leusden, F.M., de la Fonteyne, L., and Vos, J.G. 2000. *Dose response relationships for gastrointestinal pathogens in an animal model*. Rijksinstituut voor Volksgezondheid en Milieu (RIVM)

- Report 284550 007, February 2000. (Available: <http://www.rivm.nl/bibliotheek/rapporten/284550007.html>; accessed 6/8/05).
- Havelaar, A.H., Garssen, J., Takumi, K., Koedam, M.A., Dufrenne, J.B., van Leusden, F.M., de la Fonteyne, L., Bousema, J.T., and Vos, J.G. 2001. A rat model for dose-response relationships of *Salmonella* enteritidis infection. *J. Appl. Microbiol.* 91:442–452.
- Hennessy, T.W., Hedberg, C.W., Slutsker, L., White, K.E., Besser-Wiek, J.M., Moen, M.E., Feldman, J., Coleman, W.W., Edmonson, L.M., MacDonald, K.L., and Osterholm, M.T. 1996. A national outbreak of *Salmonella* enteritidis infections from ice cream. *New Engl. J. Med.* 334:1281–1286.
- Holt, P.S., and Chaubal, L.H. 1997. Detection of motility and putative synthesis of flagellar proteins in *Salmonella pullorum* cultures. *J. Clin. Microbiol.* 35(4):1016–1020.
- Kahn, H.D., and Stralka, K. 2008. Estimated daily average per capita water ingestion by child and adult age categories based on USDA's 1994–1996 and 1998 continuing survey of food intakes by individuals. *J. Exp. Sci. Environ. Epidemiol.* Advanced online publication May 2008; doi:10.1038/jes.2008.29.
- Kang, S-H, Kodell, R.L., and Chen, J.J. 2000. Incorporating model uncertainties along with data uncertainties in microbial risk assessment. *Regul. Toxicol. Pharmacol.* 32(1):68–72.
- Kasuga, F., Hirota, M., Wada, M., Yunokawa, T., Toyofuku, H., Shibatsuji, M., Michino, H., Kuwasaki, T., Yamamoto, S., and Kumagai, S. 2004. Archiving of food samples from restaurants and caterers--quantitative profiling of outbreaks of foodborne salmonellosis in Japan. *J. Food Prot.* 67(9):2024–2032.
- Kothary, M.H., and Babu, U.S. 2001. Infective Dose of Foodborne Pathogens in Volunteers: A Review. *J. Food Safety* 21:49–73.
- Laidley, R., Handzel, S., Severs, D., and Butler, R. 1974. *Salmonella* weltevreden outbreak associated with contaminated pepper. *Epidem. Bull., 111th Welf. Can.* 18(4):62.
- Lang, D.J., Kunz, L.J., Martin, A.R., Schroeder, S.A., and Thomson, L.A. 1967. Carmine as a source of nosocomial salmonellosis. *New Engl. J. Med.* 276(15):829–832.
- Latimer, H.K., Jaykus, L-A., Morales, R.A., Cowen, P., and Crawford-Brown, D. 2001. A Weighted Composite Dose-Response Model For Human Salmonellosis. *Risk Anal.* 21:295–305.
- Latimer, H.K., Jaykus, L-A., Morales, R.A., Cowen, P., and Crawford-Brown, D. 2002. Sensitivity analysis of *Salmonella enteritidis* levels in contaminated shell eggs using a biphasic growth model. *Int. J. Food Microbiol.* 75:71–87.
- Levy, M., Fletcher, M., Moody, M., and 27 others. 1996. Outbreaks of *Salmonella* serotype enteritidis infection associated with consumption of raw shell eggs – United States, 1994–1995. *MMWR.* 45:737–742.
- Lipson, A. 1976. Infecting dose of *Salmonella*. Letter to the editor. *Lancet* May 1;1(7966):969.
- Matsui, T., Suzuki, S., Takahashi, H., Ohyama, T., Kobayashi, J., Izumiya, H., Watanabe, H., Kasuga, F., Kijima, H., Shibata, K., and Okabe, N. 2004. *Salmonella* Enteritidis outbreak

- associated with a school-lunch dessert: cross-contamination and a long incubation period, Japan, 2001. *Epidemiol. Infect.* 132(5):873–9.
- Mackenzie, C.R., and Livingstone, D.J. 1968. Salmonellae in fish and food. *S. African Med. J.* 42(38):999–1003.
- McCullough, N.B., and Eisele, C.W. 1951a. Experimental human Salmonellosis. I. Pathogenicity of strains of *Salmonella meleagridis* and *Salmonella anatum* obtained from spray-dried whole egg. *J. Infect. Dis.* 88(3):278–289.
- McCullough, N.B., and Eisele, C.W. 1951b. Experimental human Salmonellosis. II. Immunity studies following experimental illness with *Salmonella meleagridis* and *Salmonella anatum*. *J. Immunol.* 66(5):595–608.
- McCullough, N.B., and Eisele, C.W. 1951c. Experimental human Salmonellosis. III. Pathogenicity of strains of *Salmonella newport*, *Salmonella derby*, and *Salmonella bareilly* obtained from spray-dried whole egg. *J. Infect. Dis.* 89(3):209–213.
- McCullough, N.B., and Eisele, C.W. 1951d. Experimental human Salmonellosis. IV. Pathogenicity of strains of *Salmonella pullorum* obtained from spray-dried whole egg. *J. Infect. Dis.* 89:259–265.
- Mellbin, T., Nordbring, F., and Bille, B. 1965. An Extensive Outbreak Of Gastroenteritis Caused By *Salmonella* Newport. II. Some Clinical Observations On 488 Hospitalized Cases And The Results Of Cultures. *Acta Med. Scand.* 177:437–444.
- Meynell, G.G., and Meynell, E.W. 1958. The growth of micro-organisms in vivo with particular reference to the relation between dose and latent period. *J. Hyg. (Lond.)* 56(3):323–346.
- Mintz, E.D., Cartter, M.L., Hadler, J.L., Wassell, J.T., Zingser, J.A., and Tauxe, R.V. 1994. Dose-response effects in an outbreak of *Salmonella* enteritidis. *Epidemiol. Infect.* 112(1):13–23.
- Moon, H., Chen, J.J., Gaylor, D.W., and Kodell, R.L. 2004. A comparison of microbial dose-response models fitted to human data. *Regul. Toxicol. Pharmacol.* 40:177–184.
- Narain, J.P., and Lofgren, J.P. 1989. Epidemic of restaurant-associated illness due to *S. Newport*. *Southern Medical Journal* 82:837–840.
- Nordbring, F., Bille, B., and Mellbin, T. 1964. An Extensive Outbreak Of Gastroenteritis Caused By *Salmonella* Newport. 3. Serum Agglutinins In 412 Hospitalized Cases. *Acta Pathol. Microbiol. Scand.* 60:131–143.
- Oscar, T. 2004. Dose-response model for 13 strains of *Salmonella*. *Risk Anal.* 24(1):41–49.
- Oscar, T.P. 2004. A quantitative risk assessment model for *Salmonella* and whole chickens. *Int. J. Food Microbiol.* 93:234–237.
- Palmer, S.R., Watkeys, J.E., Zamiri, I., Hutchings, P.G., Howells, C.H., and Skone, J.F. 1990. Outbreak of *Salmonella* food poisoning amongst delegates at a medical conference. *J. R. Coll. Physicians Lond.* 1990 24(1):26–29.
- Reitler, R., Yarom, D., and Seligmann, R. 1960. The enhancing effect of staphylococcal enterotoxin on *Salmonella* infection. *The Medical Officer* 104:181.

- Rejnmark, L., Stoustrup, O., Christensen, I., and Hansen, A. 1997. Impact of infecting dose on severity of disease in an outbreak of food-borne *Salmonella* enteritidis. *Scand. J. Infect. Dis.* 29:37–40.
- Ross, E.C., Campbell, K.W., and Ongerth, H.J. 1966. *Salmonella* Typhimurium contamination of Riverside, Calif., supply. *J. Amer. Water Works Assn.* 58:165–174.
- Roussanov, B., Hawkins, D.M., and Tatine, S.R. 1996. Estimating bacterial density from tube dilution data by a Bayesian method. *Food Microbiol.* 13:341–363.
- Slob, W. 1994. Uncertainty analysis in multiplicative models. *Risk Anal.* 14(4):571–576.
- Taylor, D.N., Bopp, C., Birkness, K., and Cohen, M.L. 1984. An outbreak of salmonellosis associated with a fatality in a healthy child: a large dose and severe illness. *Am. J. Epidemiol.* 119(6):907–912.
- Teunis, P.F., Nagelkerke, N.J.D., and Haas, C.N. 1999. Dose response models for infectious gastroenteritis. *Risk Anal.* 19(6):1251–60.
- Teunis, P.F.M., van der Heijden, O.G., van der Giessen, J.W.B., and Havelaar, A.H. 1996. *The dose-response relation in human volunteers for gastro-intestinal pathogens*. Rijksinstituut voor Volksgezondheid en Milieu (RIVM) Report 284550002, May 1996. (Available: <http://www.rivm.nl/bibliotheek/rapporten/284550002.html>; accessed 6/8/05).
- Urfer, E., Rossier, P., Méan, F., Krending, M.-J., Burnens, A., Bille, J., Francioli, P., and Zwahlen, A. 2000. Outbreak of *Salmonella* braenderup gastroenteritis due to meat pies: clinical and molecular epidemiology. *Clin. Microbiol. Infect.* 6(10):536–542.
- USDA (United States Department of Agriculture), 2000. Continuing survey of food intakes by individuals (CSFII) 1994–96, 1998. Agricultural Research Service. CD-ROM. (Also available: <http://www.ars.usda.gov/Services/docs.htm?docid=14531>; accessed 12/19/08).
- Varela, G., and Olarte, J. 1942. Infección experimental del hombre con *Salmonella* anatum. *Medicina Revista Mexicana* 22(406):57–58.
- Vought, K.J., and Tatini, S.R. 1998. *Salmonella* enteritidis contamination of ice cream associated with a 1994 multistate outbreak. *J. Food Prot.* 61(1):5–10.

Appendix A. Evaluation of Individual Outbreaks

References for this Appendix are in the general reference list in Section 11.

A.1. Outbreak data

These are discussed in order of their first appearance in Section 3.2 in Table 3, Table 4 and the subsequent discussion. Each entry or group of entries in Table 3 and Table 4 is here given an Outbreak number to distinguish them and allow easy cross-reference, so more than one Outbreak number may refer to the same outbreak. Each entry repeating an outbreak is hyperlinked to the discussion of that outbreak. The analyses described here are implemented in the workbook “Epi Attack Data.xls” (see Appendix B), primarily in the sheet “Analyses” (other locations are given where appropriate).

A.1.1. Entries in Table 3

Outbreak 1. [FSIS #1](#), [FAO/WHO #1](#) [Used here; [Serial #1](#)]

References: Boring, J.R. *et al.* 1971; Collaborative Report, 1971; Ross, *et al.* 1966.

The references describe an epidemic of salmonellosis from *S. typhimurium* in municipal drinking water in Riverside, Ca. in 1965, with an estimated 16,000 persons with symptomatic disease. The single reported quantitative drinking water analysis was of a composite sample of unequal quantities (a few ml to 500 ml) from 74 samples, each of 2 liters. The design was effectively a 4-dilution, 5-tube MPN procedure, using 1, 10, 100, and 1000 ml per tube (total water volume tested was 5,555 ml). The reported positive response pattern of (0,1,4,5) at the four dilutions corresponds to the reported result of 17 CFU/liter (Boring *et al.*, 1971). Of the 74 samples of 2 liters, 5 tested positive, suggesting a highly non-uniform contamination; however, since ingestion was over a period of days such non-uniformity would average out to some degree. The tested sample corresponds to a random sampling from the system, in some sense, and would be similar to “sampling” for individual exposures. It is assumed that this composite was representative of the water ingested by the populations exposed.

The number exposed and number ill found in a house-to-house sample survey are given in the Collaborative Report (1971). The numbers are given by age groups, with the <5 age group showing a distinctly different attack rate. There was no significant difference in attack rates by single year of age below 5 years or by age groups above 5 years of age. The <5 and 5–9 year age groups are assumed sensitive because of age (see Section 6). For water intake estimates, the U.S. age-related results reported by Kahn and Stralka (2008) for “community water ingestion” were used. The intakes by age ranges given by Kahn and Stralka (2008) were combined to age ranges <6 years, 6 to <11 years, and 11+ years and used for the Collaborative Report (1971) <5, 5–9, and 10+ year age groups respectively (these are the best matching age ranges for the two publications). Exposure for 11 days (May 21 through June 1) was assumed. Individuals would have sampled tap water many times during 11 days. With the linear dose-response curve assumed in this work as the primary dose-response curve, the total probability for illness over the 11 days would be approximately proportional to the product of total intake and average

concentration (either probabilities or doses add, assuming independence of outcomes from each exposure and relatively small probabilities).

The dose uncertainties in this outbreak are primarily due to the uncertainties in concentration, so the three age groups are evaluated as three dose-groups with known relative doses (that differ because of different water ingestion rates in different age ranges).

Outbreak 2. [FSIS #2](#), [FAO/WHO #16](#) [Used here; [Serial #41](#)]

Reference: Lipson, 1976.

A case report of dosing with contaminated pancreatin of a single infant patient (age 9 mo.) with cystic fibrosis. The pancreatin was contaminated with less than 8 CFU/g of *S. schwarzengrund*, and the patient was dosed at approximately 700 mg every 6 hours for 10 days, followed by less than 1.4 g every 6 hours in the 36 hours before onset of illness (less than a total of 291 CFU). No information is given on the method of determination of concentration of *Salmonella* in the pancreatin; the total value of 291 CFU has been treated as an upper bound on the dose. The single patient in this “outbreak” is treated here as sensitive because of age, but not because of hospitalization. Although the report (Lipson, 1976) indicates that other young patients with cystic fibrosis were treated with antibiotics, no indication is given that the reported case was treated with anything but the contaminated pancreatin.

Outbreak 3. [FSIS #3](#) & [FSIS #7](#); [FAO/WHO #6](#) [Not used here]

Reference: Fontaine *et al.*, 1978.

An outbreak associated with consumption of hamburger. There is no attack rate derivable from the reported data. FSIS (1998) arbitrarily used 45% attack rate. The two estimates of 48 or 46 cases reported for this outbreak by FSIS (1998) do not correlate with reported observations (18/26 outbreak-associated cases in Colorado were definitely primary, but of the remaining 8, 2 could not be contacted, 3 were transients, and 3 met the alternative case definition; there were 9 cases in Maryland, and 21 cases in Florida). There is very little information on dose; 6 to 23 CFU/100g is reported for some frozen samples, but it is claimed that substantial die-off was likely during freezing. FSIS (1998) suggested 60 to 2300 CFU/100g ingested based on the assumption of 1 to 2 logs die-off during freezing. FSIS (1998) does not state the estimated consumption, although the dose estimate given of 60-230 CFU dose suggests an assumption of 10 g consumption (some cases are reported to merely have nibbled raw hamburger). FAO/WHO (2002b) suggests that cooking or partial cooking would result in a die-off of 1 to 2 logs, matching the 1 to 2 logs that the original authors suggest would be the effect of freezing on the tested samples, so that the measured 6 to 23 CFU/g would correspond to what was ingested, and assumed 100g eaten. However, this is all speculation, so this material is not included in the dose-response characterization.

Outbreak 4. [FSIS #4](#); [FAO/WHO #10](#) [Not used here]

Reference: D'Aoust, *et al.*, 1975.

An outbreak in both the U.S. and Canada associated with contaminated chocolate balls. This report discusses the Canadian part of the outbreak. [Outbreak 27](#) examines the U.S. part of the

outbreak. FSIS (1998) arbitrarily assumed an attack rate of 45%, but this is almost certainly far too high. The attack rate might be estimable if the total production of chocolate balls were known, but that is not reported. The FSIS (1998) estimate of dose could be improved, but in the absence of any estimate of attack rate, this material was not incorporated in the dose response characterization.

Outbreak 5. [FSIS #5](#) & [FSIS #8](#); [FAO/WHO #2](#) [Used here; [Serial #2](#)]

Reference: Fontaine, *et al.*, 1980.

An outbreak in Denver and Pueblo associated with consumption of cheddar cheese served primarily in a well-defined way in Mexican-style restaurants. Contamination was traced to a shipment of 796 18 kg blocks of cheese (14,328 kg) received by a distributor July 20, 1976. Distribution began July 22, cases began July 23, and the cheese was collected August 9.

In employees of the Group A restaurants (those which had received this shipment) who ate cheese there and nowhere else during the appropriate period, the incidence was 21/163; but only 4/10 Group A restaurants had any illnesses. Among those 4 restaurants with cases, there was a 28% attack rate (implying 21/75 employees eating contaminated cheese and becoming ill). At 3 restaurants in Pueblo, 16/46 employees became ill, a 35% attack rate. In household contacts who ate cheese at the same restaurant as the first identified case within the household, 34/94 (36%) became ill. There are no significant differences between these attack rates, and the overall attack rate is $(21+16+34)/(75+46+94)=71/215$ (it is unclear how FAO/WHO, 2002b, obtained an attack rate of 68/205).

S. heidelberg was isolated from cheese on August 16. Six production lots showed contamination, equivalent to 4917 kg cheese (34% of total shipment), of which 2087 kg was retrieved. Measured MPN concentrations are reported as 0.36/100g and less than 1.8/100g in two determinations by different laboratories. For the standard MPN approach these imply detection patterns of (0, 0, 1) and (0, 0, 0) in three-tube, 3-dilution, tests using (1, 10, 100) gram, and (0.2, 2, 20) gram quantities respectively. [FAO/WHO (2002b) mis-reports these results as 0.36/100 g and 1.8/100 g respectively, omitting the “less than” for the second].

The probable food serving size was approximately 1 oz cheese per serving (28 g), according to restaurant records. Combining both MPN estimates (assuming the 3-tube patterns given above) gives 0.29 CFU/100 g with GSD about 2.9. But at 28 g/serving, that gives 0.08 CFU/serving, which is impossibly low; there must have been die-off before the detection was made.

