Quality of Life – Integrated Benefit and Risk Analysis.
Web-based tool for assessing food safety and health benefit
(N°022957)

Deliverable 29c

Final Qalibra Framework for Risk-Benefit assessment

Andy Hart, Jeljer Hoekstra, Helen Owen, Marc Kennedy, Marco Zeilmaker, Nynke de Jong, Helga Gunnlaugsdottir

Lead participants: RIVM and Fera

Specific Targeted Research Project
Thematic Priority 5: Food Quality & Safety
Due date of deliverable: 31.12. 2009
Actual submission date: 17.05. 2010
Start date of project: 01.04. 2006
Duration: 2006 - 2009

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Contents

Summary

This document describes a framework for quantitative assessment of health risks and benefits of dietary change, developed by the EU research project Qalibra. It also outlines how this framework has been implemented in software developed by Qalibra, which can be accessed (subject to registration) at www.qalibra.eu.

It is efficient to adopt a tiered approach to risk-benefit assessment, so that the degree of complexity and effort is adjusted to the needs of decision-making. The related EU project BRAFO has developed a tiered approach for risk-benefit assessment with 4 tiers: Tiers 1 and 2 use simpler methods to determine whether a quantitative assessment is necessary. The Qalibra framework and software are intended for quantitative assessment for risks and benefits, corresponding to Tiers 3 and 4 of the BRAFO approach.

This document begins by introducing the principles of risk-benefit assessment and the types of questions it can answer. It discusses the need for a common metric which enables different risks and benefits to be compared on the same scale, introduces the common metrics used in the Qalibra framework (disability-adjusted life years or DALYs, and quality-adjusted life years or QALYs), and explains how they are calculated in the Qalibra framework. The types of data and assumptions required for the calculations are described. The importance of addressing uncertainties is emphasised, and qualitative, deterministic and probabilistic methods for this are presented. Finally the document describes with examples the main types of output produced by the Qalibra software, and discusses its interpretation.

It is emphasised that risk-benefit assessment is inherently complex and requires a high level of expertise in the relevant fields of science including nutrition, toxicology and epidemiology. Quantitative risk-benefit assessment additionally requires significant expertise in modelling and statistics.

In this context it is hoped that the Qalibra framework will help by providing a common conceptual framework, and help users to identify important issues and data gaps. The Qalibra software is designed to provide a user-friendly environment within which users can start with a simple deterministic assessment and progressively refine it by treating key elements probabilistically (when needed). The software also helps the user to organise the large number of datasets and model runs that may be needed, and to share them with colleagues of their choosing.

1. Introduction

National and European food policy, including regulations and advice to consumers, should take account of the risks and benefits of different foods, i.e. their positive and negative effects on human health. Information on risks and benefits should also be available to other interested parties including food producers, retailers and consumers.

Usually, information on risks and benefits is presented separately. This is unsatisfactory, because it leaves the recipient uncertain as to the balance of risk and benefit. Ideally, information on risk and benefit should be combined to provide an indication of the overall effect of particular dietary choices, i.e. the net health impact.

The central goal of QALIBRA is therefore to develop improved approaches for the assessment and communication of net health impacts of dietary choices. To maximise dissemination and uptake of the project outputs, they will be implemented as web-enabled software.

Uncertainties affecting risks and benefits cause uncertainty about the magnitude and even the direction of the net health impact. Therefore the approaches developed by QALIBRA aim to take account of uncertainties and communicate them effectively to both technical users and consumers.

The specific objectives of QALIBRA were therefore as follows:

1. Develop a generalised modular approach to risk-benefit assessment,
2. Implement the approaches in web-enabled software, with different components adapted to different user groups,
3. Develop targeted risk-benefit communication strategies for integrated risk-benefit analysis, adapted to the needs of different stakeholders,
4. Use the methods and software developed by QALIBRA to carry out detailed case studies on the risks and benefits of oily fish and functional foods,
5. Establish information-sharing and joint activities with BENERIS, another EU-funded project undertaking complementary research,
6. Project management.

This report presents the final version of the QALIBRA risk-benefit framework. It describes how the computations and integration of risks and benefits are performed in the Qalibra tool. It takes into account experience gained in the case studies conducted within Qalibra and comments received from EU-reviewers, the Qalibra Scientific Advisory Panel, and participants at 2 Qalibra workshops in 2009. In addition, it is designed to be compatible with the tiered approach developed by the BRAFO project (see section 1.2), another EU project on risk-benefit analysis with which Qalibra cooperated closely.

1.1 General framework

The overall framework is similar to the quantitative risk-benefit analysis described in Hoekstra et al. (Hoekstra et al. 2008), but in more generalised form (facilitating its application to a wide range of risk-benefit questions) and adding the capability to quantify uncertainties at each stage of the process.

Since 1995 a generic risk (safety) assessment model for food standards issues had been agreed upon (FAO/WHO Expert Consultation, 1995). It is increasingly recognized, among others at a EFSA Science colloquium that a similar paradigm can – and should – be constructed for the benefit assessment or rather in an integrated risk-benefit assessment approach (EFSA, 2006a). Consequently, risk-benefit assessment can be divided into four analogous steps, i.e. 1) hazard and benefit identification, 2) hazard and benefit characterization through dose-response functions, 3) exposure assessment, and 4) risk-benefit characterization including integration of risks and benefits. To perform an integrated risk-benefit assessment new elements have to be developed. Both hazardous and beneficial effects need to be taken into account and potential risks and benefits must be balanced by use of a common measure such as disability-adjusted-life-years (DALYs), quality-adjusted-life-years (QALYs), or healthy-life-expectancy (HALE) (Havelaar et al. 2000; Ezzati et al. 2003; Ponce et al. 2000). Consequently, the Qalibra framework includes the following elements, described in detail later in this document:

- Define the question – a pre-assessment phase, sometimes referred to as problem definition or problem formulation.
- Identify the beneficial and adverse health effects to consider, the compounds that may influence these effects, and the affected population
- Assess the level of exposure to the hazardous and beneficial compounds
- Quantify the level of hazards and benefits that may result from this exposure via dose-response curves
- Integrate the adverse and beneficial health effects using a common metric
- Identify unquantified uncertainties and evaluate their implications for interpretation of the results. Also quantifiable uncertainties and variabilities must be taken into account. Although this is stated as the last step, it is an ongoing process within each of the earlier steps.

1.2 Relationship to the BRAFO tiered approach

The related EU project BRAFO has developed a tiered approach for risk-benefit assessment (Hoekstra et al., submitted). This comprises 4 Tiers, which differ principally in the way risks and benefits are integrated. At Tier1, risks and benefits are assessed separately, while in Tiers 2-4 they are integrated using increasingly sophisticated approaches. Consequently in each tier the net health impact is assessed with increasing precision. As with all tiered assessment approaches, the aim is to refine the assessment only as far as is necessary to reach a decision, in this case on whether the net health impact of a dietary change is beneficial or adverse. The 4 Tiers are as follows:

- In Tier 1, each risk and benefit is assessed independently. These assessments will often use standard screening methods, but it may be worth using more refined methods if this avoids the need to proceed to Tier 2. Thus Tier 1 comprises separate assessments of risks and benefits, each as refined as may be needed.
- In Tier 2, risks and benefits are compared in a qualitative way; no common metric is used at this tier. However, the assessment of each individual risk or benefit may be quantitative or even probabilistic.
- In Tier 3, risks and benefits are integrated quantitatively in a common metric, using deterministic methods.
- In Tier 4, risks and benefits are integrated quantitatively in a common metric, using probabilistic methods.

In practice, there is a continuum between tiers 3 and 4. Initially all parts of the assessment are treated deterministically (i.e. as fixed values), after which progressively more parameters are treated probabilistically (i.e. using probability distributions), until the net health impact is sufficiently well characterised for decision-making.

The Qalibra project focussed on developing detailed methodology for the quantitative integration of risks and benefits, i.e. Tiers 3 and 4 of the BRAFO approach. Consistent with the BRAFO approach, the Qalibra framework and software tool allows progressive refinement from Tier 3 to Tier 4, as each input may be treated deterministically or probabilistically. If one or more of the inputs is probabilistic, then the output will also be probabilistic.

Quantitative integration of risk and benefit is technically complex and requires additional types of data or assumptions (e.g. severity of effects and age of onset) that are not required in conventional assessments of risk and benefit. Therefore, Qalibra recommends that practitioners follow the BRAFO tiered approach, resolving risk-benefit questions without quantitative integration where possible, and reserve the Qalibra framework for those cases where assessment at Tiers 3 and 4 is necessary.

2. Problem formulation

This phase is essential to specify the risk-benefit problem that will be investigated. It sets the scope of the assessment, and should be defined by consultation with the risk managers or policy-makers for whom the assessment is intended. It is recommended to write the risk-benefit question in words, to help the parties involved to define it more clearly and identify the relevant scenarios (Hoekstra et al., submitted). It is also good practice to record the context or background of the risk-benefit question and the reasons why it is asked, to ensure all the involved parties have a common understanding of the problem.

As more information becomes available during the assessment, the risk-benefit question may need to be reformulated. Problem formulation can therefore be an iterative process. At each tier in the process, consultation with the risk manager/policymaker is encouraged to ensure the relevance of the risk benefit question that is being answered.

A detailed discussion of problem formulation for risk-benefit assessment is provided in the BRAFO approach (Hoekstra et al., submitted). The key elements are as follows.

2.1 Dietary scenarios

In many cases, a risk-benefit question explicitly involves comparison of two scenarios – a **reference scenario** (e.g. the current diet) and an **alternative scenario** in which the diet is changed, for example by introduction of a new policy, new product, or new advice to consumers. In other cases, the risk-benefit question may relate only to a single scenario, e.g. what is the balance of risks and benefits for a particular dietary scenario (often, the current diet). In practice, assessing a single scenario is equivalent to a two-scenario question comparing the diet in question to an implicit reference diet in which all of the risks and benefits are reduced to zero (even though such a diet may

not be realistically achievable or desirable). This point is discussed further in the BRAFO approach (Hoekstra et al., submitted). The Qalibra framework requires specification of two scenarios, but is equally applicable to one-scenario questions (by setting all intakes to zero for the reference scenario).

In assessments of new policies, products or consumer advice, it is essential to be clear what assumptions will be made regarding the effectiveness or uptake of these, and this must be taken into account in defining the dietary scenarios.

