

Model averaging in microbial risk assessment using fractional polynomials

Peer-reviewed author version

NAMATA, Harriet; AERTS, Marc; FAES, Christel & Teunis, P. (2008) Model averaging in microbial risk assessment using fractional polynomials. In: RISK ANALYSIS, 28(4). p. 891-905.

Handle: <http://hdl.handle.net/1942/8435>

Model Averaging in Microbial Risk Assessment Using Fractional Polynomials

Harriet Namata, Marc Aerts, Christel Faes

Hasselt University, Center for Statistics, Campus Diepenbeek, Agoralaan, Gebouw D,
B 3590 Diepenbeek, Belgium.

Peter Teunis

Dept IMA - private bag 86
National Institute of Public Health and the Environment, Antonie van Leeuwenhoeklaan 9
P.O. Box 1, 3720 BA Bilthoven, The Netherlands.

SUMMARY

The alleviation of food-borne diseases caused by microbial pathogens remains a great concern in order to ensure the well-being of the general public. The relation between the ingested dose of organisms and the associated infection risk can be studied using dose-response models. Traditionally a model selected according to a goodness-of-fit criterion has been used for making inferences. In this paper we propose a modified set of fractional polynomials as competitive dose-response models in risk assessment. The paper not only shows instances where it is not obvious to single out one best model but also illustrates that model averaging can best circumvent this dilemma. The set of candidate models is chosen based on biological plausibility and rationale and the risk at a dose common to all these models estimated using the selected models and by averaging over all models using Akaike's weights. In addition to including parameter estimation inaccuracy, like in the case of a single selected model, model averaging accounts for the uncertainty arising from other competitive models. This leads to a better and more honest estimation of standard errors and construction of confidence intervals for risk estimates. The approach is illustrated for risk estimation at low dose levels based on *Salmonella Typhi* and *Campylobacter jejuni* data sets in humans. Simulation studies indicate that model averaging has reduced bias, better precision and also attains coverage probabilities that are closer to the 95% nominal level compared to best fitting models according to AIC.

Keywords: Risk analysis; Dose-response models; Low-dose extrapolation; Model selection; Model uncertainty

1 Introduction

Human exposure to microorganisms such as yeasts, molds, bacteria, protozoa, helminths (worms) through food and/or water consumption dictates potential risk of food-borne infection or illness if such microbes survive in the human system. The dose exposed to is a measurement of microorganisms in relation to their ability to cause infection or illness. The unit of measurement of the dose is the colony forming unit (cfu). Ensuring food safety is directly proportional to a reduction of infectious or toxic food-borne pathogens. This process involves the identification of the microorganisms, how much of the organism an individual is exposed to and the risk associated with a given dosage. Such relations have been studied using various dose-response models to estimate risks in diverse range of problems. It is indeed very important to provide unbiased risk estimates including uncertainty as much as possible, in order to allow stakeholders (policy makers, producers, consumers) to assess risks as fairly as possible (given the current state of knowledge).

With regard to dichotomous responses, a dose-response model is a function describing the relationship between the dose administered and the probability of infection or disease. In general, any monotonic function which is bounded by zero and one is a possible function for such dose-response relationship. In the literature, several dose-response models have been proposed and they can be categorized into mechanistically and empirically oriented models. For example, Haas *et al.* (1999) developed a Beta-Poisson (BP) model from a mechanistic viewpoint where the biological process is readily taken into account. This model, owing to its biological plausibility, has received much attention in many pathogen dose-response studies (Teunis *et al.*, 1996). It should be noted however that several assumption are made in the biological process which can be questioned. Alternatively, so-called empirical models such as the log-logistic (LL), the log-normal (LN) and the extreme value (EV) models have also been used (Haas *et al.*, 1999). Whichever origin the models may have, the conventional way to making inferences has largely been dependent on a single chosen model based on some model selection criterion. Then, one typically proceeds as though that was the only model and thus discards the other possible models and model uncertainty. Using a single selected

model ignores variation that arises from other competing models and as a result leads to too small standard errors and narrow confidence intervals which are unrealistically optimistic (Burnham and Anderson 2002).

Instead of proceeding with one single model, one could argue that each model is a possible candidate dose-response model. Buckland *et al.* (1997) proposed a way of incorporating the uncertainty that arises from other competitive models by model averaging using Akaike Information Criterion (AIC) weights. The better a model, according to a certain selection criterion, the larger the weight given to this model. This approach is further discussed by Burnham and Anderson (2002). In the discipline of microbial risk assessment, model averaging has been employed. Bailer *et al.* (2005) accounted for model uncertainty for experimental studies of quantal responses using a bayesian approach and weighted the models using the bayesian information criterion (BIC). In the frequentist approach model averaging has been used by Kang *et al.* (2000) to estimate microbial risk using AIC and by Moon *et al.* (2005) to estimate effective microbial doses for infection and illness using Kullback Information Criterion (KIC). In both papers, a mix of the four aforementioned mechanistically and empirically oriented models is considered. In this paper, a further extension of dose-response modelling is investigated.

It is clear that a flexible dose-response model should be used to describe the data well. Typically, in the empirically oriented models, a linear dose trend is assumed, which might be too restrictive in the setting of microbial risk assessment. Therefore, we propose to extend the dose-response models by considering fractional polynomials within the family of empirical models. While Royston and Altman (1994) proposed the fractional polynomials as a pre-defined set of (generalized) linear models, we propose a modified set according to a biological rationale. To account for the model-uncertainty in the extended set of possible dose-response models, we estimate the risk to *Salmonella typhi* and to *Campylobacter jejuni* data sets by model averaging.

The rest of this paper proceeds with a set of plausible models considered in Section 2, a review of the model averaging approach in Section 3 followed by an application of the method

to *Salmonella typhi* and *Campylobacter jejuni* data in Section 4 and simulation studies in Section 5. A discussion and concluding remarks in Section 6 wind up the paper.

2 Microbial Dose-Response Models

The data for microbial risk assessment can be obtained experimentally where dose groups are known or as a result of outbreaks in which the exact ingested dose is not known but can be approximated. The extracted information to study, for instance, dichotomous dose-response relations involves a total number n_i of individuals or animals to which a particular dose d_i of microbes was administered and out of these it is observed that X_i become infected or ill. To translate this process statistically, X_i is assumed to follow a binomial distribution with parameters n_i and $\pi(d_i)$, the latter being the probability that a subject becomes infected (or ill). Dose-response models refer to models for $\pi(d)$ as a function of d .

Different dose-response models with one, two, and three parameters have been proposed and studied in microbial risk assessment literature (Kodell *et al.*, 2002; Moon *et al.*, 2005). The inclusion of biological processes gave rise to so-called mechanistic models of which the Beta-Poisson (BP) model is the most popular and extensively used one (e.g., Haas *et al.*, 1999). However, the adequacy of the BP model as automatic “default” model has been questioned (Marks *et al.*, 1998). Alternative models such as the so-called log-normal (LN), log-logistic (LL), and the extreme value (EV) model have been suggested (see e.g. Pinsky, 2000). These latter models are standard generalized linear models (GLM), used in statistics to analyse binary response data. At first sight, they however seem to lack any biological interpretation.