FAO/WHO (2002b) arbitrarily assumes 99% to 99.9% die-off before analysis and 30 to 300 CFU dose, while FSIS (1998) assumes 100 to 500 CFU dose. However, there is no basis for either of these estimates, beyond the hypothetical calculation by the original authors of the potential dose had there been 99.9% die-off.

There clearly must have been die-off, but the rate of die off must be consistent with the observed epidemic of illness. The pattern of that epidemic is given by the original authors in Denver and Pueblo, the differences in shape presumably being due to differences in the date and rate of introduction of the contaminated cheese into food servings. To evaluate the potential die-off, it

is here assumed that the die-off of the *Salmonella* was exponential between the start of the epidemic and the time of analysis. It has to be acknowledged that the conditions of storage could have changed after the recall, so the die-off rate could have changed then; moreover, the die-off could have been non-exponential over some of this period.

The reported epidemic observations over approximately 18 days prior to the cheese being collected were evaluated to see what ranges of exponential die-offs are consistent with the observed initial increase in cases, followed by an apparent reduction in incidence that occurred *before* the collection of the cheese (Fontaine *et al.*, 1980, Figure 1). An arbitrary-shaped ramp-up in exposure was introduced (modeled by an exponential-polynomial model with positive coefficients to allow fairly arbitrary increasing shape, with saturation), modified with an exponential die-off (calculations are performed in the sheet Font_1980 in the workbook “Epi Attack Data.xls”). This allows for an increasing quantity of contaminated cheese to be gradually introduced into the food supply, with a saturation corresponding to the maximum rate of use, with the concentration of *Salmonella* decreasing exponentially through time. The model used for the expected number of illnesses is then

$$n(t) = \alpha \left(1 - \exp \left(- \sum_{i=1}^6 \beta_i t^i \right) \right) \exp(-\lambda t) \quad (\text{A.1})$$

where $n(t)$ is the number of illnesses on day t after initial delivery of the cheese, α and β_i are adjustable parameters that allow for the number of persons at risk and the variable rate of introduction of contaminated cheese into the food supply, and λ is the die-off rate. The observed number of illnesses per day in Denver and Pueblo separately were fitted to this curve using Poisson regression (with separate α and β_i parameters for the two cities) to estimate the range of die-off rate λ consistent with the observations.

As an estimate of average dose, the average estimated concentration over the 18 days of the epidemic prior to the cheese recovery was used together with the 1 ounce serving size.

The MLE for die-off rate is 0.158 per day, corresponding to die-off of 99% between the beginning of the epidemic and the time of *Salmonella* analysis, with a ln(GSD) of about 0.96, giving an uncertainty range of 92% to 99.8% die-off (90% confidence interval) in this period. Averaging over the period of the epidemic gives an average concentration (MLE) of 6.9 CFU/100 g, corresponding to a median dose estimate of 1.93 CFU.

The die-off rate found using this analysis is consistent with die-off rates for *Salmonella typhimurium* during cheddar cheese ripening observed by Goepfert *et al.* (1968); see the sheet Font_1980 in workbook “Epi Attack Data.xls”. Figure 15 shows the die-off observed by Goepfert *et al.* (1968) during ripening of deliberately contaminated cheddar cheese, together with the range⁴⁶ of die-off estimated in this analysis over the 27 days of the outbreak. The age of the cheese associated with this outbreak is not stated; for Figure 15 it is assumed to have been

⁴⁶ The range shown corresponds to the estimates associated with MLE and 95% confidence limits on the average concentration over the 18 days of the epidemic. Each data point given by Goepfert *et al.* (1968) is the mean of measurements in 9 cheddar cheeses each made in a separate vat using the same process, but the variability and uncertainties are not stated.

approximately 60 days old at the time of distribution, corresponding to standard practice at the time according to Fontaine *et al.* (1980). The absolute number of *Salmonella* per gram in the outbreak cheese has been multiplied by 100 to bring it into the same range as the deliberate contamination introduced by Goepfert *et al.* (1968) — the figure is intended to show only the approximate correspondence of the rates of die-off (the slopes of the curves), not any correspondence in absolute numbers.

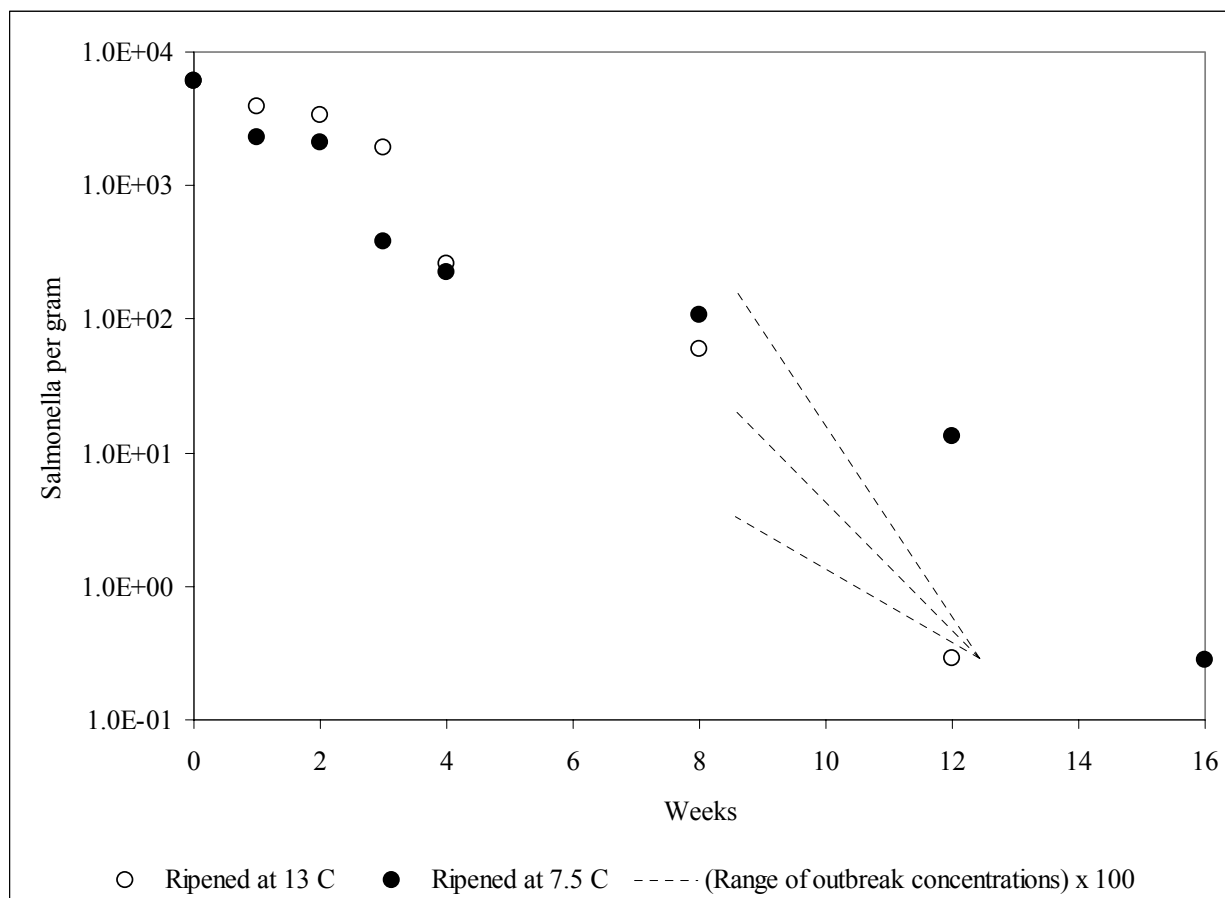


Figure 15 Die-off of *S. typhimurium* during cheddar cheese ripening (Goepfert *et al.*, 1968, Figure 2; uncertainties not listed) compared with 100 times the range of concentrations estimated in this analysis.

Outbreak 6. [FSIS # 6](#); [FAO/WHO #15](#) [Used here; [Serial #40](#)]

Reference: George, 1976.

A single (1 year old, sex not specified) patient became ill after a single dose of 0.2 g of pancreatic extract, with a concentration given as 1000 cells/g *S. schwarzengrund*. No indication is given of the method of measurement, so the uncertainty is strictly unknown; $\ln(\text{GSD})$ for this analysis has been set to a value typical for MPN methods (see Section 4.4). This patient was potentially doubly sensitive, since he or she was a young child in hospital. The patient had “failed to thrive,” but there is no indication of other medication or of any known condition that

might have made him or her susceptible. This case is therefore treated as sensitive because of age, but not because of hospitalization.

Outbreak 7. [FSIS #7](#) [Not used here]

See [Outbreak 3](#); FSIS #7 corresponds to a different point estimate of dose and different number of cases for the same outbreak.

Outbreak 8. [FSIS #8](#) [Used here; [Serial #2](#)]

See [Outbreak 5](#); FSIS #8 corresponds to a different point estimate of dose for the same outbreak.

Outbreak 9. [FSIS #9](#); [FAO/WHO #5](#) [Used here; [Serial #3](#)]

Reference: Armstrong *et al.* 1970.

Contaminated frozen chiffonade (non-dairy imitation ice cream) was implicated in outbreaks affecting 14 banquets. One banquet in particular (#3 in the paper) had a limited menu, so most people ate the chiffonade; that banquet had an attack rate of 55% based on a sample (there were 1400 people at the banquet; but the sample size was not stated). A total of 176 ill persons from the 14 banquets were interviewed. Based on “representative samples” of persons attending seven of the events gave an estimated attack rate of 52%. This suggests a total sample of about 338. Three banquets (#3, #4, and #10 in the paper) were “studied in more detail.” In #4 and #10 there was a significantly higher attack rate in those who ate the chiffonade vs. those who did not.

The attack rate for those who ate the chiffonade is here based on the observation of 55% attack rate for banquet #3, where it is fairly certain that the majority of the persons ate the chiffonade. To estimate the attack rate uncertainty requires an estimate of the sample size. “Representative samples” sufficiently large to estimate attack rates were obtained from seven banquets, including #3, with the samples being considered too small or biased for the other seven. The number of fecal samples obtained is assumed to be a minimum on the sample size for each banquet. Assuming the same sample size for the seven “representative samples” then implies a sample size less than about 45 for each banquet. An average sample size of 40 for the representative samples, and about 7 or 8 for the non-representative samples (or more where more fecal samples were collected), seems more likely. Such sizes imply a total sample of about 335 to 341 and a total number ill of 176 among that sample, for all 14 banquets (using banquet-specific attack rates where given, and the overall average attack rate for the other banquets). The attack rate of 55% for banquet #3 translates into about around 22 ill from 40 total who probably ate the chiffonade. The uncertainty for the other banquets is substantially higher because of doubts about whether individuals ate the chiffonade or not. The FAO/WHO (2002b) discussion considerably underestimates the uncertainty in attack rate, since it appears to have been assumed that all persons at the banquet were interviewed.

Four samples of chiffonade grew two distinguishable *S. typhimurium* (a total of 3 types of *Salmonella* were found, *typhimurium*, *braenderup*, and *typhimurium* var. *copenhagen*, with 3 distinct phage-susceptibilities for samples of *typhimurium*). MPN counts were “113 or fewer *Salmonellae* per 75 g serving” after storage at -20°C (up to 4 positive samples may have been quantitatively tested), but the range was not given. Inoculation studies showed “no more than two logs” reduction during such storage.

The attack rate uncertainty is adequately represented by the estimate of 22/40 ill. However, the dose uncertainty is considerable; all that is known is that doses were less than about 11300 CFU per serving; but this is the result of up to four non-null observations of 113 CFU/serving or lower, combined with an unknown reduction of less than two logs due to storage conditions. This value is used as an absolute upper bound on the dose.

Outbreak 10. [FSIS #10](#) & [FSIS #11](#); [FAO/WHO #3](#) [Not used here]

Reference: Lang, *et al.* 1967.

This paper discusses 21 total cases of *Salmonella cubana* infection, among which 12 cases followed carmine ingestion. There is no indication of the total number of carmine ingestions except that several such patients were located who were not ill. No basis is stated for the assumed number of carmine-exposed individuals (14–20) used by FAO/WHO in their analysis, and no reference anywhere for the numbers (28 cases, 100% infection) used by FSIS. Doses are less uncertain: ½ to 2 capsules, possibly repeated after several days or in subsequent studies. Measured concentration was 30,000 CFU/capsule. In the absence of any reported data that allow estimates of the infection rate, this report is not used.

Outbreak 11. [FSIS #11](#) [Not used here]

See [Outbreak 10](#). FSIS #11 corresponds to a different point estimate of dose for the same outbreak.

Outbreak 12. [FSIS #12](#) & [FSIS #14](#) [Used here; [Serial #53](#)]

Reference: Reitler, *et al.*, 1960

Two outbreaks were reported in two families. In the first family, five people (ages 8, 10, 16, 37, and 48 years) ate goat's cheese for breakfast with 10^9 *S. zanzibar*/g. All five were ill. Six days later, six people in a second family ate cheese from same batch, now with 1,500 *S. zanzibar*/g but also with *Staphylococcus aureus* (2.8×10^7 /g). All six were ill faster (onset 3–4 hrs, vs 8–12 hrs), and for longer than the first family, attributed to initial staphylococcal poisoning followed by the effect of *Salmonella*. The authors made an estimate of 100 g per meal in both cases. FSIS uses the second family (for FSIS #12) and the first family (for FSIS #14 see [Outbreak 14](#), although FSIS #14 gives the number of people affected as 8, contrary to the 5 given in Reitler *et al.*, 1960), but in view of the staphylococcal infection the attribution of illness for the second family is not clear. The outbreak in the second family is therefore not used here.

The outbreak data for the first family are used, although the extreme dose ($\sim 10^{11}$ CFU) and resulting 100% incidence provide negligible additional information⁴⁷. Reitler *et al.* (1960) state that the quantitative estimate of *Salmonella* concentration was obtained “by culturing a mixture from different parts of the specimen in increasing dilutions.” This is here interpreted as some form of MPN assay, so the generic MPN estimate of uncertainty (Section 4.4) is assumed. The

⁴⁷ The amount of added information is negligible because for such an extreme dose almost any dose-response curve (based on other available observations) would predict 100% incidence.

quantity of goat cheese eaten for breakfast was estimated as “around 100 grm.,” which estimate is adopted here with an uncertainty estimate of a factor of 2.

Outbreak 13. [FSIS #13](#); [FAO/WHO #4](#) [Used here; [Serial #4](#)]

Reference: Angelotti, *et al.*, 1961.

Eight persons in a family became ill after eating slices of raw or fried contaminated ham. The ages of the persons are not specified, although they included a mother, father, and two sons of grade school age. There is insufficient information to separate family members by age, so all are treated together as a non-sensitive group (see Section 6). The paper argues for *S. infantis* as the etiologic agent — measurements also indicated high total counts of aerobic bacteria, primarily of staphylococci and fecal streptococci, and inspection indicated unsanitary storage conditions at the supermarket from which the ham was purchased. Samples of the ham measured 23,000/g *S. infantis* using plating, with an assumed uncertainty $\ln(\text{GSD})$ of 0.2 (Section 4.4). Two patients ate three ham sandwiches (so at least 12 slices were eaten in total), the total ham size was 5.5 lb (2,500 g), and some was left over (the laboratory used a sample of 30 g). The authors estimate a slice size of 50 to 100 g, suggesting a total consumption of 600 to 1200 g. Taking these as 95% confidence limits on a lognormal uncertainty gives a median average dose of 2.4×10^6 CFU with a total $\ln(\text{GSD})$ of 0.29.

Outbreak 14. [FSIS #14](#) [Used here; [Serial #53](#)]

See [Outbreak 12](#).

Outbreak 15. [FSIS #15](#) & [FSIS #16](#); [FAO/WHO #12](#) [Used here; [Serial #5](#)]

References: Hennessy, *et al.*, 1996; Vought and Tatini, 1998.

Hennessy *et al.* (1996) describe a cross-sectional survey (that is assumed here to be representative) from an outbreak due to contaminated ice cream produced by Schwan. For those who ate Schwan ice cream products, purchased between Sept. 10 and Sept. 20, 1994 (a period of 11 days), the incidence of illness (diarrhea and fever or chills) was 30/452 (vs. 0 of 89 persons in the same households not eating the ice cream). The age range of the respondents was not given; they are here treated as a group that is not sensitive due to age or hospitalization.

Recalled Schwan ice cream products tested positive for production dates of 8/25/94 (date code 237) and 8/26/94 (date code 238), and the first cases occurred around 9/5/94, indicating a minimum delay of around 10 days from production to consumption. There were also positive isolations from product that was produced on 9/12/94 (date code 255) and 9/15/94 (date code 258),⁴⁸ and the last outbreak case was around 10/15/94 (although there were sporadic cases after that, assumed here to be unassociated with the ice cream). These observations indicate a range of about 10 to 45 days in the product pipeline (from production to consumption).

Outbreak cases occurred mostly in the period Sept 7 to Oct 15, a period of about 40 days. Based on the isolations from tested products, these cases were primarily due to about 2 days of

⁴⁸ Vought and Tatini (1998) do not mention the second of these; their table of positive isolation results is stated to come from an FDA and Minnesota Department of Health press release dated 7th October, 1994.

production. (8/25/94 to 8/26/94, date codes 237 and 238). From the reported information, production appears to have been about 5 d/wk (although this is not clear; *e.g.* 8/11 days production are shown between date codes 234 to 244 inclusive in Vought and Tatini, 1998, the period centered around the two days of principal contamination and with the same length as the cross-sectional survey period; although this could be due to lack of complete coverage of production dates in recovered material). The Sept. 10 to Sept. 20 purchase dates for the survey (Hennessy *et al.*, 1996) fall in the period when contaminated product from 8/25/94 to 8/26/94 (date codes 237 and 238) would have been on the shelves (but is likely outside the period when contaminated product from 9/12/94 and 9/15/94 would be available on store shelves). However, it is likely that only some fraction of the sampled people got ice cream from the contaminated lots. Clearly, product from the 2 days of production of contaminated lots could be present for about 40 days (the length of the outbreak). If production occurred for 8/11 days, then 2 days represents $2 \times 11/8$ calendar days of production, and the average fraction of contaminated material present on the shelves throughout the 40 day period is $2 \times 11/(8 \times 40) = 0.068$, which is essentially the same as the incidence of illness in the survey ($30/452 = 0.066$). However, the survey was for 11 days, so it is possible that the average fraction of contaminated material shopped was around $2 \times 11/(8 \times 11) = 0.25$, so that the fraction of the surveyed people eating contaminated ice cream could be about 0.25. In principle, the fraction could be higher (*e.g.* if there was preferential surveying of customers of shops that got contaminated lots), but that is unlikely. To approximate this uncertainty, the range 0.066 to 0.25 was treated (for purposes of estimating a standard deviation) as though uniformly likely on a logarithmic scale, giving a median estimate of $z = (0.25 \times 0.066)^{0.5} = 0.13$ with uncertainty $g_z = \ln(\text{GSD}) = \ln(0.25/0.066)/\sqrt{12} = 0.37$.

