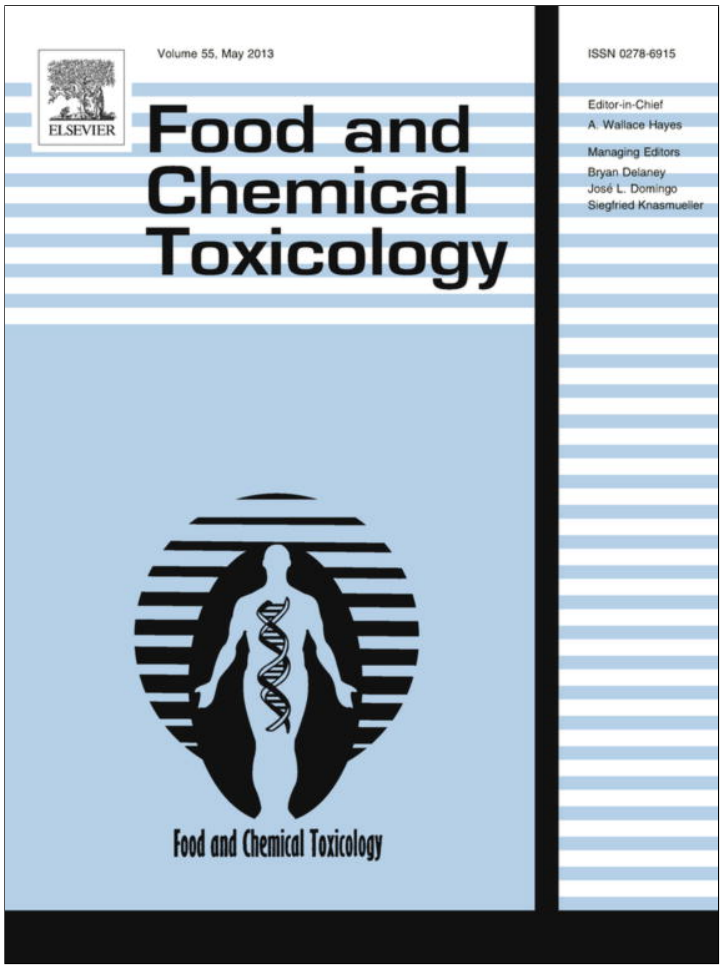

Food and Chemical Toxicology 55 (2013) 659–675

Contents lists available at SciVerse ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Critical appraisal of the assessment of benefits and risks for foods, 'BRAFO Consensus Working Group'

Alan Boobis [a], Alessandro Chiodini [b,*], Jeljer Hoekstra [c], Pagona Lagiou [d], Hildegard Przyrembel [e], Josef Schlatter [f], Katrin Schütte [h], Hans Verhagen [c,g], Bernhard Watzl [i]

[a] Imperial College London, Department of Medicine, Hammersmith, Ducane Road, London W12 0NN, UK
[b] ILSI Europe, Avenue E. Mounier 83, Box 6, 1200 Brussels, Belgium
[c] National Institute for Public Health & The Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands
[d] University of Athens Medical School, Department of Hygiene and Epidemiology & Medical Statistics, M. Asias Street 75, 11527 Athens, Greece
[e] Bolchener-Str.10, D-14167 Berlin, Germany
[f] Swiss Federal Office of Public Health, Consumer Protection Directorate, Stauffacherstrasse 101, 8004 Zürich, Switzerland
[g] University of Ulster, Cromore road, BT52 1SA Coleraine, Ireland
[h] Procter and Gamble, Temselaan 100, 1853 Strombeek-Bever, Belgium
[i] Max Rubner-Institut, Federal Research Institute for Nutrition and Food, Haid-und-Neu-Strasse, 9, 76131 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 11 April 2012
Accepted 22 October 2012
Available online 2 November 2012

Keywords:
BRAFO
Benefit–risk assessment
Tiered approach
Consensus
Risk management

ABSTRACT

BRAFO, Benefit-Risk Analysis for Foods, was a European Commission project funded within Framework Six as a Specific Support Action and coordinated by ILSI Europe. BRAFO developed a tiered methodology for assessing the benefits and risks of foods and food components, utilising a quantitative, common scale for health assessment in higher tiers. This manuscript reports on the implications of the experience gained during the development of the project for the further improvement of benefit–risk assessment methodology. It was concluded that the methodology proposed is applicable to a range of situations and that it does help in optimising resource utilisation through early identification of those benefit–risk questions where benefit clearly outweighs risk or vice versa. However, higher tier assessments are complex and demanding of time and resources, emphasising the need for prioritisation. Areas identified as requiring further development to improve the utility of benefit–risk assessment include health weights for different populations and endpoints where they do not currently exist, extrapolation of effects from studies in animals to humans, use of in vitro data in benefit–risk assessments, and biomarkers of early effect and how these would be used in a quantitative assessment.

1. Introduction and comparison of recent European activities on benefit–risk assessment of food and food compounds

BRAFO, which stands for Benefit–Risk Analysis for Foods, was a European Commission funded, Framework 6, project. The objective of BRAFO was to develop a framework for the quantitative comparison of human health benefits and risks of foods and food compounds based on a common scale of measurement. The resultant tiered approach was explored and refined by application to a series of case studies, comprising selected natural foods, dietary interventions and heat processing of foods. The proposed methodology and associated case studies are being published in Food and Chemical Toxicology (Hoekstra et al., 2010; Watzl et al. 2012, Verhagen et al. 2012a, Schütte et al. 2012). BRAFO papers have been published. See references.

It was apparent on completion of this phase of the work that there were a number of issues requiring elaboration or clarification. These are: Problem definition, Exposure assessment, Common currency, Level of evidence, Biomarkers and Use of animal data. Hence, as part of the BRAFO project, these issues have been explored in more detail and the outcome of this work is reported here.

During the period that BRAFO was underway, a number of other relevant activities on benefit–risk assessment were progressing in Europe; including sister Framework 6 projects such as QALIBRA and the development of a guidance document by the Scientific Committee of EFSA on human health benefit–risk assessment. These activities were not undertaken in isolation, and efforts were made to try to avoid unnecessary duplication of effort. However, to date, the outcomes of these activities have been reported independently of each other (Hart et al., 2013). The opportunity is therefore taken here to discuss how these developments can be integrated, to provide a broader perspective on the current state-of-the-art of benefit–risk assessment.

* Corresponding author. Tel.: +32 27710014; fax: +32 27620044.
E-mail address: publications@ilsieurope.be (A. Chiodini).

0278-6915/$ - see front matter © 2012 ILSI Europe. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fct.2012.10.028

1.1. Comparison of recent European activities on benefit–risk assessment of food and food compounds

1.1.1. BRAFO Tiered Approach

The BRAFO tiered approach assesses the benefits and risks when changing from a reference scenario to an alternative scenario (or more than one alternative scenario), resulting in a statement about the differences between the scenarios in terms of health benefits and risks for (a) specific population(s) and whether there is a net health benefit of one scenario over the other (Hoekstra et al., 2010). The tiered approach is illustrated by the flow-chart shown in Fig. 1. The assessment starts with problem formulation and the formulation of the reference and the alternative scenario. Then, a sequence of 4 tiers allows progressively refined comparisons of benefits and risks of the alternative scenario to the reference scenario, with their integration in the higher tiers.

In Tier 1, benefits and risks are assessed separately with the objective of analysing whether only benefits, only risks, both or neither will occur. Where only one type of effect on health (i.e. benefit but no risk or risk but no benefit) will occur (for example, nitrates, below ADI in vegetables (EFSA CONTAM, 2010)), then the assessment will stop at this tier, with a recommendation that a benefit only or risk only assessment be performed, as appropriate. However, for a specific scenario, there are generally both benefits and risks, generating the need for further assessment and integration, in a qualitative and/or quantitative way, with a comparison of scenarios. In Tier 2, benefits and risks are compared with each other and integrated qualitatively. Health effects are evaluated based on their impact on disease incidence (number of people affected), severity of the health effects, duration of the disease, and any additional mortality resulting from the effect. If the risk clearly dominates the benefit in the alternative scenario, then the assessment can be stopped and the reference scenario can be advised. Where there is a clear benefit over risk, the alternative scenario can be recommended. In the benefit–risk assessment of foods, Tier 2 is most often the decisive step for selecting the appropriate scenario (see Section 8).

In Tier 3, benefits and risks are calculated quantitatively in a deterministic way. Common metrics (QALYs/DALYs) (see Section 4) have been developed and are available for the process of integration (see below). Finally, in Tier 4, a probabilistic approach is applied, using the same common metrics as in Tier 3.

1.1.2. QALIBRA

The aim of the QALIBRA project was to produce software that calculates the net health effect of a benefit–risk assessment expressed in DALYs or QALYs. One of the features of the software is that it takes interindividual variability into account and can also focus on the uncertainties in the assessment by probabilistic calculations (Hart et al., 2013). The methodology used is the direct attributable health effect, which is closely related to the PIF (potential impact fraction) method (Barendregt and Veerman 2010). The user is asked to provide:

1. A characteristic population.
2. Intake (explicitly in a reference and alternative scenario).
3. Life expectations.
4. Dose–response relationships for each health effect that is associated with an intake.
5. Recovery and mortality probabilities.
6. Disease durations.
7. Disease/health state weights.

Compared to the BRAFO tiered approach, the QALIBRA instrument typically may be used in Tier 3 and 4. As in the BRAFO methodology, intake scenarios are the driving force. In using the QALIBRA software there is the explicit need to define two scenarios as input (reference and alternative, control and intake of interest), which are compared. The QALIBRA software reports the difference between the scenarios in DALYs or QALYs, together with the associated uncertainties. Also, the underlying variables that make up the DALY/QALY value, such as incidence of disease and induced mortality in relevant subpopulations, are shown.

1.1.3. BEPRARIBEAN project

Whereas benefit–risk assessment in food and nutrition is making progress, still a number of difficulties remain. It is anticipated that benefit–risk assessment in food and nutrition could benefit from looking across its borders to learn from other research areas. The BEPRARIBEAN project (BEst PRActices for RIsk BEnefit ANalysis: experience from out of food into food) was established to iden-

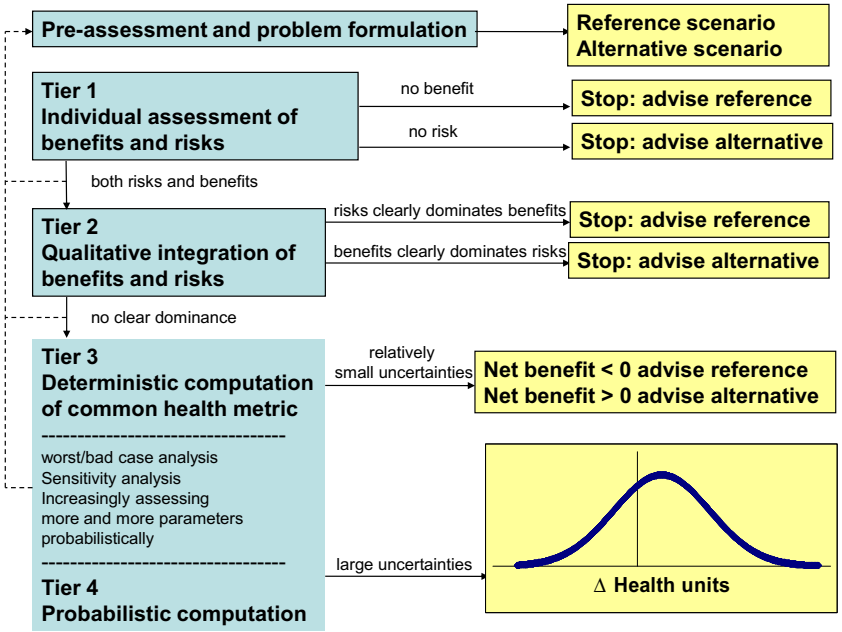

Fig. 1. A flow-chart of the BRAFO tiered approach.

tify best-practices and experiences from other areas and transpose those into the area of food and nutrition. The BEPRARIBEAN project started on the basis that risk-taking is normal in everyday life if there are associated (perceived) benefits. As such, benefit–risk assessment compares the risk of a situation to its related benefits and addresses the acceptability of the risk.

The BEPRARIBEAN project describes the state of the art in benefit–risk assessment in the areas of:

- Medicines
- Food Microbiology
- Environmental Health
- Economics & Marketing-Finance
- Consumer Perception
- Food and Nutrition

All perspectives are described in state of the art reviews and subsequently integrated to identify opportunities for further development of benefit–risk assessment for food and nutrition (and vice versa). Interesting issues that emerge are the varying degrees of risk that are deemed acceptable across the disciplines from not acceptable, via acceptable to necessary. The project started in 2009. A set of 6 'state of the art' papers covering the above areas, an introductory paper and the final integration paper were published in early 2012 (Kalogeras et al., 2012; Luteijn et al., 2012; Magnússon et al., 2012; Pohjola et al., 2012; Tijhuis et al., 2012a; Ueland et al., 2012; Verhagen et al., 2012b; Tijhuis et al., 2012b). More information can be obtained at: http://en.opasnet.org/w/Bepraribean.

1.1.4. EFSA guidance

The Scientific Committee of EFSA (2010), in its guidance on benefit–risk assessment, like BRAFO recommends a tiered approach. The EFSA approach comprises three steps: (1) initial assessment, (2) refinement, (3) comparison using a common metric. Step 1 and Step 2, at a different level of refinement, comprise Tiers 1 and 2 of the BRAFO approach. Step 3 is similar to Tiers 3 and 4 in the BRAFO approach. EFSA is less prescriptive in how an assessment should be performed than BRAFO. In both approaches, it is envisaged that the QALIBRA software could be used to perform the last step(s) or tiers. EFSA does not explicitly state that a benefit–risk assessment compares (at least) two scenarios. EFSA in its guidance explicitly identifies interaction between the (benefit–) risk manager and the assessor after each completed step, allowing further refinement of the problem formulation.

