Quantitative Methods for Food and Diet Safety Assessment

Lutz Edler

Biostatistics Unit-R0700, German Cancer Research Center
Im Neuenheimer Feld 280
D-69120 Heidelberg, Germany
edler@dkfz.de

1. Introduction

Ingredients in food have continuously raised concerns in the public about the safety of food and diet, even if their amount is very small. Therefore, measures to ensure food and diet safety have to be established and risk assessment (RA) of food is undertaken to determine the probability of the occurrence adverse (harmful) effects. Methods for the RA of food are challenged by different high level issues which stress the complexity of the problem: (1) Nature and composition of food and diet we consume influence health. Dietary factors have been postulated as major risk factors for the occurrence of diseases. (2) Human food is composed of many components and contains many ingredients. Each of them may require different consideration. (3) Food producers and regulators are succumbed to the public demand to guarantee the presence of no risk ("zero risk"). This has lead to the claim to determine "safe" levels for chemicals in food and the banning of food which can not comply with this demand. (4) Food is daily consumed and needed for maintenance of human functions and exposure to food as an exogenous "agent" is unavoidable. Therefore disadvantageous effects from food intake have to be balanced against advantageous nutritional effects. Traditionally a threshold concept has been used with the justification that non-carcinogenic and non-mutagenic effects have to be expected to follow a threshold concept of action (WHO 1999). The NOAEL (No-Observed-Adverse-Effect-Level) and the NOEL (No-Observed-Effect-Level) have been used as estimates for such threshold levels although they depend heavily on the group size, the test sensitivity and the dose spacing in a bioassay. Recently, the Benchmark dose (BMD) method, a probabilistic approach (PRA) and categorical regression (CatReg) have been proposed as new quantitative approaches which may improve food and diet safety assessment. It is the aim of this work to introduce new quantitative methods, to identify statistical challenges which have to be addressed for their better use in RA and risk quantification and to address areas of origins of uncertainties.

2. Concepts of Food and Diet Safety Assessment (FDSA)

FDSA is based on the estimation of the probability of the occurrence of an adverse effect (risk) resulting form an exposure to the causative agent contained in food. Quantitative RA involves the modeling of a **dose-response (DR) relationship** and is based on the estimation of the probability P of the occurrence of an adverse effect P = f(d), where f denotes the chosen mathematical DR function. This holds for a qualitative response endpoint. In the case of quantitative response, risk is characterized on a continuous scale as respective effect level E expressed similarly as a function E = f(d), when f is usually taken from another class of mathematical functions. The functions f are assumed monotone increasing (exact: monotone non-decreasing) in dose d. Parameters derived from the DR relationship include the shape of the dose-response curve (e.g. slope) or a fixed dose level on the dose-response curve (point of departure, PoD) from which the incidence data of the experimental range are extrapolated to incidence levels at low doses. Uncertainty estimates such as confidence limits are calculated in a statistical model. More formally: experimental toxicity data are obtained in a DR experiment where a number of dose levels $d_1, d_2, \Lambda\ , d_K$ is examined for the effect of concern when a number of n_k units is examined at each level d_k for the presence of absence of an effect (quantal dose response $Y_{ij} = 0,1$) or for the quantity of the effect (continuous

response Y_{ij}). Additionally, an untreated control group of n_0 individuals is examined at the level $d_0 = 0$.

The "safe" human exposure is usually termed "acceptable" (for a food additive) or "tolerable" (for a food contaminant) together with a time base related to the potential for accumulation, e.g. **acceptable daily intake (ADI)** or **provisional tolerable weekly intake (PTWI)** for chemicals that accumulate. The underlying concept of "safety assurance" is the assessment of a DR relationship in order to derive a critical dose below which an adverse effect specific for the investigated nutritional compound would be implausible and not expected. Therefore, all available toxicity data are reviewed for the determination of an exposure limit also called **reference dose RfD)**. To achieve this "safe" exposure limit a so-called **surrogate threshold dose** is determined from DR data. Uncertainties are factored into the resulting value of the RfD through the use of so called **safety factors** (or uncertainty factors) UF. In standard applications two safety factors have been used $UF_1 = UF_{interspecies} = 10$ for inter-species variation and $UF_2 = UF_{interindivdual} = 10$ for inter-individual variation. This lead to the usage of $UF=UF_1*UF_2=100$ with a long standing practice. Using toxicokinetic and toxicodynamic information, both, UF1 and UF2 have been further subdivided into $UF_1 = UF_1^{TK} \cdot UF_1^{TD}$ and $UF_2 = UF_2^{TK} \cdot UF_2^{TD}$ using chemical specific data available from toxicokinetic and toxicodynamic studies (Renwick, 1999).