The effective number of people at risk in the survey (*i.e.* those consuming contaminated ice cream, as opposed to consuming any Schwan ice cream) was then $z \times 452$. The observed probability of illness is $p_1 = r/n$ ($r = 30$, $n = 452$). This was treated as binomial to estimate its standard deviation (s_1), hence its coefficient of variation (c_1) and logarithm of GSD (g_1) using:

$$\begin{aligned} s_1 &= \sqrt{\frac{(r/n)(1-r/n)}{n}} \\ c_1 &= s_1 / r/n \\ g_1 &= \sqrt{\ln(1+c_1^2)} \end{aligned} \tag{A.2}$$

The observed ratio r/n was modified to $p_2 = r/(zn)$ to estimate the probability of illness in those eating contaminated ice cream, where z and r/n are treated as lognormal to estimate the median (p_2), logarithm of GSD (g_2), and coefficient of variation (c_2) of their ratio using:

$$\begin{aligned} p_2 &= p_1 / z \\ g_2 &= \sqrt{g_z^2 + g_1^2} \\ c_2 &= \sqrt{\exp(g_2^2) - 1} \end{aligned} \tag{A.3}$$

The resulting estimate of probability of illness p_2 and its coefficient of variation (c_2) was converted back to an effective (non-integer) r_2/n_2 pair of values that have identical ratio and coefficient of variation (as estimated by the formulae for a binomial sample) using:⁴⁹

$$n_2 = \frac{1-p_2}{p_2 c_2^2} \quad (\text{A.4})$$

$$r_2 = p_2 n_2$$

Hennessey *et al.* (1996) state that all positive samples for date code 238 were produced before 12:00 am, and that all implicated ice cream from date codes 237 and 238 came from a commingled premix load. They cite Vought and Tatini (1998) for this information, but that publication is not definitive, although it does state that some production with date code 237 came from a different premix. It is here assumed that all the contaminated ice cream from date codes 237 and 238 came from a single contaminated and commingled mixture, so had a similar concentration of *S. enteritidis*. This assumption is also consistent with the observation (Vought and Tatini, 1998) that only 2/15 samples from date code 237, and 6/16 from date code 238, were positive for contamination; the negative samples would have come from uncontaminated premix loads (the tested sample sizes of 100g to 300g were sufficient that detection was virtually certain at the CFU concentration, 0.06 CFU/g, estimated here). To estimate the concentration in contaminated ice cream at the time of the survey, the eight MPN results cited in Vought and Tatini (1998, Table 3)⁵⁰ for date codes 237 and 238 were combined as though a single large MPN analysis. It was demonstrated (Vought and Tatini, 1998) that there was unlikely to have been any change in *Salmonella* concentrations over the time the ice cream was stored before analysis.

Hennessey *et al.* (1996) indicate that a median of 3.2 persons consumed 1 gal of ice cream in their survey (the range given is 0 to 13 persons consuming a gallon; it is unstated how 0 people can consume a gallon of ice cream; also unstated was the interpolation scheme used to calculate a non-integer median). This value was used as the median estimate for ice cream consumption with no additional uncertainty (the sample size was sufficiently large that the uncertainty is small compared with the uncertainty in concentration). Ice cream was assumed to weigh 5.25 lb per gallon (21 CFR 135 requires ice cream to weigh at least 4.5 lb/gal, and sherbet to weight at least 6 lb/gal; the average of these two values was used). The resultant median estimate of dose was 43.4 CFU.

Outbreak 16. [FSIS #16](#) [Used here; [Serial #5](#)]

See [Outbreak 15](#). FSIS #16 corresponds to a different estimate of dose for the same outbreak.

⁴⁹ This procedure (of finding an r_2/n_2 pair with possibly non-integer values but with the correct ratio and coefficient of variation as calculated using formulae appropriate for binomial samples) maintains first and second moments correctly, ensures that probability estimates cannot be outside the range [0,1], and allows a common method of combining experiments as detailed in Sections 4.1 and 8.

⁵⁰ Vought and Tatini (1998) provide more extensive results and the paper is later than Hennessey *et al.* (1996).

Outbreak 17. [FSIS #17](#) & [FSIS #18](#); [FAO/WHO #11](#) [Used here; [Serial #6](#)]

Reference: Levy, *et al.*, 1996; FSIS (1998).

In an outbreak among persons who ate brunch at a hotel, a survey of a subset of those eating indicated 100% incidence (39/39) in those eating the suspect material (hollandaise sauce) vs. 2/25 who did not eat it. All those eating the sauce are here assumed to be adults. Among those who became ill, 20/56 were hospitalized.

The only available dose estimate is from a personal communication in FSIS (1998):

“An outbreak of salmonellosis due to SE was investigated, and the implicated food was hollandaise sauce. An informal measurement of one sample of the sauce revealed 10^3 SE bacteria per gram. However, the culture was not performed to extinction, and the specimen had been refrigerated for 72 hours after being taken home in a “doggie bag”. If two tablespoons of sauce were used, then the exposure would be about 10^4 SE per person. In this outbreak all of the 39 exposed individuals became ill and 20 were hospitalized (Levy *et al.*, 1996; bacterial count data provided by Dr. M. Moody of the DC Commission on Public Health, personal communication).”

The methodology for concentration estimation is unstated, so the uncertainty ($\ln(\text{GSD})$) in CFU/g is assumed equal to that of typical measurement methods (0.5, see Section 4.4). The Hollandaise sauce was made from the yolk of 36 extra-large eggs. Using standard estimates of egg mass (27 oz/doz. for extra large eggs), yolk mass fraction (0.3 of total egg mass), the Hollandaise sauce recipe from CSFII (USDA, 2000), and an estimated number of consumers (53, based on total 56 ill patrons and, among those surveyed, the fraction 39/41 ill who had eaten the brunch), the average serving size would have been about 48 g or less. The CSFII (USDA, 2000) lists 21 entries for Hollandaise sauce, with quantities ranging from 0.67 g to 128.5 g per serving, with an unweighted distribution that is almost lognormal (median 22.6 g, $\ln(\text{GSD})$ 1.21). In this case there is no information on whether the Hollandaise sauce was served by the hotel or by the individual. If a single person in the hotel served it, as assumed here, the uncertainty in average quantity served is similar to the between-serving variation seen in the CSFII (a single person would, approximately, serve equal sized servings to all those eating the sauce; but those servings could each have been anywhere on the distribution for individual serving size). The estimated serving quantity is thus taken to be 22.6 g with an $\ln(\text{GSD})$ of 1.21, leading to a median estimate of dose of 22,600 CFU. The average estimate of serving quantity from this distribution is 47.3 g, close to the estimated average serving size based on the Hollandaise sauce preparation information.

Outbreak 18. [FSIS #18](#) [Used here; [Serial #6](#)]

See [Outbreak 17](#). FSIS #18 corresponds to a different estimate of dose for the same outbreak.

A.1.2. Entries in Table 4**Outbreak 19. [FAO/WHO #1](#) [Used here; [Serial #1](#)]**

See [Outbreak 1](#).

Outbreak 20. [FAO/WHO #2](#) [Used here; [Serial #2](#)]

See [Outbreak 5](#).

Outbreak 21. [FAO/WHO #3](#) [Not used here]

See [Outbreak 10](#).

Outbreak 22. [FAO/WHO #4](#) [Used here; [Serial #4](#)]

See [Outbreak 13](#).

Outbreak 23. [FAO/WHO #5](#) [Used here; [Serial #3](#)]

See [Outbreak 9](#).

Outbreak 24. [FAO/WHO #6](#) [Not used here]

See [Outbreak 3](#).

Outbreak 25. [FAO/WHO #7](#) [Not used here]

References: Fontaine, *et al.*, 1978; Fazil, 1996.

This outbreak is a subset of that evaluated in FSIS #3 & FSIS #7 ([Outbreak 3](#)), with extra information obtained by Fazil on cases occurring at a Naval Station in Florida. The description in FAO/WHO (2002b) is incorrect in assuming that the implicated hamburger was used fresh. The measured concentrations were 6 to 23 MPN per gram; but these were after a freeze-thaw cycle. The hamburger used in this Florida outbreak must have been frozen before use (the outbreak was between Sept. 24 and Oct. 11, versus a probable production date for the hamburgers in late August — Table 3 of Fontaine *et al.* 1978 shows isolations from lots produced 8/21 and 8/?). FAO/WHO argues that freezing would reduce the count by 2 logs, but that partial cooking would also reduce the count by 2 logs, so that the measured count would correspond to the concentration in hamburgers as eaten. This does not apply for the Florida outbreak, where the cycle would be freeze, thaw, partial cook. Almost anything could happen to the count (decrease or increase) in this cycle during the thaw and partial cook. In addition, despite the additional information gleaned by Fazil, neither the fraction of contaminated hamburgers nor the number of hamburgers eaten is known, so the attack rate is entirely speculative.

Outbreak 26. [FAO/WHO #8](#) [Used here; [Serial #60](#)]

Reference: Narain and Lofgren, 1989

Narain and Lofgren (1989) describe an outbreak of salmonellosis due to *S. newport* contamination of meat sandwiches at a barbecue restaurant. A total of 120 cases were identified, with an age range of 6 to 52 years, median 33 years. In view of lack of further information on age, this group was treated as non-sensitive. Of the 120 cases, 105 were primary cases exposed during a four day period when the restaurant served barbecue beef, pork, chicken, and ham sandwiches, hot dogs, and cheeseburgers, together with coleslaw, chili, and soft drinks. The owner estimated a total of about 200 people served at the restaurant during that period. Narain and Lofgren (1989) provide a table of six foods eaten by 21 of the cases (those with positive stool cultures) and 25 friends and relatives of these cases who were similarly exposed to the sandwiches but did not become ill. A single measurement of a pork sandwich that had been stored refrigerated by one case gave a bacterial count of 44×10^6 *S. newport* per gram (the

methodology for the measurement was not specified, so a $\ln(\text{GSD})$ of 0.5 is assumed, see Section 4.4). Precooked ham slices from the restaurant tested positive for *S. newport*, although other materials tested negative.

FAO/WHO (2002b) omitted this outbreak from analysis, since the population exposed to the individual types of sandwiches is not given in the original report, only an estimate of the total number of people served at the restaurant. However, the table given by Narain and Lofgren (1989) provides sufficient information to estimate the fraction of those served who ate each type of sandwich, as well as the probability to become ill after each type of sandwich.

To analyze this outbreak, it is assumed that the 21 cases and 25 non-ill whose eating pattern is given are random samples from the ill and the non-ill exposed respectively, and that all the restaurant patrons ate at least one sandwich. Narain and Lofgren claim that each person documented in their table of exposures had one exposure to the restaurant and ate a single type of sandwich, although the tabular values imply that two ill persons each ate two sandwiches (or one ate three). This minor discrepancy is ignored in what follows. Suppose that the probability for a restaurant patron to eat a particular type of sandwich (labeled by $i = 1 \dots 6$, for the six types served) is p_i , with

$$\sum_{i=1}^6 p_i = 1 \quad (\text{A.5})$$

and the probability to be ill after eating sandwich type i is q_i . Then the probability for a randomly selected individual who ate a sandwich at the restaurant during the relevant period to be ill is

$$p_{\text{ill}} = \sum_{i=1}^6 p_i q_i \quad (\text{A.6})$$

and the fraction of those ill and not ill after eating a sandwich of type i are respectively

$$P_i = \frac{p_i q_i}{\sum_{j=1}^6 p_j q_j} \quad \text{and} \quad Q_i = \frac{p_i (1 - q_i)}{\sum_{j=1}^6 p_j (1 - q_j)} \quad (\text{A.7})$$

The observed numbers ill (n_i) or not ill (m_i) by food type, out of total $N = 21$ and $M = 25$ respectively ill and not ill, are then a random sample from multinomial distributions with these probabilities. The total number ill in the whole outbreak (105 primary cases = n_{tot}) is a binomial sample with probability p_{ill} (Equation (A.6)) from the total number (N_{tot}) of restaurant patrons. The latter, however, is only known approximately (about 200). It is here assumed that the uncertainty in the actual number of patrons (N_{tot}) can be approximated by a lognormal distribution with $\ln(\text{GSD})$ of $\sigma_{\text{tot}} = 0.15$ (about 15% uncertainty) and median $N_{\text{med}} = 200$. The overall loglikelihood for the observations is then the sum of components from the multinomial distribution, from the probability to be ill, and from the uncertainty in total number of patrons:

$$J_N \left(\{p_i, q_i; i = 1 \dots 6\}, N_{\text{tot}} \right) = \sum_{i=1}^6 (n_i \ln P_i + m_i \ln Q_i) + \ln \left(\binom{N_{\text{tot}}}{n_{\text{tot}}} p_{\text{ill}}^{n_{\text{tot}}} (1 - p_{\text{ill}})^{N_{\text{tot}} - n_{\text{tot}}} \right) - \frac{1}{2} \left(\frac{\ln \left(N_{\text{tot}} / N_{\text{med}} \right)}{\sigma_{\text{tot}}} \right)^2 \quad (\text{A.8})$$

Using this likelihood it is found that there was no significant difference in the probabilities to be ill after eating beef, pork, or ham and cheese sandwiches ($p = 0.35$, likelihood ratio test), so this equality was enforced. The probabilities for illness after eating hot dog, cheeseburger, or ham sandwich differed significantly, but are not of interest because no samples of such sandwiches were analyzed for *Salmonella*.

To approximate the uncertainty in the probability for becoming ill after eating a pork sandwich, the profile likelihood for this probability was computed and approximated (see sheet Narain in “Epi Attack Data.xls”) using

$$\tilde{J}_N(p) = J_B(\ln(-\ln(1-p)) + z_e, \sigma_e, \{r_e, n_e, 1\}) + \varepsilon \quad (\text{A.9})$$

where $\tilde{J}_N$ is the profile loglikelihood for the probability p of illness after eating a beef, pork, or ham and cheese sandwich, J_B is the loglikelihood given by Equation (25), and the parameters z_e , σ_e , r_e , and n_e were chosen to minimize the sum of squares of ε over 101 equally spaced values of p round the ML estimate chosen so that the change in the loglikelihood ranged up to 5. The values obtained were $z_e = 1.407$, $\sigma_e = 0$, $r_e = 29.48$, and $n_e = 30.53$, resulting in a root mean square ε of 0.14.

The single available measurement of *Salmonella* contamination was in the pork component of a sandwich refrigerated by a case. It is assumed here that the refrigeration was adequate, so the measurement was representative of the condition of the sandwich at the time of consumption. The quantity of meat in the sandwiches was not specified by Narain and Lofgren (1989). A median estimate for the quantity of meat in a sandwich was therefore obtained from the CSFII (USDA, 2000) as the unweighted logarithmic mean of the quantities of meat in barbecue beef, pork, ham, or ham and cheese sandwiches (or those meats together with other ingredients except “with gravy,” in ham salad, or in ham and egg sandwiches) eaten by those aged 6 to 52 in the survey, and the uncertainty ($\ln(\text{GSD})$) estimated as the standard deviation of the logarithm of these meat quantities (see sheet Sandwiches in “Epi Attack Data.xls”). The full variability of the observed unweighted distribution was used because the sandwiches were prepared by a single restaurant that could have standardized their sandwiches at any size.

Outbreak 27. [FAO/WHO #9](#) [Not used here]

Reference: Craven, *et al.*, 1975.

This reference reports on the U.S. component of the outbreak in Canada reported in FSIS # 4 (d'Aoust, *et al.*, 1975). As for FSIS #4 (See [Outbreak 4](#) of this discussion), the attack rate is almost entirely speculative.

Outbreak 28. [FAO/WHO #10](#) [Not used here]

See [Outbreak 4](#) of this discussion.

Outbreak 29. [FAO/WHO #11](#) [Used here; [Serial #6](#)]

See [Outbreak 17](#) of this discussion.

Outbreak 30. [FAO/WHO #12](#) [Used here; [Serial #5](#)]

See [Outbreak 15](#) of this discussion.

Outbreak 31. [FAO/WHO #13](#) [Used here; [Serial #7](#)]

Reference: Taylor, *et al.*, 1984

A report of an outbreak in which all eight persons eating some home-made ice cream were affected, and all eight were hospitalized. Four of the eight were children aged 2, 6, 8, and 9 years, the other four having ages of 13, 22, 30, and 35. The four youngest were treated as sensitive due to age (Section 6). Individual estimates of the amount of ice cream eaten are given, clearly rounded to the nearest 100 or 250 ml., so that the uncertainty in amount corresponds to a $\ln(\text{GSD})$ of about 0.29 (estimated as half the average logarithm of the ratios between successive quantity estimates of 100, 250, 500, 750, and 1000 ml). *S. typhimurium* concentration in the ice cream was obtained by serial dilution and plating, but no further detail is given, so the default uncertainty estimate for plating is applied (Section 4.4). Median dose estimates range from 10^8 to 10^9 CFU for the eight individuals. The relative dose estimates for the eight persons were treated as negligible in the main analysis, despite the larger estimated uncertainty in mass eaten ($\ln(\text{GSD})$ of 0.28) compared with that for the concentration ($\ln(\text{GSD})$ of 0.2).