It is also essential to be clear whether the scenarios take dietary substitution into account. For example, when someone eats more fish it usually means they will eat less meat or vegetables. The substitution may affect intake of contaminants or nutrients other than those present in the food of primary interest (e.g. fish), and may also cause changes in calorie intake. All of these things may affect health and ideally should be taken into account. However, often this is problematic and broadens the scope of the assessment dramatically, to include modelling foods and health effects other than those which are the primary focus of the risk-benefit question. If these are not included, their potential relevance and importance should be considered as part of the uncertainty analysis at the end of the assessment (see section 15.1).

Note that generally, it is not sufficient to consider only the change in intakes between the two assessment scenarios. Because dose-response functions are often non-linear, total exposure is important. Thus exposure through other sources than those directly described in the scenarios cannot be ignored. Therefore, it is often necessary to estimate background exposure and combine this with the sources of primary interest.

Note that changing dietary scenarios may have many other types of consequence in addition to health impacts, e.g. economic, cultural or legal consequences, and it is legitimate and appropriate to consider these in decision-making. The Qalibra framework addresses only the assessment of health impacts.

2.2 Population

The population to which the assessment refers must be defined unambiguously. This may include the target population for a new policy, product or advice, but may also need to include unintentionally exposed or special groups in the population (e.g. groups with special diets or differential sensitivity).

A complication occurs in assessments involving health effects on the next generation (i.e. on the offspring of the population experiencing the dietary change). The Qalibra framework offers two alternative approaches for addressing this (see section 10).

For practical reasons, the population considered will often refer to a single country, for which dietary surveys are available to estimate intakes. It may be possible to draw conclusions for wider general populations (e.g. the EU), especially if this can be based on parallel assessments for several countries representing a range of diets, but such extrapolations are inherently uncertain and should be considered with care.

As will be seen in section 13, the Qalibra framework and software can be applied to a single individual instead of a population. This option is intended for use in simplified

assessments, e.g. to explore the balance of potential risks and benefits for different types of individual (e.g. a 'typical' or 'worst case' individual).

3. Common metrics for integration of risks & benefits

The purpose of the Qalibra framework is to provide methodology and software for quantitative integration of health risks and benefits, as in Tiers 3 and 4 of the BRAFO tiered approach. Quantitative integration requires the use of a common metric for aggregation and comparison of different adverse and beneficial health effects.

Ideally a risk-benefit assessment should consider all relevant aspects of health, such as mortality, morbidity, and quality of life. The challenge is how to add up different kinds of health effects. How to compare, for example liver cancer with a gastrointestinal infection? Or, how to compare death with a decline in cognitive functioning? The solution is to weigh or value these different health effects using a valuation function that rates all health effects on the same scale.

Summary measures of population health combine information on mortality and non-fatal health outcomes to represent the health of a particular population as a single number. Also for the quantification of the net health effect of a food or ingredient of interest, both positive and negative effects need to be expressed in the same endpoint of interest. Several methods have been developed for evaluating the impact of illness and disability on public health, including monetary methods and methods that *quantify* disease burden as a result of (premature) death and impaired quality of life such as disability-adjusted-life-years (DALYs), quality-adjusted-life-years (QALYs), and healthy-life-expectancy (HALE) (Havelaar et al. 2000; Ezzati et al. 2003; Ponce et al. 2000). In the case of cost benefit analysis health state is often expressed in a monetary value such as Willingness To Pay (WTP) or Willingness To Accept (WTA). Ponce *et al.* (Ponce et al. 2001) and Wong *et al.* (Wong et al. 2003) provide a general review of such metrics. Qalibra Deliverable D3 also reviewed several health metrics.

The choice of the common currency and whether to apply such things as age weighting and discounting is not trivial. Ethical and equity issues play a role. Anand and Hanson (1998) nicely show that if the DALY (but also the QALY) measure is used, one would favour saving the life of a baby girl over that of a baby boy because the life expectancy of women is larger. The same goes for extending the life of a man in a wheelchair compared to extending the life of an equally old man who can walk. More disability adjusted life years will be gained when treatment is given to the able man. Nord (2005) would argue that the more weight should be given to the life years gained by the disabled man because it would be fair that both man could experience a comparable number of quality-adjusted life years.

The choice of health measure and associated parameter values such as disability weights and discount rate give, deliberate or not, priority to one age- or disease subgroup at the expense of another. Therefore, it is important to be transparent about the choices that are made. Not only the final DALY (or other summary measure) should be presented but also intermediate variables such as incidences, and these are provided by the Qalibra software.

DALYs and QALYs were chosen as the common metrics for use in the Qalibra framework and software. DALYS and QALYs are commonly used integrated health

measures. It was decided not to include economic measures such as WTP and WTA, because they are associated with additional difficulties in interpretation, particularly because they may reflect wealth of respondents more than differences in valuation of health.

The use of QALY or DALY measures in a risk-benefit assessment is illustrated in Figure 1. This illustrates how the net health impact of a dietary change is obtained as the difference between the QALY or DALY calculations for the reference and alternative scenarios. It also indicates the types of data that are required for using these metrics, including QALY or DALY weights, age of onset and duration for each health effect. In addition, information is needed on the recovery and mortality rates for each disease or health state considered (see section 14).

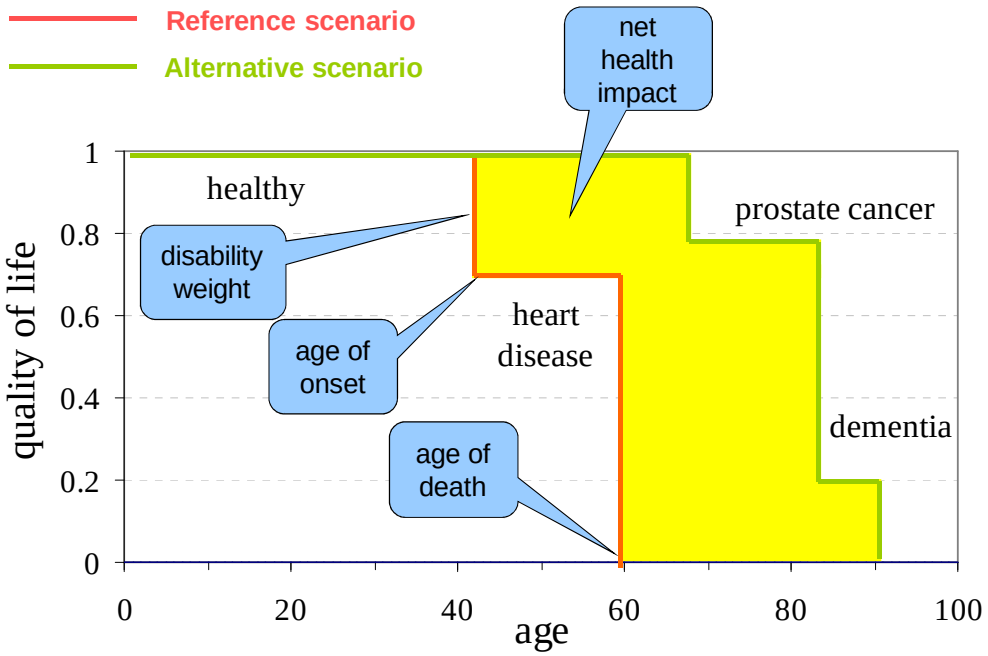

Figure 1. Diagrammatic representation of the calculation of the net health impact of a dietary change (from reference to alternative scenario) on a single individual, using quality-adjusted life years (QALYs). Calculation using disability-adjusted life years (DALYS) is equivalent but with the vertical scale reversed, so that 0 represents full health and 1 represents death.

3.1 DALY versus QALY

DALYs and QALYs are technically similar in that they both express health in time (life years) and give a weight to years lived with a disease. In the terminology of Gold et al. (2002), both measures are HALYs (Health adjusted life Years). Gold et al. give a comprehensive review of the differences between DALYs and QALYs. DALYs measure health loss and QALYs health gain so they express an inverse value. That problem is overcome by looking at differences (in absolute terms) between scenarios or interventions. More importantly the measures originate from different disciplines, which causes the disease weights to be measured in a different way with a different interpretation, resulting in different values. There is debate about the use of QALYs vs DALYs (Sassi, 2006) as they give different outcomes, but a large part of the

differences is explained by whether or not age weighting and discounting is applied. The original DALY formulation (Murray, 1994) incorporates age-weighting and discounting. In practise, the difference between a DALY and a QALY depends on whether the quality of life is expressed as a loss (DALY) or a gain (QALY), as is illustrated by the weight for death (1 for DALY, 0 for QALY). Additional differences are caused by the way disease weights are measured. Given the variability in disease weights and the influence of age weighting and discounting it seems that the choice of the common currency should be determined by the choice for disease weights and whether or not to apply age weighting and discounting (but note that specific functionality for age weighting and discounting is not provided in the current version of the Qalibra software).

3.2 Alternative approaches for calculating DALY/QALYs

The most accurate way to calculate QALYs (or DALYs) would be to obtain a record of the actual history of health changes for each individual in a population, assign the relevant QALY weight for their health state in each year of their life, and calculate the sum of years weighted by their associated QALY weights. However, risk-benefit assessment requires the estimation of QALYs for future health states for different dietary scenarios (reference and alternative).

The most obvious approach for this is to simulate future diet and the resulting health states for each individual under the two scenarios. This has the very important advantage of allowing the assessor to take account of how different diseases or health effects combine, including how effects associated with the dietary change under assessment combine with "background" diseases. How different health effects combine has an important impact on the calculation: e.g. if a person dies from one disease at 50 then another disease that would commence at 60 will have no effect.

This simulation approach has been used in DALY calculations, e.g. in the RIVM Chronic Disease Model for the Netherlands (van Kreijl et al., 2006; Hoogenveen *et al.* 2009). However, this type of modelling is extremely complex and requires data on many parameters including the normal incidences, durations and severities of major background diseases as well as the health effects of the dietary change under consideration, and the interactions between them. It also requires modelling the demographic development of the population over the period of dietary change (births, deaths, immigration, emigration).

van Kreijl et al. (2006) also present a much simpler, but less realistic, approach which estimates the "annual directly attributable health loss". This was used in a risk-benefit assessment for folate by Hoekstra et al. (2008). This approach considers only a single year, and only health effects that have their onset during that year. The potential impact of each health effect starting during the year is considered in isolation, ignoring interactions with other effects starting in the same or subsequent years, and ignoring interactions between these effects and background diseases. Because of this, the output is interpreted as measure of *potential impact per year*, rather than as an estimate of actual health outcome. Furthermore, the potential impacts of alternative disease outcomes (recover, die early as a result of the disease, or survive with the disease until the normal life expectancy) are combined as a weighted average, weighted by their respective probabilities. The result should therefore be interpreted

as estimating the *average of the potential impacts for a (large) number of similar individuals*, rather than a specific outcome for a single individual.