The purpose of this section is to indicate that many models, such as GLM’s, could have a biological interpretation, similar to that of the BP model. A basic concept in applied statistics states that the true or correct model is never known (or does even not exist) and that all dose-response models are wrong; they are merely approximations (e.g., Burnham and Anderson, 2002). So, the best one can do is to use a good approximating model. Now, depending on the setting and the willingness to rely on assumptions or on data or on both,

there are three options: i) to select the model in advance, possibly prior to data collection, only based on biological assumptions (e.g., the BP model); ii) to select a final model from a set of candidate models; and iii) to use as a final model a weighted average over all or some candidate models. For option ii) and iii) this set of candidate models may be (partly) inspired or based on biological knowledge and assumptions. In this paper we opt for the model averaging approach, since it incorporates explicitly the process of model selection and thus reflects model uncertainty. A crucial aspect of both options ii) and iii) is the use of a well-defined and rich enough set of candidate models. For that purpose, we consider the family of fractional polynomials (FP) (Royston and Altman, 1994) and propose a modified version of FP's, such that they obey some basic biological constraints. These models together with the BP, LN, LL, and EV model will define our set of candidate models.

2.1 A Generic Mechanistic Dose-Response Model

So-called mechanistic dose-response models reflect underlying biological processes involved in the kinetics of microorganisms in the body of a human or animal host, in order to determine the proportion that develop an adverse event, owing to exposure to a source containing infectious microorganisms (Haas *et al.*, 1999). The formulation of such models involves different subprocesses.

First of all, let $\pi_{ing}(j|d)$ be the probability of ingesting j organisms by an individual, from an exposure source of mean dose d . The second subprocess generates the event that k out of j organisms survive to initiate infection. Let $\pi_{surv|ing}(k|j, r)$ be the probability of such an event, where r is the probability of survival of a single pathogenic organism in a human host. The total probability that k organisms survive to initiate infectious foci is then given by

$$\pi(k|r, d) = \sum_{j=k}^{\infty} \pi_{ing}(j|d) \pi_{surv|ing}(k|j, r). \quad (1)$$

There is uncertainty about the minimal number of surviving organisms, k_{min} , that are needed to initiate infection. For $k \geq 1$, denote by $\pi_{inf,k}$ the probability that k surviving organisms initiate infection. The probability of infection, for a given mean dose d , can then be written

as

$$\pi_I(d) = \sum_{k=1}^{\infty} \left(\sum_{j=k}^{\infty} \pi_{ing}(j|d) \pi_{surv|ing}(k|j, r) \right) \pi_{inf,k}$$

which, when $\pi_{inf,k}$ equals 1 for $k \geq k_{min}$ and $\pi_{inf,k}$ is 0 for $k < k_{min}$, simplifies to

$$\pi_I(d) = \sum_{k=k_{min}}^{\infty} \sum_{j=k}^{\infty} \pi_{ing}(j|d) \pi_{surv|ing}(k|j, r). \quad (2)$$

One can expect different sources of heterogeneity in this approach. The variation in r between hosts and/or between pathogenic organisms can be described by a density $\pi_{var}(r)$ over the interval $[0,1]$, leading to the marginal distribution

$$\pi_I(d) = \int_0^1 \left(\sum_{k=k_{min}}^{\infty} \sum_{j=k}^{\infty} \pi_{ing}(j|d) \pi_{surv|ing}(k|j, r) \right) \pi_{var}(r) dr. \quad (3)$$

Model (3) can be considered as a generic mechanistic dose-response model. Different choices for π_{ing} , $\pi_{surv|ing}$, π_{var} and numerical values of k_{min} , and further assumptions, lead to different specific dose-response models. For example, taking π_{ing} to follow a Poisson distribution with mean d , a binomial distribution for $\pi_{surv|ing}$, a degenerate distribution for π_{var} (so taking r as fixed), and assuming all parameters in each of these components are different (no shared parameters), we get the dose-response model

$$\pi_I(d) = \Gamma_{k_{min}}(rd) \quad (4)$$

where $\Gamma_{k_{min}}(rd)$ is the cumulative distribution of an incomplete gamma distribution with parameter k_{min} , evaluated at rd . Taking $k_{min} = 1$ leads to the exponential model

$$\pi_I(d) = 1 - \exp(-rd). \quad (5)$$

In some situations, choices and assumptions can be verified separately, such as the Poisson assumption in dose verification studies (see e.g. DuPont *et al.*, 1995).

Choosing π_{ing} and $\pi_{surv|ing}$ and $k_{min} = 1$ as above, but now with π_{var} a beta density, the dose-response relation expressed as a complement of the confluent hypergeometric function becomes

$$\pi_I(d) = 1 - {}_1F_1(\alpha, \alpha + \beta, -d) \quad (6)$$

which Furumoto and Mickey (1967) approximated to the popular BP model

$$\pi_I(d) = 1 - \left(1 + \frac{d}{\beta}\right)^{-\alpha} \quad (7)$$

provided $\beta \gg \alpha$. Teunis and Havelaar (2000) demonstrate that the BP model can produce results similar to the hypergeometric relation, as long as the conditions given by Furumoto and Mickey (1967) are fulfilled. But it is also shown that the BP model might lead to completely different results, which can even be misleading in case very little information is available. See also Teunis *et al.* (2004).

From the general mechanistic model (3) many other models can be derived. Choosing π_{ing} as a Poisson random variable with mean $\mu(d)$, $k_{min} = 1$, π_{var} as the point mass distribution $I_{[r=1]}$, and $\pi_{surv|ing}$ as the point mass distribution $I_{[k=j]}$, we can recover the classical models and the fractional polynomials. Taking, respectively, $\mu(d) = \exp(\alpha + \beta \log d)$, $\mu(d) = \ln(1 + \exp(\alpha + \beta \log d))$ and $\mu(d) = \ln(1/(1 - \Phi(\alpha + \beta \log d)))$ leads to the EV, LL and LN models with Φ denoting the standard normal cumulative distribution function. In a similar way, by replacing the linear predictor in $\mu(d)$ the fractional polynomial models introduced in Section 2.2 can be reconstructed. Haas *et al.* (1999) and Kodell *et al.* (2002) motivated the biological plausibility of the LL, LN and EV model from a completely different angle, namely from their respective latent tolerance distribution.

We would like to emphasize that many other models can be derived from equation (3) and they can equally be useful in microbial risk assessment. The Poisson model and the binomial model for components π_{ing} and $\pi_{surv|ing}$ are based on an intrinsic assumption of independence and homogeneity. One could as well work with a zero-inflated Poisson, an overdispersed Poisson or even Poisson mixtures for π_{ing} , and a similar story holds for $\pi_{surv|ing}$. Obviously many other distributions on $[0,1]$ can replace π_{var} . Likewise, k_{min} can be a probability distribution as long as they are monotone increasing (since biologically that probability can not decrease as a function of k). In fact any cumulative distribution function on the integers, having value 0 at $k = 0$, can play the role. Many choices would however not lead to analytically tractable formula's. But any possible model, whatever choice of different distributions in (3), has some fundamental properties in common: no infection in the case

when no pathogenic organisms are ingested; the more organisms ingested, the higher the probability of infection; and an extremely high dose exposure always results in infection; or equivalently, assuming the model is a differentiable function of dose d with derivative $\pi'_I(d)$,

$$\lim_{d \rightarrow 0} \pi_I(d) = 0, \quad (8)$$

$$\pi'_I(d) \geq 0, \quad (9)$$

$$\lim_{d \rightarrow \infty} \pi_I(d) = 1. \quad (10)$$

Properties (8) to (10) are exactly the properties of a cumulative distribution function on $[0, \infty)$, and therefore they can always be written as a GLM (such as the exponential) or a generalized nonlinear model (such as the approximate BP model, with identity link and predictor $\left(1 + \frac{d}{\beta}\right)^{-\alpha}$). Other examples include again the LL, LN, EV and the fractional polynomial models (Section 2.2). In Section 4 we illustrate that, depending on the application, additional constraints can or should be added to the minimal set of properties (8) to (10).