Outbreak 32. [FAO/WHO #14](#) [Not used here]

References: D'Aoust 1985; D'Aoust, *et al.*, 1985

A large outbreak (1,500 cases) attributed to *S. Typhimurium* in cheddar cheese. D'Aoust (1985) describes 6 individual cases where the concentration in the implicated cheese was measured, and indicates that these cases were due to consumption of less than 10 CFU. However, the overall incidence of illness could not be evaluated. [Note: This summary is based on D'Aoust, 1985 and the FAO/WHO, 2002b summary. D'Aoust *et al.*, 1985 was not available for examination. FAO/WHO, 2002b, omitted this outbreak from their analysis because of the lack of information on incidence, since the total population exposed was not known; hence it does not appear in Table 4. This discussion is included here because if D'Aoust *et al.*, 1985, indicates total production of the implicated cheese, there is a possibility of evaluating the incidence.]

Outbreak 33. [FAO/WHO #15](#) [Used here; [Serial #40](#)]

See [Outbreak 6](#).

Outbreak 34. [FAO/WHO #16](#) Used here; [Serial #41](#)]

See [Outbreak 2](#).

Outbreak 35. [FAO/WHO #17](#) [Used here; [Serial #8](#)]

Reference: Greenwood and Hooper, 1983.

The outbreak affected three people who ate chocolate bars: one boy (age 10, so treated here as sensitive due to age) ate 2 bars/day on two consecutive days and became ill, while his mother and brother (age 13) each ate 2 bars on a single day, but were symptom free (although the brother was infected). The paper evaluates 48 chocolate bars (16 g each) from 8 packets (6 bars per packet) from a box supplied by the same retailer who sold the chocolate bars to the cases. Each bar was weighed, homogenized in peptone water, and subjected to MPN analysis; sub-

samples were used in a 3-tube, 3-dilution MPN analysis, and the remaining homogenate also incubated and tested. Re-analysis of the results shows that if each packet is evaluated separately the concentrations are consistent with being equal for all 6 bars ($p > 0.05$, likelihood ratio test), although if the same analysis is applied globally to all 8 packets this result is somewhat unlikely ($p = 0.029$, likelihood ratio test). If a lognormal distribution of concentrations is assumed between bars within each packet, the standard deviation estimates are all consistent with zero. These results were considered sufficient to assume that the within-packet concentrations were equal, with packet concentrations distributed in a manner consistent with a lognormal variation between packets ($p = 0.6$, Shapiro-Wilk test⁵¹). The parameters for this concentration model were evaluated by accumulating all results to the packet level, and these are used to define the uncertainty distribution for concentration in any given packet of chocolate bars. It is assumed that all bars eaten by the family came from the same packet, so doses are proportional to the number of bars eaten. (FAO/WHO (2002b) misses that 3 persons were exposed, but only one became ill. As a consequence, this outbreak was not incorporated in the analysis by FAO/WHO, and does not appear in Table 4.

Outbreak 36. [FAO/WHO #18 through #33](#) [Not used here]

Reference: Ministry of Health and Welfare [Japan], 1999.

These outbreak reports are superseded by those reported by Kasuga *et al.*, 2004 (some of the reports in FAO/WHO, 2002a,b, are not reported in Kasuga *et al.*, 2004, and vice versa). See [Outbreak 37](#).

A.1.3. Additional outbreak entries

Outbreak 37. [Additional 1](#) [Used here; [Serial #9 through #39](#)].

Reference: Kasuga, *et al.*, 2004.

Kasuga *et al.* (2004) report on 31 outbreaks in Japan (#6 is reported as two separate outbreaks, 6a and 6b in children and their teachers, but is analyzed here as a single outbreak with two exposed groups, the children being sensitive; both are given the [Serial #14](#) in this work). The number given as “sick” was used as the incidence of illness, and the number given as “exposed” as the exposed population. Numbers with various diagnoses are given (diarrhea, stomach ache, fever, vomiting, nausea, chill, fatigue, headache, and other), but the precise definition used for “sick” is unclear (in one case, #26, here [Serial #34](#), the number with the “other” diagnosis exceeds both the number “sick” and the number “exposed”).

Outbreaks 1 through 4 were associated with school lunches in schools with pupils in the age range 6 to 12 years. Outbreaks 5, 6b, and 7 were in daycare centers, with children ages 0 to 6 years (outbreak 6a was in teachers eating the same food as the children in outbreak 6b). The children in these outbreaks were assumed sensitive due to age, and placed in the appropriate age categories (Section 6). Outbreaks 8 through 12 were in hospitals, so the exposed populations

⁵¹ This test is used heuristically here — it underestimates the probability for deviations from lognormal because the values tested are themselves uncertain, contrary to the assumptions of the test.

were considered sensitive due to hospitalization. All other reported outbreaks were in general populations, considered non-sensitive.

For food quantities, the estimate required for the analysis is the mean value in the exposed population. The value given in the tables of Kasuga *et al.* (2004) was used as the central estimate. For school lunches, daycare centers, and hospitals, the value was assumed exact, except where it was stated to be estimated from other situations, in which case an uncertainty of a factor of 1.1 to 1.5 was incorporated.⁵² For restaurants, it was assumed that the uncertainty in mean food quantity was a factor of 2 about the given value. For confectioneries, the uncertainty in mean food quantity was assumed to be a factor of 1.2 about the given value.

Uncertainties in concentration estimates were obtained using approaches described in Sections 4.2 through 4.6, using the descriptions given in Kasuga, *et al.* (2004) to infer the methods used for concentration measurement. Where MPN methods were reported (outbreaks 2, 4, 7, 10, 11, 14, 27, and 31 of Kasuga *et al.*, 2004) the ln(GSD) estimate was based on the ratio of upper and lower 95th percentiles from an MPN likelihood profile analysis (see Section 4.2). It was assumed that three sequential decade dilutions were used in the MPN method, and that where not stated otherwise 5-tube MPN was used (for outbreaks 4 and 7 this is confirmed by Abe *et al.*, 2004, Table 1, footnotes *b* and *c*; see [Outbreak 39](#)). For outbreaks 1 and 8 of Kasuga *et al.* (2004), *Salmonella* colonies were detected in single 25 g incubated samples, but not when plating solutions with 10 mg of sample. These results were treated as an MPN analysis with 4 dilutions, and ln(GSD) estimated from ± 1 SD likelihood profile estimates (see Section 4.5). Table 3 of Kasuga *et al.* (2004) is self-contradictory for outbreak 11, stating that plating was used but giving details of MPN methodology. The latter is assumed here, using the MPN details given — the concentration estimated in this way corresponds with that reported. For outbreak 27, details are provided of MPN analyses for two components (sponge and mousse) of a 3-layer cake. As in Kasuga *et al.* (2004), equal quantities of the two components were assumed to be eaten, so the concentration estimates and their uncertainties were averaged by estimating means and standard deviations from the medians and ln(GSD) estimates obtained from the MPN analyses (assuming lognormal distributions), averaging the results, then converting back to a logarithmic scale (again assuming a lognormal distribution).

Where reported CFU were based on plate counts (outbreaks 3, 5, 6, 9, 12, 13, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, and 30), an effort was made to reconstruct the number of colonies counted, and the ln(GSD) estimated as described in Section 4.3. Kasuga *et al.* (2004) report the maximum amount of the original food material analyzed, but in many cases such a quantity implies very high plate counts; in such cases, it is assumed serial decade dilutions were made and that the concentration estimate came from the first dilution with a plate count of 400 or fewer.

For outbreak 3, the reported plating details are inconsistent with the reported concentration (since the initial stated dilution and concentration estimate gives a plate count of 0.4); it was assumed that the quantity plated was 10 times larger than stated (1 ml in place of 0.1 ml) with a plate count of 4. Outbreak 5 is used as reported for bacterial concentration and food intake — the “Dose ingested” given by Kasuga *et al.* 2004, 44,000 CFU/person, differs slightly from the product of

⁵² The uncertainties were not consistently estimated; however, results are extremely insensitive to the values used.

concentration and food intake, 40,000 CFU/person. For outbreak 6 no plated quantity is given, so a plate quantity of 1 ml is assumed, giving a plate count of 27 colonies. Outbreak 16 is reported as an MPN count, but with no footnote confirming MPN or direct plate count analysis. The details provided correspond to a plate count analysis, and that has been assumed here. For outbreak 17 the concentration estimate indicated a plate count, but additional information was presented corresponding to a 1-tube MPN analysis; both sets of data have been used (see Section 4.6 for the methodology). Outbreak 18, 21, 22, and 26 have no reported “maximum amount analyzed”; it has been assumed to be 0.01 g in every case (this corresponds to the value given in the majority of the other reports). Outbreak 30 has a reported concentration “<3,000” with no enumeration method given, but the description of laboratory methods indicates plating. The specified maximum amount of food plated is 0.01 g, which would give a count of 30 CFU at 3,000 CFU/g. It appears likely that the reported concentration was an upper confidence bound based on small numbers of observed CFU. The reported concentration has here been assumed to correspond to an observation of 22 CFU from a plated amount of 0.01 g, which gives a Poisson-based upper 95th percentile confidence limit of 3,064 CFU/g.

For all the reported outbreaks, it has been assumed that the storage conditions of the food prior to measurement of *Salmonella* concentrations were such that no growth or death of *Salmonella* occurred. Most reports indicated storage either frozen or refrigerated, with many facilities following a non-mandatory directive suggesting storage of 50 gram portions of each raw food material for more than 2 weeks. Storage conditions were unknown for some outbreaks ([Serial #10](#), [#25](#), [#29](#), and [#30](#)), and [Serial #39](#) indicated storage for 5 days at ambient temperature.

Outbreak 38. [Additional 2](#). [Used here, [Serial #51](#)]

Reference: Matsui *et al.*, 2004. Personal communication, 2005.

Matsui *et al.* (2004) report an outbreak of *S. enteritidis* due to contaminated desert buns served in schools in Japan. The attack rates were 92/18571 in elementary schoolchildren (age range 6–12), 3/4265 in junior high schoolchildren, and 0 in teachers among those served the implicated buns (the number of teachers so served is not stated). Elementary schoolchildren were assumed to be a sensitive population due to age (see Section 6). One of 13 buns tested was positive for *S. enteritidis* using 25 g samples of the 40 g buns (weights provided in personal communication, T. Matsui, Sept. 2005), but MPN analysis (5-tube, starting with a dilution containing 1 g, according to Abe *et al.*, 2004) showed non-detect. Treating the 25 g sample testing together with the standard MPN analysis as a 4-dilution MPN analysis gives the median estimate of dose per bun as 0.12 CFU, but with a ln(GSD) estimate of 1.08, so that many buns would not have been contaminated. However, with the linear-exponential dose-response adopted here the calculations remain accurate even for (average) doses of less than 1 CFU provided the incidence is small, as is true here.

Outbreak 39. [Additional 5](#). [Used here, [Serial #55 through #57](#)]Reference: Abe *et al.*, 2004.

Abe *et al.*, (2004) describe outbreaks due to *Salmonella enteritidis* in Japan that overlap the outbreaks reported by Kasuga *et al.*, (2004; [Outbreak 37](#)) and Matsui *et al.* (2004; [Outbreak 38](#)), see Table 21.

Table 21 Correspondence between outbreaks reported by Abe *et al.* 2004, Matsui *et al.* 2004, and Kasuga *et al.* 2004.

Abe <i>et al.</i> 2004 Outbreak number	Reference	Kasuga <i>et al.</i> 2004 Outbreak number
1	Matsui <i>et al.</i> 2004	–
2	Kasuga <i>et al.</i> 2004	7
3	Kasuga <i>et al.</i> 2004	4
4	Kasuga <i>et al.</i> 2004	3
5	Not reported elsewhere	
6	Not reported elsewhere	
7	Kasuga <i>et al.</i> 2004	18
8	Not reported elsewhere	
9	Kasuga <i>et al.</i> 2004	23

The otherwise unreported outbreaks 5, 6, and 8 are used here. Abe *et al.* (2004) summarize details of food item, number of illnesses, attack rate, concentration of *Salmonella* in the food item, quantity of food item ingested, and method used to estimate *Salmonella* concentration. These are used in exactly the same way as for the outbreaks reported by Kasuga *et al.* (2004), see [Outbreak 37](#) for further details. Outbreaks 1, 3, 4, 5 and 8 are stated as occurring in facilities of elementary and junior high schools, although Kasuga *et al.* (2004) described outbreaks 3 and 4 as occurring only in pupils of elementary school age range (6–12 years). Average ages of patients for all outbreaks examined in nursery schools (2 outbreaks) was 4.6 years, and for other schools 10.6 years. Outbreak 6 was in a nursery school, so these children are also considered sensitive due to age in the <5 year class. The children exposed in outbreaks 5 and 8 are assumed to be in the elementary school age range (6–12 years) and sensitive due to age (see Section 6).

For outbreak 5 the reported concentration (10 CFU/g) corresponds to the precision given only with a 5-tube MPN result that is highly unlikely (a 5,0,5 pattern); the nearest high probability result (10.9 CFU/g with 5,3,1 pattern) has been assumed (the differences in results are negligible compared with the uncertainties involved). For outbreak 8 the reported MPN details are consistent with the result only if the minimum dilution for the MPN result was 100 times more than specified in the footnote specifying the method (such further dilution may be considered inherent in the MPN method that is stated to be used). In this case the nearest high-probability result (631 CFU/g with 5,1,2 pattern) has been used since there is no MPN pattern that rounds to the given result (6.0×10^2 CFU/g) to the given precision.

Outbreak 40. [Additional 6](#) [Used here, [Serial #58](#)]

Reference: Urfur *et al.* 2000.

An outbreak in the Canton de Vaud, Switzerland was initially discovered from surveillance (mandatory reporting of *Salmonella* isolates). Subsequently, all hospitals and practitioners were queried and there was a “systematic search during the case interviews for persons who may have been exposed to the incriminated food.” It is assumed here that the 215 cases identified represent the total number of cases. Insufficient age information is provided to allow separation by age groups (29/156 patients interviewed were children less than 16, and the age range of patients was 1–80 years), so the population affected was considered non-sensitive (Section 6).

The implicated foods were individual meat pies or terrines produced by a single small food factory; a total of 2,000 pies and 1,500 terrines were distributed, 500 items recalled, and an unknown number destroyed by retailers and consumers. Quantitative culture of unsold food items collected during the peak of the epidemic indicated a concentration of 3×10^6 CFU/ml in homogenized pies, with each pie weighing 70 g. Contamination in the jelly of the pies was higher (3.2×10^6 CFU/ml) than in the meat (1.5×10^6 CFU/ml). The jelly of terrines contained 1.1×10^6 CFU/ml. The measurement method was not specified; it is assumed here that MPN methodology was used.

The initial case identified occurred on a Thursday, with a stop-production order being issued 9 days later on a Saturday. Local newspapers advised against consumption of the pies another 6 days later on Friday, and cases apparently ceased after another 3 days.⁵³ Incubation time in the 156 cases interviewed ranged from 6 to 72 hours. The meat pies were produced daily (by inference from the statement that the jelly was used daily), and terrines only on Mondays or Tuesdays, so the outbreak was apparently caused by two weeks of production, primarily the second week based on the epidemic curve. The jelly considered the main contributing factor was produced once per week and added to the leftover jelly from the previous week. It was melted daily at low temperature for re-use.

The description of jelly handling indicates that concentrations of *Salmonella* in it were likely to vary over the week as it was re-melted and cooled. Food products for quantitative cultivation were taken “during the peak of the epidemic” thus late in the second week, corresponding for pies to production at the end of that week, whereas the terrine jelly measurement would correspond to jelly produced on the Monday or Tuesday of that week. The average of the two jelly measurements is taken to be the best estimator of the average over the epidemic, and this is scaled to a whole pie or terrine by the ratio of measured concentrations in jelly and whole pie. The uncertainty in concentration is taken to be the generic MPN uncertainty estimate (Section 4.4). The mass of pie or terrine is taken to be 70 g with no uncertainty (this is the mass cited in the paper for a pie; terrines are assumed identical), and cases are assumed to have consumed a single pie or terrine.

⁵³ Figure 1 of Urfur *et al.* (2000) showing the date of onset for cases contains 137 cases in total, although the paper indicates interview of 156 cases and a total identified number of 215 cases.

A.2. Human feeding studies

Human 1. [Additional 3](#) [Used here, [Serial #52](#)]

Reference: Hormaeche, *et al.*, 1936 (in Spanish).

See Section 3.1 for a summary description of the experiment. Five adult individuals were administered 2×10^9 cells (two individuals) or 4×10^9 cells (three individuals) of *S. typhimurium*. The method of quantification is not stated, so the default value of 0.5 (Section 4.4) for $\ln(\text{GSD})$ is used. The two doses were administered simultaneously from the same stock, so it is assumed that they were obtained by dilutions and were in an exact ratio of 1:2. Over the next few days, one of those administered the lowest dose experienced a mild elevation in temperature (not treated here as illness), and the other four experienced more severe symptoms (treated here as illness).

Human 2. [M&E](#) [Used here, [Serial #42 through #50](#)]

References: McCullough and Eisele, 1951a,c.

See Section 3.1 for a summary description of the experiment. This analysis uses illnesses, not infection, as the end point. McCullough and Eisele (1951 a,c) often dosed the same individuals more than once, introducing the possibility that earlier exposures to the same or different *Salmonella* isolate might affect subsequent probability of illness. However, using the linear dose-response assumption for all subjects versus just naive (never previously dosed) subjects gives adequate fits in all cases and no significant difference in parameter estimates, so this analysis uses all subjects. As previously mentioned, this analysis uses only illness diagnoses, not the diagnoses of infection. In the experiments evaluated here, illness diagnosis required subjective complaints, recovery of the organism from stool, and one or more of: diarrhea or vomiting, fever, rise in specific agglutination titer, or other observable signs.