A comparison of these two approaches showed that health impacts estimated by the Chronic Disease Model were somewhat lower than those estimated by the directly attributable health impacts approach (van Kreijl et al., 2006). This occurs because the Chronic Disease Model takes account of combination effects (comorbidity, substitution and delayed effects) that prevent the full potential impacts of effects being realised.

Because of the complexity and resource and data requirements of simulating health effects over whole lifetimes, the Qalibra framework and software is based instead on the simpler approach of calculating annual directly attributable health loss. A practical view is to consider this as a basic option for Tiers 3 and 4 of the Brafo tiered approach, with simulation of health over whole lifetimes being reserved as a higher tier option for use only with risk-benefit questions that cannot be satisfactorily resolved by the simpler, directly attributable health impacts method. The assessor would then start with the directly attributable health impacts approach, interpreting the results very carefully to take account of its limitations. In particular, the assessor will need to consider how the overall health impact might be affected by the way the individual effects combine (see interpretation of results, section 19).

4. Main elements of the Qalibra framework

The Qalibra framework is constructed as a logical sequence of steps through the calculation of DALYs or QALYs. The steps in the framework comprise:

1. Problem formulation including specification of the dietary scenarios and the population to be considered.
2. Identification of the adverse and beneficial health effects to be assessed.
3. Estimation of the intakes or exposures that cause those health effects.
4. Modelling of the dose-response relationship for each effect, including the probability of onset at the current age and, if the effect is continuous, the magnitude of the effect.
5. Estimation of the probabilities of recovery and mortality for affected individuals.
6. Selection of a common currency (DALY or QALY)
7. Specification of the severity and duration of the effect.
8. Calculation of the net health impact.

These steps are illustrated graphically in Figure 2. Note that steps 3-5 and 7 must be repeated for each adverse and beneficial effect considered. In principle, any number of effects can be included.

The Qalibra software includes a “wizard” function which guides the user through steps 1-8 in sequence. However, steps 3-5 and 7 can actually be done in any order and the software allows this as an alternative, which may be more convenient for experienced users.

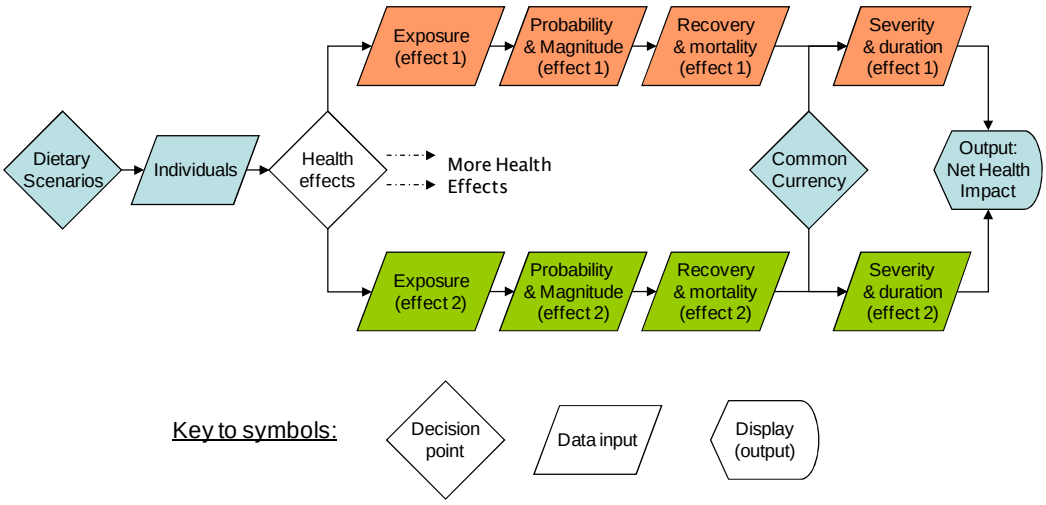

Figure 2. Diagrammatic representation of the Qalibra framework for risk-benefit assessment using DALYs or QALYs. See text for explanation.

The following sections describe in more detail the DALY/QALY calculations, and then discuss the data required for steps 1-8.

5. When is use of the Qalibra framework appropriate?

As indicated in section 1.2, the Qalibra framework is intended for assessments requiring quantitative integration of risks and benefits, corresponding to Tiers 3 and 4 in the BRAFO tiered approach (Hoekstra et al. submitted). The need for assessment at Tier 3 is identified by assessment at Tier 2.

The essential difference between Tiers 2 and 3 of the BRAFO approach is that in Tier 2, the balance of beneficial and adverse effects is assessed qualitatively, whereas in Tier 3, it is assessed quantitatively.

Tier 3 is needed when Tier 2 shows that both beneficial and adverse effects are present, but neither is clearly dominant when considered qualitatively. This situation is made more common by the fact that health effects have multiple dimensions, including the incidence of the effect (number of people affected), its severity, its duration, the rate of mortality caused by the effect, and the number of life years lost due to early mortality from the effect. A change in dietary scenario affects these dimensions to differing extents, for different effects, so it will often be difficult to judge the overall adversity or benefit of the health change unless the different dimensions can be integrated in some way. Tier 2 attempts to integrate the dimensions qualitatively, but this can only be done with confidence where either the beneficial or adverse changes clearly dominate. Where this is not possible, approaches that integrate risk and benefit quantitatively are required, such as those provided by the Qalibra framework and in Tiers 3 and 4 of the Brafo approach.

6. Calculation of DALYs for quantal health effects

The principles of the directly attributable health loss calculation are introduced first for DALYs and quantal effects (effects that are modelled as either absent or present, e.g. cancer). The following sections describe how the calculation differs for

continuous effects (effects expressed as a change in a continuous variable, such as a change in body weight), and for QALYs.

The calculation takes account of three alternative outcomes of each effect: an individual may recover, die early as a result of the disease, or survive with the disease until the normal life expectancy. It allows for the possibility that the severity of the disease (represented by DALY weights, *w*) may differ between individuals who recover, individuals who die from the disease, and those who continue living with the disease until their normal life expectancy, although in many cases these are the same.

For an individual that recovers, the DALY loss is calculated as:

$$YLD_{rec} \cdot w_{rec}$$

where:

YLD_{rec} = duration of disease for those who recover
w_{rec} = DALY weight for disease, for those who recover.

For an individual that does not recover, but survives with the disease until their normal life expectancy, the DALY loss is:

$$(LE - CA) \cdot w_{live}$$

where:

CA = current age of individual in year of disease onset
LE = normal life expectancy[1] (generally a function of current age)
w_{live} = DALY weight for disease, for those who continue living with it until their normal life expectancy.

For an individual that dies from the disease, the DALY loss is:

$$YLD_{die} \cdot w_{die} + LE - CA - YLD_{die}$$

where:

YLD_{die} = duration of disease (years lived with disease) for those who die of it
w_{die} = DALY weight for disease, for those who die from the disease.

Note that the loss for those who die comprises two parts: the DALY loss for the period prior to death ($YLD_{die} \cdot w_{die}$) and the loss of years due to dying earlier than would be expected without the disease.

The average DALY loss for individuals who get the disease can be obtained as a weighted average of the three contributions:

$$p_{rec} \cdot YLD_{rec} \cdot w_{rec} + p_{die} \cdot (YLD_{die} \cdot w_{die} + LE - CA - YLD_{die}) + (1 - p_{die} - p_{rec}) \cdot (LE - CA) \cdot w_{live}$$

where

p_{rec} = probability of recovery from the effect
p_{die} = probability this effect causes death.

The expected total DALY loss due to this disease can then be estimated as:

[1] Note: life expectancy is sometimes defined as the number of years of life remaining, but here it is defined as the expected age at death.

$$DALY = I_{sf} \cdot p_{effect} \cdot [p_{rec} \cdot YLD_{rec} \cdot w_{rec} + p_{die} \cdot (YLD_{die} \cdot w_{die} + LE - CA - YLD_{die}) + (1 - p_{die} - p_{rec}) \cdot (LE - CA) \cdot w_{die}]$$

(Equation 1)

where

p_{effect} = probability of onset of the disease in the current year
I_{sf} = individual scaling factor.

In assessments considering only a single individual (e.g. typical or worst case representative), I_{sf} is set to 1.

To calculate DALYs for a group of N similar individuals with the same age (CA) and other attributes (e.g. gender), I_{sf} should be set to N.

In order to obtain an appropriate estimate for the average annual DALY loss of a whole population, the calculation needs to be repeated for individuals of different ages, in proportion to the age structure of the population. This is done in the Qalibra software by repeating the calculation for individuals of different ages and setting I_{sf} for each calculation equal to the number of individuals of that age in the population.

In many assessments, the individuals for whom DALYs are calculated (i.e. to whom equation 1 is applied) will derive from a sample in a dietary survey that has been used to estimate the intakes of the relevant nutrients and contaminants. In some cases, such surveys include scaling factors that are provided by the survey authors to correct biases in the sample compared to the national population. When available, these scaling factors may be incorporated in I_{sf} in the Qalibra calculation.

Note that p_{effect} depends on intake, that LE and p_{effect} will generally differ between age groups and genders, and that other parameters may also depend on age, gender and other attributes (e.g. the severity of a disease might depend on the age at which it occurs).

When assessing the impact of a dietary change or intervention, its effect on the parameters in Equation (1) need to be quantified. Most obviously, p_{effect} is related by a dose-response relationship to the intake of a nutrient or contaminant affected by the change in diet. In principle, however, other parameters in Equation (1) could also be functions of intake.

7. Calculation of DALYs for continuous health effects

The treatment of a continuous effect will depend on whether its severity as a disease (as measured by a DALY weight) is a step function, applying only above or below some critical effect size. If so, then it may be simpler to model the effect as quantal, e.g.. as a dose-response relation for exceeding the critical effect size, and calculate the DALY loss using Equation (1) above.

If the severity of the disease is a continuous or multi-step function of the effect size, then it will be necessary to use Equation (2):

$$DALY = I_{sf} \cdot p_{effect} \cdot [p_{rec} \cdot YLD_{rec} \cdot w_{rec}(effect) + p_{die} \cdot (YLD_{die} \cdot w_{die}(effect) + LE - CA - YLD_{die}) + (1 - p_{die} - p_{rec}) \cdot (LE - CA) \cdot w_{live}(effect)]$$

Equation (2)

Where the new terms are:

effect = magnitude of health effect

w(*effect*) = DALY weight expressed as a function of effect magnitude.