A crucial concept in the theory of multimodel inference is the fact that a “correct model” does not exist. Any model is incorrect and merely tries to approximate the true process. It is rather a matter of selecting a good approximating model or, in the approach of multimodel inference, to average over a certain subset of good approximating models, assigning higher weights to better approximating models. In this philosophy it is obviously important to define a rich enough family of candidate models. Here, as indicated in Section 2.2, the family of fractional polynomials have shown to be a well-defined and rich family of models. All models considered in this paper as candidate models are listed in Table 1. They are the typical models used in the literature, the aforementioned BP, LL, LN, EV, extended with a new family of fractional polynomial models, as introduced in Section 2.2.

2.1.1 Infection versus Illness

Often the data refer to illness rather than to the infection status of the individual. This is the case for the *Salmonella typhi* data (Hornick *et al.*, 1970) in Table 2. These are experimental data on healthy adult volunteers, not exposed previously. Infection was not

reported separately, only illness, which was described as developing fever (higher than 103 °F) followed by headaches and abdominal pain. Nevertheless the BP model is typically used for analyzing such illness data. But since

$$P(\text{illness}|d) = P(\text{illness}|\text{infection}, d)\pi_I(d),$$

the BP model loses its direct biological interpretability for estimating the illness probability $P(\text{illness}|d)$. Moreover in most cases no data are available on $P(\text{illness}|\text{infection}, d)$ and little is known in order to build a biologically meaningful parametric model. In most cases however, one can assume the probability $P(\text{illness}|d)$ to share the same fundamental properties or constraints (8), (9) and (10), namely being monotone increasing from 0 to 1, when dose d varies from 0 to ∞ (see e.g. Teunis et al., 1999). This again, even more strongly, motivates the point of view that many statistical models are biologically plausible.

2.2 Fractional Polynomials

Other more flexible models like the fractional polynomials can be competitors of the BP, LN, LL and EV commonly used models in microbial risk assessment. Unlike conventional polynomials that take on positive integer powers (up to the degree being considered), FP's use powers from a predefined set, $\mathcal{P} = \{-2, -1, -0.5, 0, 0.5, 1, 2, \dots, \max(3, m)\}$, with m the degree of the FP (see below). In principle, other fractional (negative and positive) powers can be considered, but Royston and Altman (1994) illustrated that the above restricted set is sufficient for most practical purposes. The family of fractional polynomials has been shown to be useful in several other, somewhat related fields of application, see e.g. Faes *et al.*, (2003, 2006ab), Shkedy *et al.*, (2006), Hens *et al.*, (2007).

Fractional polynomials of degree m with powers $p_1 \leq p_2 \leq \dots \leq p_m$, and in the GLM framework for binary response data, are defined as

$$g(\pi(d)) = \beta_0 + \sum_{j=1}^m \beta_j H_j(d), \quad (11)$$

where for $j = 1, \dots, m$,

$$H_j(d) = \begin{cases} d^{p_j} & \text{if } p_j \neq p_{j-1}, \\ H_{j-1}(d) \log(d) & \text{if } p_j = p_{j-1} \end{cases}$$

with $p_0 = 0$ and $H_0(d) \equiv 1$. In (11), π denotes the probability on the adverse effect of interest (infection, illness) and g some link function (such as the logit or probit link). As shown by Royston and Altman (1994), FP's of degree m higher than 2 are rarely needed in practice.

The FP model (11) does not automatically satisfy properties (8) to (10). First of all, first order FP's seem to be less appropriate as a family of candidate models. For example, the model $g(\pi(d)) = \beta_0 + \beta_1 d^{p_1}$ with $p_1 > 0$, or with $p_1 < 0$ and $\beta_1 > 0$, does not obey property (8). Property (10) does not hold for $p_1 > 0$ and $\beta_1 < 0$, or for $p_1 < 0$. Similar issues for the model $g(\pi(d)) = \beta_0 + \beta_1 \log(d)^{p_1}$. Since no unambiguous or clear and consistent constraints on parameters β_k and powers p_1 guarantee properties (8) to (10), we do not further consider modifications of first degree FP's. Second degree FP's however can be modified to satisfy the fundamental properties as follows

$$g(\pi(d)) = \beta_1(\log(d+1))^{p_1} + \beta_2(\log(d+1))^{p_2} \quad \text{with} \quad \beta_1, p_1 < 0 \quad \text{and} \quad \beta_2, p_2 > 0 \quad (12)$$

for a given link function g . So, as compared to the original definition of FP's, there is no intercept, d is replaced by $\log(d+1)$ and coefficients and powers of both terms have to be opposite in sign. The first term on the rhs of (12) guarantees property (8), while the second one guarantees property (10), and both terms are automatically monotone. We will use the typical GLM links: the logit, the probit and the complementary log-log link. In the applications and the simulations in the next sections, model (12) will be fitted by constrained maximum likelihood, to ensure that $\beta_1, p_1 < 0$ and $\beta_2, p_2 > 0$.

An overview of all candidate models considered in this paper and the functions to derive them are displayed in Table 1 (for all admissible powers in the set \mathcal{P}). We include the “classical” models (Beta-Poisson, log-logistic, log-normal, and extreme-value) and the family (12) of modified fractional polynomials of degree 2 with logit, probit and complementary log-log link. Note that all models have the same degree of complexity (two parameters). This set of candidate models contains a total of $M = 40$ models: 4 classical models (BP, LL, LN, EV) and three times 12 FP models (for three different link functions: 3 negative powers, each combined with 4 positive powers).

Table 1: *Set of candidate dose-response models: BP (approximate Beta-Poisson), LL (log-logistic), LN (log-normal), EV (extreme-value), FPL (FP with logit link), FPN (FP with probit link), FPEV (FP with complementary log-log link). Φ denotes the standard normal cumulative distribution function. $\pi(d)$ refers to the illness or infection probability. The column Subset \mathcal{P} shows the admissible values for the powers p_1 and p_2 from the set \mathcal{P} , defining the FP candidate models.*

$\pi(d)$	Parameters	Subset \mathcal{P}	Model
$1 - (1 + d/\beta)^{-\alpha}$	$\alpha > 0, \beta > 0$		BP
$1/(1 + \exp[-(\alpha + \beta \log(d))])$	$\alpha < 0, \beta > 0$		LL
$\Phi(\alpha + \beta \log(d))$	$\alpha < 0, \beta > 0$		LN
$1 - \exp[-\exp(\alpha + \beta \log(d))]$	$\alpha < 0, \beta > 0$		EV
$1/(1 + \exp[-(\beta_1(\log(d+1))^{p_1} + \beta_2(\log(d+1))^{p_2})])$	$\beta_1 < 0, \beta_2 > 0$	$p_1 < 0, p_2 > 0$	FPL
$\Phi(\beta_1(\log(d+1))^{p_1} + \beta_2(\log(d+1))^{p_2})$	$\beta_1 < 0, \beta_2 > 0$	$p_1 < 0, p_2 > 0$	FPN
$1 - \exp[-\exp(\beta_1(\log(d+1))^{p_1} + \beta_2(\log(d+1))^{p_2})]$	$\beta_1 < 0, \beta_2 > 0$	$p_1 < 0, p_2 > 0$	FPEV

The SAS procedure NLMIXED has been used to fit all models (SAS code is available from the authors on request). To improve computational stability, the BP model is reparameterized in terms of mean and variance related parameters, using

$$\frac{e^u}{1 + e^u} = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad e^v = \alpha + \beta. \quad (13)$$

The approach used to calculate confidence intervals for the predicted probabilities was to calculate the confidence intervals on the scale of the underlying continuous variable (i.e., on the log-odds scale or the probit scale or the cloglog scale) and the resulting upper and lower confidence bounds are then converted into probabilities using the logistic distribution, the cumulative distribution function of the standard normal distribution or a Gompertz distribution. For the BP model, which does not use any of the three link functions, the confidence intervals for the estimated parameters are obtained using a logit-based transformation (shown in the next section).