The BRAFO and EFSA approaches do not contradict each other, but rather they are complementary, as they are with the QALIBRA software. Each provides more detailed guidance than the other in certain aspects of how a benefit–risk assessment can be performed. Both the BRAFO and the EFSA approaches suggest starting by investigating the problem in a simple way and then refining the assessment if the simple approach does not provide a satisfactory outcome.

2. Problem definition

Prior to undertaking a risk assessment or a benefit assessment it is important to define the problem. This is particularly true in benefit–risk assessment, in view of the resources necessary to perform such an assessment.

A change in intake of a food or a food item may be associated with some degree of risk as well as benefit. In most cases, it is implicit in the problem formulation that the risks outweigh the benefits or vice versa, so only one is formally addressed in any detail. If neither benefits nor risks dominate a priori, it becomes important for policy makers and benefit–risk managers[1] to be provided with information on the net health impact, which can only be obtained by undertaking a benefit–risk assessment. Risk management is defined as: "... the process, distinct from risk assessment, of weighing policy alternatives in consultation with interested parties, considering risk assessment and other legitimate factors, and if need be, selecting appropriate prevention and control options".[2] In general, this means that within the context of the intake of interest, the question that a benefit–risk manager has to address will be one of two types, which are closely related:

(1a) Is there a public health issue?
(1b) If so, is there an alternative scenario that can improve public health?
(2) What is the impact of a proposed policy or alternative scenario on public health?

Both questions can be answered by comparing two or more scenarios and indeed may require such a comparison, but this is not always necessary. Clearly it is important to define the public health goal: Does this mean improvement in the health of all members of the population, or does it mean an improvement in net health at the population level, but at the cost of a possible decrease in the health of certain subgroups?

Problem formulation is essential to specify the benefit–risk problem that will be investigated. It describes the purpose, scope and limitations of the assessment to ensure that the outcome is relevant. The benefit–risk manager, for whom the assessment is intended, is responsible for problem formulation in consultation with the assessor. It is good practice to record the context or background of the benefit–risk question and the reasons why it has been asked, to ensure that involved parties and relevant stakeholders have a common understanding of the problem.

Some issues that should normally be considered during problem formulation are:

- The intake of interest, i.e. which substances and/or food item(s) should be part of the assessment, and is substitution with other substances and/or food items part of the problem.
- Which health endpoints are included or which inclusion criteria are used to decide which health endpoints are included.
- Which (sub)populations are considered (e.g. children, men, pregnant women,...)
- Do the consequences of a potential policy decision have to be evaluated.
- Does the assessment also need to explore which intake scenarios could improve, or optimise, health and what the limitations for an alternative intake scenario are.

2.1. Scenarios

Benefit–risk assessment, like environmental health impact assessment, often focuses on interventions and policies rather than on substances or agents in isolation (Briggs, 2008). Its aim is to support decision-making on, for example, what dietary advice should be given, whether foods should be fortified, or if a new production process is acceptable. Therefore the benefit–risk question is generally a choice between two, or a series of, alternative policies or courses of action, described in the form of scenarios. Accordingly, in the BRAFO approach, the problem formulation

[1] This document will use the term benefit–risk manager for risk managers and all other policymakers for whom the benefit–risk assessment is intended, to advise on decisions on food policies.

[2] Quote by Frans Verstraete (European Commission-DG SANCO) at the BRAFO workshop in Barcelona, 16 November 2010-12-02.

stage generally involves definition of a reference and an alternative scenario for comparison. The benefit assessment and risk assessment in Tier 1 of the BRAFO approach may focus only on guidance values, such as the ADI and RDA. If everyone's intake is below the toxicological (e.g. ADI) and within an acceptable margin of the beneficial (e.g. RDA) guidance value then the conclusion could be that there are no appreciable risks and benefits cannot be increased, and hence public health could not be improved by changing the intake. Thus the assessment stops without the need to define two scenarios. Therefore explicit scenarios are needed only if it is necessary to progress to the second tier. If benefits as well as risks exist the benefit–risk manager may want to determine whether public health could be improved by actions that increase the benefits and decrease the risks. Also, actions that decrease both benefits and risks may result in a net increase of health, as could actions that increase both benefits and risks. The benefit–risk manager may want to explore these possibilities by considering a series of different scenarios all of which may increase net health. Such an assessment can serve to derive scenarios or policy actions that would result in an improvement in public health.

A scenario is a narrative that describes a hypothetical or real situation, in this context referring to some intake of food(s) or food constituent and its consequences. An important part of problem formulation when the assessment progresses beyond Tier 1, is to identify the relevant scenarios. Benefits, i.e. the probability of an increase in beneficial health effects or a decrease in adverse health effects, and risks, i.e. the probability of an increase of adverse health effects, are assessed against a reference scenario. From Tier 2 and beyond the benefit–risk question must be formalised as two (or more) scenarios, a reference or control and one, or a series of, alternative(s). The reference scenario is usually the current situation or a hypothetical zero exposure scenario. The alternative scenario(s) to be compared with the reference scenario could be some change that would result from a policy that is perceived to improve health. A series of alternative scenarios can be used to examine the scope of potential policies or to optimise a potential policy.

The definition of the alternative scenario will not always be clear or cannot be made clear in the problem formulation stage. Then, the formulation of the alternative scenario or a series of alternatives becomes part of the assessment. An example would be when health is perceived to be improved by some extra intake (compared to the current situation), but not by too much, because that would introduce risks. In such cases, the benefit–risk assessment is designed to identify those exposure scenarios that improve or optimise the net health benefit.

3. Exposure assessment

The following issues should be considered in the exposure assessment for a benefit–risk assessment:

- Habitual intake
  - The benefits and risks of concern in a benefit–risk assessment are likely to be a consequence mainly of long-term exposures, rather than of acute (peak) exposure. In such situations, measured intake needs to be transformed into long-term average intake – so-called habitual intake. Various ways of performing such a transformation are available (Dodd et al., 2006; Souverein et al., 2011).
- Background exposure
  - The food or dietary substance of interest is not always the only source of exposure. The total exposure needs to be considered when estimating the health effects resulting from exposure to that substance. Therefore, exposure from other sources must also be taken into account
- Subpopulations
  - Some health effects are specific for certain subpopulations (e.g. pregnant women). Clearly then it is the intake of that specific sub-population that is of concern
- Concentrations in food items
  - When a specific substance triggers a benefit–risk assessment, in addition to data on consumption of food items, occurrence data for that substance in food items are important (i.e. in what food items is the substance present and at what concentrations)
- Substitution
  - Sometimes one is concerned about the risks of a certain food item and investigates scenarios with a reduced intake of that food item. Usually that will mean that some other food is eaten (more) to compensate for the reduced consumption of the item of concern. This food may also be associated with benefits and risks and if substitution is part of the terms of reference, then consumption (and the associated benefits and risks) of this food will need to be estimated as well.
- Internal exposure
  - Concentration of a beneficial/harmful substance in specific target organs can provide invaluable insight into the occurrence of beneficial or adverse health effects. Accumulation and elimination in the body over time can be estimated using physiologically-based pharmacokinetic (PBPK) models

4. Common currency

In a benefit–risk assessment, generally, different health states or diseases associated with the benefits and risks, respectively, must be compared with each other. Each health state may have completely different characteristics from other health states in terms of severity, morbidity and induced mortality. Therefore, in higher tier assessments, it is necessary to express each health state in a common metric or index, which captures all important aspects of a health state. Two such metrics are the DALY (Disability Adjusted Life Years) and QALY (Quality Adjusted Life Years), which are both suitable for use in benefit–risk assessment, although in specific cases other metrics (e.g. mortality) can be more appropriate. Both metrics include morbidity and mortality as dimensions. They are both expressed in life years and assume that quality of life can be exchanged by length of life. Disability weights can be established, by using questionnaires that ask how many life years an individual is willing to trade for not suffering from a particular disease or health state. Fig. 2 shows a graph of the Q(D)ALYs of one individual during his/her lifetime. It illustrates the metric and how it can be calculated. The Q(D)ALYs for a (sub)population are simply the sum of the Q(D)ALYs for each individual in the population. Note that DALYs and QALYs are conceptually the same in the sense that they both measure health as life years adjusted for years with disease but there are differences. DALYs measure the loss of adjusted life years thereby assuming an expected age of death if a disease would not occur, whereas QALYs measure the number of adjusted life years accumulated since birth. In the graph the DALY and QALY weights for quality of life are presented as similar; i.e. $w_Q = 1 - w_D$. For benefit–risk assessments, the difference between two scenarios calculated either in DALYs or in QALYs needs to be determined. The difference between two scenarios in QALYs or DALYs will be the same provided that the weights are equal i.e. $w_Q = 1 - w_D$. In practice the weights are derived in different ways, e.g. by interviewing different groups (practitioners and patients, respectively) and by using different questionnaires. As a result they will have different values. However, both weights are a measure of how severe a disease is. Therefore, for the final implication of the

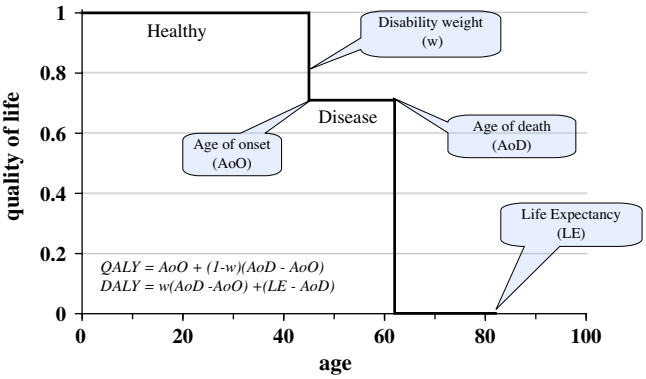

Fig. 2. Graphical representation of quality of life metric [Q(D)ALY] in an individual.

assessment it will rarely make a difference whether the calculations are performed using DALYs or QALYs.

The difficulty in computing Q(D)ALYs is not only in valuing the health state but also in estimating if, and at which age, exposed individuals will experience a different health state. That is why dose–response functions that link exposure to disease or health state are so important. For the accurate calculation of QALYs or DALYs one needs to estimate the duration of the disease and, for DALYs, the years of life lost due to induced mortality. Data for these parameters are often not readily available, which could result in estimates with large uncertainties. In particular, when risks are estimated using data from studies in laboratory animals, assumptions are needed to convert the health effects seen in those studies to a human health effect (see below). The time that an animal prematurely dies in an experiment is seldom measured, let alone the duration an animal suffers from the health effect. Both are necessary parameters to calculate a Q(D)ALY. Hence, these parameters often need to be estimated from human disease statistics. However, this ignores the dose dependencies that usually exist. Therefore, careful consideration of the necessary assumptions is required and often large uncertainties are introduced if animal data are used to compute QALYs or DALYs of human health effects. The case study on fish carried out in the QALIBRA project provides an example of where animal data were used to calculate DALYs (Hoekstra et al. 2012a,b).

5. Level of evidence

Evaluation of the strength of evidence is an integral part of the identification of the benefits and risks in a benefit–risk assessment. Beneficial health effects should be identified primarily from intervention or observational studies in humans. In contrast, adverse health effects are often deduced from studies in experimental animals or from human observational studies. They are rarely identified in intervention studies because, for ethical reasons, it is obviously of importance to try to avoid such effects. Judgement on the level of evidence for both benefits and risks should be performed using the same agreed criteria, e.g. the WHO criteria (WHO, 2003). However, in the context of benefit–risk assessment, the strength of evidence accepted to establish an effect on health may be different when assessing risks from that when assessing benefits.

With respect to benefit assessment, meta-analyses of randomised controlled trials provide the highest level of evidence for the documentation of the existence of a benefit. The evidence is considered as convincing when results from good quality trials consistently show an association between exposure and a favourable health outcome. The evidence should preferably be supported by the results of observational studies and the effect should be biologically plausible.

Evidence can be judged as probable when the available epidemiological (observational and intervention) studies show fairly consistent associations between exposure and effect, but there are shortcomings in the studies or some evidence to the contrary. Evidence is judged as possible when it relies on inconsistent results from intervention and observational studies, but the evidence from non-epidemiological studies is supportive. Insufficient evidence is based on inconsistent findings from a limited number of studies.

The quality of the evidence (Table 1) is closely related to the quality of the studies as evaluated e.g. in the GRADE approach (Grades of Recommendation, Assessment, Development and Evaluation Working Group, (2004)), which has been adopted by the Cochrane Collaboration (Schünemann et al., 2008) (http://www.igh.org/Cochrane/tools/Ch12_Interpreting.pdf).

The quality level is highest for randomised trials, but randomised trials are downgraded when one or more of the factors in Table 2 are present. The presence of each one of the factors means one quality level less and the overall quality of a randomised trial can finally be rated less than that of observational studies.

Observational studies are usually rated as of low quality of evidence, but may be upgraded to moderate or even high quality if the observed effects are large and cannot be explained by bias (Table 3).

In general, it has been recommended, in the PASSCLAIM project (Howlett and Shortt, 2004), that only those benefits for which the evidence is considered to be convincing or at least probable should be included in assessing benefit for accepting health claims of foods and food ingredients, and this is in accord with the approach of the EFSA Panel on Dietetic Products, Nutrition and Allergies (NDA) to the scientific substantiation of health claims (EFSA NDA, 2011a,b; Verhagen et al., 2010). However, health effects for which the level of evidence is judged as lower, but which are expected to be substantial in terms of incidence and magnitude, should not be ignored in the benefit–risk assessment.