Calculated doses for each feeding were determined turbidimetrically using a photoelectric colorimeter, but final determination of the concentrations of each dose were determined by serial dilution and plating in duplicate (McCullough and Eisele, 1951a). We use the generic plating estimate (Section 4.4) of 0.2 for $\ln(\text{GSD})$ for the dose at each individual dose level. The likelihood given by Equation (25) applies in the case where the relative doses in a multi-dose experiment are known exactly (although the absolute values of the doses may be uncertain). It was found that the effect on the loglikelihood of incorporating independent uncertainties at each dose could be accurately approximated by using Equation (25) with a modified uncertainty estimate, a slight shift in the parameter estimate, and a slight shift in the value of the loglikelihood (this last is unimportant in the analysis, since only changes in loglikelihood are relevant). That is, using the notation of Equation (25), for the N dose groups

$$\prod_{k=1}^N J_B(x, \sigma_k, \{r_j, n_j, d_j; j = k\}) = J_B(x + \gamma, \tau, \{r_j, n_j, d_j; j = 1 \dots N\}) + \eta + \varepsilon \quad (\text{A.10})$$

where σ_k is the $\ln(\text{GSD})$ for dose group k , and γ , τ , and η are estimated empirically by demanding that the error ε is small and has zero mean over some range of values for x . In practice the parameters γ , τ , and η were estimated by minimizing the sum of squares of ε over 101 equally spaced values of x around the maximum likelihood value, with the spacing chosen so

that the reduction of the likelihood at the extreme values was about 10; then ε was ignored in the main analyses. Table 22 shows the estimated values obtained for Serial #42 through #50, together with Serial #59 ([Human 4](#)), with root mean square (RMS) ε small enough to be unimportant (by comparison, changes of order 0.5 in the loglikelihood may be considered important). The calculations leading to these values are in sheet “MultiDose” in the workbook “Epi Attack Data.xls,” together with plots showing the accuracy of the approximation.

For the alternative dose-response model of Section 8.7, exactly the same analysis was applied (all that is required is replacement of the dose-response model in Equation (27) with the alternative dose-response model in Equation (36) with $S_i = 1$, and re-interpretation of the m parameter of Equation (25) as the logarithm of the median d_{50} rather than of the median virulence. There is one complication, and that is the evaluation of the slope parameter β . This was obtained by simultaneously maximizing the sum of the exact loglikelihoods (left side of Equation (A.10) using Equation (36) with $S_i = 1$) for the nine McCullough and Eisele (1951a,c) experiments to obtain maximum likelihood estimates for the nine d_{50} parameters and the single β parameter. The resulting estimated slope parameter $\beta = 0.967194$ was left fixed for the rest of the analysis. Once again, a set of estimates γ_P , τ_P , η_P , and ε_P were obtained empirically to use in the rest of the analysis (Table 23). Again, the small size of the RMS ε_P indicate an adequate approximation. The calculations leading to the values shown in Table 23 are in sheet “MultiDoseP” in the workbook “Epi Attack Data.xls,” together with plots showing the accuracy of the approximation.

Table 22 Empirical estimates for γ , τ , and η , and root-mean-square value of ε attained, for the linear-exponential dose-response model.

Serial #	Identity	Shift in ln(Virulence) γ	ln(GSD) τ	Likelihood shift η	RMS ε
42	M&E A_I	0.031	0.099	-0.020	0.003
43	M&E A_II	0.033	0.112	0.013	0.004
44	M&E A_III	0.010	0.141	-0.036	0.011
45	M&E M_I	0.029	0.125	-0.009	0.006
46	M&E M_II	0.027	0.109	-0.014	0.007
47	M&E M_III	0.016	0.129	-0.024	0.001
48	M&E N	0.014	0.143	-0.005	0.001
49	M&E B	0.003	0.132	-0.053	0.008
50	M&E D	0.027	0.127	0.024	0.003
59	Addl. 7	0.044	0.154	0.000	0.004

Table 23 Empirical estimates for γ_P , τ_P , and η_P , and root-mean-square value of ε_P attained, for the probit-log-dose dose-response model.

Serial #	Identity	Shift in $\ln(d_{50})$ γ_P	$\ln(\text{GSD})$ τ_P	Likelihood shift η_P	RMS ε_P
42	M&E A_I	0.031	0.099	-0.020	0.003
43	M&E A_II	0.033	0.112	0.013	0.004
44	M&E A_III	0.010	0.141	-0.036	0.011
45	M&E M_I	0.029	0.125	-0.009	0.006
46	M&E M_II	0.027	0.109	-0.014	0.007
47	M&E M_III	0.016	0.129	-0.024	0.001
48	M&E N	0.014	0.143	-0.005	0.001
49	M&E B	0.003	0.132	-0.053	0.008
50	M&E D	0.027	0.127	0.024	0.003
59	Addl. 7	0.044	0.154	0.000	0.004

Human 3. [Additional 4](#). [Used here; [Serial #54](#)]

Reference: Varela and Olarte, 1942 (in Spanish).

Varela and Olarte (1942) fed one 37 year old male 20 c.c. of an emulsion containing 5×10^8 CFU *S. anatum* per c.c., causing stomach pains and diarrhea on the day of administration and altered fecal consistency for some days thereafter. There is no indication of how the concentration of *Salmonella* in the emulsion was estimated. The generic estimate (Section 4.4) of 0.5 for $\ln(\text{GSD})$ for dose is used here.

Human 4. [Additional 7](#). [Used here, [Serial #59](#)].

Reference: Greenberg, 1964

Greenberg (1964) reports an experiment in which laboratory-hatched flies were contaminated with *S. typhimurium* by exposure to contaminated dog feces, then allowed to contaminate 10 batches, 100 ml each, of atole (a Mexican drink made with corn flour, sugar, boiled milk, and flavoring). After 10.5 hours incubation, each batch of atole was sampled for measurement and consumed by an adult volunteer, one batch to each of the ten volunteers. Eight of the ten atole specimens had measurable quantities of *S. typhimurium*, but none of the volunteers became ill (although *S. typhimurium* was recovered from the feces of six).

Measured doses of *S. typhimurium* in the eight contaminated atole batches ranged from 5×10^3 to 6.4×10^5 CFU, with measurement by plating of dilutions of the atole. Measured doses per 100 ml of atole are published in a form that is here assumed to indicate the actual plate counts and dilutions, which values are used to estimate uncertainties for the individual doses assuming

Poisson statistics. The two batches of atole with no measurable *S. typhimurium* are assumed to be uncontaminated, and are not included in the analysis. The loglikelihood for the combined set of data was approximated using the technique described in [Human 2](#), that is by using Equation (A.10) and neglecting the ε terms. The values obtained for the approximations are shown in Table 22 and Table 23.

Appendix B. Analyses in the workbook “Epi Attack Data.xls.”

All calculations discussed in the text are performed in the workbook “Epi Attack Data.xls.” The calculations make extensive use of Visual Basic for Application (VBA) functions written specifically for this application and included in the workbook, together with utility functions and tools encoded in the add-in function library EACC_funs (Crouch, 2006). This add-in function library should be installed according to the directions given in the manual prior to opening the workbook “Epi Attack Data.xls.”

This appendix gives a short summary of where to find the calculations described in the text, and in places indicates how they are performed. However it is not intended as a complete description of the calculations and numerical methods used, nor is it a complete prescription of how to make the workbook perform the calculations. Section B.1 describes the structure of the workbook, while the discussion after Section B.1, generally in the order of calculations in the text, summarizes what is done where and how it is done.

B.1. Structure of the workbook “Epi Attack Data.xls”

The workbook contains the following worksheets (in the order in which they appear in the workbook):

All_data	The main analysis of the workbook; fitting of the linear-exponential dose-response to all suitable data (see Section 8.3 through 8.6, B.9 and B.11). Also contains the sensitivity analyses (Section 8.8), and construction of Figure 7.
Distrib	Tables of values copied from worksheet All_data, sorted, and used to construct Figure 5 and Figure 6, and to test for lognormality.
Probit	The analysis of the alternative (probit-log-dose) dose-response curve (see Section 8.7 and compare with Sections 8.3 through 8.6, B.9 and B.11).
Distrib_p	Tables of values copied from worksheet Probit, sorted, and used to construct Figure 12 and to test for lognormality. Also contains an analog of Figure 6 that is not shown in the text.
DR-pict	Construction of the dose-response graphs shown as Figure 8, Figure 9, Figure 10, Figure 11 and Figure 14.
Attack-pict	Construction of Figure 3 and Figure 4 from the data in worksheet All_data.
FSIS_WHO	Summary tables of data used in FSIS (1998) and FAO/WHO (2002b), and Figure 17.
Analyses	Evaluation of uncertainties for various of the outbreaks included in All_data, except those reported by Kasuga <i>et al.</i> (2004) (see Section B.3).
Kasuga	An augmented table of results from Kasuga <i>et al.</i> (2004), together with evaluations of the dose uncertainties for each case (see Section B.4).
McCullough	Raw data from McCullough and Eisele (1951a,c), summary tables, and fits of dose-response models to those data.
Font_1980	Implementation of the model described in Outbreak 5 (Fontaine <i>et al.</i> , 1980) of Appendix A, together with the data of Goepfert <i>et al.</i> (1968).
Narain	Implementation of the analysis described in Outbreak 26 (Narain and Lofgren, 1989) in Appendix A.

MPN tables	Tables of MPN values used for identification of the pattern of results in MPN analyses, based on the reported concentrations.
Sandwiches	Analysis of CSFII (USDA, 2000) data on sandwiches described in Outbreak 26 (Narain and Lofgren, 1989) in Appendix A.
Hollandaise	Analysis of CSFII (USDA, 2000) data on hollandaise sauce described in Outbreak 17 in Appendix A.
Multidose	Implementation of the approximation used for multi-dose experiments as described in Human 2 in Appendix A, giving the results shown in Table 22 for the linear-exponential dose-response model.
MultidoseP	Implementation of the approximation used for multi-dose experiments as described in Human 2 in Appendix A, giving the results shown in Table 23 for the probit-log-dose dose-response model.
h_coeffs	Evaluation of the coefficients h_0, h_1, h_2, h_3 described in Appendix C.2.2 and given in Table 25.

In addition to the 18 worksheets described above, the workbook contains a Visual Basic for Applications module named Bin_Ran, the contents of which are visible in the Visual Basic Editor. This module contains the following 18 functions (in the order in which they appear):

maxloc	The location of the maximum of the integrand of the integral in Equation (C.7), <i>i.e.</i> the solution of Equation (C.23).
logintegrand	The logarithm of the integrand in Equation (C.7) (omitting multiplying constants).
binrandloglik	The logarithm of the likelihood given in Equation (C.7).
logUnnormBeta	The incomplete Beta integral given by Equation (C.30).
maxloc_2	The location of the maximum of the integrand of the integral in Equation (C.29), <i>i.e.</i> the solution of Equation (C.33).
logintegrand_2	The logarithm of the integrand in Equation (C.29).
binrandloglik_2	The logarithm of the likelihood given in Equation (C.29).
MPN_prob	The probability for a particular pattern of results in an MPN analysis.
PLDLNmax	The location of the maximum of the integrand of the integral in Equation (C.44), <i>i.e.</i> the solution of Equation (C.54).
PLDLNlogintegrand	The logarithm of the integrand in Equation (C.44) (omitting multiplying constants).
PLDLNloglik	The logarithm of the likelihood given in Equation (C.44).
Aux_func	The two solutions for z , given t , of $z - 1 - \ln(z) = t \quad (\text{B.1})$ This function is not currently used.
Subtr_exp	See Section 4.3. The solution for z , given t , of $e^z - z - 1 = t^2/2 \quad (\text{B.2})$ that analytically continues the solution $z \sim t$ for small t.
logphiterm	The logarithm of the expression within brackets in the last line of Equation (C.61).
sumlogs	Accurate calculation of $\ln(X + Y)$ given $\ln(X)$ and $\ln(Y)$.
PLDLNmax_2	The location of the maximum of the integrand in Equation (C.59) using the approximation of Equation (C.66).

PLDLNlogintegrand_2 The logarithm of the integrand in Equation (C.59) using the approximation of Equation (C.66) (omitting multiplying constants).

PLDLNloglik_2 The logarithm of the likelihood given in Equation (C.59), using the approximation of Equation (C.66).

B.2. Attack rate summary pictures — worksheet “Attack-pict”

Figure 3 and Figure 4 are constructed in this worksheet, using the data summarized in worksheet “All_data.”

B.3. Analyses of individual outbreaks — worksheet “Analyses”

The analyses of several outbreaks included in Table 5 and described in Appendix A are performed in this worksheet, in approximately the order of Appendix A. Where necessary, there are cross-references to other worksheets where more detailed modeling is performed. The results obtained for dose, dose uncertainty, numbers ill, and number exposed for each outbreak are used in workbook All_data (columns J, L, M, and N). The analyses included are of the outbreaks shown in Table 24, and can be located in the worksheet by locating the abbreviated reference given in Column A of “Analyses”(which contains only such references). Table 24 also shows other worksheets that are referenced by these analyses.

Table 24 Outbreaks evaluated in the worksheet “Analyses”

Appendix A ^a	Entry in Column A of worksheet “Analyses”	See also worksheet
Outbreak 1:	Boring et al. 1971; and Collaborative, 1971	
Outbreak 5:	Fontaine et al., 1980.	Font_1980
Outbreak 9:	Armstrong 1970	
Outbreak 13:	Angelotti, 1961	
Outbreak 15:	Hennessey et al. 1996; Vought & Tatini, 1998	
Outbreak 17:	Levy 1996	Hollandaise
Outbreak 31:	Taylor 1984	
Outbreak 35:	Greenwood 1983	
Outbreak 38:	Matsui 2004	
Outbreak 12:	Reitler, 1960	
Outbreak 39:	Abe et al., 2004 Outbreak 5	
Outbreak 39:	Abe et al., 2004 Outbreak 6	
Outbreak 39:	Abe et al., 2004 Outbreak 8	
Outbreak 40:	Urfur et al. 2000	
Human 4:	Greenberg 1964	
Outbreak 26:	Narain and Lofgren, 1989	Narain Sandwiches
^a Hyperlinks are to the discussions in Appendix A.		

The analyses in this worksheet are generally not standardized, although some conform to the approaches described in Appendix B.4. Mathematical descriptions, where necessary, are given in the appropriate entry in Appendix A (hyperlinked in Table 24). The additional worksheets referenced, “Font_1980,” “Hollandaise,” “Narain,” and “Sandwiches,” provide additional information or analysis, as described in the corresponding entries in Appendix A (see Table 24).

B.4. Analysis of Kasuga *et al.* (2004) — worksheet “Kasuga”

This worksheet continues the analyses of outbreaks listed in Table 5 and described in Appendix A with analyses of the outbreaks listed in Kasuga *et al.* (2004). Kasuga!A6:AJ38 is an extension of the tables of Kasuga *et al.* (2004), including cross-references to the FAO/WHO (2002b) analysis, and including estimates of concentration, food intake, dose, and their uncertainties. The values for each outbreak in this summary table for dose, dose uncertainty, numbers ill, and number exposed are used in workbook All_data (columns J, L, M, and N). Uncertainties are estimated as described in [Outbreak 37](#) in Appendix A; the calculations are performed in rows Kasuga!A46:A539. The methodologies used for evaluation of concentration uncertainties for MPN and plating analyses respectively are mathematically described in Sections 4.2 and 4.3. The layout of the numerical methods for MPN analyses is summarily described in Kasuga!G54:G77. The MPN analyses use the eacc_MPN function from the EACC_funs addin to compute MPN estimates, and eacc_binomial_LL to compute the loglikelihood for binomial observations. For estimates of confidence limits where the concentration was measured by plating, the layout is similar but simpler; the function Subtr_exp (see Appendix B.1 and Section 4.3) is used rather than the numerical optimizer.

B.5. Approximating loglikelihoods — worksheets “MultiDose” and “MultiDoseP”

Worksheets “MultiDose” and “MultiDoseP” are used to construct the approximations to the loglikelihood for the multiple-dose experiments of McCullough and Eisele (1951a,c) and the effective multiple-dose experiment of Greenberg (1964). These approximations are described in [Human 2](#) in Appendix A. In addition, “MultiDoseP” is used to obtain an estimate of the slope parameter for the alternative dose-response model (Section 8.7). The summary results given at Table 22 and Table 23 are from the ranges C10:H21 in the respective worksheets, the results of standardized analyses in column ranges C through R, AC through AR, BC through BR, etc.

As an example, columns MultiDose!C:R analyze Serial # [42](#) in Table 5. MultiDose!L32:N39 show the observations, MultiDose!P32:P39 the ln(GSD) for each dose, and MultiDose!Q32:Q39 the loglikelihood components for uncoupled uncertainties at each dose for a log(virulence) given by MultiDose!Q29. Column D shows the range of log(virulence) evaluated (min. and max. set by cells MultiDose!D46 and MultiDose!D47), column F the loglikelihood for uncoupled uncertainties at each dose, and column G the approximating loglikelihood with no relative uncertainty between doses. The required range for min. and max. may be found by first finding the maximum likelihood estimate for log(virulence) by maximizing MultiDose!Q45 with respect to MultiDose!Q29 (the maximum loglikelihood and MLE are in MultiDose!R45 and MultiDose!R29 respectively), then finding the range of log(virulence) that alters the loglikelihood by about 5.

Column G is calculated automatically by the binrandloglik function entries, but Column F has to be calculated for each log(virulence) in column D. This is done by copying the entry in Column D to MultiDose!Q29, then copying the result calculated in MultiDose!Q45 back to Column F (this process may be automated using the Profiler macro in the EACC_funs menu added by the EACC_funs addin, provided the Solver is set to do nothing; the Profiler will automatically put the results next to the corresponding entries in Column D [*i.e.*, in Column E], so those results should be copied to Column F; Column E has been left empty for this purpose). Subsequently, the entries for the shift in log(virulence), ln(GSD), and loglikelihood in MultiDose!G44:G46 are calculated by minimizing MultiDose!H48.

The worksheet MultiDoseP is substantially identical to worksheet MultiDose, except that it uses the probit-log-dose model rather than the linear-exponential model, and one extra initial step has to be taken. The initial step is to calculate the slope parameter of the probit-log-dose model (see Section 8.7) by setting MultiDoseP!N10 to TRUE and maximizing the loglikelihood in MultiDoseP!M23 with respect to the slope and $\ln(d_{50})$ parameters listed in MultiDoseP!N12:N21. Subsequently, with MultiDoseP!N10 set to FALSE, worksheet MultiDoseP is used in the same way as worksheet MultiDose.