3 Model Averaging Approach

This section recapitulates the necessary formulae for model averaging according to Burnham and Anderson (2002). The set of M plausible candidate models presented in Section 2 is fit

to the data and the probability of infection (or illness) $\pi(d)$ at some low dose d is estimated for each of the models. Let us denote the estimate for, or better conditional on, model m as $\hat{\pi}_m(d)$. The ‘unconditional’ estimate for the probability of infection is defined as (suppressing the specification of the dose level d)

$$\hat{\pi}_a = \sum_{m=1}^M w_m \hat{\pi}_m,$$

with weights

$$w_m = \frac{\exp(-\frac{1}{2}\Delta_m)}{\sum_{h=1}^M \exp(-\frac{1}{2}\Delta_h)}.$$

So $\hat{\pi}_a$ is a weighted average of the conditional estimates with weights quantifying the relative importance of the different conditional models. These weights are based on the AIC differences Δ_m , defined as the difference between the AIC value of model m and the AIC value of the ‘best’ model with the lowest AIC value.

From Buckland *et al.* (1997) the large sample approximation of the variance around the averaged risk estimate is taken as:

$$\widehat{\text{var}}(\hat{\pi}_a) = \left[\sum_{m=1}^M w_m \sqrt{\widehat{\text{var}}(\hat{\pi}_m) + (\hat{\pi}_m - \hat{\pi}_a)^2} \right]^2.$$

This variance estimator is clearly the sum of two components: the conditional sampling variance $\widehat{\text{var}}(\hat{\pi}_m)$ of $\hat{\pi}_m$ given model m and a term for the variation in the estimates across the M models $(\hat{\pi}_m - \hat{\pi}_a)^2$. The square root of this sum is then weighted by the weights w_m . This formulation is useful since it accounts for both the within and between model variability in estimating the variance of the averaged risk estimate.

In order to guarantee that the confidence interval for $\pi(d)$ is part of the eligible (0,1) range, we first construct a confidence interval $[l, u]$ for the logit transformed parameter $\text{logit}(\pi(d)) = \log(\pi(d)/(1 - \pi(d)))$ and next transform it back to the probability scale using the expit transformation, leading to $[e^l/(1 + e^l), e^u/(1 + e^u)]$. Since both transformations forth and back are monotone one-to-one, the coverage probability remains exactly the same (see also Burnham *et al.*, 1987). Using the delta-method to calculate the standard error of $\text{logit}(\hat{\pi}_a)$ for the construction of the interval $[l, u]$, the final large sample confidence interval for $\pi(d)$

Table 2: *Results of the dose-response experiment (Hornick et al., 1970) for Salmonella typhi Quailles in healthy human subjects. Dose: number of organisms ingested. Total: number of subjects at a given dose. Ill: number of subjects with symptoms of typhoid fever.*

Dose (cfu)	Total	Ill
10^3	14	0
10^5	116	32
10^7	32	16
10^8	9	8
10^9	42	40

can be written as

$$\left[\frac{\hat{\pi}_a}{\hat{\pi}_a + (1 - \hat{\pi}_a)C}, \frac{\hat{\pi}_a}{\hat{\pi}_a + (1 - \hat{\pi}_a)/C} \right] \quad \text{with} \quad C = \exp \left[\frac{z_{\alpha/2} \widehat{\text{se}}(\hat{\pi}_a)}{\hat{\pi}_a(1 - \hat{\pi}_a)} \right].$$

4 Application to Data

In this section, we illustrate the use of the 36 modified fractional polynomials (12) along with the four classical models and their model average in order to estimate the risk of illness or infection. In a first study, we use data from Hornick *et al.*, (1970) where the volunteers ingested wild-type *Salmonella typhi* in 45 mL of milk. When clinical illness developed, the volunteers were promptly treated with effective antibiotics to which the challenge strain was highly susceptible. The data are presented in Table 2. It includes the dose of pathogenic *Salmonella typhi*, the total number of individuals exposed and the individuals who eventually became ill. Thus, in the first study we investigate the probability of illness due to *Salmonella typhi* in relation to the ingested dose. A second study, investigates the risk for infection due to *Campylobacter jejuni* (Black *et al.*, 1988), with the data presented in Table 3. It includes the dose of *Campylobacter jejuni*, the total number of individuals exposed, the individuals infected and the individuals who eventually became ill. Both data sets are also discussed in Teunis *et al.* (1996).

Table 3: *Results of the dose-response experiment (Black et al., 1988) for Campylobacter jejuni in healthy volunteers. Dose: ingested number of C. jejuni A3249. Total: number of subjects exposed to a given dose. Infected: number of subjects infected (excretion of C. jejuni). Ill: number of subjects with gastro-enteric symptoms (fever, vomiting, diarrhea).*

Dose (cfu)	Total	Infected	Ill
8×10^2	10	5	1
8×10^3	10	6	1
9×10^4	13	11	6
8×10^5	11	8	1
1×10^6	19	15	2
1×10^8	5	5	0
1×10^8	4	4	2

4.1 Salmonella Typhi

Table 4 shows, for *Salmonella typhi*, the estimated risks at a dose of 100cfu, their standard errors and 95% confidence intervals for 9 out of the 40 fitted models to keep the size of the table reasonable. The BP, LL, LN and the EV models as well as the 5 best fitting FPs are shown. The estimated probability of illness due to *Salmonella typhi* ranges from 2.1×10^{-9} to 0.01986. The question is on which model inference should be based. Using the Akaike information criterion, the relative importance of each model is calculated. One way is to use the estimates from a model that has AIC-weight greater or equal to 0.9 (Haas *et al.*, 1999). However, none of our models meets that criterion and it is unlikely in reality that such a model would be found. For this data example, the 5 best fitting FPs have somewhat higher weights but still far below 0.9. Instead of selecting one final model, model averaging based on all available models (or a selection) can be used. The averaged risk estimate is shown on the last line of Table 4 together with its standard errors and 95% Wald confidence intervals. The confidence intervals are wider than for individual models but this is expected because model averaging incorporates variability between competing models. This uncertainty indicates the importance of model averaging, especially at low doses where we do not have data. This uncertainty is also clearly visualised in Figure 1.

Table 4: *Salmonella typhi*. Estimated probability of illness at dose 100cfu. For the Beta-Poisson and the model averaged(MA) estimates logit-back transformed CIs are used while the other model estimates use CIs back transformed according to their corresponding link function. Fractional polynomials are denoted as $FPlink_{(powers)}$.