With respect to risk assessment, risks are often identified in studies in experimental animals. The quality of such data should be assessed according to criteria such as those proposed by Klimisch et al. (1997). These authors identified three key aspects of a study that should be taken into account when considering whether the results should be included in a risk assessment. These are reliability – the inherent quality of a study and the clarity with which the methodology and results are described, providing evidence for plausibility of the findings; relevance – the extent to which the results are appropriate for the specific risk assessment; adequacy – the utility of the data for the assessment. Klimisch et al. (1997) proposed 4 categories for the reliability of the experimental evidence. These were category 1: reliable without restrictions, where studies were conducted according to established guidelines or standards; category 2: reliable with restrictions, where the study does not comply with accepted guidelines but sufficient information is provided to enable the quality of the study to be judged as acceptable; category 3: not reliable, where either there were methodological deficiencies or where insufficient informa-

Table 1
Levels of quality of a body of evidence in the GRADE approach (Schünemann et al., 2008).

Underlying methodology	Quality rating
Randomised trials; or double-upgraded observational studies	High
Downgraded randomised trials; or upgraded observational studies	Moderate
Double-downgraded randomised trials; or observational studies	Low
Triple-downgraded randomised trials; or downgraded observational studies; or case series/case reports	Very low

Table 2
Factors that may decrease the quality level of a body of evidence in the GRADE approach (Schünemann et al., 2008).

1. Limitations in the design and implementation of available studies suggesting high likelihood of bias
2. Indirectness of evidence (indirect population, intervention, control, outcomes)
3. Unexplained heterogeneity or inconsistency of results (including problems with subgroup analyses)
4. Imprecision of results (wide confidence intervals)
5. High probability of publication bias

Table 3
Factors that may increase the quality level of a body of evidence (Schünemann et al., 2008).

1. Large magnitude of effect
2. All plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results show no effect
3. Dose-response gradient

tion is provided to enable the adequacy of the method to be assessed; category 4: not assignable, where insufficient information is provided on the methods and results to enable any evaluation of the quality of the study. Schneider et al. (2009) have recently developed a software-based tool (ToxRTool) to assist in the assignment of Klimisch criteria for reliability. Assessment of relevance and adequacy is distinct from assessment of reliability. A study may be judged to be highly reliable, but it may not be relevant, or it may not be adequate, for the needs of a specific risk assessment. Hulzebos et al. (2010) have recently provided some guidance on these additional issues in the quality assessment of toxicological data.

In risk assessment, the highest dose at and below which there is no effect (the so-called "critical no observed adverse effect level" – NOAEL), or other suitable reference point such as the benchmark dose, lower 95% confidence limit (e.g. BMDL10) (see below), is chosen for the determination of the risk which is then extrapolated to a presumed sensitive human subpopulation, using safety factors (uncertainty factors). Not all hazards identified in experimental animals have their counterpart in humans. Therefore, all hazards identified in such studies should be evaluated for the level of evidence that they have relevance to humans. (For extrapolation of hazards identified in animals to humans see Section 7).

If, as is possible and indeed likely, the level of evidence differs for the benefit and the risk assessment, it is up to the benefit–risk-manager to decide whether or not risks for which the level of evidence is rated only to be possible or even insufficient should be included in the benefit–risk analysis, to ensure a sufficiently high level of public health protection (Commission of the European Communities, 2000). In fig. 3a and b, this asymmetrical consideration of risks and benefits in benefit–risk management versus benefit–risk assessment is illustrated.

In risk assessment, risks across all levels of evidence are taken into account; risks at lower levels of evidence may be compensated by lower safety factors or lower margins of safety. In benefit assessment (i.e. for the purpose of health claims assessment under E.U. Regulation 1924/2006), only benefits at the higher levels of

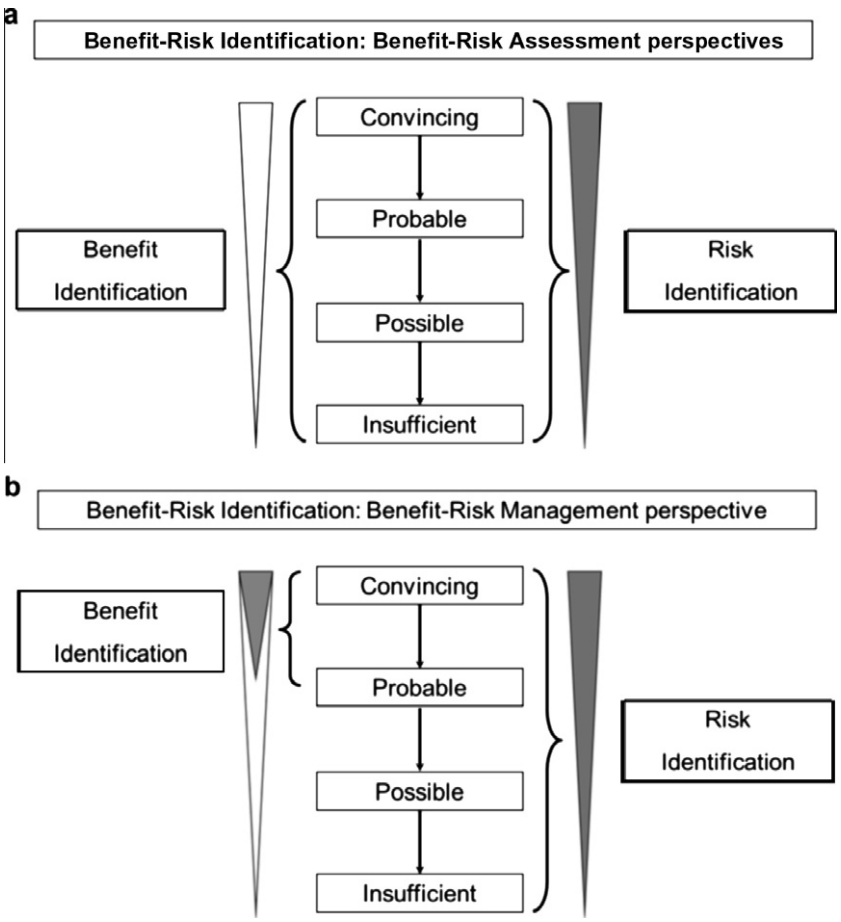

Fig. 3a and b. Benefit–risk assessment versus benefit–risk management, respectively.

evidence are included in considering evidence for a cause and effect relationship. For the purpose of benefit–risk assessment and benefit–risk management it seems prudent also to consider studies with lower grades of evidence in an overall assessment, since a low level of evidence which potentially impacts appreciably on a large number of people may contribute to public health. Here the purpose of the assessment makes the difference as to whether or not to take a particular grade of evidence into account. Weighting factors could be introduced to address differences in the uncertainties related to judging the evidence.

From a benefit–risk assessment perspective, a scientific activity, it is possible to assign grades of evidence for benefits and risks (white arrows; Fig. 3a). From a benefit–risk management perspective, a policy activity, a decision has to be made as to the acceptance of benefits and risks across grades of evidence; as such the E.U. accepts all grades of evidence from the perspective of potential health risks (right-hand grey arrows), whereas the E.U. under E.U. Regulation 1924/2006 accepts benefits only in cases where a cause and effect relationship has been established (grey arrow; Fig. 3b) whereas lower grades of evidence are not accepted for the scientific substantiation of health claims (left-hand white arrow; Fig. 3b).

6. Biomarkers

6.1. General introduction to biomarkers for benefit–risk assessment

Clinical endpoints defined as "a characteristic or variable that reflects how a patient (or consumer) feels, functions or survives" (Institute of Medicine, 2010) can be measured directly for benefits and risks in studies designed for this purpose and such studies would be the most suitable to use as the basis for a benefit–risk assessment. However, it is not always feasible to measure clinical outcome for several reasons. For example, there could be a long time-period between the introduction of an intervention and the outcome and it may not be ethical to start an intervention with an expected adverse outcome. Furthermore, whilst it may be possible to measure some clinical endpoints in large-scale intervention studies, this may not be appropriate or practical.

Therefore, biomarkers, defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention" (Institute of Medicine, 2010), which are more readily measurable than clinical endpoints, might be used as proxies or surrogates for the outcomes of interest. The use of biomarkers in benefit and risk assessment was discussed separately within the PASSCLAIM (benefits) and FOSIE (risks) projects (Aggett et al., 2005; Renwick et al., 2003). Biomarkers can consist of results from chemical analyses of blood or other body fluids or of tissues, as well as genetic, genomic, metabolomic or imaging data, for example. Biomarkers can be used to assess exposure and/or the effects of exposure on biological processes or mechanisms related to clinical outcomes. When a biomarker has been demonstrated to have the capacity to substitute for a clinical endpoint and can predict health outcomes it is called a surrogate endpoint.

In the FOSIE project (Renwick et al., 2003) it was concluded that biomarkers as measures of exposure, effect or susceptibility are important tools for risk assessment. The FOSIE project defined two types of biomarker:

(1) Biomarkers of exposure including those for external dose, internal dose and biologically effective dose.
(2) Biomarkers of effect including those of early effects and of sub-clinical disease.

In the PASSCLAIM project (Aggett et al., 2005) biomarkers for beneficial effects were classified into three categories.

(1) Biomarkers that reflect exposure to the food component being studied, for example a serum or tissue concentration. These markers can give an indication of the bioavailability or the presence of the food component or that of a functional derivative or metabolite in the body.
(2) Biomarkers that relate to the target function or biological response. These can reflect a change in a given function (e.g. induction of enzymes, release of hormones), or a change in body fluids.
(3) The final category concerns an appropriate intermediate endpoint of an improved state of health and well-being or of a reduction of risk of disease or both (e.g. blood pressure). For example, such a biomarker could be measurement of a biological process that is an intermediate leading to the health endpoint.

There is clearly appreciable overlap between the types of biomarker discussed in the FOSIE project and those addressed in the PASSCLAIM project, which is illustrated in Fig. 4.

6.2. Biomarkers of exposure and of effect can be used in several steps of the benefit–risk assessment

A *biomarker of exposure* can be either the food chemical or one or more of its metabolites (Kroes et al., 2002). Such biomarkers

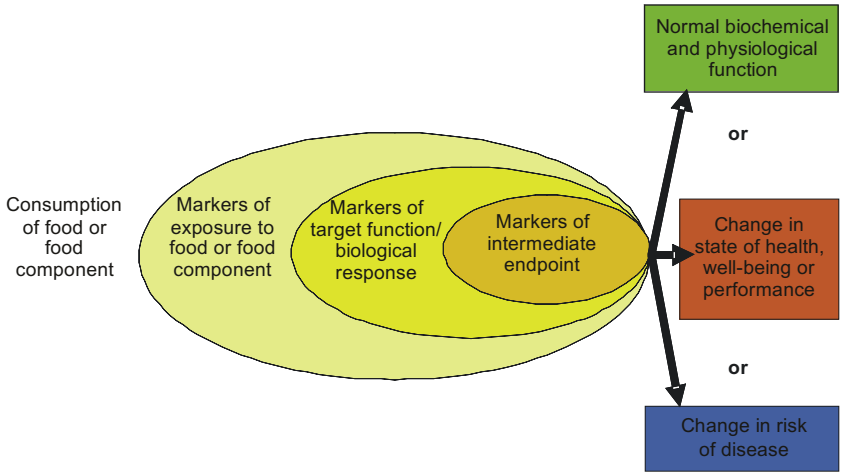

Fig. 4. Classification of biomarkers of either benefit or risk, based on PASSCLAIM criteria (Aggett et al., 2005).

are determined mostly in the urine or blood, but may also be measurable in breast milk, hair, skin, adipose tissue, buccal swabs, exhaled air or faeces. The main advantage of such biomarkers is that, when validated, they can provide an accurate estimate of the amount of a nutrient or food chemical that has been consumed and this estimate is not biased by over- or underreporting of intake. They may also provide an estimate of absorbed dose. However, the period of time over which the biomarker reflects exposure will depend on the metabolism and excretion of the nutrient or food chemical. Urinary biomarkers can reflect both short-term and long-term exposure and, in general, quickly become immeasurable when exposure stops, whilst the presence of lipid soluble biomarkers in adipose tissue may continue for a considerable time after exposure has ceased. Moreover, whilst biomarkers of exposure are, as a rule, specific for one chemical, they are also influenced by other than dietary sources of exposure (e.g. cosmetics or smoking) and hence reflect total exposure (Kroes et al., 2002). In most cases, discrimination between periods of exposure is not possible, unless e.g. biomarkers of both long- and short-term exposure are combined. (Example: fluoride in urine or saliva for short-term exposure and fluoride in nails, dentine or bone for long-term exposure).

Some examples of biomarkers of exposure to nutrients include nitrogen measured in 24 h urine samples over several days, as a biomarker of dietary protein consumption in adults. Valid estimates of sodium exposure may be obtained from levels in urine. Vitamin E and fatty acids can be assessed in adipose tissue, including blood lipids. For chemical substances, biomarkers of exposure include urinary levels of nitrate, cadmium and heterocyclic aromatic amines (HAAs) (Kroes et al., 2002).

Biomarkers of effect can be used in the identification of both beneficial effects and hazards. In the case of hazards this will mostly be done in animal studies, although human observational studies can also provide evidence for associations between diets or dietary constituents and biomarkers of effect, particularly if supported by biomarkers of exposure. The development of early biomarkers of adverse effects was identified as a key issue for further work (Barlow et al., 2002). The use of biomarkers of effect in hazard characterization is still an area in which additional research is needed, especially when a level of exposure without significant adverse effects (NOAEL) cannot be determined (Edler et al., 2002). Validated biomarkers would be very valuable when they can be applied to both epidemiological and animal studies, as they can greatly reduce the uncertainties related to the extrapolation of the dose–response curve based on animal studies to humans (Renwick et al., 2003). The same can be said for the use of biomarkers of effect in beneficial effect characterization.