When assessing the risks and benefit of a dietary change, *effect* will be related by a dose-response function to the intake of some contaminant or nutrient. It is important to note that the current version of the Qalibra software *assumes the response is the same for all individuals at a given dose*. It does not allow variation around the dose-response relationship to be treated as real variation in the response of individuals at the same dose (it does allow this variation to be treated as uncertainty, see section 15.2). An alternative equation, which allows for (and integrates over) variation in response between individuals at the same dose, is given by Hoekstra et al. (submitted). The potential impact of this limitation of the Qalibra software should be taken into account when considering unquantified uncertainties affecting the assessment (see section 15.1).

8. Calculation of QALYs for quantal and continuous health effects

Whereas DALYs are intended for representing health losses, QALYs are intended for representing remaining health, after allowing for any disease(s) that may be present.

The directly attributable health loss approach does not attempt to model the net health of the individual taking account of all concurrent diseases. However, it can be used to estimate potential health losses in terms of lost QALYs, subject to the same limitation as before, that it considers each disease in isolation from others.

For a quantal effect, the QALY loss can be calculated as:

$$QALYloss = I_{sf} \cdot p_{effect} \cdot [p_{rec} \cdot YLD_{rec} \cdot q_{rec} + p_{die} \cdot (YLD_{die} \cdot q_{die} + LE - CA - YLD_{die}) + (1 - p_{die} - p_{rec}) \cdot (LE - CA) \cdot q_{live}]$$

Equation (3)

For a continuous effect, the QALY loss can be calculated as:

$$QALYloss = I_{sf} \cdot p_{effect} \cdot [p_{rec} \cdot YLD_{rec} \cdot q_{rec}(effect) + p_{die} \cdot (YLD_{die} \cdot q_{die}(effect) + LE - CA - YLD_{die}) + (1 - p_{die} - p_{rec}) \cdot (LE - CA) \cdot q_{live}(effect)]$$

Equation (4)

where

q = 1- QALY weight for a quantal effect

$q(effect)$ = 1 – QALY weight expressed as a function of continuous effect size,

and where the QALY weights can differ between individuals who recover, die or live with the disease (as shown by the subscripts in equation 4).

9. Recurrent effects

The direct health loss method of equations 1-4 is not readily applicable to recurrent effects. It may also be possible to represent effects that can recur multiple times a year, by adapting the inputs to the calculation, e.g. by expressing intake not as the

level of intake but as the frequency of intakes in the current year that exceed an effect level, setting the probability of effect to 1, and making the duration of the effect (YLD) and, if appropriate, its severity (the DALY or QALY weight) functions of the intake measure (i.e. of the number of episodes of the disease in the year).

As the direct health loss method only considers effects that begin in the current year, the calculations do not require consideration of recurrence in future years. However, the probability, magnitude and severity of recurrent events in future years (which are represented in the calculation by individuals of older ages) would be dependent on whether effects occurred in the current year. The existence of such dependencies is not considered in the direct health loss model and, where relevant, must be taken into account when evaluating unquantified uncertainties affecting results (see section 15.1).

10. Effects on the next generation

Some assessments involve health effects on the next generation (i.e. on the offspring of the population experiencing the dietary change). There are two alternative ways to represent such effects when using the Qalibra framework and software, as described below. Both of them involve additional complications for the user when preparing model inputs and require additional care when interpreting results.

One method is to consider effects on offspring and their mothers together in the calculation of net health impact. In this approach, the offspring are not included as separate individuals in the calculation; instead, the effect on them is included in the calculation for their mothers. This requires the assessor to specify p_{effect} for females, as the probability of giving birth during the current year to a child that has the effect[2], which will be a function of the mother's current age. If the effect is gender-specific (e.g. an effect on sperm counts of male offspring) then p_{effect} needs to be the probability of giving birth to a child of the relevant gender, that has the effect. When using this approach for next-generation effects, other parameters in the calculation (CA, LE, disease weights, etc.) should in principle be set to the values appropriate to the attributes of the offspring at the age when the effect occurs (e.g. zero for effects in newborn infants, puberty for reproductive effects, etc.). However, it was not feasible to provide for this in the current version of the Qalibra software. Instead, Qalibra uses values based on the attributes of the mother for all parameters except current age CA, which is set to zero automatically when the user indicates (by ticking a check box in the inputs) that the effect is on offspring. Consequently, other parameters will be set to the values appropriate for individuals of age zero, but with other attributes equal to those of their mothers. This means, for example, that LE (life expectancy) will be set to the female value for all newborns, which will normally be higher than the value that should be used for male newborns, leading to some over-estimation of the health impact. It also means health impacts will be over-estimated for effects that actually begin after age zero (e.g. at puberty). This and any other consequences of this approach need to be borne in mind when interpreting the results: if there is concern that the result will be materially misleading (e.g. enough to change the balance of risk and benefit), then it would be advisable to try using the alternative approach for

[2] The possibility of twins, triplets etc. may be incorporated by specifying p_{effect} not as a probability of giving birth, but as the average number of offspring born with the effect during the current year for a woman of the relevant age.

representing next-generation effects (see next paragraph), which allows all parameters to be based on the attributes of the offspring. Note that, when using the above method, treating mothers and their offspring together, the total number of individuals represented in the overall assessment will be greater than the number of individuals specified as input (the current generation), due to addition of offspring generated in modelling the next-generation effect. Also, when examining variation between individuals in contribution to the population's net health impact, the contribution of mother and offspring will be shown together, although their relative importance may be seen by examining the breakdown of effects produced by Qalibra, or by rerunning the assessment without the next-generation effect. Significant additional care is therefore required when interpreting the results of next-generation assessments using this method.

The alternative method for representing next-generation effects avoids the difficulties of the first method, but presents a different challenge. This method requires the user to include an appropriate proportion of newborn children in the population for the assessment. This allows parameters including CA and LE for offspring to be set according to their age and other attributes at the point of disease onset. It also means that these individuals will be represented separately in the assessment results, simplifying interpretation. The challenge for this method relates to the inputs for intake or exposure to the substances causing the next-generation effects, as these should be intakes of the mother and not of the newborn individuals themselves. These intakes could be based on a dietary survey data for women of child-bearing age, however it may be necessary to include a substantial number of newborns in the population to provide a sufficient sample to adequately represent variation in the intakes of mothers, especially if the effect is associated with unusual or extreme intakes. This in turn will either require that the overall population is very large (in order to include an adequate number of newborns), leading to long model run times, or that newborns are over-represented in the calculations but that other age groups are scaled up to compensate (using the scaling factor, I_{sf})[3].

11. Multiple effects on the same health endpoint

Some assessments may involve multiple nutrients and/or contaminants that affect the same endpoint.

Where the combined effect of different substances can be modelled using toxic equivalency factors, e.g. for dioxin-like substances, this can be incorporated by converting the intakes to toxic equivalents outside the Qalibra framework and then using the combined intake as the "dose" for a single effect in the Qalibra model.

In other cases, combining intakes using toxic equivalency factors may be inappropriate, or the nature of the interaction between different effects on the same endpoint may be unknown. For example, oily fish contain nutrients that may increase IQ of the next generation but also methylmercury, which may decrease next generation IQ. There is no special provision for such effects in the Qalibra framework. A practical approach is to include the different nutrients and/or contaminants

[3] If scaling factors are used, care should be taken to ensure that the number of individuals modelled explicitly (before scaling up) is sufficient to represent adequately the variability of intakes amongst individuals of similar attributes.

separately, and then take special care in interpreting the results. If the health endpoint is continuous (e.g. IQ) and if its relationship to the severity weight (DALY or QALY weight) is linear, then the additive aggregation of DALYs/QALYs in the Qalibra framework should give an appropriate result. Where the endpoint is quantal, or if the relationship to severity weights is non-linear, additive aggregation may give misleading results. To guard against this, the user should inspect very carefully the contributions of the individual substances: if either the negative or positive effects are very dominant, the result may be reasonable, whereas if both negative and positive effects are non-negligible then additive aggregation may either over- or under-estimate the net impact; in the latter case, it may be necessary to develop a bespoke model (outside Qalibra) to model the interaction of effects more appropriately.

12. Calculating the net health impact of a dietary change

In a risk-benefit assessment the focus is on evaluating the net health impact of a dietary change or intervention, represented as a change from a reference scenario to an alternative scenario.

The net health impact in DALYs may be calculated as:

$$\Delta DALY = \sum DALY_{alt} - \sum DALY_{ref}$$

where

$\sum DALY_{ref}$ = sum of DALY losses for reference scenario

$\sum DALY_{alt}$ = sum of DALY losses for alternative scenario

The summation in these equations is over both health effects (within individuals) and individuals (to provide population totals).

As DALYs represent health loss, a positive result for $\Delta DALY$ implies that changing to the alternative scenario has an adverse health impact overall, whereas a negative $\Delta DALY$ implies the alternative scenario is beneficial.

Analogous calculations can be done with the QALY losses from Equations (3) and (4), leading to an estimate for $\Delta QALYloss$. Then as a final step, the net health impact can be expressed as a change in QALYs (rather than QALY losses):

$$\Delta QALY = 1 - \Delta QALYloss$$

As QALYs represent health loss, a positive result for $\Delta QALY$ implies that changing to the alternative scenario has an beneficial health impact overall, whereas a negative $\Delta QALY$ implies the alternative scenario is adverse.

The results for both DALYs and QALYs are subject to the same limitations as before, that the effects have been considered in isolation. How these uncertainties can be taken into account in interpreting the results is discussed in section 15.1.

13. Direct health loss calculations for a single individual

In general the directly attributable health loss approach to risk-benefit assessment is intended for application to a population or subpopulation comprising many individuals, and the result is an estimate of the total annual impact for that population.

It is also possible to apply Equations (1)-(4) for a single individual, carrying out the calculation only once for a single age *CA* and setting $I_{sf} = 1$. When this is done, the output should be interpreted as the average annual net health for individuals of the type represented by the modelled individual (e.g. with the same age, gender, life expectancy, etc.). However, it must be remembered that the result is specific for individuals of that age and type, and that net health impacts for other ages and types of individuals will generally be different. In special cases, calculations for single individuals may be interpretable if all the health effects of the dietary change are thought to have the same age profile, e.g. if the diseases that increase or decrease all occur in the same age range. If the age profiles of the diseases differ, then conducting calculations for a range of ages may help to build a picture of how the net health impact varies with age. Usually, it will be preferable to carry out calculations for a representative sample of individuals covering all ages of interest (e.g. as done by Hoekstra et al., 2008). Nevertheless, calculations for single individuals may be helpful when developing the assessment to explore the consequences of different assumptions or data, and to help the user understand the operation of the model.