Model	AIC	Model			wald CIs	
		Weights	$\hat{\pi}_i(100)$	$se(\hat{\pi}_i(100))$	lower	upper
FPEV _(-2,3)	20.43	0.07474	0.00010	0.00012	1.0E-05	0.00103
FPN _(-2,3)	20.93	0.05811	2.1E-09	1.1E-08	2.4E-14	1.3E-05
FPEV _(-1,3)	21.12	0.05294	0.01986	0.00935	0.00787	0.04969
FPN _(-1,3)	21.17	0.05160	0.00489	0.00492	0.00055	0.02851
FPEV _(-1,2)	21.22	0.05026	0.00964	0.00532	0.00326	0.02829
EV	22.10	0.03246	0.04779	0.01626	0.02442	0.09242
LN	22.79	0.02294	0.01159	0.00852	0.00239	0.04275
LL	23.79	0.01388	0.02329	0.01128	0.00894	0.05930
BP	26.00	0.00461	0.00070	0.00030	0.00030	0.00162
MA estimate			0.01285	0.01603	0.00109	0.13404

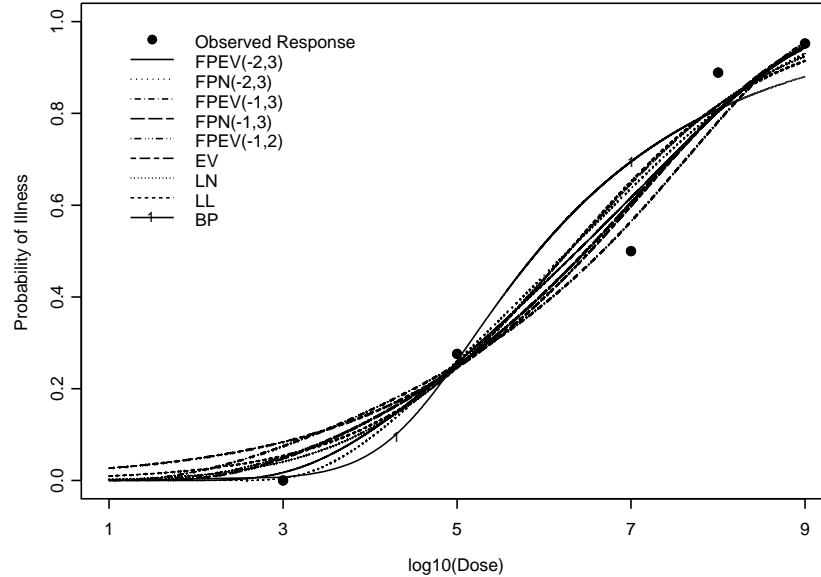


Figure 1: Estimated probabilities of illness for the *Salmonella typhi* data of Table 2.

4.2 Campylobacter Jejuni

Applying model averaging over all 40 models on the *Campylobacter jejuni* data example, we obtained an averaged risk estimate, at the dose=10cfu, of 0.21281 with a standard error equal to 0.29038 and confidence intervals from 0.00896 to 0.88986. The five best fitting models extended with the “classical” models BP, LL, LN and EV (if not included in the top 5) are shown in Figure 2(a). This figure shows some curves with a peculiar pattern left from the data range. They can be characterised by a steep increase at low dose levels. Based on biological knowledge or expertise such models might be considered as not plausible. Extending the minimal set of criteria (8), (9) and (10), such models might be excluded from the initial set of candidate models as follows.

In addition to (8), (9) and (10), the constraint

$$\pi'_I(d) \leq C, \text{ for } d \text{ in some low dose range,} \quad (14)$$

controls the increase of the models in the low dose range by excluding all models which derivatives exceed a certain threshold C in the range. The choice of C depends on the particular application and should be governed by additional biological expertise. Here we illustrate this idea by applying the threshold $C = 0.2$ on the dose range from 1e-20 to 5 cfu (transformed to log base 10 scale). As a result 14 models are excluded and averaging is restricted to 26 models. Figure 2(b) shows the fitted curves of the five best fitting models, extended with the BP, LN, LL and EV model. Since now the EV model is ranked on the fourth position, there is one curve less as compared to Figure 2(a)).

Table 5 shows the estimated risks at a dose of 10cfu, their standard errors and 95% confidence intervals for *Campylobacter jejuni*. Eight models, the five best fitting models and the classical models, out of the 26 fitted models after the rule-out are shown. In this example, only small differences in AIC are seen, and all models get about the same weight. The probability of infection due to *Campylobacter jejuni* at a dose of 10cfu ranges from 0.0137 to 0.3236. By model averaging we obtain an averaged risk estimate of 0.089, which is about two times less than the averaged risk estimate over all 40 models. The confidence interval based on the

Table 5: *Campylobacter jejuni*. Estimated probability of infection at dose 10cfu. For the Beta-Poisson and the model averaged(MA) estimates logit-back transformed CIs are used while the other model estimates use CIs back transformed according to their corresponding link function. Fractional polynomials are denoted as $FPlink_{(powers)}$.

Model	AIC	Weights	$\hat{\pi}_i(10)$	$se(\hat{\pi}_i(10))$	Wald CIs	
					lower	upper
FPEV _(-2,3)	19.8482	0.05985	0.04804	0.13982	0.00014	1.00000
FPEV _(-2,2)	20.1956	0.05031	0.01373	0.04392	2.5E-05	0.99951
FPN _(-2,2)	20.2796	0.04824	0.04627	0.28467	5.1E-14	0.99998
EV	20.3201	0.04727	0.32355	0.14222	0.12735	0.67428
FPEV _(-0.5,1)	20.4400	0.04452	0.20325	0.15167	0.04303	0.69081
LN	20.5414	0.04232	0.26749	0.17085	0.05072	0.65434
LL	20.6902	0.03929	0.26161	0.16551	0.06198	0.65514
BP	20.8785	0.03576	0.08640	0.16325	0.00164	0.84489
MA estimate			0.08900	0.15178	0.00248	0.79301

model average is still very wide [0.00248,0.79301] (averaging over all 40 models led to [0.00896, 0.88986]), but it has to account for the (inevitable) high degree of model uncertainty when extrapolating at low dose levels.

5 Simulation Study

In a small simulation study we explore and compare the performance of the estimated risk based on three model approaches: each of the 40 candidate models individually, the best fitting model according to AIC (varying from run to run), and by model averaging. The model chosen most often by AIC across the simulations was also monitored.

A total of $S = 1000$ datasets were generated. For each run, we assume the same dose levels (d) and total number of individuals exposed (n) as those in the *Salmonella typhi* data set, and generate the number of ill individuals based on the BP model or based on a fractional polynomial EV model, with parameters based on the estimates from these models

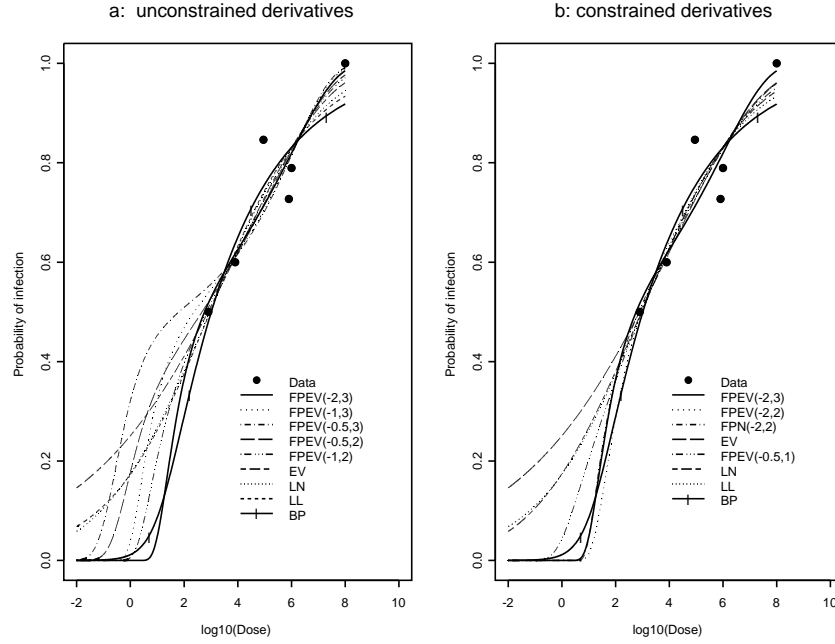


Figure 2: *Estimated probabilities of infection for the Campylobacter jejuni data of Table 3. Left panel: the fitted curves of the classical models (BP, LL, LN, EV) together with the five best fitting FP's for the full set of 40 candidate models. Right panel: the fitted curves of the five best models together with the classical models (BP, LL, LN) for the reduced set of 26 candidate models. The reduced set is based on the exclusion criterion: derivatives in the low dose region from dose=5cfu and below should not exceed 0.2.*