All biomarkers should be methodologically and biologically valid. Methodological validity relates to the validation of measurement, which should be accurate, reproducible and precise. Biological validity relates to the understanding of the role of the biomarker in the biological mechanism and the necessary changes concurrent with the biological mechanism. This refers to the sensitivity (probability of correctly identifying the effect) and the specificity (probability of incorrectly identifying the effect) of a biomarker, which should be known. It is important that the role of the biomarker within the mechanisms and pathways leading to the health effect is characterised. Furthermore, the effects of other factors that might influence the biomarker, such as other dietary factors, type of sample, lifestyle or physiological factors, and genetic variability should be determined before the biomarker can be used in benefit or risk assessments (Jenab et al., 2009).

It has been suggested that a combination of biomarkers may better reflect the biological endpoint and provide greater certainty than a single biomarker (Elliott et al., 2007). In this respect, -omics technologies could be of particular value. Indeed, using such approaches, the development of multiple biomarkers has improved substantially (Albers et al., 2005; Palou et al., 2009). For example, a recent publication studying anti-inflammatory effects induced by nutritional intervention in overweight men was able to identify multiple physiological changes in genes, proteins and metabolites (Bakker et al., 2010). This study showed that the application of -omics techniques provided insight into the interaction of the several processes involved. Still, the interpretation and validation of the large-scale output of -omics technology remain difficult. The value of -omics based biomarkers in benefit or risk assessments depends not so much on the reliability of an assay to qualitatively and quantitatively assess the biomarker, but more crucially, on the association between the biomarker and the biological endpoint (Vlaanderen et al., 2010).

6.3. Use of biomarkers in benefit–risk assessment

Validated biomarkers would be very useful in benefit–risk assessment. The biomarker to be used should be biologically and methodologically validated, including an assessment of its range of variation, and biologically relevant changes therein.

The use of biomarkers becomes a challenge in the quantitative phase (Tier 3 and 4) of a benefit–risk assessment, when the benefits and risks are compared using common metrics, such as DALYs and QALYs, and presented as an integrated result using these metrics. Currently, DALYs or QALYs are unavailable for intermediate endpoints (e.g. cholesterol level). In this phase, benefits and risks based on biomarkers can be quantified separately but cannot be expressed as an integrated measure without making arbitrary assumptions.

7. Use of animal data

7.1. Introduction

As discussed above, data on adverse health effects are usually derived from studies in experimental animals, which requires animal to human extrapolation before a benefit–risk assessment can be undertaken. Whilst health based guidance values, such as the tolerable daily intake (TDI), based on default assumptions, can be used in lower tiers of a benefit–risk assessment, in higher tiers, a more quantitative estimate of risk in human (sub-) populations is necessary. In Hoekstra et al. (2010) and EFSA (Scientific Committee, 2010) some guidance is included on the use of data from studies in experimental animals in benefit–risk assessment. However, few details are provided. This aspect is therefore covered in greater detail here, focusing on hazard, as this will be the situation most frequently encountered in a benefit–risk assessment when there is a need to extrapolate from studies in experimental animals to humans. However, a number of the issues covered will also be applicable to extrapolation of benefit from studies in experimental animals to humans, in the rare cases where this is necessary.

In extrapolating data on toxicological hazards from studies in experimental animals to humans, several issues need to be considered.

Hazard identification starts with identifying those hazards that can result from exposure to the substance, i.e. what adverse effects does the substance cause in experimental animals, followed by determining which of these might be relevant to humans. Recent reviews (Guzelian et al., 2005; Boobis et al., 2008) have emphasised the need for evaluating study quality and applying a weight of evidence approach when assessing the information obtained from studies in experimental animals (see Section 5, above). This is particularly important where the database is complex, with a variety of study types and replicate studies. An effect may be

equivocal, or unrelated to the substance itself, but rather to some aspect of study design, such as dosing vehicle. Results of one study may be contradicted by those of another. Careful, objective evaluation of study quality and biological consistency, for example with the results of studies on other relevant biological effects, is critical to reaching a conclusion on causation. Where there are contradictory results of studies of equal and acceptable quality, it is necessary to clearly point this out and identify data needs to resolve this uncertainty so that the benefit–risk manager can decide whether to assume cause and effect, or whether further study is required before the benefit–risk assessment can proceed.

Once substance-dependent hazards have been established in experimental animals, the relevance of these hazards to humans should be addressed, before considering them in benefit–risk assessment. The WHO International Programme on Chemical Safety (IPCS) has developed a systematic, structured approach to the evaluation of the human relevance of toxicological hazards, based on mode of action (Seed et al., 2005; Boobis et al., 2006, 2008). A mode of action comprises a series of key events that are quantifiable and necessary, though individually usually not sufficient, for the toxicological effect. The key events are assessed on the basis of a number of considerations, first outlined by Bradford Hill for evaluating cause and effect in epidemiological studies of association (Hill, 1965). The IPCS approach is essentially a two-step procedure. The first step is to determine whether it is possible to establish a mode of action with reasonable confidence for the toxicological effect in the experimental studies (Sonich-Mullin et al., 2001). If it is not possible to establish a mode of action with confidence, the default assumption is that the hazard is relevant to humans. However, during consideration of the mode of action, critical insights into the characteristics of the hazard are often obtained, for example the nature of the dose–response relationship, and temporal pattern of effects. It is important that this knowledge is incorporated into hazard characterisation as discussed below.

If it is possible to establish a mode of action for the toxic effect observed in experimental animals, the next step is to determine whether the mode of action is relevant to humans. This is achieved by qualitative and quantitative comparisons of the key events between experimental animals and human, taking into account both toxicokinetic and toxicodynamic factors (Preston and Williams, 2005; Boobis et al., 2006, 2008). This evaluation may result in a clear conclusion that the mode of action is or is not relevant to humans. In the latter case, the hazard is not considered further in the benefit–risk assessment. In the former, the hazard may need to be considered in the benefit–risk assessment and relevant information obtained during the evaluation, for example on dose–response relationships, potentially susceptible sub-populations, and human interindividual variability, should be utilised later in the benefit–risk assessment (see below). If no clear conclusion can be reached on human relevance of the mode of action, the normal default position would be to assume that the mode of action is relevant, as this is more health protective.

The above approach is applicable to a diverse range of hazards, including non-nutrient chemicals, nutrients, allergens and pathogenic microorganisms (Julien et al., 2009), and endpoints (Seed et al., 2005; Boobis et al., 2006, 2008, 2009; Preston and Williams, 2005).

In benefit–risk assessment, as noted above, it is often the case that the risk will have been identified in studies in experimental animals, whereas the benefit will have been identified in studies in humans. The prerequisite for establishing beneficial effects in humans was identified in the course of the PASSCLAIM project (Aggett et al., 2005) and was also described as a necessity by EFSA NDA (EFSA, 2011b; Verhagen et al., 2010). This introduces substantial inequality into the uncertainty associated with effect identification. Whereas there may be reasonable confidence that there is a benefit, because it was determined in studies in humans, there may be appreciable uncertainty as to whether there is any risk to humans, as the effect observed in experimental animals may not be relevant to humans, but there is insufficient information to exclude this possibility. This should be clearly reflected in the evaluation of uncertainty in the benefit–risk assessment (e.g. see Linkov et al., 2009).

7.2. Effect Characterisation

Once substance-dependent hazards have been identified and concluded or assumed to be relevant to humans, their dose–response relationships need to be characterised in order to obtain a quantitative estimate of the magnitude or incidence of effect caused by a given exposure level. Much has been written about dose–response assessment (Dybing et al., 2002; Slikker et al., 2004; O'Brien et al., 2006), so only a brief summary is provided here. Broadly, two approaches have been used in evaluating dose–response relationships for toxicological effects, and to some extent these overlap. The first is to identify a so-called point of departure (POD) or reference point. This may be one of the experimental doses used, or it may represent the dose producing a defined response, for example 5% or 10%. In the former, the most widely used value is the no observed adverse effect level (NOAEL), which has been defined as the highest dose of those tested at which the response is not statistically significantly different from that in the concurrent controls. The advantages and disadvantages of the NOAEL have been reviewed previously (Vermeire et al., 1999; Renwick et al., 2003). It should be noted that the NOAEL (and LOAEL) is, by definition, one of the doses used in the study, and hence is dependent on study design, e.g. dose spacing.

The POD may also be a dose at which there is a defined response. The most commonly used such POD is the benchmark dose (BMD), for a 5% response for continuous data or 10% for quantal data (BMD5 and BMD10), or the statistical lower bound estimate of the 95% confidence interval of the BMD (BMDL5 and BMDL10). The BMD, BMDL5 and BMDL10 are obtained by dose–response modelling, fitting curves with appropriate statistical and/or biological characteristics to the experimental data. The types of curve suitable for this purpose have been reviewed by a number of authors (Filipsson et al., 2003; Sand et al., 2008) and are incorporated into the software packages most widely used for BMD modelling (BMDS and PROAST). Unlike the NOAEL, the BMD is not constrained directly by study design issues such as choice of dose spacing. The BMDL5/10 is often numerically similar to the NOAEL for the study (Bokkers and Slob, 2005).

7.3. Interpretation of point of departure (POD)

Mode of action has little impact in the outcome of a POD determination per se, although one might argue that it would be appropriate to exclude certain dose–response models on the basis of mode of action information (see below). Where mode of action does have a marked consequence in dose–response assessment is in how the relationship is extrapolated from the range of doses used experimentally to the level of human exposure, which may be many orders of magnitude lower. The shape of the dose–response curve at such exposures is not known and is therefore assumed. A major distinction is often made between carcinogenicity resulting, or presumed to result, from DNA-reactive genotoxic compounds and most other effects. In the former case, it is assumed that there is no threshold in the dose–response curve. Some authorities advocate linear low-dose extrapolation as a plausible reasonable worst case for the nature of the dose–response at low exposure levels (Subramaniam et al., 2006). This provides an estimate of a "virtually safe dose" below which there is considered to

be a very low probability of an appreciable risk. Others argue that the uncertainty is so great that such extrapolation is not warranted and that the risk management strategy should be to ensure that exposure is as low as reasonably achievable (ALARA) (Pratt et al., 2009). However, whilst designed to protect public health, this strategy does not provide any indication of whether ALARA will be sufficiently protective, nor does it allow prioritisation of different carcinogenic risks for risk management action. Hence, EFSA and JECFA have proposed that for this purpose the margin of exposure (MOE) could be used (Barlow et al., 2006; Benford et al., 2010), which is the ratio of the POD (BMDL10 is preferred when using data from experimental animals) and a conservative estimate of human exposure. It has been suggested that an MOE of >10,000 would correspond to a low risk and would not be a priority for risk management action (Barlow et al., 2006).

Where the mode of action is such that it would be reasonable to assume a threshold in the dose–response curve, the conventional approach, used by almost all authorities, is to determine the human equivalent exposure in a sensitive subpopulation to the POD determined experimentally (e.g. see Boobis, 2007). In essence, assuming parallel dose–response curves, at least at low responses, the POD from the experimental studies is extrapolated to the population average human by assuming 10-fold greater sensitivity of humans than the experimental animals and then by extrapolating from the average human to a hypothetical sensitive sub-population by assuming a further 10-fold greater sensitivity. In practice this means dividing the POD by 100 (10×10). The factor used has been described as a "safety factor" or "uncertainty factor". It allows for uncertainty regarding inter-species sensitivity and for interindividual variability in human sensitivity. The resulting exposure value has been described generically as a health based guidance value. Examples include the acceptable daily intake (ADI), the (provisional) tolerable daily intake ((P)TDI) and the acute reference dose (ARfD), where exposure is for shorter durations, for example up to 24 h (Solecki et al., 2005). Obviously, the hazard used to determine the POD should be appropriate to the duration of exposure that is of concern.

It should be noted that in establishing health based guidance values from a POD, it is not assumed that the response at the POD is necessarily zero. The assumption is that there is sufficient conservatism in the risk assessment process overall that the health based guidance value will protect the vast majority of the population and that the risk from exposure at or below this value would be negligible. However, the level of protection provided by health based guidance values (percentile of the population in whom there is no risk) has rarely been defined (Hattis et al., 2002) (see below). The level of protection assumed is critically dependant on the assumptions made on the nature of the dose–response relationship for the critical toxicological effect, and in particular whether it can be assumed to exhibit a threshold. Risk analysis policy, for example by Codex Alimentarius, is based on the premise that for most toxicological endpoints of concern to human health it can be assumed there is a threshold dose at and below which there is no response (WHO, 2009). It is perhaps worth commenting that the insights into mechanism being provided by e.g. toxicogenomics in general support this view (e.g. Daston and Naciff, 2010).

In addition to assessing the dose–response relationship, it is important to consider the temporal aspects of the response. Just as there may be a minimum dose necessary to elicit a response, so there may also be a minimum duration of exposure. However, whereas there are accepted (though less than ideal) methods for extrapolating dose from experimental animals to humans, the means of extrapolating duration of exposure is less well established. The most commonly used method is for broad percentages of lifetime, for example 2 years in rats is equivalent to 70 years in humans (100%). A related problem occurs when trying to extrapolate early life exposures, due to differences in comparative developmental stages between experimental animals and humans, such as the appropriate age in rodents with which to compare toddlers, pre-pubescent children, etc. (Makris et al., 2008). Knowledge of mode or mechanism of action can be useful in extrapolating duration of exposure (Luke et al., 2010).