14. Data needed as inputs to the Qalibra framework

The following sections outline the types of data needed as inputs to calculations using Equations (1)-(4), and briefly identify some of the main issues and complications that may be encountered. They are presented in the order shown in the diagram of the framework (Figure 2).

Note that the Qalibra software optionally allows quantification of uncertainty for most of the model inputs (see sections 15.2 and 15.3). In addition, many of them may be functions of age, intake, or other individual characteristics (e.g. gender).

14.1 Individuals and their attributes

The population relevant to the dietary change or intervention under assessment should already have been identified in earlier stages of assessment (i.e. at Brafo Tiers 1 and 2). For quantitative integration of risks and benefits using the Qalibra framework (Brafo Tiers 3 and 4) it is necessary to know the age structure of this population, i.e. the proportions of individuals in different age groups, unless the assessment is for a single individual (see section 13). If the calculation will be based on data for individuals from a dietary survey, I_{sf} in Equations (1)-(4) can be used to scale the dietary survey population up to the total population of interest.

The current age *CA* of each individual is a required input, shown explicitly in Equations (1)-(4). Data on other attributes of each individual (e.g. sex, body weight, etc.) may also be required if they are known to influence any of the other input parameters, e.g. if they appear as covariates in dose-response relationships.

When conducting an assessment for a population (as opposed to for a single individual), the calculations are performed for a sample of individuals from that population (and then scaled up to the full population using the scaling factor I_{sf} if desired). These individual characteristics can be data on real individuals (e.g. from a dietary survey) or simulated from a model based on summary statistics for the population (e.g. age distribution). If the sample of individuals is too small, then the results may be unduly influenced by chance inclusion of different individuals or different combinations of parameter values (a form of sampling uncertainty), and to adequately represent variability of intakes among individuals with similar attributes. It is therefore important to ensure that the sample of individuals (prior to scaling up) is large enough to provide stable estimates of the assessment outputs, i.e. the estimates should not change materially if the size of the sample is increased.

14.2 Life expectancy

Data on life expectancy *LE* is also required as a separate input matrix by the Qalibra software. *LE* will generally be a function of age and gender, and may also be a function of other individual attributes. *LE* should be expressed as the *average* life expectancy for individuals with each contribution of age, gender and other attributes.

14.3 Health effects to be quantified

The health effects to be considered for a quantitative integration of risks and benefits should be selected by a process of screening. This should start by identifying all adverse and beneficial effects that might potentially occur (including a comprehensive search of relevant literature) and assessing them individually to determine which may be expected to occur in the scenarios under assessment (Tier 1 of the Brafo approach, Hoekstra et al., submitted). Those effects that are expected to occur may then be considered together qualitatively (Tier 2 of the Brafo approach), to evaluate their relative contributions to net health impact. Quantitative assessment of net health impact (Qalibra framework and Tiers 3 and 4 of Brafo) should potentially consider all of the effects assessed at Tier 2. It may not be necessary to quantify all of these effects: if the difference between scenarios after quantifying the larger effects is big enough, it may be possible to conclude by qualitative evaluation that it would not be outweighed by the sum of the smaller effects, without quantifying them. Therefore, it may be efficient to start Tier 3 by quantifying only the effects that appear from Tier 2 to be largest, and then quantify the smaller effects only if this is necessary to reach a conclusion.

14.4 Dietary intakes

Estimates are required of the intakes or exposures associated with each health effect, in each scenario (reference and alternative). Intake must be expressed in the same units as the dose-response relationship for the health effect. Often this will be expressed as daily intake of a specified nutrient or contaminant, e.g. in mg/kg bw/day. In other cases, the dose-response relationship may instead be expressed in terms of weight or portions per day of a specified food type (e.g. fish), in which case intakes must be expressed in the same units.

Note that the estimates of dietary intake should refer to the period of exposure that is relevant to the occurrence of the effect at the current age of the individual under consideration. This is not necessarily the intake at the current age. For chronic effects,

it is the habitual long-term intake for each dietary scenario (ignoring the period of transition between the two, which is not considered in the directly attributable health loss method). For effects in offspring, it will be the intake of the mother during the relevant time period. For acute effects it might be, for example, the maximum one-day intake experienced during the year, but this will require consideration case-by-case.

For simple calculations based on a single individual, only a single value for intake is required (e.g. average intake, or other choices depending on the needs of the assessment, e.g. best case or worst case). For calculations with more than one individual, a separate exposure value is required for each individual (e.g. this could be estimated intakes generated by probabilistic exposure models such as those provided on the MCRA platform[4]). Note that the intakes for the reference and alternative scenarios should relate to the same sample of individuals. The Qalibra software requires that these individuals and their intakes appear in the same order in the input data for both scenarios, so that the change in health between the two scenarios can be calculated for each individual.

Often, intakes of different foods or substances in the same assessment will be correlated (e.g. fish intake may be correlated with methyl mercury intake). Where such correlations exist, these should be represented appropriately in the input data (e.g. the fish intake for individual X should be consistent with the MeHg intake for the same individual).

14.5 Probability of quantal effects

Normally the dose-response function for a quantal effect represents the probability of the effect occurring as a function of the intake of the relevant contaminant or nutrient. Generally the function is continuous, e.g. probit, and can be summarised by 2 or more parameters (e.g. intercept and slope) relating the probability of effect to an appropriate measure of intake.

Equations (1) and (3) for quantal effects are readily applicable to chronic effects, or to acute effects that may not recur for the same individual within the same year. The direct health loss method is not well suited for dealing with recurrent effects: possible work-arounds for these are outlined earlier in section 9.

Effects which occur in future generations, e.g. the children of mothers exposed to a substance, also require modifications to the calculations (see section 10).

Dose-response relationships may include covariates, e.g. the dose-response relationships can be different for different population groups, e.g. sex or age classes. If so, this needs to be taken into account in the calculations. The Qalibra software allows the use of discrete covariates (e.g. sex); continuous covariates have to be represented by discretising them (e.g. age must be divided into age classes). The number of classes is not specifically limited: the spacing and number of classes should be chosen by the user so as to adequately represent the influence of each covariate.

A key issue concerning the representation of dose-responses is that the directly attributable health loss calculation requires the probability of effect to be expressed as

[4] www.biometris.wur.nl/UK/Software/MCRA+Monte+Carlo+Risk+Assessment/

the *probability of onset per year for specified ages or age groups*. Dose-response epidemiological studies may report dose-response relationships in this form (e.g. relative risk per year, as a function of age) but animal studies generally estimate lifetime probabilities.

When dose-responses are expressed as lifetime probabilities, it will be necessary to convert them to annual probabilities, which will often vary with age. This requires data or assumptions on age of onset of effect. Age of onset may be available from epidemiological or intervention studies of humans. Age of onset is rarely determined in animal studies, and even when it is, there would be substantial uncertainty if it is assumed the same relative age would apply for humans. In some cases, the age of onset may be clear, e.g. for effects manifested in offspring at birth. For other cases, a possible approach is to assume age of onset when the effect is caused by the dietary change follows the general distribution for the same or similar diseases in the human population. However it must be taken into account that such assumptions are likely to involve considerable uncertainty, as it is very likely that the two distributions will differ (e.g. if part of the incidence of the disease is due to causes other than the contaminant or nutrient being assessed, with different ages of onset). When the appropriate treatment for age of onset is very uncertain, it will be prudent to try different assumptions to explore the sensitivity of the assessment to this choice.

Equations (1) and (3) also require the probability of effect to be expressed in absolute not relative terms. When the dose response is relative, e.g. a relative risk from an epidemiological study, it will be necessary to convert it to absolute probability of effects by combining the relative risk with data on baseline effect probabilities: this may be available from medical statistics for the population under assessment, or for another population which is considered similar.

Equations (1) and (3) also require that the probability of effect must relate to humans. If the dose-response relationship comes from animal studies, it will be necessary either to assume the human dose-response is the same, or to apply a suitable form of extrapolation (e.g. apply an adjustment consistent with the inter-species extrapolation component of uncertainty factors used in risk assessment).

Generally, dose response relationships are continuous, and need to be specified by a mathematical equation containing one or more parameters. The original plan for the Qalibra software was to provide a drop-down menu of commonly-used dose-response models, but it was considered desirable to provide more flexibility to allow use of any dose-response relationship the user considered appropriate. This is achieved by requiring the user to input the dose-response in discretised form, i.e. as a series of paired values for dose and response. The Qalibra software then forms the continuous relationship by linear interpolation between the successive pairs of values. Users need to check that the number and spacing of the pairs of values is sufficient for linear extrapolation between them to represent the continuous relationship adequately (this can be checked by repeating the assessment with more or less points to see if it alters the outcome).

The Qalibra software also provides the option of specifying the dose-response relationship for probability of effect using a single value. Qalibra treats this value as a threshold dose, below which the probability of the effect is zero, and above which it is

one. This allows the user to carry out simplified assessments, e.g. using NOAEL or LOAEL values as thresholds, which may be useful for exploratory or screening purposes. Note that Qalibra offers this option only for increasing relationships for probability of effect; if the probability of effect decreases with increasing dose, then this must be specified as a series of pairs of values (see previous paragraph).

It will be clear from the issues discussed above that deriving appropriate dose-response relationships for use in risk-benefit assessment requires specialised expertise in toxicology, epidemiology & modelling, and involves substantial uncertainties. The impact of all identified uncertainties should be considered when interpreting the results of the assessment (see section 15.1).

14.6 Probability and magnitude of continuous effects

Equations (2) and (4) require two types of input for continuous effects: first, a dose-response relationship describing the magnitude of the effect, and secondly, the probability that the effect will occur (i.e. begin) in the current year, as a function of age. All of the issues discussed above for quantal effects apply in similar ways to one or both of these two parameters: the effect may be chronic or acute, may depend on covariates which must be expressed in discrete form, probabilities must relate to onset in the current year, and both probability and magnitude must be expressed in absolute not relative terms and relate to occurrence of the effect in humans (extrapolated if necessary from animals).

The probability and magnitude of effect may both be functions of the same measure of intake and therefore correlated with one another. This can be represented in the Qalibra software by specifying both as functions of the same measure of intake, and using the same intake data.

The Qalibra software uses linear interpolation between points specified by the user to construct continuous dose-response relationships for both probability and magnitude of effect (as for the probability of quantal effects, see section 14.5). Alternatively, a single value can be used to represent a threshold for probability of effect (with the same restriction as for quantal effects), but not for magnitude of effect.