The presence **of biological thresholds** has been postulated and suggested for a number of toxic mechanisms including food ingredients. However, this experience could be distorted by an observational bias in the case of small incidences of adverse effects which are impossible to be seen in a population, and especially not in a study population of a limited size. It is often overlooked that the biological arguments (e.g. homeostasis) address only the individual level (at a certain point in time), whereas DR data relate to the population level, with all other experimental factors (noise) involved. Questions of scaling the dose metric (linear versus logarithmic) have often not been addressed adequately (Edler et al.1995). For a detailed discussion see Slob (1999).
Formally, a **threshold dose response model** can be defined such that the existent but unknown threshold dose d_0 can in principle be estimated from the data. The model is

(1) $f(d) = c$ if $d \leq d_0$ and $f(d) = c + f_1(d)$ if $d > d_0$

where $f_1(d)>0$ denotes the dose response function of the range of doses of effects above background c. Although, statistical methods for fitting threshold models to dose-response data have been developed, models of this type have not been considered in the practice of FDSA. An intrinsic statistical problem in fitting is model selection and model identifiability.

3. Methods used for FDSA

3.1 Threshold of Toxicological Concern (TTC)

The TTC is an exposure value determined through qualitative chemical and toxicological arguments which is so low compared to actual human exposure levels that it is of no toxicological concern, e.g. sufficiently lower than toxic doses of structural analogues, such that there exists an adequate safety margin. Its determination is usually based on quantitative structure-activity relationship analysis. Chemicals are characterized by a set of quantitative and qualitative chemical characteristics (independent variables) and another set of biological activity and potency values (dependent variables). The empirical distribution of NOAELs of previously examined chemicals and low quantiles of it is used (WHO 2000) together with safety factors.

3.2 Threshold Methods

If the existence of a biological threshold can be assumed on the basis of biological and toxicological reasoning RA proceeds in two steps: (i) determination of a surrogate threshold value ST and (ii) application of safety factors UF_i to determine a dose level D* which would be without an adverse health effect ("no biologically significant adverse effect") to humans:

(2) $$D^* = ST \Big/ \prod_i UF_i$$

The ST is estimated normally by the no-adverse-effect level (NOAEL) defined as that largest dose level at which the effect is not significantly different from the background effect at level d_0. Significance has been usually defined as statistical significance at some predetermined $\alpha = 0.05$. However, this must not be the case and significance was in practice also defined as biological significance even if there is no statistical significance. The determination of the ST through the NOAEL has been criticized, especially when defined through statistical significance, because of the dependence of the NOEAL (a) on the sample sizes and the power of the statistical test, (b) on the design of the DR experiment and especially on the dose-spacing, and (c) on the sensitivity of the toxicological endpoint, the sensitivity of its measurement, and the sensitivity of the statistical test procedure. The NOAEL faces a further problem if the effects are less well-defined and occur both in the dose group and also in the untreated group (non-zero background) and increases continuously with increasing dose. Then a definition of a NOEAL might be quite arbitrary.

3.3 Low Dose Extrapolation - Non-threshold

The threshold dose approach was not applied to carcinogens since it was supposed that carcinogens are unlikely to have a threshold. Assuming no biological threshold, the slope of the incidence data in the experimental range can be extrapolated by a mathematical model to low dose levels in order to provide a direct quantitative risk estimate. A number of DR models have been suggested for cancer incidence data. Some of them were based on oversimplified carcinogenic concepts (one-hit, multi-hit, mulit-stage, Weibull) whereas others were purely empirical (probit, logit). Since most mechanisms of food substances are completely unknown and much less investigated than environmental and occupational carcinogens it seems quite reasonable to use simple generic DR models for FDSA. The low-dose extrapolation suffers from the fact that the estimated risk at low doses depends much more on the model chosen to extrapolate than on the experimental data. In other words, the class of plausible and possible models contains a large number of models among which none can be identified as best fitting in the sense of statistical model discrimination e.g., based on likelihood methods.