There is evidence that for some effects the response is proportional to dose $*$ time. This has been termed Haber's law (Gaylor, 2000). The difficulty is as before, that it is not easy to extrapolate quantitatively the time component from experimental animals to humans. For many toxicological responses it is now recognised that the dose $*$ time relationship is more complex, and is more proportional to dose $*$ time^n or to even more complex functions (Miller et al., 2000).

7.3.1. Chemical specific information on toxicokinetics

Whereas ideally, extrapolation from experimental animals to humans would be based on detailed toxicokinetic and toxicodynamic information on the compound, there are a number of situations where such information is not available and it would not be feasible to obtain it. However, even without such detailed knowledge it may be possible to modify the extrapolation factor, at least for toxicokinetics, on the basis of partial information. The resultant factors have been described as categorical or pathway-derived, depending on the basis of the information used (Meek, 2001; Dorne and Renwick, 2005). An example of the former would be where knowledge is available that the effect of a compound was dependent on maximum concentration, influenced primarily by rate and extent of absorption, rather than area under the curve (a measure of systemic exposure), which is dependent on clearance processes. As evidence from a variety of compounds suggests that there is less variability in absorption than in clearance, the use of a reduced toxicokinetic factor might be appropriate. The WHO/FAO Joint Meeting on Pesticide Residues (JMPR) has suggested that both the inter-species and inter-individual toxicokinetic factors could be reduced for such a compound by 50% (with the proviso that the compound has a short half-life and that its duration of action is short) (Solecki et al., 2005).

The second type of extrapolation, pathway-derived, utilises information on a number of compounds sharing some toxicokinetic property that is an important determinant of systemic exposure. This is best exemplified by the work of Renwick and colleagues on the routes of elimination of compounds (Dorne and Renwick, 2005). The interindividual and/or inter-species differences in the activities of enzymes of drug metabolism and of major routes of excretion, such as renal, are characterised in sufficient detail that where a compound depends primarily on one of these processes for its elimination, the respective adjustment factor for that route can be obtained from the database.

Whilst extrapolation of information obtained in vitro to the in vivo situation is often difficult and uncertain, in the area of enzyme specificity and variation in metabolic fate, in vitro approaches are now very mature (e.g. Pelkonen and Turpeinen, 2007). A range of in vitro systems is available, including hepatic sub-cellular fractions (particularly the microsomal fraction for cytochrome P450-dependent reactions), isolated hepatocytes, liver-derived cell lines and recombinant expressed enzymes. The utility of such systems has been reviewed several times in recent years (Vermeir et al., 2005; Li, 2007; Donato et al., 2008; Gómez-Lechón et al., 2008; Laine, 2008; Chiba et al., 2009). Using such systems it is possible to obtain an estimate of quantitative inter-species differences and of the contribution of a given enzyme pathway to the elimination of a compound. Such information may be supplemented by limited in vivo studies, either in experimental animals or in humans. This could then form the basis for an

adjustment to the factor used for extrapolating from the POD to obtain a health based guidance value.

Whilst much useful information on human toxicokinetics can be obtained using in vitro systems, information obtained in vivo in humans is obviously invaluable, when available. Such information may arise from experimental studies in humans, for example in clinical trials, or through incidental observation, for example following occupational exposure or in rare cases from epidemiological or other observational studies of the general population (Hays and Aylward, 2009). In the former, the conditions of exposure can be very well defined and an optimal sampling plan to characterise the kinetics of the compound is often possible. In the latter, exposure is often not well defined and the sampling plan may well be sub-optimal. Nevertheless, it may still be possible to utilise such information to derive quantitative chemical-specific toxicokinetic adjustment factors, supported by suitable studies in experimental animals or in vitro, if necessary.

7.3.2. Chemical specific information on toxicodynamics

As indicated above, chemical-specific information on toxicodynamics is most likely to arise from studies in vitro. Hence, if the toxicological target, such as a specific cell type, receptor, enzyme or ion channel is available for study in vitro, it may be possible to obtain quantitative data on species differences in response (WHO, 2005). Systems suitable for this purpose may be from human subjects, if feasible and ethically acceptable or more likely from recombinant technology or from cultured cell lines. Such studies can be enhanced if the mode of action of the compound has been established. This enables quantitative comparison of key events between experimental animals and humans, either in vitro or in vivo (Boobis et al., 2009; Luke et al., 2010). Information on the quantitative contribution of such processes to adverse health outcomes may be available from the general medical literature on the consequences of specific genetic polymorphisms, disease conditions of known aetiology or the effect of defined exposures such as dietary factors or therapeutic drugs, although in this last case there will be likely confounding due to pharmacokinetic differences.

7.4. Extrapolation of dose–response relationship to humans

As discussed above, for effects for which it is reasonable to assume a threshold, extrapolation of the dose response relationship obtained from studies in experimental animals to humans (i.e. to establish a health based guidance value) can be achieved by applying an uncertainty or safety factor to a fixed point on the dose response curve, the point of departure (e.g. the NOAEL). The factor applied may be modified for a number of reasons, including absence of a clear NOAEL, duration of dosing did not match the scenario of concern in humans, quality of, or deficiencies in, the database, and for policy driven reasons (e.g. Boobis, 2007). If human exposure is estimated to be below this health based guidance value, the conclusion is that there is no appreciable human health risk. However, when exposure exceeds the health based guidance value, this approach is not suitable for estimating the associated risk.

When dose–response modelling is used to determine the point of departure (i.e. the BMD(L)), the response at any point on the dose–response curve can be calculated. If one assumes, as when establishing health based guidance values, that the dose–response curve in an average and in a sensitive human is parallel to that in experimental animals, the response in humans exposed above the health based guidance value can be determined by shifting the dose–response curve by the appropriate uncertainty factor. However, this would only apply when exposure was less than 10% of the POD, otherwise one would be at the top of the extrapolated dose–response curve for a sensitive subpopulation. Hence, in practice, where exposure exceeds a health based guidance value, the potential human response would be estimated by multiplying the human exposure level by the appropriate uncertainty factor, adjusted where possible using categorical, pathway-derived or chemical specific information, and using the dose–response model in experimental animals to determine the response at the calculated dose.

Where no threshold can be assumed, essentially for DNA-reactive genotoxic carcinogens, either low dose linear extrapolation or the margin of exposure is used to estimate relative risk. In the former case, it is a simple matter to calculate the risk at any given exposure level above the virtually safe dose, although this will most likely be a very conservative estimate. Nevertheless, it could serve as the basis of a lower tier benefit–risk assessment. Where the margin of exposure is preferred, this cannot be readily converted into a quantifiable risk. However, it would be possible to use the MOE to compare different benefit–risk scenarios, particularly when the MOE is >10,000, i.e. the risk can be assumed to be below that which would be a priority for risk management.

Extrapolation of the dose response curve could be refined by using some estimate of internal dose, for example the area under the plasma concentration–time curve (AUC), as the basis of the dose–response curve (e.g. Valcke and Krishnan, 2011). This would remove some of the uncertainty associated with extrapolating external doses from experimental animals to human exposure levels. Factors such as the extent of absorption and pre-systemic metabolism may affect the accuracy of such extrapolation. However, all of these methods require a number of assumptions and they are associated with appreciable uncertainty. In higher tier assessments, more 'sophisticated' approaches may be used to reduce the uncertainty in extrapolating dose–response relationships from experimental animals to humans. Modelling approaches could be used for either, or both, toxicokinetics and toxicodynamics. In the case of toxicokinetics, physiologically-based toxicokinetic approaches (PBTK) can be very powerful, allowing an estimate of target tissue exposure, based on chemical-specific and biological knowledge, for example on tissue volumes, local perfusion rates, tissue partitioning, absorption and clearance processes (Clewell and Clewell, 2008; Kim and Nylander-French, 2009). PBTK enables a number of additional factors that might influence target tissue exposure to be explored and quantified, including the impact of interindividual variability in kinetic parameters, the consequences of different patterns of exposure, with respect to duration and frequency, and lifestage differences. It is possible to construct PBTK models on the basis of in vitro and in vivo data from studies in experimental animals and in vitro data from humans.

Whilst PBTK modelling can reduce uncertainty associated with inter- and intra-species differences in kinetics, it does not overcome the uncertainties associated with the toxicodynamic response. For this purpose, toxicodynamic (TD) models are necessary, preferably linked to a PBTK model, so that the target tissue response to target tissue levels of the compound can be evaluated (El-Masri, 2007). These are often known as biologically based dose–response models. However, it is also possible to model the toxicodynamic response at the level of the whole organism, although this will be associated with greater uncertainty. Data for TD models can be obtained from characterisation of key events, where the mode of action is known (Conolly et al., 2004; Luke et al., 2010). Such models will be mechanistically-based. Otherwise TD models will need to be empirical, based on the observed relationships between exposure, perhaps target tissue dose determined using PBTK, and response (see Conolly and Andersen, 1991). Where the mode of action is known or suspected, the key events can be used to help in the identification and selection of biomarkers of exposure, effect and susceptibility (Gundert-Remy et al., 2005). Such biomarkers may be of use in studies of the human population or for comparison across species for differences in either (or both)

internal exposure or intermediate effects, following exposure to a given external dose. There is considerable interest in the development of systems-based TD models, based on detailed knowledge of the signalling pathways and other molecular events responsible for a toxicological effect, using findings from e.g. toxicogenomics studies (e.g. Zhang et al., 2010). However, it will be some time before such approaches are sufficiently reliable that they will be able to form the basis of a benefit–risk assessment.

The possibility of using PBTK to estimate interindividual variability was mentioned above. This is best achieved using a probabilistic approach. The variables of interest in the PBTK model, such as enzyme-specific metabolic disposition, are represented by distributions rather than by discrete values. The range of possible outcomes is then determined by random sampling of the various distributions, using Monte Carlo methods, to construct a distribution of, for example, target tissue levels for a given external dose (Jamei et al., 2009). Similar approaches would be applicable to toxicodynamics and to combined TK-TD models. Whilst human specific information will be needed to define the distributions for some of the parameters used in probabilistic modelling, for others it may be possible to use one or a small number of discrete values, obtained in vivo or in vitro in humans or perhaps in experimental animals. This would be appropriate where a parameter does not contribute much to overall variability, as determined by sensitivity analysis, or where a parameter shows little variation, either between individuals or between species. It should be stressed that whilst such models can reduce uncertainty in dose–response characterisation and in inter-species extrapolation, currently their use does not appreciable reduce the uncertainty associated with extrapolation to very low doses (i.e. extrapolation over many orders of magnitude). This is because the biological determinants of the dose–response curve in this range are not known (see Crump et al., 2010).

Probabilistic approaches can also be used to obtain estimates of the range of risk rather than a discrete estimate of risk. Hence, the uncertainty factors used in the deterministic approaches can be replaced by distributions, and rather than using worst case by multiplying them together, they can take a range of values, alone and in combination (Slob and Pieters, 1998; Kodell and Gaylor, 1999). Such modelling allows an estimate of the level of protection, but this requires a number of assumptions for which there is currently little experimental support. Hence, their use in risk assessment, and in benefit–risk assessment, is not yet widespread.

8. Case studies

The BRAFO tiered approach, previously described in this paper, was applied to three series of case studies – on natural foods, dietary interventions and heat processing of foods, respectively. The key characteristics and conclusion of these case studies are summarised in Table 4.

The first series of case studies, on natural foods, focused on farmed salmon and soy protein. Nutritional recommendations emphasise the importance of consumption of oily fish as well as of plant protein sources such as soy (Watzl et al., 2012). However, due to the presence of contaminants in oily fish (e.g., PCBs) and anti-nutritive constituents in soy, these natural foods represent cases of relevant benefit–risk issues. These two cases of natural foods demonstrated that the BRAFO tiered approach enabled a qualitative benefit–risk assessment at lower tiers, indicating that the benefit outweighs the risks, highlighting the need to consider both risks and benefits in order to establish the net health impact associated with the consumption of specific food products.

The second group of case studies consisted of an assessment of the number and specific groups of people affected by the potential beneficial and adverse effects across actual intake levels of folic acid, including a description of the severity and the probability of the effects occurring when consuming food subject to dietary fortification (Verhagen et al., 2012a). This case study required assessment using tiers 3/4, as quantitative data were needed to come to an overall conclusion on benefits and risks. Another case study was on the replacement of mono- and disaccharides in sugar-sweetened beverages with low calorie sweeteners in which a conclusion could be reached at tier 2, i.e. benefits outweighed possible risks. There were two case studies on macronutrient replacement/food substitution i.e. the isocaloric replacement of saturated fatty acids with carbohydrates and the replacement of saturated fatty acids with monounsaturated fatty acids. However, these stopped at tier 1 because, in all cases, the endpoint was the risk of cardiovascular disease and hence these were not true benefit–risk questions, emphasising the importance of problem formulation. A final test case was on the chlorination of drinking water, which was followed to tiers 3/4, without being able to reach a clear conclusion, due to the unavailability of suitable data.

The third series of case studies comprised three different examples of heat processing of food that provided interesting benefit–risk questions (Schütte et al., 2012).

The first of these was on the formation and possible reduction of acrylamide in potato and cereal based products. In tier 1, resolution of the benefit–risk assessment was not possible. Tier 2 of the BRAFO methodology enabled tentative conclusions on the desirability of mitigation efforts, but it was not possible to characterise the overall benefits and risks with sufficient certainty to enable a final conclusion. There were insufficient data to enable the corresponding DALY changes for the effects of concern to be calculated and hence higher tier assessments were not possible.