14.7 Probabilities of recovery and death

Equations (1)-(4) required data on overall recovery and mortality rates for each effect, obtained for example from national health statistics. Note that these are overall probabilities for recovery or death within the remaining normal life expectancy, and not probabilities for recovery or death in the current year. This is because the directly attributable health loss method considers the total impact (over remaining life) of diseases that begin in the current year.

In some cases the probabilities of recovery or death may be dependent on personal attributes such as age or sex. It should also be considered whether they may be influenced by the intake of the substance causing the effect (e.g. a high intake might lead not only to a higher probability of getting the disease, but also to a lower probability of recovery and higher probability of death).

Note that the approach described here makes the simplifying assumption that individuals who recover regain full health, whereas in reality for some diseases there

may be only partial recovery. Where this applies, its influence on the outcome may be explored by trying alternative assumptions, e.g. setting the probability of recovery to zero, and/or changing the disease weight to approximate an average of the disease severity over the remaining life.

An important uncertainty affecting the estimation of probabilities of recovery and death is that they may differ depending on the cause of the disease, in which case the probabilities for the general population (e.g. from national health statistics) may differ from those that would apply when the disease is caused by the dietary change under assessment. Further uncertainty is introduced if the dose-response for the effect is derived from animal studies, because it is then necessary to make an assumption about which human diseases the effect corresponds to, in order to make use of human data on recovery and mortality. The impact of all identified uncertainties should be considered when interpreting the results of the assessment (see section 15.1).

14.8 Duration of disease

Equations (1) to (4) require data on the duration of each health effect in years. Separate estimates are required for individuals who recover (YLD_{rec}) and those who die (YLD_{die}). In both cases, the duration should be the *average* for individuals with that outcome. Both types of data may be derived from national health statistics. Both may depend on personal attributes, such as age or sex, and they may also depend on the intake of the substance causing the effect.

Disease duration may depend on the cause of the disease, in which case the durations for the general population (e.g. from national health statistics) may differ from those that would apply when the disease is caused by the dietary change under assessment. Another important source of uncertainty is introduced if the dose-response for the effect is derived from animal studies, because it is then necessary to make an assumption about which human diseases the effect corresponds to, in order to make use of human data on disease duration. The impact of all identified uncertainties should be considered when interpreting the results of the assessment (see section 15.1).

14.9 Severity of effect (disease weights)

DALY or QALY weights are available in the literature (e.g. WHO). In many cases a single weight is given for each disease or health state, although in reality the severity of a disease may differ depending whether the individual eventually recovers, dies from the disease, or continues with the disease until their normal life expectancy. The Qalibra framework provides the option to represent this, using different weights for different stages of the disease, if these are available in the published literature or if the user wishes to explore the effect of different assumptions. It is also possible that disease severity may depend on personal attributes such as age or sex, or on the intake of the substance causing the effect. The weights should represent the *average* severity of the effect for individuals with the relevant set of attributes.

It is conceivable that disease severity may depend on the cause of the disease, in which case the weights published for the general population may differ from those that would apply when the disease is caused by the dietary change under assessment. Another important source of uncertainty is introduced if the dose-response for the effect is derived from animal studies, because it is then necessary to make an

assumption about which human diseases the effect corresponds to, in order to make use of human disease weights. For continuous effects, published weights may be presented as step functions of the effect (e.g. IQ), whereas in reality the severity of the effect is more likely to be a continuous function of effect size. The impact of all identified uncertainties should be considered when interpreting the results of the assessment (see section 15.1).

Disease weights for QALYs are usually measured with standardised questionnaires like EQ-5D, SF-6D, HUI3 etc. in which generic health states are valued. As is to be expected different measurement instruments (questionnaires) will show different values (e.g. (Fryback et al. 2007). Kopec and Willison (2003) show that depending on the measurement method disease weights (QALY) can differ substantially for the same disease e.g. from 0.26 to 0.71 (depression) or from 0.58 to 0.72 for arthritis. Also different populations value health states differently (Havranek and Steiner, 2005).

Schwarzinger et al. (2003) show differences in disability weights (DALY) depending on the population and on the method with which weights are measured. Disability weights for stroke e.g. differ between 0.17-0.68. Differences also exist when patients, the general public or health practitioners are asked (De Wit et al. 2000, Hoekstra et al. 2008, Schwarzinger et al. 2003). Essink-Bot et al. (2007) conclude that health state valuations may be sensitive to individual response patterns that could not be explained by characteristics such as age, sex or educational level. (Fryback et al. 2007) points out that the disease weights of the one who takes the decision should be used and that disease weights and consequently the valuation of health states is only a part of the general societal utility function.

The use of DALYs and especially age weighting and discounting is an area of open debate (e.g. Arnesen and Kapiriri, 2004; Murray and Lopez, 2000; Paalman et al. 1998; Murray and Acharya, 1997). Murray and Lopez (1996) published discounted as well as non-discounted values because of the difficulties of choosing a discount rate.

In the light of these issues, it is clearly important to maximise transparency when common health measures used (Arnesen and Kapiriri, 2004). Care should be taken that disease weights are relevant for the population of interest. Furthermore, each disease should be measured with the same instrument and the group valuing the disease should be similar also, not the general public for one disease and patients for the other.

When animal experiments are performed with toxic substances, endpoints are measured that have no direct clinical endpoint. Sometimes this is also true for human epidemiological studies. Examples are sperm counts in rats that are exposed to dioxins or IQ tests for children whose mother has been exposed to methyl mercury via fish consumption. In order to compare these effects with other (beneficial) health effects they need to be converted in a common health measure such as DALYs. Obviously, this cannot be done without crude assumptions and the introduction of large uncertainties that would be difficult to quantify. Presumably, the measured endpoint will have some relation with human health (if it does not, there is no need to include it in the assessment). Because the conversion to a clinical endpoint for which DALY or QALY weights exist is not straightforward the conversion needs to be explained and

sensitivity analysis must be performed to investigate the influence of the assumptions on the final results. The assumptions have to be made clear and the argumentation must be given in a narrative.

15. Addressing uncertainty in risk-benefit assessment

Risk-benefit analysis is affected by many potential sources of uncertainty which may all contribute to uncertainty in the estimated net health impact. This uncertainty may have important implications for decision-makers. For example, the median estimate for net health impact may be positive but if the probability interval is wide there may be a large chance that the actual impact is negative.

The importance of considering uncertainty in risk assessment is recognised in the Codex Working Principles for Risk Analysis, which state: "Constraints, uncertainties and assumptions having an impact on the risk assessment should be explicitly considered at each step in the risk assessment and documented in a transparent manner. Expression of uncertainty or variability in risk estimates may be qualitative or quantitative, but should be quantified to the extent that is scientifically achievable." (Codex 2007). Logically, if these principles apply to risk assessment they should also be applicable to risk-benefit analysis.

The Codex principle quoted above implicitly recognises that is not feasible to quantify all sources of variability and uncertainty. It implies that all those that are quantifiable should be quantified, but in fact this is not necessary if a qualitative consideration of uncertainty is sufficient for decision-makers to reach a decision with adequate confidence (e.g. if it is clear that the assessment is conservative). Therefore what is needed is a flexible or tiered approach which:

- Documents all identifiable sources of uncertainty
- Evaluates all of them at least qualitatively
- Quantifies them to the extent that is necessary for decision-making.

An approach of this sort has been described in guidance published by EFSA (2006b) for dealing with uncertainty in exposure assessment, and is outlined in the following sections below. Although aimed at exposure assessment, the approach is sufficiently general that it can be applied equally to the assessment of adverse and beneficial effects and net health impact.

Qualitative evaluation of uncertainties may be sufficient for decision-making in some cases, e.g. if it provides sufficient confidence that the net health impact of a proposed intervention is positive. In cases where more confidence is required, one option is to obtain additional data to reduce uncertainty. The alternative is to refine the characterisation of the uncertainties, by quantifying one or more of the most important uncertainties.

EFSA (2006b) suggests an iterative process which progressively quantifies more of the uncertainties, either deterministically or probabilistically, until there is sufficient confidence for decision-making. Ideally, priority would be given to quantifying the uncertainties thought to have most influence on the outcome, as indicated by the qualitative evaluation; probabilistic analysis may be targeted on those uncertainties identified as most influential by sensitivity analysis.

Generally only a small proportion of all uncertainties will be quantified, so it is essential that this is always accompanied by qualitative evaluation of the unquantified uncertainties, so as to provide a comprehensive characterisation of the overall uncertainty affecting the assessment.

The overall magnitude of uncertainty associated with a risk-benefit assessment may often be large. This should not be regarded as implying a failure of the assessment process; on the contrary, it provides essential information for decision-making (Codex 2007, Madelin 2004).

Common sources of uncertainty affecting risk-benefit analysis are summarised in Table 1. This list may be helpful as an *aide memoire* when identifying the uncertainties relevant in individual assessments, but is not comprehensive.

Table 1. Common sources of uncertainty affecting risk-benefit analysis.

Uncertainties affecting problem formulation
• Specification of reference and alternative intake scenarios
• Specification of relevant population
• Identification of sensitive or otherwise important sub-populations
Uncertainties affecting hazard and benefit identification
• Identification of relevant nutrients and contaminants
• Identification of relevant health endpoints for each nutrient and contaminant
Uncertainties affecting intake assessment
• Measurement uncertainty in concentration data
• Measurements below the limit of detection, quantification or reporting
• Sampling uncertainty due to limited number of concentration measurements
• Bias due to intentional targeting of monitoring for contaminants
• Uncertainty about correlations between concentrations of different contaminants/nutrients
• Extrapolation of concentrations from measured to unmeasured foods
• Future changes in levels of chemical use or contamination
• Uncertainty in recording of foods and food weights in dietary surveys
• Sampling uncertainty due to limited numbers of persons and days in dietary surveys
• Measurement uncertainty and bias in body weight data (usually minor)
• Uncertainty about degree of uptake of dietary recommendations
• Uncertainty about level of background exposure (other foods, or other routes of exposure)
• Uncertainty about compensatory changes in existing diet when taking up dietary recommendations
• Assumptions about how the diets of individuals change over long time periods
Uncertainties affecting dose/response relationships estimated from animal data
• Within- and between-laboratory variation
• Sampling uncertainty due to limited number of subjects
• Experimental errors e.g. in administration of treatments (generally minor)
• Choice of dose-response model (goodness of fit)
• Extrapolation to low doses (when required)
• Extrapolation from animals to humans
• Provision for within-species variation
Uncertainties affecting dose/response relationships estimated from epidemiological studies
• Sampling uncertainty due to limited number of subjects
• Estimation of intakes
• Choice of dose-response model (goodness of fit)
• Extrapolation to low doses (when required)
• Combination of multiple studies (meta-analysis)
• Relevance of study population to target population
Uncertainties affecting conversion to a common health currency (e.g. DALY, QALY)
• Estimation of age of onset for health effects (especially those modelled from animal data)
• Characterisation of disease severity (e.g. DALY or QALY weights)
• Difficulty integrating effects if more than one stressor affects the same health endpoint
• Variation of disease severity and duration, including any relation with dose
• Estimation of recovery rate and time to recovery
• Estimation of mortality rate and time to death
• Recovery, death, severity, duration, age of onset may depend on cause of effect, and therefore differ from national statistics
• Interactions between different diseases or health endpoints (including background diseases not directly affected by the dietary change or intervention), including comorbidity and substitution
Uncertainties in the probabilistic treatment of uncertainties
• Choice of distributions to represent uncertainties
• Dependencies between parameters
• Uncertainty introduced by computational methods, e.g. number of simulations/iterations.
Uncertainties due to factors not considered in the assessment