fitted on the *Salmonella typhi* data. In the first setting, the data sets were generated using a BP model taking parameters $u = -11.8739$ and $v = 10.2810$ in formulas (13). In the second setting, the data sets were generated using the cloglog fractional polynomial with powers $(p_1, p_2) = (-1, 3)$, as one of the best models in the *Salmonella typhi* example, with parameters $\beta_1 = -18.1425$ and $\beta_2 = 22.5300 \times 10^{-5}$. The dose response curves corresponding to both settings are shown in Figure 1. To the 1000 generated samples, we fit the set of 40 candidate models described earlier, and estimate the risk to *Salmonella typhi* illness π at dose=100cfu, for each of the candidate models but also keeping track of the best fitting model according to AIC. Finally, an averaged risk over the candidate models is calculated. We summarize the performance of the different methods by reporting, across the simulations, the average variance of the estimated probability of *Salmonella typhi* illness $\overline{(se(\hat{\pi}_*(100)))^2}$ using model approach *, the average length of the 95% confidence intervals and the coverage probability of these intervals. The variability of the risk estimate $\hat{\pi}_*$ about its average risk $\bar{\pi}_* = \sum_{s=1}^S \hat{\pi}_{*s}/S$ over the S simulations was calculated as $\hat{\sigma}^2(\hat{\pi}_*(100)) = \sum_{s=1}^S \{\hat{\pi}_{*s} - \bar{\pi}_*\}^2/(S-1)$, while $bias = \{\bar{\pi}_* - \pi\}$ gives the difference between the average risk estimate and the true risk based on the data generating model. The mean-squared error is given by $MSE = bias^2 + \hat{\sigma}^2(\hat{\pi}_*(100))$.

5.1 First Setting

Table 6 shows the results of the first setting, which uses the BP model as the true model. Not all 40 models are shown but the BP, LL, LN and EV, and the most frequently selected best fitting model. For this latter model it was observed which model was selected most often by AIC after all runs were finished. The results from the best fitting models (BFM) and by model averaging (MA) are shown as well. Note that the BFM changes from run to run. Estimates based on the BP model have very small bias and variance characteristics, as compared with the other models, and a coverage close to 95%, also illustrated by Figure 3. Since this model is the true simulation model, this is not unexpected. But the extremely poor behaviour of most other models is surprisingly low. Further note that the performances of the LL, LN and EV model are quite similar, indicating that the choice of link function

Table 6: Simulation I results for estimated risk at dose 100cfu. \dagger represents the model considered for data generation while \ddagger shows the most frequently selected best AIC model. BFM denotes the best fitting models and MA denotes model averaging.

Model	$(se(\hat{\pi}_*(100)))^2$	Cllength	coverage	(sign)bias ²	$\hat{\sigma}^2(\hat{\pi}_*(100))$	MSE
\dagger BP	2.2E-07	0.00172	95.60	(+)7.2E-09	1.5E-07	1.6E-07
LL	0.00020	0.05833	0.00	(+) 0.00079	0.00017	0.00097
LN	0.00017	0.05206	5.50	(+) 0.00031	0.00013	0.00045
EV	0.00042	0.08205	0.00	(+) 0.00399	0.00037	0.00437
\ddagger FPL _(-2,0.5)	1.5E-12	1.3E-05	0.00	(-) 4.8E-07	2.8E-13	4.8E-07
BFM	9.2E-06	0.00398	6.60	(+) 8.5E-07	3.8E-05	3.8E-05
MA	7.8E-05	0.08390	87.10	(+) 1.9E-05	2.2E-05	4.1E-05

has only minor influence on the estimated risk. Figure 3 helps to understand these results. The order of the simulation runs on the horizontal scale corresponds to the order of the corresponding point estimates. So runs leading to smaller estimates are more to the left side of the horizontal scale (as shown by the solid curve of estimates). The right upper panel of Figure 3 shows that the LN tends to overestimate the true value. The BFM (lower right panel of Figure 3) estimates show smaller bias and smaller variances, showing in some way its adaptive nature, but it also tends to underestimate the true value for most of the runs. At the right end of the horizontal scale, it exhibits an extremely high variable pattern. This all combines to a very low coverage of 6.6%. The model which was chosen most often as the best model is the fractional polynomial logit-model with powers (-2,0.5). Also this model performs surprisingly poorly in terms of coverage. The averaged risk estimate, based on all considered models, shows small bias and but relatively high variance properties. But this model accounts for the variability introduced by the model selection procedure, resulting in wider confidence intervals and a coverage probability of 87% (see Figure 3). Actually this is also a somewhat disappointing result, but compared to all other misspecified models it is outstanding.

5.2 Second Setting

In a second setting, we study the performance of the different models and of the model selected and averaged risk estimates again, now with the $\text{FPEV}_{(-1,3)}$ model being the true underlying simulation model. Results are summarized in Table 7. In this setting, the MSE of the BP model is no longer the smallest, and also the coverage probability is very small as it underestimates the true value most of the times (see left upper panel of Figure 4). Again, as expected, the data generating model $\text{FPEV}_{(-1,3)}$ has the smallest MSE and its coverage probability nicely reaches the nominal as expected (see right upper panel of Figure 4). The best fitting model has an average bias similar to that of the BP model, but again, an additional source of variability enters the estimation process. The fractional polynomial EV model with powers $(-0.5,3)$ is chosen most often as the best model but as in the previous setting it exhibits a very low coverage, partly also reflecting a similarly low coverage for the BFM. The averaged risk estimate has a small bias combined with a larger variance leading to a coverage probability of 97.90.

Both simulation settings show that model averaging has a beneficial effect in reducing bias, in accounting for the variability induced by the model selection process, and consequently in better coverage characteristics. In the philosophy that a true and correct model does not exist, or that you will never know it exactly (which seems the only realistic situation), this simulation also shows that not just one single model is appropriate to describe the dose-response relationship of microbial risks, but there exists a whole set of possible models. Fractional polynomials are very flexible to estimate the low-dose risk. However, the model selection procedure induces extra variability that should be accounted for. The averaged risk estimate gives larger weights to better-fitting models, resulting in a smaller bias. The model selection uncertainty is accounted for in this approach, and a better coverage probability obtained as illustrated in Figure 4. However, from both simulation settings it became evident that for this particular risk assessment application the coverage percentages for most of the model fall far below the chosen level of confidence (0.95). The magnitude of inflation of the coverage was certainly surprising to us and is in itself a clear warning to be very careful

Table 7: Simulation II results for estimated risk at dose 100cfu. † represents the model considered for data generation while ‡ shows the most frequently selected best AIC model. BFM denotes the best fitting models and MA denotes model averaging..

Model	$(se(\hat{\pi}_*(100)))^2$	Clength	coverage	$(\widehat{\text{sign}}\text{bias})^2$	$\hat{\sigma}^2(\hat{\pi}_*(100))$	MSE
BP	1.5E-06	0.00309	2.40	(-)0.00036	1.1E-06	0.00036
LL	0.00015	0.05118	95.00	(+)2.0E-05	0.00013	0.00015
LN	0.00011	0.04357	89.70	(-)3.2E-05	9.4E-05	0.00013
EV	0.00028	0.06764	34.50	(+)0.00077	0.00026	0.00104
$^\dagger\text{FPEV}_{(-1,3)}$	0.00011	0.04284	95.60	(+)8.7E-07	9.9E-05	9.9E-05
$^\ddagger\text{FPEV}_{(-0.5,3)}$	0.00049	0.08874	4.50	(+)0.00281	0.00047	0.00328
BFM	0.00023	0.05146	35.20	(+)0.00022	0.00101	0.00123
MA	0.00048	0.12179	97.90	(+)0.00002	0.00025	0.00028

and thoughtful when restricting the analysis to one single model. It certainly motivates the use of different plausible models, at least as a type of sensitivity analysis, but even better in a multimodel approach such as model averaging.