B.6. Adequacy of fit of dose-response models — worksheet “McCullough”

The adequacy of fit of the dose-response models (Section 8.2, Table 6 and Table 7 for the linear-exponential model, and Section 8.7, Table 16 and Table 17 for the probit-log-dose model) are calculated in worksheet McCullough. Rows 8 through 308 are copies of the individual tabular data provided in McCullough and Eisele (1951a,c). Model parameters for both the two dose-response models, assuming both exact doses and relatively uncertain doses, are listed in rows 310 to 323. Summaries of the data are constructed in Rows 333 to 386, and illness probabilities, loglikelihoods, and chi-squared components for the various models are also calculated in these rows. The models are fitted by maximizing the loglikelihoods in row 328 with respect to the parameters in rows 315 to 323. Table 6, Table 7, Table 16 and Table 17 are then calculated at rows 389 to 400.

Comparison of naive subjects alone with all subjects for the linear-exponential model is performed at McCullough!BM312:BV386.

B.7. Individual ML estimates of virulence — worksheet “All_data”

Worksheet All_data collects together summary information on all the experiments and outbreaks, and implements the fitting of the linear-exponential dose-response model. All_data!A29:N151 summarize the available information, with dose and response information (columns J:N) cross-referencing other worksheets where necessary. Table 5 is constructed from these summary data at rows All_data!A171:K292.

The point estimates for virulence and their uncertainties for individual outbreaks and experiments (Section 8.3 and Table 8) are derived in All_data!P29:W151. Column P contains the loglikelihood at the current value of ln(virulence) in column S, and column T holds the current value of the virulence parameter (derived from the entry in column S). The maximum likelihood estimate is obtained by maximizing the loglikelihood in column P with respect to the

$\ln(\text{virulence})$ in Column S. The value of the maximum loglikelihood is copied into column Q, and the maximum likelihood value for the parameter is copied from column T into column U. The confidence limits in columns V and W are calculated by minimizing or maximizing the $\ln(\text{virulence})$ parameter in column S subject to the p-value in column R being greater than or equal to 0.1 (using the Solver); the calculated values are then copied from column T into columns U or V respectively. These maximum likelihood estimates and confidence limits take account of both the binomial uncertainty due to finite population sizes, and the uncertainties in dose. Table 8 is then constructed in All_data!N171:V240, and Table 9 at All_data!Z171:AB187 (the latter was constructed using the add-in function eacc_rem_dups, and is not automatically updated).

B.8. Distribution of individual ML estimates of virulence — worksheet “Distrib”

The ML estimates of virulence for individual outbreaks and experiments have been copied from worksheet All_data into worksheet Distrib and sorted, allowing construction of Figure 5 and Figure 6. An estimate for the average size of the uncertainties (where they are finite and measurable) is constructed at Distrib!B29:C29.

B.9. Estimates for the distribution of virulence — worksheet “All_data”

To estimate the parameters of the lognormal virulence distribution (Section 8.4), it is necessary to compute expressions like Equation (25) for each outbreak or human feeding study, form their sum, and maximize the result with respect to the parameters S , m and Σ (called S , μ and sd in the worksheet). This is done in columns Y through AQ of worksheet All_data. Initially, separate estimates for the parameters of Equations (25) and (27) were assumed for *S. enteritidis* (indicated by an appended _se to names in the worksheet), *S. typhimurium* (an appended _t), and all other serotypes combined (an appended _o) since only *S. enteritidis* and *S. typhimurium* are associated with more than four outbreaks each; see All_data!Y10:Z18 where the parameters are given names. These parameters are assigned to each serotype in All_data!T10:V24 (which is given the label “serotypes”). The correct values for each outbreak are obtained in columns All_data!AB32:AC151 using a lookup function, and the total variance ω (see Equation (26)) obtained in All_data!AD32:AD151 (the terms U_multiply and U_add in the spreadsheet, appearing in the calculation of the total variance, are used in the sensitivity analysis, see Appendix B.13). Then All_data!AE32:AE151 contain the loglikelihoods given in Equation (25), computed with the function binrandloglik,⁵⁴ a VBA function included in the spreadsheet (see Section C.1 for the method of computation). The total loglikelihood for all outbreaks is obtained in All_data!Z20. This is maximized with respect to the parameters in All_data!Z10:Z18.

All_data!AB6:AP24 summarizes the likelihood ratio tests described in Section 8.4 and used to evaluate whether various parameters may be set equal or should be kept distinct. The unconstrained MLE estimates were copied to All_data!AB10:AB18, with the value of the ML copied from All_data!Z20 to All_data!AB20. Likelihood ratio tests are implemented in All_data!Z20:Z23 by constructing twice a difference in loglikelihoods in All_data!Z21 (it is necessary to alter the expression in this cell to find the difference of All_data!Z20 from the

⁵⁴ Where only an upper on dose is available the function binrandloglik_2 is used.

appropriate comparison entry among All_data!AB20:AP20). As each constraint on the parameters is implemented, the corresponding parameters were copied from All_data!Z10:Z18 into the appropriate column in All_data!AB10:AP18 (using copy and paste), and the values obtained in All_data!Z10:Z23 were copied (using copy and paste values) into the corresponding column in All_data!AB20:AP20.⁵⁵

The final set of parameters described in Table 10 is given at All_data!AN10:AN18. Some of these have been given names with _mle appended to allow easy reference elsewhere in the workbook, and Table 10 itself is constructed at All_data!Q9:R14. All_data!AO21:AO23 performs a test of the final selection against the unconstrained MLE (since the final values were obtained by a stepwise process), and All_data!AP10:AP23 estimates a p-value for $\Sigma = 0$ (as requested by a referee).

The shape of the loglikelihood surface shown in Figure 7 is constructed at All_data!DQ4:FR62 using the GetArray item on the EACC_funs addin menu.

B.10. The random isolate dose-response — worksheets “All_data” and “DR-pict”

Figure 8 and Figure 9 (the maximum likelihood estimate for the virulence distribution for *S. enteritidis*, and selected dose-response curves with virulences selected from that distribution) are constructed in worksheet DR-pict at columns AT through BM. Figure 10 and Figure 11 showing the random isolate dose-response curve (Section 8.5) are also constructed in DR-pict, although the confidence limits on this curve are computed in worksheet All_data at columns CW through DC (the practical method of construction is explained in All_data!CW6:CW10).

B.11. Uncertainties in the parameter estimates — worksheet “All_data”

The uncertainties in parameter estimates (Section 8.6) were obtained using standard likelihood methods. First, the profile likelihoods for S , m and Σ were obtained using the Profiler tool of the add-in EACC_funs (columns AW through BT of All_data) over the ranges of parameter values described in Section 8.6. Then a transformation was applied (if necessary) to provide a practically linear relationship between the signed square root of twice the deviation of the loglikelihood from the maximum and the transformed parameter. These relationships, using the transformations of Table 11, are shown in the graphs located approximately at All_data!AW8:BV26. The transformations were found empirically; the closeness to normality of them is shown by the very high correlations between the transformed parameter and the signed square root of twice the change in loglikelihood. The variance-covariance matrix in Table 12, the standard deviation/correlation matrix of Table 13, and the Cholesky decomposition of Table 14 were obtained numerically by computing the information matrix using the add-in tool InformationMatrix and performing the necessary matrix manipulations using standard and add-in functions (MINVERSE, eacc_cholesky, and eacc_sd_correl). These calculations are performed in All_data!BX10:CC49. A short description of the practical methodology is given in All_data!BX6:BX8; the parameter setup required has been constructed in All_data!AQ10:AQ18. The numerical calculations of the information matrix were performed using two step sizes that

⁵⁵ Using copy and paste for the parameters entries (and not “paste values”) allows the entries in each column in All_data!AB10:AP18 to be copied back into All_data!Z10:Z18 while preserving any constraints as formulas.

differed by a factor of 100. The values used are listed in All_data!CE12:CF17, and the required set may be copied into the “Delta_actual/Delta_vals” entries at All_data!CC12:CC17. The larger set is designed to correspond in size roughly to the standard deviations of each parameter (see the main diagonal of Table 13, also shown at All_data!BX45:CC49).

The standard deviation estimates for the untransformed parameters (Table 15) are obtained by using the inverse transformations at All_data!CD45:CE49.

B.12. Alternative dose-response curve — worksheet “Probit” and “Distrib_p”

The analysis of the probit-log-dose dose-response model (Section 8.7) is performed in the worksheet Probit, with the distributions constructed in worksheet Distrib_p. Most of the calculations are exactly analogous to the analysis of the linear-exponential dose-response model (see Appendix B.7 through Appendix B.10). The integrals are performed in the VBA function PLDLNloglik (see Section C.2 for the method of calculation). The dose-related parameters in the worksheet are all natural logarithms of doses, since this gives values that are more conveniently handled for this dose-response curve (e.g. in using the Solver); so the d_0se, d_0t, and d_0o of the worksheet correspond to $\ln(D_{50})$ of Section 8.7 for the corresponding serotypes, and sig_se, sig_t, and sig_o correspond to ω in the same way. No uncertainty analysis is performed, since none is needed. The parameter β (labeled the “slope” parameter in the worksheet) is evaluated once by finding the ML estimate for it using only the human feeding data (see Appendix B.5).

B.13. Sensitivity analysis — worksheet “All_data”

The sensitivity analyses of Section 8.8 are implemented in worksheet All_data. The modification of the dose uncertainty estimates is performed by changing U_multiply and U_add in All_data!I7:I8, re-optimizing, and saving the changes in the parameter estimates (All_data!H10:H18). Table 19 is then constructed in All_data!K8:O14.

The evaluation of the hospital patient subgroup corresponds to manipulation of the parameter S_H (All_data!Z10) and re-optimization.

Omission of the experimental data is implemented by constructing a loglikelihood in All_data!AA20 that omits the loglikelihood components contributed by the experimental data, and performing the analyses using that loglikelihood (rather than the total loglikelihood in All_data!Z20). The results used in Section 8.8 are stored in columns CL through CP of All_data (these are analogous to, and obtained in the same way as, the results discussed in Appendix B.9). Table 20 is then constructed at All_data!CM27:CN32.

B.14. Calculation of approximations for an integral — worksheet “h_coeffs”

The coefficients h_0, h_1, h_2, h_3 given in Table 25 are calculated in worksheet h_coeffs in columns A through K using the approximation described in Appendix C.2.2. The numerical investigation described below Equation (C.66) can be performed using the setup of columns Q through Y.

B.15. Reference tables — worksheet “MPN tables”

This worksheet contains reference tables of MPN results that were used in the interpretation of published results discussed in Appendix A. The standard MPN tables (MLE values obtained using the add-in function in EACC_funcs) or modified tables (Roussanov *et al.*, 1996) were used to impute patterns of detections from reported MPN results, to allow estimates of confidence limits on the reported values. In most cases the reported value implied a unique pattern; in those cases where more than one pattern agreed with the reported value, the most likely pattern was selected.

B.16. Reference specific to Appendix B

Crouch, E.A.C., 2006. Quattro Pro and Excel Add-In Function Library EACC. Cambridge Environmental Inc., 58 Charles Street, Cambridge, MA 02141.
www.CambridgeEnvironmental.com. (Unpublished; attached to this work and available from the author).

Appendix C. Evaluation of the integrals.

Symbols used in this mathematical appendix may have different meanings from those in the text, and the same symbols may be re-used with different meanings in different contexts; however, the meanings should be clear from the context.

C.1. Binomial response, linear-exponential dose-response with lognormal parameters.

C.1.1. Multiple doses, all doses known

We assume sets of observations $\{r_j, n_j, d_j\}$ of r_j positive events from n_j trials at a dose measure d_j , for $j = 1 \dots N$, with a probability model for event occurrence at dose d of

$$p = 1 - \exp(-(a + bd)) \quad \text{with} \quad a \geq 0, b \geq 0, \text{ and } ab > 0 \quad (\text{C.1})$$

where a , b , or d is lognormally distributed with a distribution common to all the observations and outcomes that are binomially distributed with probability p .⁵⁶ The likelihood for the set of such observations is then

$$L = \frac{H}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} dz \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) \prod_{j=1}^N \left\{ \left[1 - \exp(-(a_j + b_j e^z))\right]^{r_j} \exp(-(n_j - r_j)(a_j + b_j e^z)) \right\} \quad (\text{C.2})$$

where suitable specialization of the a_j and b_j correspond to a , b , or the d_j having a lognormal multiplier, the logarithm of which has mean μ and standard deviation σ , and H is a normalization constant that is usually of no particular interest in likelihood analysis. For convenience, here H is defined by

$$\frac{1}{H} = \prod_{j=1}^N \left\{ \left(\frac{r_j}{n_j}\right)^{r_j} \left(1 - \frac{r_j}{n_j}\right)^{n_j - r_j} \right\} \quad (\text{C.3})$$

(where the terms are appropriately interpreted as unity if $r_j = 0$ or n_j) so that $L = 1$ if the observations match the model exactly; this replaces the exact binomial normalization

$$\frac{1}{H} = \prod_{j=1}^N \binom{n_j}{r_j}. \quad (\text{C.4})$$

This integral may be accurately and rapidly evaluated, with a known error estimate, using a standard quadrature method (Crouch and Spiegelman, 1990; Crouch, 1991). Determination of the error estimate requires investigation of the location of the maximum of the integrand, and the analytic and numerical behavior off the real axis of the integrand omitting the first exponential term. To put the integral in standard form, write

$$z = \mu + st \quad \text{with} \quad s = \sigma\sqrt{2} \quad (\text{C.5})$$

and put

$$c_j = b_j e^{\mu} \quad (\text{C.6})$$

to get

$$L = \frac{H}{\sqrt{\pi}} \int_{-\infty}^{\infty} dt e^{-t^2} \prod_{j=1}^N \left\{ \left[1 - \exp(-(a_j + c_j e^{st}))\right]^{r_j} \exp(-(n_j - r_j)(a_j + c_j e^{st})) \right\} \quad (\text{C.7})$$

⁵⁶ Since the dose terms enter the likelihood only through p , the sensitivity factor S may be ignored in this analysis; introducing a sensitivity factor simply corresponds to increasing all doses uniformly by that sensitivity factor.

Now examine the behavior of the product term in the integrand (*i.e.* the integrand excluding the $\exp(-t^2)$ term) off the real axis, with $t \rightarrow t + i\kappa$ where t and κ are real. Then one may write

$$\exp\left(-\left(a_j + c_j e^{st+i\kappa t}\right)\right) = q_j e^{-i\varphi_j} \quad (\text{C.8})$$

where

$$0 \leq q_j = \exp\left(-\left(a_j + c_j e^{st} \cos \kappa s\right)\right) < 1 \quad \text{and} \quad \varphi_j = c_j e^{st} \sin \kappa s \quad (\text{C.9})$$

To ensure convergence of the off-axis integral (as required by the method) requires

$$\cos \kappa s \geq 0 \quad (\text{C.10})$$

which may be ensured by imposing

$$\kappa s \leq \pi/2 \quad (\text{C.11})$$

Then, off the real axis, the product term in the integrand satisfies

$$\begin{aligned} & \left| \prod_{j=1}^N \left\{ \left(1 - q_j e^{-i\varphi_j}\right)^{r_j} q_j^{n_j - r_j} e^{-i(n_j - r_j)\varphi_j} \right\} \right| \\ & \leq \prod_{j=1}^N \left\{ \left(1 + q_j\right)^{r_j} q_j^{(n_j - r_j)} \right\} \leq \prod_{j=1}^N \left\{ \left(1 + e^{-a_j}\right)^{r_j} e^{-(n_j - r_j)a_j} \right\} \leq 2^{\sum_{j=1}^N r_j} \end{aligned} \quad (\text{C.12})$$

Setting

$$M = \prod_{j=1}^N \left\{ \left(1 + e^{-a_j}\right)^{r_j} e^{-(n_j - r_j)a_j} \right\} \quad (\text{C.13})$$

gives an error term $E(h)$ for the standard quadrature method with step size h of (ignoring terms that will be much smaller)

$$E(h) \leq 2MH \exp(\kappa^2 - 2\pi\kappa/h) \quad (\text{C.14})$$

To select the step size, suppose that the maximum allowable error is $L_a\eta = IH\eta$, where η is an approximate relative error, L_a an approximation to L , and I an approximation to L/H . The inequality (C.14) shows that the step size must be selected so that

$$2MH \exp(\kappa^2 - 2\pi\kappa/h) \leq IH\eta \quad (\text{C.15})$$

which requires

$$\kappa^2 - 2\pi\kappa/h - \ln\left(\eta I / 2M\right) \leq 0 \quad (\text{C.16})$$

There are now two cases. First, choose κ to minimize the expression on the left. That gives

$$\kappa = \pi/h \quad (\text{C.17})$$

and the inequality (C.16) requires

$$h \leq \pi / \sqrt{-\ln\left(\eta I / 2M\right)} \quad (\text{C.18})$$

and the largest such h is most efficient (choose the equality sign). However, the limitation (C.11) implies that this value can only be used if

$$s \sqrt{-\ln\left(\eta I / 2M\right)} \leq \frac{\pi}{2} \quad (\text{C.19})$$

In the second case, where this inequality is violated, choose κ as large as possible, so

$$\kappa = \pi/2s \quad (\text{C.20})$$

and then (C.16) requires

$$h \leq \frac{4\pi^2 s}{\pi^2 - 4s^2 \ln\left(\frac{\eta I}{2M}\right)} \quad (\text{C.21})$$

and again the equality should be chosen for highest efficiency.

To obtain an appropriate step size, an estimate of the size of the integral is required. In addition, the standard quadrature method performs best numerically when the quadrature is started at the maximum of the integrand. Furthermore, complete numerical evaluation also requires proof of a single maximum of the integrand, or requires that steps be taken to ensure that all the maxima are taken into account.