15.1 Qualitative evaluation of uncertainties

The EFSA approach starts with a systematic examination of all parts of the assessment to list all identifiable sources of uncertainty. Initially, every uncertainty is evaluated qualitatively. This could be done in various ways, but the EFSA (2006b) document suggests using scores in the form +, ++, +++, -, - -, - - - and combinations such as - / ++ to represent a subjective assessment of the direction and magnitude of the influence of each source of uncertainty on the outcome of the assessment (i.e. to indicate how different the true outcome might be). A similar score is then assigned to represent a subjective evaluation of the combined influence of all the identified uncertainties. EFSA (2006b) suggests presenting this evaluation in a tabular format: an adaptation of this format for risk-benefit assessment is illustrated in Table 2.

Table 2. Tabular approach for qualitative evaluation and expression of uncertainties affecting risk-benefit assessments (adapted from EFSA, 2006b). The +/- symbols indicate whether each source of uncertainty has the potential to make the true health impact of changing from the reference to alternative scenario more beneficial (+) or adverse (-). The number of symbols provides a subjective relative evaluation of the magnitude of the effect (e.g. +++ indicates an uncertainty that could make the true benefit much higher). If the effect is uncertain, or could vary over a range, lower and upper evaluations are given (e.g. - / ++ or + / ++). Finally, the combined impact of all the uncertainties is evaluated subjectively.

Source of uncertainty	**Magnitude and direction of influence on true net health impact**
Concise description of source of uncertainty (e.g. intake of contaminant in alternative scenario may be under-estimated due to treating non-detects as zero concentrations)	*Symbols to show evaluation of influence* (e.g.: -/- -)
Insert one row for each source of uncertainty affecting the assessment	
Overall evaluation of uncertainty affecting the assessment outcome *Add narrative text here, describing the assessor's subjective evaluation of the overall degree of uncertainty affecting the assessment outcome, taking account of all the uncertainties identified above.*	*Evaluation of overall uncertainty* (e.g. - - - / +)

15.2 Quantitative evaluation of uncertainties

EFSA (2006b) distinguishes between 2 levels of quantitative evaluation of uncertainty: deterministic and probabilistic.

- Deterministic methods include "worst case" analysis, interval analysis, and sensitivity or scenario analysis. These explore the impact on the assessment of alternative inputs or assumptions, without attempting to quantify their relative probability. They therefore generate a range of possible values for the outcome (e.g. the net health impact), without quantifying their relative likelihood. This approach is used in Tiers 2 and 3 of the Brafo framework (Hoekstra et al. submitted) and can be implemented with the Qalibra framework and software by performing repeat assessments with different inputs to explore alternative assumptions.

- Probabilistic methods go beyond this by using probability distributions to represent the relative likelihood of alternative inputs or assumptions. These distributions are then "propagated" through the assessment, to generate a probability distribution for the outcome (e.g. net health impact) which represents the combined effect of all the quantified uncertainties. Probabilistic methods are included in Brafo Tier 4, and are available as an option for quantifying uncertainty in the Qalibra software.

In probabilistic modelling generally, distributions are used to represent variability (real differences in a parameter, e.g. in the intake for different individuals) and/or uncertainty (due to lack of precise knowledge about the parameter, e.g. due to measurement errors or sampling variation). Distributions for input parameters are "propagated" through the model or assessment, e.g. using Monte Carlo simulation, to generate a probability distribution for the net health impact that represents the combined effect of all the quantified uncertainties.

Monte Carlo simulation repeats the whole assessment calculations thousands of times (until the output distributions stabilise), each time with different parameter values drawn from their distributions. The results from each iteration are combined to form probability distributions for the outputs. These output distributions may be presented graphically, and/or used to put confidence intervals on numerical outputs. They can also be used to derive estimates for the probability of achieving particular outcomes, e.g. the probability that the change in net health impact is positive, or the probability it exceeds some critical level of interest to the policy-maker.

Probabilistic modelling requires a high level of statistical expertise in addition to the other disciplines that are already required for risk-benefit assessment. Probabilistic methods are increasingly used for exposure assessment (e.g. Cullen and Frey 1999, EFSA 2007) and are beginning to be used for risk characterisation (e.g. van der Voet and Slob 2007) and risk-benefit analysis (e.g. van der Voet et al. 2007, www.beneris.eu and this project). Some general principles for Monte Carlo risk assessment have been published by the US EPA (1997).

15.3 Probabilistic treatment of uncertainty in Qalibra

The Qalibra software is designed to provide maximum flexibility and control to users in how uncertainty is quantified. This is achieved by allowing the user to enter a sample of values representing uncertainty for any or all of the input parameters. A separate input file is required for each input parameter: for parameters for which uncertainty is not quantified, this file contains only one column of estimates for that parameter; whereas for parameters for which uncertainty is quantified, the input file contains multiple columns containing the sample of multiple estimates for the parameter, representing its uncertainty (the rows in each file relate to different combinations of covariates, e.g. age, intake, gender, etc.). This sample is then used to represent the uncertainty for that parameter in the Qalibra calculations.

The flexibility provided by this approach allows the analyst to select any suitable, or convenient, method to quantify uncertainty. Samples for uncertain parameters may be generated by frequentist or Bayesian statistical methods, as appropriate, or by sampling from distributions specified by expert judgement (preferably using formal methods of expert elicitation). All methods used should be fully justified and

documented, and any unquantified uncertainties associated with them (e.g. relating to underlying data and how they are modelled) should be taken into account as part of the qualitative evaluation of uncertainties (see section 15.1).

The Qalibra software executes multiple iterations of the risk-benefit calculation (currently 10,000), each taking different combinations of uncertain values for those inputs where the user has provided samples to represent uncertainty. If the user provides input samples smaller than 10,000 per parameter then Qalibra resamples the input values to obtain the larger sample. If the user provides input samples greater than 10,000 then Qalibra uses this larger value for the number of iterations.

Intakes for different effects (for both reference and alternative scenarios) are resampled together, i.e. in any given iteration, the same column of the uncertain estimates in the input matrix is taken for the reference and alternative intakes for all of the effects in the assessment. This is done because intakes of different contaminants and nutrients in the diet will generally be inter-dependent: e.g. different substances in the same food will be positively correlated, substances in different foods may be negatively correlated (e.g. if a person eats more fish they may tend to eat less meat), and if an individual's intake is higher than expected in the reference scenario it may also be higher in the alternative scenario. If the user represents these dependencies in the input matrices, then they will be maintained by the Qalibra resampling process. Note that this makes it important that the user provide sufficiently large samples for the intake parameters to adequately explore the different combinations of the uncertain values, as the linked resampling of these parameters in Qalibra will not explore additional combinations.

Other inputs to the calculation (i.e. parameters other than dietary intake) are resampled *independently* by the Qalibra software, to explore different combinations of uncertain values amongst these inputs and between them and the intake parameters. If it is thought these other inputs may be significantly interdependent, then this should be taken into account when considering the unquantified uncertainties affecting the assessment (see section 15.1).

Special care is required in the treatment of uncertainty for any parameters which are expressed as a function of individual attributes (age, gender, etc.). For such inputs, values relating to subpopulations with different combinations of attributes (e.g. juvenile females, adult males) are provided by the user as separate blocks of rows in the input matrix. If the user quantifies uncertainty, this is represented as a series of columns for each block of rows. In each iteration of the overall population calculation, Qalibra uses the same column for all rows. *This makes the user responsible for ensuring that any dependency of the parameter between subpopulations is represented appropriate in the input matrix.* For example, in the example presented later in this document (section 18), there are two effects for which p_{effect} is a function of fish intake. For both effects, the dose-response differs between age-gender classes, and uncertainty in the dose-response is quantified in the model inputs. It seems probable that the dose-responses for different age-gender classes are interdependent – e.g. if the true slope of the dose response for juvenile females is actually towards the upper end of its confidence interval, then the same is likely to apply to the slope for other age-gender classes, e.g. adult females, adult males, etc. To represent this dependency, the user needs to ensure that the uncertainty realisations for each

subpopulation are suitably correlated. This can be done rigorously by explicitly modelling the dependency when generating the input values, or more approximately (as in our example) by generating the input values independently and then sorting the columns for each subpopulation in the same order (e.g. so that the right-most column contains the most positive realisation of the dose-response for every age-gender group). If this is not done, then the uncertainty of the population health impact will be underestimated because, in many iterations of the calculation, upper-percentile values for one subpopulation will be offset by lower-percentile values for other subpopulations. In the example presented in section 18, the 95% uncertainty interval for the net health impact of increasing fish intake was -1.4 to -7.5 DALYs per year per 999 people when dependency of dose-responses for different age-gender groups was included, and -3.3 to -5.4 when they were sampled independently[5]. *These issues can have a large impact on the apparent precision of results, and must be considered for any parameter which varies between subpopulations and for which uncertainty is quantified.*

It is important to ensure that the samples of values provided by the user to represent uncertainty are large enough for the calculations to produce stable outputs. This can be checked by generating a series of different samples, or smaller and larger samples: if the sample statistics (mean and percentiles) differ materially, then the sample sizes should be increased. Similarly, the user may check whether the 10,000 iterations executed by the Qalibra software is sufficient to explore the different combinations of the uncertain inputs in the same way, by repeating the run several times and comparing the results. If the results differ materially, then it would be advisable to increase the number of iterations (by providing a larger number of samples for at least one input, see above) or download and combine the output from multiple runs.