The second case study, on benzo(a)pyrene, looked at alternative food processing options besides grilling and smoking of meat and fish and whether they would change the benefit–risk ratio with regard to benzo(a)pyrene levels formed. Tier 1 assessment enabled the conclusion that two of the three alternative scenarios considered would provide a potential benefit whilst minimising the potential risks. A higher tier evaluation of the third scenario was not possible, due to lack of data.

The third of these case studies was on the heat-treatment of milk. Tier 1 assessment led to the conclusion that pasteurisation and UHT treatment of milk provide the benefit of a microbiologically safe product, whilst the risks from chemical changes through high thermal impact were low to negligible. Hence, it was concluded that heat treatment of milk has appreciable benefits that outweigh potential risks. The lack of physiological data with relevant dose–response-relationships would not have allowed higher tier assessments in this case study.

The case studies demonstrated the applicability of the BRAFO Tiered approach to various real benefit–risk issues in the field of public health nutrition. The respective examples illustrated how the BRAFO Tiered approach resulted in various outcomes, ranging from a quick stop as the result of an early, clear conclusion that benefit outweighs risk, to continuation through the tiers into deterministic/probabilistic calculations. Well-defined reference and alternative scenarios, as well as target populations, were essential in applying the tiers of the framework.

The results on the applicability of the BRAFO approach on the case studies were presented and discussed at a workshop in October 2009 in order to adapt the methodology according to the findings of the case studies. The methodology appeared to be applicable to all case studies, in many cases enabling conclusions to be reached on the balance of risks and benefits. However, areas for further improvement were identified. For example it was considered that more guidance on problem definition and specification of scenarios, for which there must be at least two to enable a comparison beyond tier 1, were needed, and of the various tiers,

Table 4
Summary of different case studies developed in the BRAFO project.

Case study	Reference (R) and Alternative (A) Scenario	Genuine benefit-risk question	Problem formulation	T1	T2	T3/4	Comments
Farmed salmon	R No intake A 200 g of farmed salmon (weekly intake)	Yes	√	√	√		Benefits and risks identified. Benefits (of omega-3 fatty acids, vitamin D) outweigh risks (of methyl mercury, PCB's and dioxins) since levels of contaminants remain below safe intake levels. Stop after Tier 2
Soy protein	R No intake A 25 g of soy protein (daily intake)	Yes	√	√	√		Benefits and risks identified. Beneficial effects predominantly due to soy protein. These clearly exceed adverse (and beneficial) effects of isoflavones. Stop after Tier 2
Addition of folic acid to flour/bread	R No fortification A fortification of bread with folic acid at 70 μg/100 g	Yes	√	√	√		Benefits and risks identified. In Tier 3, a scenario was identified in which the benefits (prevention of neural tube defects) substantially outweigh the risks (masking of vitamin B12 deficiency, cancer). Deterministic computation sufficient. Stop after Tier 3
Replacement of saturated fatty acids by mono-unsaturated fatty acids	R Two levels of intake of SAFA – high and low A substitution of 5% of calories of SAFA	No	X				Not a genuine benefit-risk question (only one endpoint involved). Stop after Tier 0
Replacement of saturated fatty acids by carbohydrates (isocalorically)	R Two levels of intake of SAFA – high and low A substitution of 5% of calories of CHO	No	X				Not a genuine benefit-risk question (only one endpoint involved). Stop after Tier 0
Replacement of mono-and disaccharides by low calorie sweeteners (LCS's)	R Sugar sweetened beverages A substitution with artificial sweetened beverages	Yes	√	√	√		Benefits and risks identified. As intake of LCS's remains below ADI levels essentially no risks at intake levels. Stop after Tier 2
Addition of chlorine to water	R Water not treated A water chlorination at different level	Yes	√	√		(√)	Benefits and risks identified. Quantitative comparison falls short because of lack of suitable scenarios and underlying data. Stop in Tier 3
Acrylamide	R Cooking practices before AA-mitigation measures were introduced A AA-mitigation methods	Yes	√	√	√	(√)	Benefits and risks identified. Quantitative comparison falls short because of complexity of case study. Stop in Tier 3
Benzo(a)pyrene	R Smoking and grilling as per current practice in home A addition of artificial smoke flavours	Yes	√	√			Benefits and risks identified. Any alternative scenario will decrease levels of B(a)P. Stop after Tier 1
Heat treatment of milk	R Raw cow's milk A UHT milk	Yes	√	√			Benefits and risks identified. Potential risks very low/negligible. Stop after Tier 1

particularly Tier 2 and adding a specific section on the potential role for uncertainty factors.

9. Output

The results of a benefit–risk assessment should be given in both a detailed and a concise narrative report. The narrative report should contain the question posed by the benefit–risk manager to the benefit–risk assessor and the final formulation of the problem that was the subject of the assessment, the description of the scenarios to be compared and the population group(s) to be considered specifically. The source, quantity and quality of data used in the assessment should be given and data gaps should be stated as well as the nature of the assumptions made at the different steps of the assessment to compensate for insufficient human data, uncertainties in the assessment, including in the extrapolation of animal data to humans and classification of weights for disability and quality of life, respectively. In the case of a quantitative benefit–risk assessment (Tier 3 or 4), where benefits and risks are integrated by using common metrics, translation of the outcome into language that states in what way and to what extent benefits dominate the risks (or vice versa) for the population under consideration in a defined scenario is needed. This may require presentation of several different metrics, not just QALYs or DALYs. An indication of which scenario has the greater beneficial health impact should be stated at the end of the report.

10. Discussion and conclusions

In the course of the BRAFO project, a range of benefit–risk assessments of foods and food compounds was undertaken, covering relatively diverse scenarios. In some, it was possible to reach a conclusion in the earliest tier, but in others it was necessary to continue to higher tiers to enable satisfactory resolution of the problem. These case studies established that the benefit–risk methodology proposed (Hoekstra et al., 2010) was applicable to

many different scenarios, including natural foods, but that considerable expertise and judgment were often required in the conduct of such assessments, with significant demands on both time and resources, such as the personnel involved. As in risk assessment, interpretation of the outcome of the different tiers of a benefit–risk assessment requires both knowledge in a number of areas and experience. Obviously, the range of expertise necessary will vary with the question being addressed, as was apparent in developing the case studies, where teams with specific expertise relevant to the cases were assembled. It would have been very difficult, if not impossible, to have completed the assessments without the availability of such expertise. This has significant implications for problem formulation and the identification of necessary and available resources for the conduct of a benefit–risk assessment.

The tiered approach was invaluable in maximising the use of resources and in helping avoiding unnecessary effort. In some scenarios it was possible to stop very early in the assessment, at Pre-Assessment/Problem Formulation, thus obviating the need for detailed data retrieval and time-consuming evaluation. In others, problem resolution could be achieved at Tier 1, which requires only modest effort in the assessment. In contrast, application of the tiered approach in some scenarios clearly established that problem resolution was not currently possible and that this would require computationally-intensive higher tier assessment for which data were not available. The tiered assessment enabled clear identification of data needs to enable problem resolution (see Table 4).

Benefit–risk assessment methodology such as that developed within the BRAFO project enables many scenarios to be explored. This could be used to optimise the trade-off between benefit and risk. However, this is not always the policy question of concern. For example, it might be that the interest is in a proposed specific level of food fortification or in the current level of chemical contamination of a food. In the former, the question would be to compare the benefits and risks resulting from such fortification with those arising from current consumption. In the latter, the question might be whether the benefits from current consumption outweigh the risks.

Problem formulation is therefore critical in benefit–risk assessment, as emphasised by EFSA in their recent guidance on this topic (EFSA Scientific Committee, 2010). During this stage, policy managers should identify the key benefit–risk question, for example is this optimisation of benefit versus risk or comparison of benefit and risk for a specified scenario. The scope of the benefit–risk assessment therefore needs to be clearly defined and understood by both assessors and managers, such that the output of the agreed assessment will provide information relevant to the policy question of concern. This will require dialogue between assessors and managers to ensure that the terms of reference for the assessment are appropriate and transparent. Benefit–risk assessment should address all of the scenarios covered in problem formulation and should clearly identify each scenario and the benefits and risks associated with these. The output of a benefit–risk assessment is not a single number, but is multi-dimensional. Hence, information should be provided not only on common currency (composite metrics), such as QALYs or DALYs, for example when these have been used in higher tier assessment, but also on the nature of the effects, the affected populations and the numbers of individuals likely to be affected, for both benefits and for risks. The uncertainty associated with the various dimensions of benefit and of risk should also be clearly reported. The decision on the relative acceptability of benefits versus risks is a policy decision, taken by managers, and hence the assessor should avoid direct estimates of net benefit or net risk without explaining the underlying data and assumptions.

As emphasised by the EFSA Scientific Committee (2010), the above demonstrates the importance of dialogue between assessors and managers, not only during problem formulation, but also when moving from one tier of the assessment to another. This interaction with managers was one aspect of benefit–risk assessment that could not be tested during the conduct of the BRAFO case studies, but its importance was readily apparent from the nature of the decisions that had to be made during the tiered assessment.

Whilst it was possible to address all of the case studies developed with the proposed methodology (Table 4), it was apparent that there are areas where additional methodological development and agreement are required. Often, there was a lack of suitable data, either on benefit or on risk, to enable other than a lower tier assessment to be undertaken, which was insufficient for problem resolution but simply highlighted the data gaps. In many of the case studies addressed, there was a lack of human data, the availability of which would have substantially reduced uncertainty and would have improved the benefit–risk assessment considerably. In part, the difficulty is that in higher tiers, benefit–risk assessment requires quantitative estimates of benefit and of risk. Hence, information on dose–response is necessary, which is frequently lacking, many human intervention studies involving only a single exposure level. In the absence of information on biological gradient, it is not possible to conduct a quantitative assessment, such as those involving QALYs or DALYs.

An additional data gap identified in some of the case studies was the absence of health weights for some effects. Indeed, whilst existing health weights for QALYs and DALYs have been used in benefit–risk assessments to date, it became apparent that there are a number of issues that need to be resolved if they are to be used more widely in benefit–risk assessment of foods and compounds in food. Existing health weights were developed for purposes other than benefit or risk assessment and are often not ideally suited to this purpose. The basis of determining health weights for calculating QALYs differs from that used for determining health weights for calculating DALYs, most notably in those upon whose judgment the weights are based. In general, health weights are applied to the general population, with little or no consideration of sub-population, geographical or socio-economic differences. Health weights have not been determined for a number of health effects, including many positive health effects. There is no general agreement on how to apply health weights to data obtained from experimental animals, much less from in vitro or in silico methods. Hence, there is clearly a need for the issue of health weights, as applied to benefit–risk assessment, to be reviewed in detail and updated as necessary.

The BRAFO methodology has a high demand for human data on disease incidences, which may not be available for the target population subject to a benefit–risk assessment. The BRAFO approach argues for comprehensive health statistics data in all (or at least adequately representative) European countries.

A difficulty of obtaining human data for benefit–risk assessment of food and chemicals in food is the resource that would be required to conduct the necessary studies, particularly for potential long-term effects. It is also ethically unacceptable to obtain information on potential risks by intentionally exposing subjects at levels such that risk would be anticipated. For these, and other, reasons, there is considerable potential in the use of suitably qualified biomarkers in benefit–risk assessment. Such biomarkers should have an established relationship to health effects of interest, should provide information earlier than the health outcome itself and should be more sensitive than conventional health effects measures. Development of such biomarkers is not trivial, but their availability would greatly improve benefit–risk assessment. It is recognised that use of biomarkers, particularly in higher tier assessments, would require agreement on the appropriate health weights to be utilised. Suitable biomarkers of exposure and of susceptibility could also improve benefit–risk assessment.

An often encountered difficulty in the case studies was that whereas data on benefit were available from studies in humans,

those on risk were from studies in experimental animals. Whilst such information is widely used for risk assessment, this is usually for the purposes of establishing a level of exposure below which there is considered to be a negligible risk of harm, e.g. a health based guidance value. Such assessments often include conservative assumptions, for example in the authorisation of a new pesticide or food additive. However, in benefit–risk assessments beyond Pre-Assessment/Problem Formulation, the nature of the problem is such that it is axiomatic there is no exposure associated with benefit that is without some risk. Hence, the conservatism of the assumptions used in estimating risk should not be markedly disproportionate from those used for assessing benefits, otherwise it is not possible to perform a meaningful comparison. When using data from experimental animals to estimate potential human risk, there are several major uncertainties, including time of onset of disease, disease severity, dose–response relationship in humans and of course relevance to humans. Whilst methods are available, or are being developed, to address some of these issues, there is still a lack of agreement of the underlying approaches and assumptions that should be used when utilising such data in benefit–risk assessment. With the increasing interest in the use of non-animal methods, such as in silico and in vitro, in toxicity testing, these difficulties are likely to be compounded in the near future. Methods such as physiologically-based toxicokinetics can help in the extrapolation of dose, but quantitative extrapolation of response remains problematic, particularly with respect to the time-dependent aspects, discussed above.

In conducting a benefit–risk assessment, the same criteria should be applied in assessing the strength of the evidence for benefits as for risks. Hence, uncertain benefits should be given equal weight to uncertain risks and similarly more certain benefits should be considered with the same weight as more certain risks. However, this will rarely apply in benefit–risk management, where the evidence for the presence of beneficial effects usually needs to be convincing, whereas the evidence for the presence of adverse effects may be anything from convincing to probable, possible or even insufficient, because of the need to consider a high level of protection of public health (Tijhuis et al., 2012a).