3.4 The Benchmark Dose Approach

The BMD was introduced by Crump (1984) as lower confidence limit on the dose that produces a predetermined small increase in the rate of the occurrence of the adverse effect considered over the background rate. Let $P(d|\theta)$ denote the quantal DR model with unknown model parameter(-vector) θ ; let p denote the Benchmark response level (BMR; usually taken as 0.1. 0.05 or 0.01). The original definition of the BMD for BMR = p is as follows: Set the BMR p equal to the effect change compared to the background effect of the untreated control. Then calculate

(3) $$d_p(\theta) = Q\left[P(0|\theta) + p(1 - P(0|\theta)\right]$$

BMD = lower 95% confidence interval of $d_p(\theta)$, if Q denotes the inverse of P. The BMD for a quantitative DR model is obtained similar. In contrast, Murrell et al (1998) argue strongly in favor of using the point estimate $\hat{d} = d_p(\hat{\theta})$ as point of departure, where $\hat{\theta}$ is the maximum likelihood estimate of θ, and to calculate then standard confidence intervals, not for determining a point of departure but for assessing the statistical variability of the point estimate. The BMD is not a low-dose extrapolation procedure but provides "only" an estimation of the quantity of the effect at the lower end of the DR curve. For the determination of an ADI from the BMD one has to apply safety factors as for the NOEAL.

3.5 New Methods Based on Statistical Reasoning

Calculation of an ADI with the generic equation (2) yields a single value with no possibility to assess variability and uncertainty. In order to obtain a statistical estimate of the ADI a probabilistic approach was implemented which assigns to the right hand side of (2) elements of probability distributions (Slob&Pieters, 1998). In this approach of PRA, distributions are calculated using resampling techniques such as the boostrap. Categorical regression on the severity of overall toxicity (Dourson at al, 1997) can be applied to studies with different endpoints and dosing schedules. The method assigns severity categories to the toxicity endpoint and it allows all types of endpoints analyzed together given that each one has an assigned severity category. The generalized effects model or ordinal regression

(4) $$P(Y \geq k | d,t) = H(a_k + b_1 d + b_2 t)$$

is applied where d and t denote the dose and the exposure time metric. This model allows the determination of $ED_{x\%}^{(k)}$ = dose causing x% effect of at least severity category k.

4. Statistical Challenges of FDSA

There are a number of statistical issues and open problems of the design and the analysis which have to addressed for better use and for better understanding in quantitative FDSA. Overall remains the challenging question on the absence or presence of thresholds which in the short run is a question of change-point analysis, but in the long run one of model identification and model comparison. Since the models are in generally not nested, standard methods originally based in linear model theory are not straight applicable. Approximate comparisons and simulation-based methods have been proposed. Predictive models of QSAR data are needed to improve the TTC and change points methods could made applicable for the threshold methods. Mechanistic modeling, model identifiability, optimal experimental design and software development is a challenge for non-threshold low-dose extrapolation and for the further development of the BMD approach. The new statistical methods PRA and CatReg need more research in evaluation e.g. using historical data bases. Nonlinear generalized models in categorical data analysis and the examination of the sensitivity of the assignment of categories is needed for better use of CatReg. Although it is not believed that this list is comprehensive, progress in some of the mentioned areas may improve FDSA and lead to new useful statistical thinking

Acknowledgments: The author is greatly indebted to the EU Concerted Action of Food an Diet Safety Assessment in Europe (FOSIE) and the colleagues of its working groups.

REFERENCES

Crump K.S. (1984). A new method for determining allowable daily intakes. Fund. Appl. Toxicol. 4: 854-871.

Dourson M.L., Teuschler L.K., Durkin P.R., Stiteler W.M. (1997).Categorial regression of toxicity data: a case study using Aldicarb.Regul. Toxicol. Pharm. 25: 121-129.

Edler L. Kopp-Schneider A.,Portier J.(1994). Zur Existenz von Schwellenwerten: Wissenschaftliche Methode oder statistisches Artefakt in der Risikoabschätzung. Zbl für Arbeitsmed 44, 16-21.

Murrell J.A., Portier J.C., Morris R.W. (1998). Characterizing dose-response I: Critical assessment of the benchmark dose concept. Risk Analysis 18, 13-26.

Renwick A.G. (1999). Subdivision of uncertainty factors to allow for toxicokinetics and toxicodynamics. Hum. Ecol. Risk Assess. 5: 1035-1050 (1999).

Slob W. (1999). Thresholds in toxicology and risk assessment. Int. J. Toxicol. 18: 259-268.

Slob W., Pieters M.N. (1998). A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. Risk Anal. 18: 787-798.