The exponent in the integrand of (C.7) may be written as

$$f(t) = -t^2 + \sum_{j=1}^N \left\{ r_j \ln p_j + (n_j - r_j) \ln(1 - p_j) \right\} \quad (\text{C.22})$$

with $p_j = 1 - \exp\left(-\left(a_j + c_j e^{st}\right)\right)$

which has its maximum at the solution of

$$f'(t) \triangleq -2t + \sum_{j=1}^N \left(\frac{r_j}{p_j} - n_j \right) s c_j e^{st} = 0. \quad (\text{C.23})$$

Multiplying by $s/2$, using (C.5), and substituting $x = e^{st}$ this is equivalent to

$$g(x) \triangleq -\ln x + \sigma^2 x \sum_{j=1}^N c_j \left(\frac{r_j}{p_j} - n_j \right) = 0 \quad (\text{C.24})$$

Differentiating gives

$$g'(x) = -\frac{1}{x} - \sigma^2 \sum_{j=1}^N c_j \left\{ (n_j - r_j) + \frac{r_j(1 - p_j)}{p_j^2} (x c_j - p_j) \right\} \quad (\text{C.25})$$

where the term in braces ($\{\}$) has been written so it is easy to see it is positive for all $x \geq 0$ and $0 \leq r_j \leq n_j$ if $a_j = 0$. In that case, therefore, g' , hence also f'' , is negative definite, the integrand has a single maximum, and the Newton-Raphson step direction for iteratively solving (C.24) will always be correct.

To estimate the size I of the integral for the purpose of setting the step size, it is easiest to find the maximum of the integrand, find the second derivative of the logarithm of the integrand at that maximum, and use the standard saddle point approximation.

C.1.2. Upper bound on single dose

In this case, the likelihood is proportional to

$$L = \int_0^d \left\{ \frac{H}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[\exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) (w(x, z))^r ((1-w(x, z)))^{n-r} \right] dz \right\} dx \quad (\text{C.26})$$

with

$$w(x, z) = 1 - \exp(-Sxe^z) \quad (\text{C.27})$$

where the nomenclature is the same as in Section C.1.1 (but for a single dose with known upper bound d) and S is a constant (the relative sensitivity).⁵⁷ Setting

$$z = \mu + st \quad s = \sigma\sqrt{2} \quad \tilde{\mu} = \mu + \ln S \quad (\text{C.28})$$

gives

$$L = \frac{H}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left\{ e^{-t^2 - \tilde{\mu} - st} B(r+1, n-r, 1 - \exp(-d \exp(\tilde{\mu} + st))) \right\} dt \quad (\text{C.29})$$

where B is an incomplete beta integral

$$B(a, b; x) = \int_0^x z^{a-1} (1-z)^{b-1} dz \quad (\text{C.30})$$

The integral in (C.29) may be evaluated using the same method as in Section C.1.1. The off-real-axis bound is obtained in exactly the same way (the bounds obtained on the integrand in Section C.1.1 are uniform with respect to the integration variable x in Equation (C.26)), with the result analogous to Equation (C.13) being

$$M = d2^r \quad (\text{C.31})$$

and the error analysis otherwise proceeds as before.

As before, to obtain an appropriate step size, an estimate of the size of the integral is required. In addition, the standard quadrature method performs best numerically when the quadrature is started at the maximum of the integrand. Furthermore, complete numerical evaluation also requires proof of a single maximum of the integrand, or requires that steps be taken to ensure that all the maxima are taken into account.

The integrand of the integral in Equation (C.29) can be written as the exponential of

$$f(t) = -t^2 - \tilde{\mu} - st + \ln B(r+1, n-r, p) \quad (\text{C.32})$$

with $p = 1 - \exp(-de^{\tilde{\mu} + st})$ so $0 \leq p \leq 1$

and it is straightforward to see that $f \rightarrow -\infty$ for $t \rightarrow \pm\infty$, so that the maximum of the integrand occurs at a solution of

$$f'(t) \triangleq -2t - s + \frac{1}{B} \frac{dB}{dt} = 0 \quad (\text{C.33})$$

Multiplying by $s/2$, setting $\xi = e^{st}$, and using Equation (C.28), this is equivalent to

$$g(\xi) \triangleq -\ln \xi - \sigma^2 - \sigma^2 \frac{p^r (1-p)^{n-r} \ln(1-p)}{B(r+1, n-r, p)} = 0 \quad (\text{C.34})$$

Differentiating gives

$$g'(\xi) = -\frac{1}{\xi} - \sigma^2 S de^{\mu} (1-p) \frac{d}{dp} \left[\frac{p^r (1-p)^{n-r} \ln(1-p)}{B(r+1, n-r, p)} \right] \quad (\text{C.35})$$

Now

⁵⁷ The dose now enters the expression for the likelihood both through p and through the limit of an integral; it is therefore necessary to explicitly account for the sensitivity factor S , in contrast to Section C.1.1.

$$\begin{aligned} \frac{d}{dp} \left[\frac{p^r (1-p)^{n-r} \ln(1-p)}{B(r+1, n-r, p)} \right] = \\ - \frac{\ln(1-p) p^{r-1} (1-p)^{n-r-1}}{B^2} \left\{ p^{r+1} (1-p)^{n-r} + \left(np - r + \frac{p}{\ln(1-p)} \right) B \right\} \end{aligned} \quad (C.36)$$

so that if

$$u(p) \triangleq np - r + \frac{p}{\ln(1-p)} \geq 0 \quad (C.37)$$

then g' is trivially negative, whereas if $u(p) < 0$ (which necessarily occurs for small enough p) then consider

$$h(p) = \frac{p^{r+1} (1-p)^{n-r}}{-u(p)} - B \quad (C.38)$$

Elementary considerations show that

$$h(0) = 0 \quad \text{for } 0 \leq r \leq n$$

$$\begin{aligned} h'(p) = \frac{p^r (1-p)^{n-r-1}}{u^2} \left[r(1-p) \left(1 + \frac{p}{\ln(1-p)} \right) - (n-r) \frac{p}{\ln(1-p)} \right] \\ \geq 0 \quad \text{for } 0 \leq r \leq n \end{aligned} \quad (C.39)$$

so that $h \geq 0$, and it follows that g' is everywhere negative, so that Equations (C.34) and (C.33) have exactly one solution, the integrand of Equation (C.29) has a single maximum, and the Newton-Raphson step direction for iteratively solving Equations (C.34) and (C.33) will always be correct.

As before, to estimate the size of the integral for the purpose of setting the step size, it is easiest to find the maximum of the integrand, find the second derivative of the logarithm of the integrand at that maximum, and use the standard saddle point approximation.

C.2. Binomial response, probit-log-dose dose-response with lognormal parameters

C.2.1. Multiple doses, all doses known

We assume sets of observations $\{r_j, n_j, d_j\}$ of r_j positive events from n_j trials at a dose measure d_j , for $j = 1 \dots N$, with a probability model for event occurrence at dose d of

$$p = \Phi(\beta \ln(d/d_{50})) \quad \text{with } \beta > 0, d_0 > 0 \quad (C.40)$$

where Φ is the standard cumulative normal function,

$$\Phi(x) = \int_{-\infty}^x Z(t) dt \quad \text{with } Z(x) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) \quad (C.41)$$

d_{50} is lognormally distributed with a distribution common to all the observations and outcomes that are binomially distributed with probability p .⁵⁸ The likelihood for the set of such observations is then

$$L = \frac{H}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(z - \ln D_{50})^2}{2\sigma^2}\right) \prod_{j=1}^N \left\{ \left[\Phi\left(\beta(\ln(d_j) - z)\right) \right]^{r_j} \left[1 - \Phi\left(\beta(\ln(d_j) - z)\right) \right]^{n_j - r_j} \right\} dz \quad (\text{C.42})$$

where again H is defined by (C.3) (see the discussion of that definition), and $\ln(d_{50})$ is normally distributed with mean $\ln(D_{50})$ and standard deviation σ . This may be put in standard form by the substitutions

$$s = \beta\sigma\sqrt{2} \quad \text{and} \quad \xi_j = \beta \ln(d_j / D_{50}) \quad (\text{C.43})$$

to obtain

$$L = \frac{H}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} \prod_{j=1}^N \left\{ \left[\Phi(\xi_j + st) \right]^{r_j} \left[1 - \Phi(\xi_j + st) \right]^{n_j - r_j} \right\} dt \quad (\text{C.44})$$

This integral may also be evaluated using the standard quadrature method (Crouch and Spiegelman, 1990; Crouch, 1991). It is necessary to examine the behavior of the product term in the integrand (*i.e.* the integrand excluding the first exponential term) off the real axis, with $t \rightarrow t + i\kappa$ where t and κ are real. Now for x, ε real,

$$\begin{aligned} \Phi(x + i\varepsilon) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x+i\varepsilon} \exp(-t^2/2) dt \\ &= \frac{\exp(\varepsilon^2/2)}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-t^2/2 - i\varepsilon t) dt \end{aligned} \quad (\text{C.45})$$

so

$$|\Phi(x + i\varepsilon)| \leq \exp(\varepsilon^2/2) \Phi(x) \quad (\text{C.46})$$

and in an exactly similar fashion

$$|1 - \Phi(x + i\varepsilon)| \leq \exp(\varepsilon^2/2) [1 - \Phi(x)] \quad (\text{C.47})$$

so that

$$\begin{aligned} &\prod_{j=1}^N \left\{ \left[\Phi(\xi_j + s(t + i\kappa)) \right]^{r_j} \left[1 - \Phi(\xi_j + s(t + i\kappa)) \right]^{n_j - r_j} \right\} \\ &\leq \exp(s^2 \kappa^2 N_t / 2) \prod_{j=1}^N \left\{ \left[\Phi(\xi_j + st) \right]^{r_j} \left[1 - \Phi(\xi_j + st) \right]^{n_j - r_j} \right\} \end{aligned} \quad (\text{C.48})$$

where

$$N_t = \sum_{j=1}^N n_j \quad (\text{C.49})$$

⁵⁸ As in Section C.1.1, it is not necessary to explicitly consider the sensitivity factor S , because doses occur in the analysis only through p (see Footnote 56).

In this case, there is no limit on the size of κ — the integral is well-defined and finite for all finite κ . The error term $E(h)$ for a step size h using the standard quadrature approach (again ignoring terms that will be small) is bounded by

$$E(h) \leq 2 \exp(\kappa^2 - 2\pi\kappa/h) L \exp(s^2 \kappa^2 N_t / 2) \quad (C.50)$$

Thus for a relative error η the step size must be chosen to satisfy

$$\kappa^2 (1 + N_t s^2 / 2) - 2\pi\kappa/h \leq -\ln(\eta/2) \quad (C.51)$$

for which the numerically optimum solution (*i.e.* the largest h) is obtained by setting κ at the value minimizing the left side of (C.51), leading to

$$h \leq \frac{\pi}{\sqrt{(-\ln(\eta/2))(1 + N_t s^2 / 2)}} \quad (C.52)$$

The maximum of the integrand of the integral in (C.44) occurs at the maximum of

$$f(t) = -t^2 + \sum_{j=1}^N \left\{ r_j \ln \Phi(\xi_j + st) + (n_j - r_j) \ln(1 - \Phi(\xi_j + st)) \right\} \quad (C.53)$$

or equivalently at the zero of

$$g(t) = f'(t) = -2t + s \sum_{j=1}^N \left\{ \frac{r_j Z_j}{\Phi_j} - \frac{(n_j - r_j) Z_j}{1 - \Phi_j} \right\} \quad \text{where} \quad Z(x) = \Phi'(x) \quad (C.54)$$

and the subscripts j on Z and Φ indicate evaluation at the point $\xi_j + st$. This has only one zero since every term of

$$g'(t) = -2 + s^2 \sum_{j=1}^N \left\{ r_j \left(\frac{Z}{\Phi} \right)'_j - (n_j - r_j) \left(\frac{Z}{1 - \Phi} \right)'_j \right\} \quad (C.55)$$

is negative. This follows from

$$\left(\frac{Z}{\Phi} \right)' = -\frac{Z}{\Phi} \left(x + \frac{Z}{\Phi} \right) \quad (C.56)$$

which is obviously negative for $x \geq 0$. For $x < 0$, write $y = -x$ so y is positive. Then we may integrate by parts in

$$\Phi(x) = 1 - \Phi(y) = \int_y^\infty Z(t) dt = \frac{Z(y)}{y} - \int_y^\infty \frac{Z(t)}{t^2} dt < \frac{Z(y)}{y} \quad (C.57)$$

giving

$$x + \frac{Z}{\Phi} > 0 \quad (C.58)$$

for $x < 0$ also. The two terms multiplying the counts in the sum in (C.55) are symmetric around zero, so this proof applies also to the second such term.

The location of the maximum may then be found by Newton-Raphson iteration, and the negativity of g' ensures that the Newton direction will always be correct in the iteration.

C.2.2. Upper bound on single dose

In this case the required integral may be put in the form (cf. Equation (C.44))

$$L = \frac{H}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} \int_0^d \left\{ [\Phi(\xi + st)]^r [1 - \Phi(\xi + st)]^{n-r} \right\} dx dt \quad (C.59)$$

where

$$s = \beta\sigma\sqrt{2} \quad \text{and} \quad \xi = \beta \ln(Sx/D_{50}) \quad (C.60)$$

using the same nomenclature as in Section C.2.1, with the addition of S , the relative sensitivity.⁵⁹ The analysis proceeds in exactly the same way as in Section C.2.1, except for the incorporation of only one dose in this case (so that no subscripts on r , n , and ξ are required). The presence of the integration with respect to x in Equation (C.59) does not affect the analysis of the required step size in the integration over t , which is the same in this case as in Section C.2.1 for the same relative error in the integral. Proceeding in this fashion requires evaluation of the inner integral. However, only two cases are required with the current data, corresponding to $(r,n) = (22,40)$ and $(1,1)$ (see Table 5, Serial # [3](#) and [41](#) respectively).

All the $(r,n) = (1,1)$ integrals can be obtained analytically. In this case $H = 1$ and

$$\begin{aligned} L &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^d \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \frac{(z - \ln D_{50})^2}{\sigma^2}\right) \Phi(\beta(\ln(Sx) - z)) dz dx \\ &= \int_0^d \Phi(\gamma \ln(Sx/D_{50})) dx \quad \text{with} \quad \gamma = \frac{\beta}{\sqrt{1 + \beta^2 \sigma^2}} \\ &= d \left[\Phi(\gamma \ln(Sd/D_{50})) - \frac{D_{50}}{Sd} \exp\left(\frac{1}{2\gamma^2}\right) \Phi(\gamma \ln(Sd/D_{50}) - 1/\gamma) \right] \end{aligned} \quad (C.61)$$

Analytic evaluation of numerical methods for the general case has not been performed to the same depth as in Section C.1.2 for the linear-exponential dose-response case. Instead, approximations have been used that are adequate for the $(r,n) = (22,40)$ case. For the general case, write

$$\begin{aligned} w &= \xi + st = \beta(\ln(Sx/D_{50}) + t\sigma\sqrt{2}) \\ \theta &= \Phi(w) \end{aligned} \quad (C.62)$$

Then

$$\begin{aligned} &\int_0^d [\Phi(\xi + st)]^r [1 - \Phi(\xi + st)]^{n-r} dx \\ &= \frac{D_{50}}{\beta S} \exp(-t\sigma\sqrt{2}) \int_{-\infty}^{\beta(\ln(Sd/D_{50}) + t\sigma\sqrt{2})} [\Phi(w)]^r [1 - \Phi(w)]^{n-r} \exp(w/\beta) dw \\ &= \frac{D_{50}}{\beta S} \exp(-t\sigma\sqrt{2}) \int_0^{\Phi(\beta(\ln(Sd/D_{50}) + t\sigma\sqrt{2}))} \theta^r (1 - \theta)^{n-r} \frac{\exp(w/\beta)}{Z(w)} d\theta \end{aligned} \quad (C.63)$$

⁵⁹ See Footnote 57. Again, the sensitivity factor must be explicitly included because the dose enters the likelihood both through p and through a limit on an integral.

In the last expression in Equation (C.63), the outer integration variable t occurs only in the pre-multiplying factor and in the limit of the integral; the integrand depends only on r , n , and β . Numerical investigation shows that the term

$$\frac{\exp(w/\beta)}{Z(w)} \quad (\text{C.64})$$

may be approximated as a function of θ by

$$\frac{\exp(w/\beta)}{Z(w)} \approx h_0 \theta^{h_1} (1 - \theta)^{h_2} (1 + h_3 \theta) \quad (\text{C.65})$$

where the coefficients h_0, h_1, h_2, h_3 may be estimated numerically. For $\beta = 0.976194$ (see Section 8.7), the choice of coefficients given in Table 25 give a relative error of less than 0.5% over the range where the integrand exceeds 10^{-10} of its maximum value for $(r, n) = (22, 40)$.⁶⁰ The limit of the integral in the last expression of Equation (C.63) sweeps through the whole range of 0 to 1 as t varies in the outer integral in Equation (C.59), so the larger relative errors outside this range are unimportant. The error in this approximation corresponds to assuming that the uncertainty distribution for the observation of r ill out of n exposed is not quite binomial, so that the likelihood function does not quite follow the binomial form particularly at extreme deviations from its maximum.

Table 25 Coefficients for Equation (C.65)

Coefficient	Value
h_0	0.48358
h_1	-0.38637
h_2	-1.31076
h_3	1.19939

With the approximation given by Equation (C.65), the inner integral is given by

$$\begin{aligned} & \int_0^d [\Phi(\xi + st)]^r [1 - \Phi(\xi + st)]^{n-r} dx \\ & \approx \frac{D_{50} h_0}{\beta S} \exp(-t\sigma\sqrt{2}) \left[B(r+1+h_1, n-r+1+h_2, \Phi(\beta(\ln(Sd/D_{50}) + t\sigma\sqrt{2}))) \right. \\ & \quad \left. + h_3 B(r+2+h_1, n-r+1+h_2, \Phi(\beta(\ln(Sd/D_{50}) + t\sigma\sqrt{2}))) \right] \end{aligned} \quad (\text{C.66})$$

where B is again defined by Equation (C.30). Numerical experiments with $(r, n) = (22, 40)$ indicate the integrand in Equation (C.59) is unimodal, so calculation of that integral follows the same pattern as in Section C.2.1.

⁶⁰ The calculation of these coefficients is performed in sheet h_coeffs of the workbook “Epi Attack Data.xls”.

C.3. References specific to Appendix C

Crouch, E.A.C. and Spiegelman, D. 1990. The evaluation of integrals of the form

$\int_{-\infty}^{\infty} f(t) \exp(-t^2) dt$: application to logistic-normal models. *J. Am. Stat. Assoc.* 85(410):464–469.