The BRAFO methodology was designed for assessing the direct effects of benefits and risks on health and it was used for this purpose in all of the case studies investigated. However, the scope of a benefit–risk question is often wider than direct health impact, and can include socio-economic, psychological and/or environmental dimensions. Comparison of these additional dimensions would need a common currency, different from those related to health impact, such as the DALY or QALY. However, the advantage of the methodology proposed by BRAFO is that the health metrics can be readily converted into other currencies, for example Euros. Once conversion factors have been agreed, the methodology could be adapted to indirect health impacts and other effects relevant to a benefit–risk assessment. Obviously, such assessments would be more complex than those involving direct health impacts, and appropriate consideration would need to be given to problem formulation.

The benefit–risk assessment methodology developed by BRAFO (Hoekstra et al., 2010) is fit for purpose. It is suitable for a range of benefit–risk problems that are of current concern. However, a number of issues remain to be resolved. Given the value of benefit–risk assessment when applied appropriately, there is need for further work to address these concerns and enhance the utility of the methodology.

11. Conflict of Interest

For those experts affiliated with academic institutions, the Commission of the European Communities covered, through ILSI Europe, the travel and accommodation costs related to their participation in the BRAFO project. H.P., J.S. and P.L. received an honorarium for writing this paper. None of the authors declared any interest that may conflict with the provision of their solely scientific input to this manuscript.

Acknowledgements

This study has been carried out with partial financial support by the Commission of the European Communities, Priority 5 Food Quality and Safety, within the Sixth Framework Programme (Contract No: 031731 BRAFO: Benefit Risk Analysis of Foods). This manuscript does not necessarily reflect the views of the Commission and in no way anticipates the future policy in this area. The preparation of this manuscript was coordinated by ILSI Europe. ILSI Europe would like to thank all of the contributors to the BRAFO Scientific Supported Action. Overall, we would like to thank the ILSI Europe Risk Assessment of Chemicals in Food Task Force members and the BRAFO Steering Committee members for their support and guidance. Finally we would like to thank the European Commission for their financial support. As co-ordinator of the Scientific Support Action, ILSI Europe would like to express its profound gratefulness to all of them.

References

Aggett, P.J., Antoine, J.M., Asp, N.G., Bellisle, F., Contor, L., Cummings, J.H., Howlett, J., Müller, D.J., Persin, C., Pijls, L.T., Rechkemmer, G., Tuijtelaars, S., Verhagen, H., 2005. PASSCLAIM: consensus on criteria. Eur. J. Nutr. 44 (Suppl. 1), 15–30.

Albers, R., Antoine, J.M., Bourdet-Sicard, R., Calder, P.C., Gleeson, M., Lesourd, B., Samartin, S., Sanderson, I.R., Van Loo, J., Vas Dias, F.W., Watzl, B., 2005. Markers to measure immunomodulation in human nutrition intervention studies. Br. J. Nutr. 94, 452–481.

Bakker, G.C., van Erk, M.J., Pellis, L., Wopereis, S., Rubingh, C.M., Cnubben, N.H., Kooistra, T., van Ommen, B., Hendriks, H.F., 2010. An antinflammatory dietary mix modulates inflammation and oxidative and metabolic stress in overweight men: a nutrigenomics approach. Am. J. Clin. Nutr. 91, 1044–1059.

Barendregt, J.J., Veerman, J.L., 2010. Categorical versus continuous risk factors and the calculation of potential impact fractions. J. Epidemiol. Commun. Health 64 (3), 209–212.

Barlow, S.M., Greig, J.B., Bridges, J.W., Carere, A., Carpy, A.J., Galli, C.L., Kleiner, J., Knudsen, I., Koëter, H.B., Levy, L.S., Madsen, C., Mayer, S., Narbonne, J.F., Pfannkuch, F., Prodanchuk, M.G., Smith, M.R., Steinberg, P., 2002. Hazard identification by methods of animal-based toxicology. Food Chem. Toxicol. 40, 145–191.

Barlow, S., Renwick, A.G., Kleiner, J., Bridges, J.W., Busk, L., Dybing, E., Edler, L., Eisenbrand, G., Fink-Gremmels, J., Knaap, A., Kroes, R., Liem, D., Müller, D.J., Page, S., Rolland, V., Schlatter, J., Tritscher, A., Tueting, W., Würtzen, G., 2006. Risk assessment of substances that are both genotoxic and carcinogenic report of an international conference organized by EFSA and WHO with support of ILSI Europe. Food Chem. Toxicol. 44, 1636–1650.

Benford, D., Bolger, P.M., Carthew, P., Coulet, M., DiNovi, M., Leblanc, J.C., Renwick, A.G., Setzer, W., Schlatter, J., Smith, B., Slob, W., Williams, G., Wildemann, T., 2010. Application of the Margin of Exposure (MOE) approach to substances in food that are genotoxic and carcinogenic. Food Chem Toxicol. 48 (Suppl. 1), S2–S24.

Bokkers, B.G., Slob, W., 2005. A comparison of ratio distributions based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. Toxicol. Sci. 85, 1033–1040.

Boobis, A.R., 2007. Risk assessment of dietary supplements. Novartis Found Symp. 282, 3–25.

Boobis, A.R., Cohen, S.M., Dellarco, V., McGregor, D., Meek, M.E., Vickers, C., Willcocks, D., Farland, W., 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit. Rev. Toxicol. 36, 781–792.

Boobis, A.R., Doe, J.E., Heinrich-Hirsch, B., Meek, M.E., Munn, S., Ruchirawat, M., Schlatter, J., Seed, J., Vickers, C., 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Crit. Rev. Toxicol. 38, 87–96.

Boobis, A.R., Daston, G.P., Preston, R.J., Olin, S.S., 2009. Application of key events analysis to chemical carcinogens and noncarcinogens. Crit. Rev. Food Sci. Nutr. 49, 690–707.

Briggs, D.J., 2008. A framework for integrated environmental health impact assessment of systemic risks. Environ. Health 7, 61.

Chiba, M., Ishii, Y., Sugiyama, Y., 2009. Prediction of hepatic clearance in human from in vitro data for successful drug development. AAPS J. 11, 262–276.

Clewell, R.A., Clewell 3rd, H.J., 2008. Development and specification of physiologically based pharmacokinetic models for use in risk assessment. Regul. Toxicol. Pharmacol. 50, 129–143.

Commission of the European Communities, 2000. White paper on food safety. Brussels, 12 January 2000. Available from: http://ec.europa.eu/dgs/health_consumer/library/pub/pub06_en.pdf.

Conolly, R.B., Andersen, M.E., 1991. Biologically based pharmacodynamic models: tools for toxicological research and risk assessment. Annu. Rev. Pharmacol. Toxicol. 31, 503–523.

Conolly, R.B., Kimbell, J.S., Janszen, D., Schlosser, P.M., Kalisak, D., Preston, J., Miller, F.J., 2004. Human respiratory tract cancer risks of inhaled formaldehyde: dose-response predictions derived from biologically-motivated computational modelling of a combined rodent and human dataset. Toxicol. Sci. 82, 279–296.

Crump, K.S., Chen, C., Chiu, W.A., Louis, T.A., Portier, C.J., Subramaniam, R.P., White, P.D., 2010. What role for biologically based dose-response models in estimating low-dose risk? Environ. Health Perspect. 118, 585–588.

Daston, G.P., Naciff, J.M., 2010. Predicting developmental toxicity through toxicogenomics. Birth Defects Res. C Embryo Today 90, 110–117.

Dodd, K.W., Guenther, P.M., Freedman, L.S., Subar, A.F., Kipnis, V., Midthune, D., Tooze, J.A., Krebs-Smith, S.M., 2006. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. J. Am. Diet. Assoc. 106, 1640–1650.

Donato, M.T., Lahoz, A., Castell, J.V., Gómez-Lechón, M.J., 2008. Cell lines: a tool for in vitro drug metabolism studies. Curr. Drug Metab. 9, 1–11.

Dorne, J.L., Renwick, A.G., 2005. The refinement of uncertainty/safety factors in risk assessment by the incorporation of data on toxicokinetic variability in humans. Toxicol. Sci. 86, 20–26.

Dybing, E., Doe, J., Groten, J., Kleiner, J., O'Brien, J., Renwick, A.G., Schlatter, J., Steinberg, P., Tritscher, A., Walker, R., Younes, M., 2002. Hazard characterisation of chemicals in food and diet. Dose response, mechanisms and extrapolation issues. Food Chem. Toxicol. 40, 237–282.

Edler, L., Poirier, K., Dourson, M., Kleiner, J., Mileson, B., Nordmann, H., Renwick, A., Slob, W., Walton, K., Würtzen, G., 2002. Mathematical modelling and quantitative methods. Food Chem. Toxicol. 40, 283–326.

EFSA Contam, 2010. Statement on possible public health risks for infants and young children from the presence of nitrates in leafy vegetables. EFSA J. 8 (12), 1935. http://dx.doi.org/10.2903/j.efsa.2010.1935, www.efsa.europa.eu/en/efsajournal/pub/689.htm.

EFSA NDA, 2011a. Nutrition and health claims. Avaliable from: http://www.efsa.europa.eu/en/nda/ndaclaims.htm. Accessed 16/08/11.

EFSA NDA, 2011b. General guidance for stakeholders on the evaluation of Article 13.1, 13.5 and 14 health claims. EFSA J. 9, 2135–2159, http://www.efsa.europa.eu/en/efsajournal/doc/2135.pdf.

EFSA Scientific Committee, 2010. Guidance on human health risk–benefit assessment of food. EFSA J. 8, 1673–1713.

Elliott, R., Pico, C., Dommels, Y., Wybranska, I., Hesketh, J., Keijer, J., 2007. Nutrigenomic approaches for benefit–risk analysis of foods and food components: defining markers of health. Br. J. Nutr 98, 1095–1100.

El-Masri, H.A., 2007. Experimental and mathematical modeling methods for the investigation of toxicological interactions. Toxicol. Appl. Pharmacol. 223, 148–154.

E.U. Regulation, 2006. REGULATION (EC) No 1924/2006 Of The European Parliament and of the Council of 20 December 2006. Official Journal of the European Union. L 404/9.

Filipsson, A.F., Sand, S., Nilsson, J., Victorin, K., 2003. The benchmark dose method – review of available models, and recommendations for application in health risk assessment. Crit. Rev. Toxicol. 33, 505–542.

Gaylor, D.W., 2000. The use of Haber's law in standard setting and risk assessment. Toxicology 149, 17–19.

Gómez-Lechón, M.J., Castell, J.V., Donato, M.T., 2008. An update on metabolism studies using human hepatocytes in primary culture. Expert Opin. Drug Metab.Toxicol. 4, 837–854.

Gundert-Remy, U., Dahl, S.G., Boobis, A., Kremers, P., Kopp-Schneider, A., Oberemm, A., Renwick, A., Pelkonen, O., 2005. Molecular approaches to the identification of biomarkers of exposure and effect – report of an expert meeting organized by COST Action B15. November 28, 2003. Toxicol. Lett. 156, 227–240.

Guzelian, P.S., Victoroff, M.S., Halmes, N.C., James, R.C., Guzelian, C.P., 2005. Evidence-based toxicology: a comprehensive framework for causation. Hum. Exp. Toxicol. 24, 161–201.

Hart, A., Hoekstra, J., Owen, H., Kennedy, M., Zeilmaker, M., de Jong, N., Gunnlaugsdottir, H., 2013. Qalibra: a general model for food risk–benefit assessment that quantifies variability and uncertainty. Food Chem. Toxicol. http://dx.doi.org/10.1016/j.fct.2012.11.056.

Hattis, D., Baird, S., Goble, R., 2002. A straw man proposal for a quantitative definition of the RfD. Drug Chem. Toxicol. 25, 403–436.

Hays, S.M., Aylward, L.L., 2009. Using biomonitoring equivalents to interpret human biomonitoring data in a public health risk context. J. Appl. Toxicol. 29, 275–288.

Hill, B., 1965. The environment of disease: association or causation? Proc. R. Soc. Med. 58, 295–300.

Hoekstra, J., Hart, A., Boobis, A., Claupein, E., Cockburn, A., Hunt, A., Knudsen, I., Richardson, D., Schilter, B., Schütte, K., Torgerson, P.R., Verhagen, H., Watzl, B., Chiodini, A., 2010. BRAFO tiered approach for benefit–risk assessment of foods. Food Chem. Toxicol.. http://dx.doi.org/10.1016/j.fct.2010.05.04.

Hoekstra, J., Hart, A., Owen, H., Zeilmaker, M., Bokkers, B., Thorgilsson, B., Gunnlaugsdottir, H., 2012a. Fish, contaminants and human health: quantifying and weighing benefits and risks. Food Chem. Toxicol. 50, S684–S698.

Hoekstra, J., Hart, A., Owen, H., Zeilmaker, M., Bokkers, B., Thorgilsson, B., Gunnlaugsdottir, H., 2012b. Fish, contaminants and human health: quantifying and weighing benefits and risks. Food Chem. Toxicol.. http://dx.doi.org/10.1016/j.fct.2012.01.013.

Howlett, J., Shortt, C., 2004. PASSCLAIM – report of the Second Plenary Meeting: review of a wider set of interim criteria for the scientific substantiation of health claims. Eur. J. Nutr. 43 (Suppl. 2), 174–183.

Hulzebos, E., Gunnarsdottir, S., Rila, J.P., Dang, Z., Rorije, E., 2010. An integrated assessment scheme for assessing the adequacy of (eco)toxicological data under REACH. Toxicol. Lett. 198, 255–262.

Institute of Medicine, 2010. Evaluation of biomarkers and surrogate endpoints in chronic disease. The National Academies Press, Washington, D.C..

Jamei, M., Dickinson, G.L., Rostami-Hodjegan, A., 2009. A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of 'bottom–up' vs 'top–down' recognition of covariates. Drug Metab. Pharmacokinet. 24, 53–75.