5.3 To include fractional polynomials or not

As pointed out before, and clear from the nature of the model averaging approach, the set of candidate models should be rich enough. This section tries to illustrate this point by comparing the results from the data analyses in Section 4 and the simulation results in the previous sections to the corresponding results of model averaging over the four classical models only. Let us denote by set one the restricted set of classical models BP, LL, LN and EV and by set two the modified flexible fractional polynomials added to set one. For *salmonella typhi* example, the averaged risk estimates are 0.029 (se 0.021 and confidence interval [0.007,0.114]) and 0.01285 (se 0.016 and confidence interval [0.001,0.134]) respectively for set one and two. The risk estimate from set two is reduced to about half that of set one. For *campylobacter jejuni* example, the averaged risk estimates are 0.2429 (se 0.180 and confidence interval [0.045,0.686]) and 0.089 (se 0.152 and confidence interval [0.002,0.793]) for set one and two respectively. Including fractional polynomials reduces the estimate to

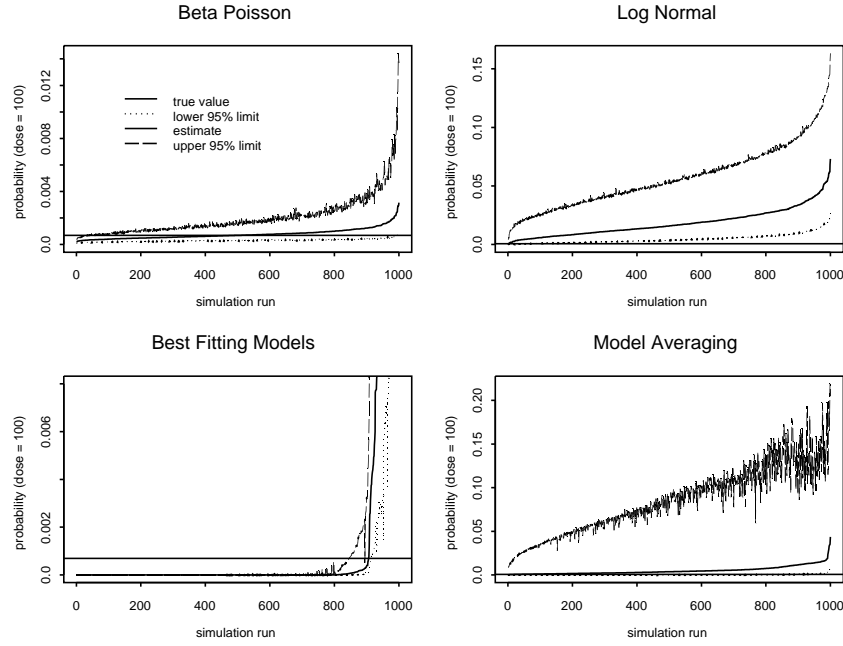


Figure 3: *Simulation I: True probability of *S. Typhi* based on the Beta Poisson generating model and the estimated probabilities for the Beta Poisson (data generation model), the log-normal model, the best fitting models across simulations and the averaged risk. The corresponding confidence limits of the estimated risks are shown to illustrate how best they capture the true value.*

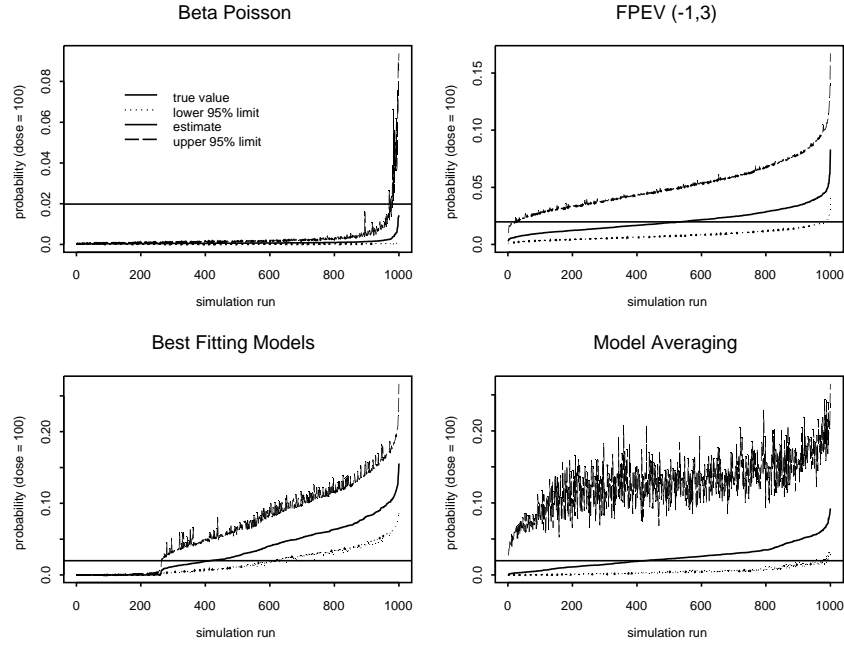


Figure 4: *Simulation II: True probability of $S. Typhi$ based on $FPEV(-1,3)$ the generating model and the estimated probabilities for the Beta Poisson (taken as standard model), the $FPEV(-1,3)$ model, the best fitting models across simulations and the averaged risk. The corresponding confidence limits of the estimated risks are shown to illustrate how best they capture the true value.*

about 2.5 times from not including them.

For the first simulation setting, with the restricted set one, the true risk estimate was captured in 38.5% of the runs by the confidence intervals of the averaged risk compared to 87.1% for set two. Also, the square bias and the variance were higher for set one models than for set two models. In the second simulation setting, averaging over the four models attained a coverage probability of 87.3% versus 97.9% for set two. To this end, we see that using the richer set two, which includes the flexible fractional polynomial models, yielded coverage probabilities closer to the nominal 95% level, less biased and more precise risk estimates than set one.

6 Discussion

In quantitative risk assessment for microbial pathogens, dose-response assessment is a critical issue. It must be included, health effects are the end-point in any risk assessment. Data are scarce, however. Only few pathogens have been used in clinical studies, and even these have been done on a small scale with few volunteers exposed per dose, due to the high costs involved. As a consequence, microbial dose-response data often do not contain a great deal of information on the shape of the dose-response relation, and model choice based entirely on experimental data is not feasible (Teunis and Havelaar 2000). This problem has raised many discussions and several solutions have been proposed. One could use animals, as these allow, at least in principle, collection of more extensive data sets. However, this paper also has demonstrated that this may merely give a better insight into the variability involved, not necessarily decreasing the uncertainty about the shape of the relation. Model averaging has been used before, weighting over an arbitrary collection of mathematical relations considered useful. Reasons for inclusion may range from biological plausibility to flexibility or mere tradition (like threshold models). Here it has been shown that fractional polynomials can be derived from the general equation (3) thereby potentially extending the family of eligible models a great deal, and providing a natural choice for the collection of models to include in model averaging.