Crouch, Edmund. 1991. The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(t) \exp(-t^2) dt$ and the

numerical evaluation of $I(x, y; n) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dt \frac{e^{-t^2}}{(1 + e^{x+yt})^n}$ with particular attention to

$I(x, y) = I(x, y; 1)$. Cambridge Environmental Inc., 58 Charles Street, Cambridge, MA 02141. November 17, 1991. www.CambridgeEnvironmental.com. (Unpublished; attached to this work and available from the author).

Appendix D. The effects of a distribution of doses

Primary dose-response curve

To examine the effects of a distribution of doses on the probability for illness, consider the primary dose-response model

$$p = 1 - \exp(-\lambda d) \quad (\text{D.1})$$

This is defined to be the population dose-response relationship that would be observed if every individual in the population received the same dose d (necessarily an integer number of CFU).

Suppose instead the doses to individuals are random and distributed as a Poisson variate (as would be the case for contamination with uniform concentration in food consumed in equal-sized portions by the whole population exposed), then the probability for a dose of exactly d organisms would be

$$\Pr(d) = e^{-D} \frac{D^d}{d!}$$

at a population mean dose of D (d is necessarily an integer number of organisms, but D need not be). Since the doses are discrete, there are sub-populations receiving each dose d . Applying Equation (D.1) to each sub-population with equal dose, and summing over such sub-populations, the population dose-response (the expected fraction of the population ill) would be

$$\Pr(\text{illness}) = \sum_{d=0}^{\infty} e^{-D} \frac{D^d}{d!} (1 - e^{-\lambda d}) = 1 - \exp(-(1 - e^{-\lambda})D) \quad (\text{D.2})$$

so that in this case a population dose-response curve of the form of Equation (D.1) could be applied using the population mean dose, but the multiplier of that dose would be bounded above by unity. However, every outbreak included in the analysis has $\lambda \leq 0.21$ (at the MLE), so that for every outbreak included in the analysis

$$1 - \exp(-\lambda) \approx \lambda$$

to an approximation (<10%) substantially better than the available information⁶¹ (the average standard deviation on individual estimates of λ is approximately a factor of 2).

As another example, suppose that some fraction α of the population receives a dose D/α (an integer number of CFU), while the remainder of the population is not exposed, so the average dose in the population is again D (not necessarily an integer). The probability of illness for a randomly selected member of the whole population would then be

$$\Pr(\text{illness}) = \alpha (1 - \exp(-\lambda D/\alpha)) \quad (\text{D.3})$$

and the relative error of the approximation using just the average dose will be about γ or less if

$$\lambda D < \frac{2\gamma\alpha}{1-\alpha} \quad (\text{D.4})$$

e.g. for $\alpha = 0.1$, for a relative error less than about 10% the probability of illness for the exposed part of the population should be less than about 0.22.

⁶¹ The next largest is ~ 0.07 , so the approximation for all but the largest value is better than 4%.

In a similar fashion, other distributions of doses among sub-populations will give different dose-response curves at the required population level. Thus the analyses of exposures should strictly take account of the distribution of doses; although that is unknown in every outbreak analyzed. Insofar as doses to all sub-populations lie on the approximately linear part of the dose-response curve, only the mean dose matters. That is, to the extent that the approximation

$$1 - \exp(-\lambda d) \approx \lambda d \quad (\text{D.5})$$

holds for all doses to all sub-populations, then using the mean dose will be at least as accurate, since then

$$\begin{aligned} \Pr(\text{illness}) &= \sum_{d=0}^{\infty} \Pr(d) (1 - \exp(-\lambda d)) \\ &\approx \sum_{d=0}^{\infty} \Pr(d) \lambda d \\ &= \lambda D \end{aligned} \quad (\text{D.6})$$

In general, with the linear-exponential dose-response model and an arbitrary distribution of doses with mean D ,

$$\Pr(\text{illness}) = \sum_{d=0}^{\infty} \Pr(d) (1 - \exp(-\lambda d)) \quad (\text{D.7})$$

so that

$$\begin{aligned} 1 - \Pr(\text{illness}) &= \exp(-\lambda D) \sum_{d=0}^{\infty} \exp(-\lambda(d-D)) \\ &= \exp(-\lambda D) \left(1 + \sum_{k=2}^{\infty} \frac{(\lambda D)^k}{k!} \frac{M_k}{D^k} \right) \end{aligned} \quad (\text{D.8})$$

where M_k is the k^{th} moment about the mean of the dose distribution (the first moment is zero by definition, so the sum starts at $k = 2$). Thus if the doses are reasonably clustered around the mean (so the higher moments of the dose distribution are small) the terms in the sum on the right will be small — *e.g.* the first term in the sum is

$$\frac{(\lambda D)^2}{2} (CV)^2 \quad (\text{D.9})$$

where CV is the coefficient of variation of the dose distribution. For a CV of 10%, and if the fraction of people unaffected is more than about 10% (so $\lambda D < 2.3$), the error in estimating the unaffected fraction by using only the mean dose in the dose-response equation will be less than about 3%.

Alternative dose-response models

For an arbitrary dose-response model

$$p = f(d) \quad (\text{D.10})$$

the population dose-response expected given exposure of sub-populations to different doses is given by

$$\begin{aligned}
\Pr(\text{illness}) &= \sum_{d=0}^{\infty} \Pr(d) f(d) \\
&= f(D) + \sum_{k=2}^{\infty} \frac{f^{(k)}(D) M_k}{k!}
\end{aligned} \tag{D.11}$$

where $f^{(k)}(x)$ is the k^{th} derivative of f evaluated at x , and as before M_k is the k^{th} moment about the mean of the dose distribution. Approximating the probability of illness by $f(D)$ therefore requires the sum to be small, which will happen if the higher derivatives of the dose-response model evaluated at D are sufficiently small, the moments about the mean of the dose distribution are sufficiently small, or some combination of both. Generally, if the dose-response model is reasonably linear in the region of large probability for dose, the approximation may be adequate; but if the dose-response model is strongly curved in that region the approximation may fail badly.

The alternative dose-response model used here (a probit-log-dose model) is locally fairly linear except for small doses. Figure 16 shows an example with the slope parameter ($\beta = 0.967194$) used here and with $d_{50} = 100$ CFU. For distributions of doses with small spread and with mean above about 25 CFU, approximation by the value of the dose-response model at the mean dose will be relatively accurate for this example. For small mean doses, however, the error may get large even with fairly small spread of doses. For example, for a Poisson distribution of doses with mean 25 CFU, the relative error is about 1%, whereas at a mean dose of 5 CFU the relative error in the estimate of the probability of illness grows to 35%. For wider distributions than the Poisson, the error can be large even at larger doses; for example, for a (discretized) lognormal distribution of doses with mean 100 CFU and $\ln(\text{GSD})$ of 1.0 (that is, a factor of 2.7), the error is about 40% in the probability of illness.

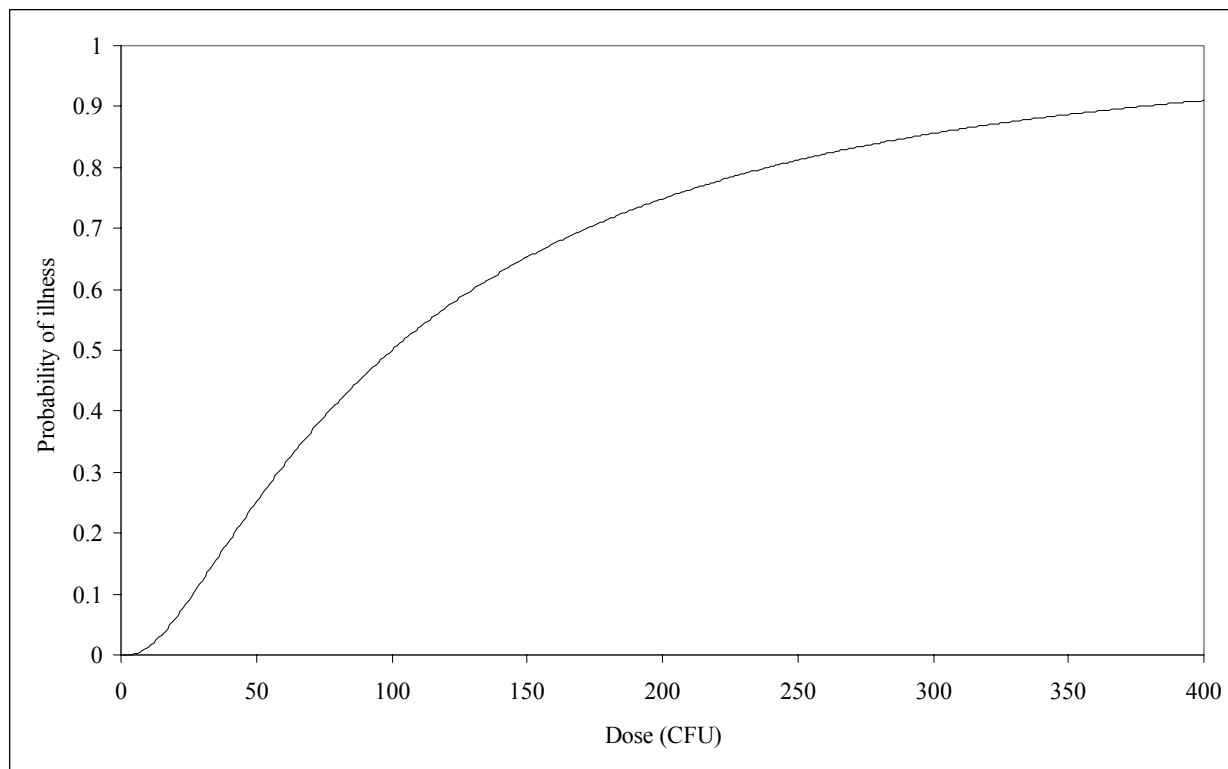


Figure 16 Alternative dose-response model (Equation (32)) with $S = 1$, $d_{50} = 100$ CFU.

Appendix E. Conceptual difficulties with FSIS (1998) and FAO/WHO (2002a,b)

FSIS (1998) used a dose-response model of beta-Poisson functional form, with parameter estimates derived from data on infection by surrogate species, *Shigella dysenteriae*. The parameter estimates were “adapted” for use for *Salmonella* spp. by using two versions of the dose-response model, one with a virulence parameter (the ID₅₀, the dose required to cause infection in 50% of the population exposed) reduced by a factor of 10, corresponding to an decrease by a factor of 10 of the dose required to cause a given response. In addition, uncertainty distributions were added to the ID₅₀ parameter estimate, based on the uncertainty bounds obtained for fitting *Shigella dysenteriae* data. The only comparison with epidemiological data on *Salmonella* was a graphical depiction of these two dose-response curves together with the point estimates of attack rate listed here in Table 3 (see Table E-3 and Figure E-24 of FSIS, 1998). It is clearly implied that the same dose-response function (*i.e.* the same functional form, and the same parameter values) apply to all the isolates obtained from *Salmonella* illness outbreaks. The uncertainty distributions added to the ID₅₀ parameters (one for a normal population, one for a sensitive population) of the dose-response curve appear to be intended to express the uncertainty of these parameter estimates.

FAO/WHO (2002b) similarly uses the approach of assuming the same dose-response function for all the outbreaks listed here in Table 4. Each point estimate of attack rate and dose given in Table 4 was accompanied by estimates of uncertainty, expressed as uncertainty distributions for dose, number of persons ill, and number of persons exposed (Table 26). No basis was provided for the choice of log-uniform (for doses) or Pert (for numbers ill and exposed) distributions. Uncertainty distributions for the parameters of the dose-response curve were obtained by a Monte Carlo procedure. To obtain each Monte Carlo sample for the parameters, each of the distributions shown in Table 26 was sampled, and the resulting samples fitted to a beta-Poisson model for probability of illness using maximum likelihood with the assumption of binomial sampling statistics for the number ill. No account appears to have been taken that some of the outcomes sampled were impossible for Case # 19,⁶² although this would have a small effect on the distribution of values obtained.

The principal difficulty with this approach is that none of the dose-response curves obtained in the Monte Carlo simulation adequately fit the data for that simulation, and the relative likelihood for the fits obtained in the simulation vary substantially. To illustrate, Figure 17 shows the fit obtained for central estimates of dose (on a logarithmic scale) and maximum likelihood estimates for the number ill and number exposed (*i.e.* the middle numbers of the Pert distribution specifications in Table 26). Also included on Figure 17 as error bars round each point are the full ranges of doses, and a range of probabilities of illness corresponding to the range of minimum to maximum ratio of ill to exposed shown in Table 26 augmented by 99% confidence limits on binomial samples (*e.g.* for Case # 22 the minimum is a 99% binomial lower confidence limit on 60/115, and the maximum and upper 99% binomial confidence limit on 66/104).

⁶² The minimum and maximum attack rates provided in Table 3.15 of FAO/WHO (2002b) appear to be those obtained among a particular set of 5,000 Monte Carlo samples, although that is not explicitly stated. The given maximum of 103.51% for Case # 19 is clearly impossible. The theoretical maximum attack rate obtainable for Case # 19 from the distributions shown would be 107.21%.

Table 26 Uncertainty distributions used in FAO/WHO (2002b)

Case #	Dose range (log-uniform)	Ill	Exposed
1	37–375	Pert(983,1035,1087)	8788
2	30–300	Pert(57,68,74)	205
3	15000–60000	12	Pert(14,17,20)
4	1150000–4600000	8	8
5	1130–11300	Pert(732,770,809)	1400
7	4–30	19	Pert(725,1813,3627)
11	10000–100000	39	39
12	10–235	Pert(29,30,32)	Pert(407,451,452)
13	1.00E+08–7.50E+08	7	7
18	135000–375000	3	5
19	1.08E+06–3.00E+06	Pert(107,113,119)	Pert(111,122,123)
20	495–1375	967	Pert(2638,3517,5276)
22	1057500–2937500	Pert(60,63,66)	Pert(104,114,115)
23	337500–937500	11	13
24	28–77	644	Pert(2660,3990,5320)
25	2278–6328	53	133
30	3375–9375	Pert(530,558,586)	Pert(797,884,885)
31	236–656	Pert(1302,1371,1440)	Pert(4593,5102,5103)
32	13–37	42	156
33	4.22E+09–1.17E+10	11	11

With the uncertainties specified by FAO/WHO (2002b), therefore, the four observations shown circled on Figure 17 would be essentially impossible if the fitted dose-response curve were correct. Moreover, it is clear that no single dose-response curve can adequately fit all these data for any sample from the specified uncertainty distributions; and this conclusion would remain true even if the uncertainty distributions for individual points were substantially wider. The same conclusion is even more apparent from the data used in this report (see Figure 4).

Thus both the FAO/WHO (2002b) and FSIS (1998) dose-response analyses are based on the untested hypothesis that a single dose-response curve is adequate to explain all the data; while testing that hypothesis shows it to be false.

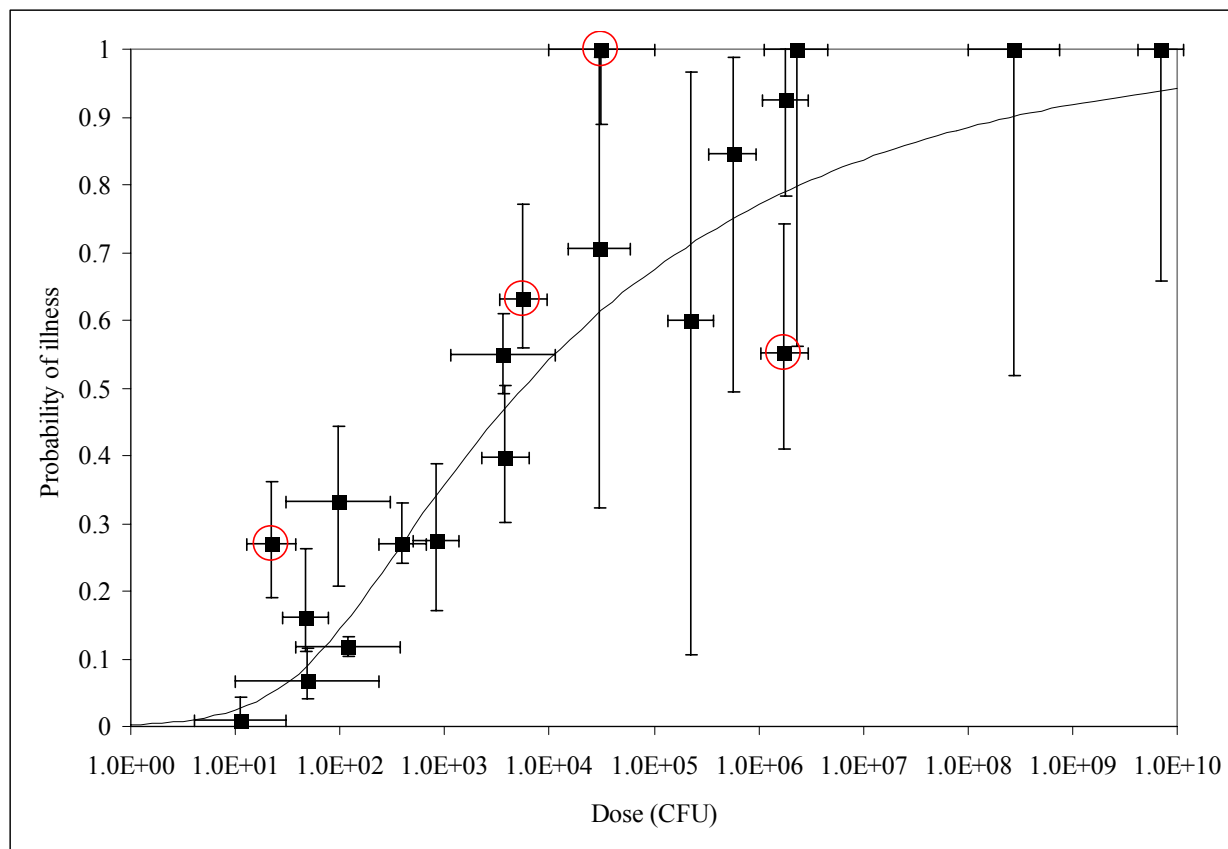


Figure 17 FAO/WHO (2002b) dose-response curve fit for central estimates, showing (circled) the non-fitting points.