Jenab, M., Slimani, N., Bictash, M., Ferrari, P., Bingham, S.A., 2009. Biomarkers in nutritional epidemiology: applications, needs and new horizons. Hum. Genet. 125, 507–525.

Julien, E., Boobis, A.R., Olin, S.S., 2009. ILSI research foundation threshold working group. The key events dose-response framework: a cross-disciplinary mode-of-action based approach to examining dose-response and thresholds. Crit. Rev. Food Sci. Nutr. 49, 682–689.

Kalogeras, N., Odekerken-Schröder, G., Pennings, J.M.E., Gunnlaugsdóttir, H., Holm, F., Leino, O., Luteijn, J.M., Magnússon, S.H., Pohjola, M.V., Tijhuis, M.J., Tuomisto, J.T., Ueland, Ø., White, B.C., Verhagen, H., 2012. State of the art in benefit–risk analysis: economics and marketing-finance. Food Chem. Toxicol. 50, 56–66.

Kim, D., Nylander-French, L.A., 2009. Physiologically based toxicokinetic models and their application in human exposure and internal dose assessment. EXS 99, 37–55.

Klimisch, H.J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. 25, 1–5.

Kodell, R.L., Gaylor, D.W., 1999. Combining uncertainty factors in deriving human exposure levels of noncarcinogenic toxicants. Ann. NY Acad. Sci. 895, 188–195.

Kroes, R., Muller, D., Lambe, J., Kroes, R., Müller, D., Lambe, J., Löwik, M.R., van Klaveren, J., Kleiner, J., Massey, R., Mayer, S., Urieta, I., Verger, P., Visconti, A., 2002. Assessment of intake from the diet. Food Chem. Toxicol. 40, 327–385.

Laine, R., 2008. Metabolic stability: main enzymes involved and best tools to assess it. Curr. Drug Metab. 9, 921–927.

Li, A.P., 2007. Human hepatocytes: isolation, cryopreservation and applications in drug development. Chem. Biol. Interact. 168, 16–29.

Linkov, I., Loney, D., Cormier, S., Satterstrom, F.K., Bridges, T., 2009. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. Sci. Total Environ. 407, 5199–5205.

Luke, N.S., Sams 2nd, R., DeVito, M.J., Conolly, R.B., El-Masri, H.A., 2010. Development of a quantitative model incorporating key events in a hepatotoxic mode of action to predict tumor incidence. Toxicol. Sci. 115, 253–266.

Luteijn, J.M., White, B.C., Gunnlaugsdóttir, H., Holm, F., Kalogeras, N., Leino, O., Magnússon, S.H., Odekerken, G., Pohjola, M.V., Tijhuis, M.J., Tuomisto, J.T., Ueland, Ø., McCarron, P.A., Verhagen, H., 2012. State of the art in benefit–risk analysis: medicines. Food Chem. Toxicol. 50, 26–32.

Magnússon, S.H., Gunnlaugsdóttir, H., van Loveren, H., Holm, F., Kalogeras, N., Leino, O., Luteijn, J.M., Odekerken, G., Pohjola, M.V., Tijhuis, M.J., Tuomisto, J.T., Ueland, Ø., White, B.C., Verhagen, H., 2012. State of the art in benefit–risk analysis: food microbiology. Food Chem. Toxicol. 50, 33–39.

Makris, S.L., Thompson, C.M., Euling, S.Y., Selevan, S.G., Sonawane, B., 2008. A lifestage-specific approach to hazard and dose-response characterization for children's health risk assessment. Birth Defects Res. B Dev. Reprod. Toxicol. 83, 530–546.

Meek, M.E., 2001. Categorical default uncertainty factors – interspecies variation and adequacy of database. Hum. Ecol. Risk Assess 7, 157–163.

Miller, F.J., Schlosser, P.M., Janszen, D.B., 2000. Haber's rule: a special case in a family of curves relating concentration and duration of exposure to a fixed level of response for a given endpoint. Toxicology 149, 21–34.

O'Brien, J., Renwick, A.G., Constable, A., Dybing, E., Müller, D.J., Schlatter, J., Slob, W., Tueting, W., van Benthem, J., Williams, G.M., Wolfreys, A., 2006. Approaches to the risk assessment of genotoxic carcinogens in food: a critical appraisal. Food Chem. Toxicol. 44, 1613–1635.

Palou, A., Pico, C., Keijer, J., 2009. Integration of risk and benefit analysis – the window of benefit as a new tool? Crit. Rev. Food Sci. Nutr. 49, 670–680.

Pelkonen, O., Turpeinen, M., 2007. In vitro–in vivo extrapolation of hepatic clearance. Biological tools, scaling factors, model assumptions and correct concentrations. Xenobiotica 37, 1066–1089.

Pohjola, M.V., Leino, O., Kollanus, V., Tuomisto, J.T., Gunnlaugsdóttir, H., Holm, F., Kalogeras, N., Luteijn, J.M., Magnússon, S.H., Odekerken, G., Tijhuis, M.J., Ueland, Ø., White, B.C., Verhagen, H., 2012. State of the art in benefit–risk analysis: environmental health. Food Chem. Toxicol. 50, 40–55.

Pratt, I., Barlow, S., Kleiner, J., Larsen, J.C., 2009. The influence of thresholds on the risk assessment of carcinogens in food. Mutat. Res. 678, 113–117.

Preston, J.R., Williams, G.M., 2005. DNA-reactive carcinogens: mode of action and human cancer hazard. Crit. Rev. Toxicol. 35, 673–683.

Renwick, A.G., Barlow, S.M., Hertz-Picciotto, I., Boobis, A.R., Dybing, E., Edler, L., Eisenbrand, G., Greig, J.B., Kleiner, J., Lambe, J., Müller, D.J., Smith, M.R.,

Tritscher, Tuijtelaars, S., van den Brandt, P.A., Walker, R., Kroes, R., 2010. Risk characterisation of chemicals in food and diet. Food Chem. Toxicol. 41, 1211–1271.

Sand, S., Victorin, K., Filipsson, A.F., 2008. The current state of knowledge on the use of the benchmark dose concept in risk assessment. J. Appl. Toxicol. 28, 405–421.

Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., Hartung, T., Hoffmann, S., 2009. "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicol. Lett. 189, 138–144.

Schünemann, H.J., Oxman, A.D., Vist, G.E., Higgins, J.P.T., Deeks, J.J., Glasziou, P., Guyatt, G.H., 2008. Interpreting results and drawing conclusions. In: Higgins, J.P.T., Green, S., (Eds.), Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1, Chapter 12. The Cochrane Collaboration, 2008. Available from: www.cochrane-handbook.org.

Schütte, K., Boeing, H., Hart, A., Heeschen, W., Reimerdes, E.H., Santare, D., Skog, K., Chiodini, A., 2012. Application of the BRAFO tiered approach for benefit–risk assessment to case studies on heat processing contaminants. Food Chem. Toxicol. 50, S724–S735.

Seed, J., Carney, E.W., Corley, R.A., Crofton, K.M., DeSesso, J.M., Foster, P.M., Kavlock, R., Kimmel, G., Klaunig, J., Meek, M.E., Preston, R.J., Slikker Jr., W., Tabacova, S., Williams, G.M., Wiltse, J., Zoeller, R.T., Fenner-Crisp, P., Patton, D.E., 2005. Overview: Using mode of action and life stage information to evaluate the human relevance of animal toxicity data. Drug Chem. Toxicol. 35, 664–672.

Slikker Jr., W., Andersen, M.E., Bogdanffy, M.S., Bus, J.S., Cohen, S.D., Conolly, R.B., David, R.M., Doerrer, N.G., Dorman, D.C., Gaylor, D.W., Hattis, D., Rogers, J.M., Woodrow-Setzer, R., Swenberg, J.A., Wallace, K., 2004. Dose-dependent transitions in mechanisms of toxicity. Toxicol. Appl. Pharmacol. 201, 203–225.

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.

Solecki, R., Davies, L., Dellarco, V., Dewhurst, I., Raaij, M., Tritscher, A., 2005. Guidance on setting of acute reference dose (ARfD) for pesticides. Food Chem. Toxicol. 43, 1569–1593.

Sonich-Mullin, C., Fielder, R., Wiltse, J., Baetcke, K., Dempsey, J., Fenner-Crisp, P., Grant, D., Hartley, M., Knaap, A., Kroese, D., Mangelsdorf, I., Meek, E., Rice, J.M., Younes, M., 2001. International programme on chemical safety. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul. Toxicol. Pharmacol. 34, 146–152.

Souverein, O.W., Dekkers, A.L., Geelen, A., Haubrock, J., de Vries, J.H., Ocké, M.C., Harttig, U., Boeing, H., van 't Veer, P., 2011. Comparing four methods to estimate usual intake distributions. Eur. J. Clin. Nutr. 65 (Suppl. 1), S92–S101.

Subramaniam, R.P., White, P., Cogliano, V.J., 2006. Comparison of cancer slope factors using different statistical approaches. Risk Anal. 26, 825–830.

Tijhuis, M.J., de Jong, N., Pohjola, M.V., Gunnlaugsdóttir, H., Hendriksen, M., Hoekstra, J., Holm, F., Kalogeras, N., Leino, O., van Leeuwen, F.X., Luteijn, J.M., Magnússon, S.H., Odekerken, G., Rompelberg, C., Tuomisto, J.T., Ueland, O., White, B.C., Verhagen, H., 2012a. State of the art in benefit–risk analysis: food and nutrition. Food Chem. Toxicol. 50, 5–25.

Tijhuis, M.J., Pohjola, M.V., Gunnlaugsdóttir, H., Kalogeras, N., Leino, O., Luteijn, J.M., Magnússon, S.H., Odekerken-Schröder, G., Poto, M., Tuomisto, J.T., Ueland, O., White, Holm, F., Verhagen, H., 2012b. Looking beyond borders: integrating best practices in benefit–risk analysis into the field of food and nutrition. Food Chem. Toxicol. 50, 77–93.

Ueland, Ø., Gunnlaugsdottir, H., Holm, F., Kalogeras, N., Leino, O., Luteijn, J.M., Magnússon, S.H., Odekerken, G., Pohjola, M.V., Tijhuis, M.J., Tuomisto, J.T., White, B.C., Verhagen, H., 2012. State of the art in benefit–risk analysis: consumer perception. Food Chem. Toxicol. 50, 67–76.

Valcke, M., Krishnan, K., 2011. Evaluation of the impact of the exposure route on the human kinetic adjustment factor. Regul. Toxicol. Pharmacol. 59, 258–269.

Verhagen, H., Vos, E., Francl, S., Heinonen, M., van Loveren, H., 2010. Status of nutrition and health claims in Europe. Arch. Biochem. Biophys. 510, 6–15.

Verhagen, H., Andersen, R., Antoine, J.M., Finglas, P., Hoekstra, J., Kardinaal, A., Nordmann, H., Pekcan, G., Pentieva, K., Sanders, T.A., van den Berg, H., van Kranen, H., Chiodini, A., 2012a. Application of the BRAFO tiered approach for benefit–risk assessment to case studies on dietary interventions. Food Chem. Toxicol. 50, S710–S723.

Verhagen, H., Tijhuis, M.J., Gunnlaugsdóttir, H., Kalogeras, N., Leino, O., Luteijn, J.M., Magnússon, S.H., Odekerken, G., Pohjola, M.V., Tuomisto, J.T., Ueland, Ø., White, B.C., Holm, F., 2012b. State of the art in benefit–risk analysis: introduction. Food Chem. Toxicol. 50, 2–4.

Vermeir, M., Annaert, P., Mamidi, R.N., Roymans, D., Meuldermans, W., Mannens, G., 2005. Cell-based models to study hepatic drug metabolism and enzyme induction in humans. Expert Opin. Drug Metab.Toxicol. 1, 75–90.

Vermeire, T., Stevenson, H., Peiters, M.N., Rennen, M., Slob, W., Hakkert, B.C., 1999. Assessment factors for human health risk assessment: a discussion paper. Drug Chem. Toxicol. 29, 439–490.

Vlaanderen, J., Moore, L.E., Smith, M.T., Lan, Q., Zhang, L., Skibola, C.F., Rothman, N., Vermeulen, R., 2010. Application of OMICS technologies in occupational and environmental health research; current status and projections. Occup. Environ. Med. 67, 136–143.

Watzl, B., Gelencsér, E., Hoekstra, J., Kulling, S., Lydeking-Olsen, E., Rowland, I., Schilter, B., van Klaveren, J., Chiodini, A., 2012. Application of the BRAFO-tiered approach for benefit–risk assessment to case studies on natural foods. Food Chem. Toxicol. 50, S699–S709.

WHO, 2003. Diet, nutrition and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation. WHO Technical Report Series 916. World Health Organization, Geneva, Switzerland. Available from: http://www.fao.org/DOCREP/005/AC911E/AC911E00.HTM (accessed 06/04/2011).

WHO, 2005. Chemical-specific adjustment factors for interspecies differences and human variability: Guidance document for use of data in dose/concentration–response assessment, WHO, Geneva. Available from: http://www.inchem.org/documents/harmproj/harmproj/harmproj2.pdf (accessed 30/10/10).

WHO, 2009. Principles and methods for the risk assessment of chemicals in food. EHC 240, ISBN 978 92 4 157240 8, WHO, Geneva. Available online at http://www.who.int/ipcs/food/principles/en/index1.html.

Zhang, Q., Bhattacharya, S., Andersen, M.E., Conolly, R.B., 2010. Computational systems biology and dose-response modeling in relation to new directions in toxicity testing. J. Toxicol. Environ. Health B Crit. Rev. 13, 253–276.