As observed in the simulation study, the coverage probabilities of the confidence intervals of the individual models, the best fitting models and the model average highly depend on the generating (true) model. When bias values are high relative to the variance, the coverage of these intervals considerably fall below the nominal level of confidence. The bias values for the averaged risk were 10 times higher when only averaging over the four commonly used dose-response models than when fractional polynomials were included and as a result the coverage probabilities in the former case were always below nominal. Our findings corroborate with the findings of Wheeler and Bailer (2007). Though Wheeler and Bailer (2007) employed model averaging based on the “average-model” versus the “average-risk” method that we use, they pointed out the importance of an adequate model space. In their extensive simulation study involving the 3-model space of flexible models (the weibull, the multistage and the log-probit) and the 7-model space covering the quantal linear, quantal quadratic, the logistic and the probit models in addition to those in the 3-model space, they conclude a better performance, in terms of bias and coverage across most simulations conditions, for the three-model space than for the seven-model space. Therefore, the consideration of more flexible models in applying model averaging to risk assessment is an issue that can be extended with the set (or biologically more plausible subset) of the modified fractional polynomials that we propose in this paper.

An essential improvement in dose-response assessment is its extension to a hierarchical framework: like most biological problems, data can frequently be organized hierarchically. For example several isolates (pure strains of microorganisms separated from a mixed bacterial culture) of a single pathogen species in the same host species, different but related pathogen species in the same host, or a single pathogen isolate in hosts with different levels of immunity. As the purpose of risk assessment usually is to predict risks for an exposed population, dose-response relations should be translated from the special, experimental setting to a more general, unspecified situation. For instance, given the different responses to a limited collection of isolates of a pathogen, what would be the response to a newly isolated specimen, of the same type? Fixed effects models are easy to implement and represent a first solution, but the random models, although more complex, use all available data and are more suitable for

explicative studies. The generalized linear mixed models (Agresti, 2002; Molenberghs and Verbeke, 2005) allow prediction by using the (joint) distribution of the random parameters to make exactly this generalization. Our future interest is to extend these proposed fractional polynomials to investigate dose-response relationships between isolates of a single pathogen using generalized linear mixed models.

Further research is needed to go into some issues, like the estimation at very low dose levels, especially in case of limited information. Since low dose extrapolation is crucial in risk assessment and the AIC method selects a model regardless of its intended use, we wish to further investigate selecting a model looking at the (lower dose) region of interest, by using e.g. the focused information criterion (Claeskens and Hjort 2003).

7 Acknowledgments

This study has been carried out with the financial support of the Research Project BOF04N03 of Hasselt University. The authors also gratefully acknowledge the financial support from the IAP research Network P6/03 of the Belgian Government (Belgian Science Policy). Marc Aerts acknowledges support from FWO-Vlaanderen Research Project G.0151.05.

8 References

- Agresti, A., 2002. Categorical Data Analysis. Second edition. Wiley, New York.
- Bailer, A.J, Noble, R.B., and Wheeler, M.W. (2005). Model uncertainty and risk estimation for experimental studies of quantal responses. *Risk Analysis* **25**, 292–299.
- Black, R.E., Levine, M.M., Clements, M.L., Hughes, T.P., and Blaser, M.J. (1988). Experimental *Campylobacter jejuni* infections in humans. *Journal of Infectious Diseases*, **157**(3), 472 – 479.

- Buckland, S. T, Burnham, K. P, and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics*, **53**, 603–618.
- Burnham, K. P, and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Second Edition. Springer Verlag, New York.
- Burnham, K. P., Anderson, D. R., White, G. C., Brownie, C., and Pollock, K. H. (1987). Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society, Monograph **5**.
- Claeskens, G. and Hjort, N.L. (2003). The Focussed Information Criterion, *Journal of the American Statistical Association*, **98**, 900-916.
- DuPont, H. L., Chappell, C. L., Sterling, C. R., Okhuysen, P. C., Rose, J. B. and Jakubowski, W. (1995). The infectivity of *Cryptosporidium parvum* in healthy volunteers. *The New England Journal of Medicine*, **332(13)**, 855-859.
- Faes, C., Geys, H., Aerts, M. and Molenberghs, G. (2003). On the use of fractional polynomial predictors for quantitative risk assessment in developmental toxicity studies. *Statistical Modelling*, **3**, 109–126.
- Faes, C., Hens, N., Aerts, M., Shkedy, Z., Geys, H., Mintiens, K., Laevens, H. and Boelaert, F. (2006a). Random-effects models for clustered binary data using monotone fractional polynomials; with application to the estimation of a herd-specific force of infection. *Applied Statistics*, **55**, 595-613.
- Faes, C., Aerts, M., Geys, H., and Molenberghs, G. (2006b). Model Averaging using Fractional Polynomials to Estimate a Safe Level of Exposure. *Risk Analysis*, **27**, 111–123.
- Furumoto, W. A, and Mickey, R. (1967). A mathematical model for the infectivity-dilution

curve of tobacco mosaic virus: Theoretical considerations. *Virology*, **32**, 216–223.

Haas, C. N, Rose, J. B, and Gerba, C. P. (1999). *Quantitative Microbial Risk Assessment*. Wiley, New York.

Hens, N., Faes, C., Aerts, M., Shkedy, Z., Mintiens, K., Laevens, H. and Boelaert, F. (2007). Handling missingness when modelling the force of infection from clustered zeroprevalence Data. *Journal of Agricultural, Biological and Environmental Statistics*, **12**, 1–16.

Hornick, R.B., Greisman, S.E., Woodward, T.E., DuPont, H.L., Dawkins, A.T., and Snyder, M.J. (1970). Typhoid fever: pathogenesis and immunological control. *The New England Journal of Medicine* **283(13)**, 686 – 691.

Kang, S., Kodell, R. L, and Chen, J. J. (2000). Incorporating Model uncertainties along with Data Uncertainties in Microbial Risk Assessment. *Regulatory Toxicology and Pharmacology*, **32**, 68–72.

Kodell, R.L., Kang, S., and Chen, J.J (2002) Statistical models of health risk due to microbial contamination of foods. *Environmental and Ecological statistics*, **9**, 259–271.

Marks, H.M., Coleman, M.E., Lin, C.J., and Roberts, T. (1998) Topics in microbial risk assessment;dynamic flow tree process. *Risk Analysis*, **18**, 309–328.

Molenberghs, G., and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer Verlag, New York.

Moon, H., Kim, H., Chen, J. J, and Kodell, R. L. (2005). Model averaging using Kullback information criterion in estimating effective doses for microbial infection and illness. *Risk Analysis*, **25(5)**, 1147–1159.

- Pinsky, P.F. (2000) Assessment of risk from long term exposure to waterborne pathogens. *Environmental and Ecological statistics*, **7**, 155–175.
- Royston, P, and Altman, D. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). *Applied Statistics*, **43**, 429–467.
- Shkedy, Z., Aerts, M., Molenberghs, M., Beutels, Ph. and Van Damme, P. (2006). Modeling age-dependent force of infection from prevalence data using fractional polynomials. *Statistics in Medicine*, **25**, 1577–1591.
- Teunis, P. F. M., and Havelaar, A. H. (2000). The Beta Poisson model is not a single hit model. *Risk Analysis*, **20(4)**, 511–518.
- Teunis, P. F. M., Nagelkerke, N. J. D., and Haas, C. N. (1999). dose-response models for infectious gastroenteritis. *Risk Analysis*. **19(6)**, 1251–1260.
- Teunis, P. F. M., Takumi, K. and Shinagawa, K. (2004). dose-response for infection by *Escherichia coli* O157:H7 from outbreak data. *Risk Analysis*, **24(2)**, 401–408.
- Teunis, P. F. M., Van der Heijden, O. G, Van der Giessen, W. B, and Havelaar, A. H. (1996). The Dose-Response Relation in Human Volunteers for Gastro-intestinal Pathogens, Report Nr. 284550002. National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
- Wheeler, M.W., and Bailer, A.J. (2007). Properties of Model-Averaged BMDLs: A study of Model Averaging in Dichotomous Response Risk Estimation. *Risk Analysis*, **27(3)**, 659–670.