16. Treatment of variability in the Qalibra framework

The Qalibra framework and software model variability between individuals in their expected health outcomes. This variability is driven by two sources. The first source is variability between individuals in their diets and hence their intakes of the contaminants and nutrients considered in an assessment. This is represented in the Qalibra software by the user entering intakes for each individual as separate values. The second source of variability is individual differences in the other parameters of the risk-benefit calculation (dose-response, severity and duration of disease, probabilities of recovery and death, normal life expectancy). In the Qalibra software, these are not entered as separate values for every individual; however, the user has the option to make them a function of discrete covariates representing individual attributes that influence them. In principle it would be possible to specify a covariate that takes different values for every individual, but usually the covariates will be used to represent variability between groups or classes of individuals. Age is a required covariate, because it is a variable in equation 1. In effect, it is used as the age of onset in the annual directly attributable health loss approach. In general, both p_{effect} and life expectancy are a function of age. Other covariates are optional. It may often be appropriate to specify gender as a covariate, as it is common that different dose-response relationships are reported for males and females. Qalibra requires covariates to be discrete: when a covariate is actually continuous (e.g. age), the user will need to

[5] The median estimates with and without dependency were almost identical: -4.32 and -4.29.

divide it into a sufficiently large number of classes to adequately represent its influence on the risk-benefit calculation.

17. Treatment of dependencies in Qalibra framework

The Qalibra framework allows the user to represent of some types of dependencies in the risk-benefit calculation:

- Dependencies between intakes of different foods, contaminants or nutrients (e.g. individuals who eat less meat are likely to eat more vegetables or dairy products).
- Dependency of effects on intakes, including the familiar dose-response relationships for the probability of quantal effects and the magnitude of continuous effects, and also any dose-dependency that may exist for other aspects of effects (e.g. age of onset, duration, severity, recovery rate, mortality rate).
- Dependencies between uncertainty distributions for the same parameter in different subpopulations (e.g. different age/gender groups, see section 15.3).

Users of Qalibra are responsible for correctly representing these dependencies within the input formats accepted by the software.

Some types of dependency that may be relevant when considering health risks and benefits cannot readily be accommodated within the directly attributable health impacts model adopted for Qalibra. These include:

- Dependencies or interactions between different health effects (e.g. the severity of one disease, or the probability of recovery or mortality from it, may depend on whether the individual suffers another disease concurrently).
- Dependencies in the between-individual variation of different parameters, except for intakes. This is because the directly attributable health impact calculations uses averages across individuals (of the same subpopulation) for all parameters except intakes, and because the result of a calculation based on averages may differ from the average result of the same calculation repeated for different individuals.

The potential impact of such dependencies is difficult to evaluate qualitatively, but should be considered as part of the overall evaluation of unquantified uncertainties affecting the assessment (see section 15.1). In cases where such dependencies might have a material effect, it may be possible to explore them quantitatively, either by using a different modelling approach or by representing them very approximately in Qalibra, e.g. by using the functionality for individual attributes to define separate population subgroups in different quantiles for parameters where individual variation may be important. Any assessment of this type would require very careful planning and interpretation.

18. Presentation of results

The primary output of the directly attributable health loss approach described above (based on Equations (1)-(4)) is the *potential annual change in health effect for the assessed population* (expressed in QALYs or DALYs) that results from implementing the alternative scenario instead of the reference scenario. However, it is also very informative to show also intermediate results such as the contributions of different health effects to the total, and the severity, incidence and years of life lost for each effect.

The format used by the Qalibra software for displaying these primary numerical outputs is illustrated in Table 3. This shows results for 2 health effects from an

assessment on consumption of oily fish done as a case study in the Qalibra project (the full assessment includes additional adverse and beneficial effects). The dietary change (increasing fish consumption to 200g/week for those individuals whose current consumption is lower) has a beneficial impact on both health effects so the individual and total changes in DALYs are negative. The assessment relates to 999 individuals (the number specified by the user who created the assessment), and the results are estimates of the annual average health impact for that population.

Table 3 is divided into several sections, in descending order:

- The change in DALYs between scenarios, for each effect and overall
- The DALYs for each scenario (reference and alternative)
- The incidence of the effects in each scenario
- The average magnitude of effects in each scenario (this applies only to continuous effects, and shown as zero for quantal effects)
- Results for those individuals who recover from each effect (the number of individuals in this group, the average duration of disease YLD, and the total DALYs)
- Results for individuals who continue with the effect to their normal life expectancy (as above)
- Results for individuals who die from each effect (as above plus the average years of life lost YLL).

The Qalibra software generates 95% uncertainty intervals for every output, although due to limitations of space these are shown only in the top row in Table 3. These intervals represent the combined effect of the uncertainties the user has quantified for the various inputs (uncertainty in the dose-response relationships for probability of effects).

Numerical results are shown to 6 decimal places, to enable the user to check the stability of the estimates by comparing different model runs. Given the many uncertainties affecting such assessments, it is recommended to truncate the results to fewer significant figures for general presentation.

The numerical results shown in Table 3 are aggregated over the population of individuals specified by the user (this can be any positive number). The Qalibra software also generates 4 types of graphical output that show how contributions to the overall health impacts vary across the population. This may be useful both in helping the user to understand the way the totals are built up, and for identifying situations where part of the population experiences a net health gain while another part experiences a net loss.

Three of the graphical outputs generated by Qalibra are illustrated in Figures 3-5, and the fourth is a variant of Figure 3 that uses a pie chart to show the proportions of individuals with zero and non-zero DALY changes. Figures 3-5 all relate to the same example assessment as Table 3.

Figure 3 is a histogram showing the distribution of individual contributions to the total annual DALY change for the population, i.e. how the total shown in the top right hand corner of Table 3 is distributed between the 999 individuals modelled in that assessment. It can be seen there is a large peak of individuals with close to zero

DALY change, and a proportion of individuals with varying degrees of expected benefit (negative DALY change, to the left of the figure). Note that even though the actual impact of stroke or fatal CHD is not small (e.g. a total of around 15 years of life lost for fatal CHD) the *expected* impact per person *per year* is very small because (a) the probability of a particular individual getting CHD in a given year is small, and (b) the change in that probability between the reference and alternative scenarios is still smaller.

Figure 4 shows the same distribution as Figure 3, but plotted as a complementary cumulative distribution function. This provides more detail on the shape of the modelled distribution than the histogram, and also allows the uncertainty intervals to be plotted. The curve shows the percentage of the population (on the vertical axis) with DALY changes more positive than each point on the horizontal axis; for example, it can be seen that about 35% of the population have no change in health: this is because these individuals already consume over 200g fish per week in the reference scenario, and it is assumed their consumption is unchanged in the alternative scenario.

Figure 5 shows the same DALY changes as Figures 3 and 4, plotted against the two covariates used in this assessment: age (a mandatory covariate in every assessment) and gender. The stepped pattern in the left hand graph reflects a combination of two factors. First, the probabilities of stroke and CHD increase with age, so the benefit from avoided stroke and CHD becomes greater at older ages. Second, within each age group, a reverse trend occurs, with larger benefits at younger ages within the age group. This is because the probabilities of CHD do not change within age groups, but younger individuals within an age group can benefit more from avoided fatal CHD because they have more remaining life years to lose (and hence, avoid losing). In reality, probability of CHD is probably a continuous rather than step function of age. If it were considered this might materially alter the principal results (total DALY change for population), then narrower age groups could be used when modelling probability of CHD.

The Qalibra software also allows the user to download the DALY estimates for the individual persons for each cell in the top three rows of Table 3. This enables the user to examine, outside the Qalibra software, the variation between individuals within the population and to generate statistics and graphs in the format of their choosing.

It is emphasised that results should always be accompanied by a clear explanation of how they should be interpreted, including that they represent an *indication* of the *potential annual change in health impact for the population* (see next section). In addition, the quantitative results should always be accompanied by evaluation of the many uncertainties that inevitably affect risk-benefit assessment, some of which were indicated in the preceding sections and Table 1. It is recommended to use the format illustrated in Table 2 to summarise the uncertainties together with an evaluation of their potential impact on the assessment outcome. It is also recommended that each assessment should be concluded with an overall narrative characterisation of the net health impact, which takes account of both the quantitative and qualitative evaluations and also any other relevant evidence regarding the health impact of the scenarios under assessment, e.g. comparisons with epidemiological data that have not been used as model inputs.

Table 3. Example of tabular risk-benefit assessment results generated by Qalibra software for effects of oily fish consumption on incidence of stroke and fatal coronary heart disease (CHD). Reference scenario = current Dutch diet, alternative = oily fish consumption increased to 200g/week for those individuals whose current consumption is lower. Results are estimates of annual directly attributable health effects for a sample population of 999 individuals, expressed in DALYs. 95% uncertainty intervals are shown in brackets (produced for all results but shown only in top row here). See text for more explanation.

	Incidence of stroke	Incidence of fatal CHD	Total:
Change in TOTAL DALY from Reference scenario to Alternative scenario	-2.561087 (-5.293814, 0.051080)	-1.769932 (-3.245847, -0.284764)	-4.301226 (-7.429154, -1.273373)
TOTAL DALY, Reference scenario	34.799440	16.673234	51.473170
TOTAL DALY, Alternative scenario	32.238353	14.900700	47.172092
Incidence (per year) in 999 individuals, Ref scenario	3.298551	1.103577	
Incidence (per year) in 999 individuals, Alt scenario	3.065471	0.991378	
Average magnitude of effect, Reference scenario	0.000000	0.000000	
Average magnitude of effect, Alternative scenario	0.000000	0.000000	
Recover, Reference scenario	0.000000	0.000000	
Recover, Alternative scenario	0.000000	0.000000	
Total YLD if recover, Reference scenario	0.000000	0.000000	
Total YLD if recover, Alternative scenario	0.000000	0.000000	
TOTAL DALY if recover, Reference scenario	0.000000	0.000000	
TOTAL DALY if recover, Alternative scenario	0.000000	0.000000	
Survive (with effect), Reference scenario	3.298551	0.000000	
Survive (with effect), Alternative scenario	3.065471	0.000000	
Total YLD if survive (with effect), Ref scenario	57.048262	0.000000	
Total YLD if survive (with effect), Alt scenario	52.849759	0.000000	
TOTAL DALY if survive (with effect), Ref scenario	34.799440	0.000000	
TOTAL DALY if survive (with effect), Alt scenario	32.238353	0.000000	
Die (from effect), Reference scenario	0.000000	1.103577	
Die (from effect), Alternative scenario	0.000000	0.991378	
Total YLD if die (from effect), Reference scenario	0.000000	0.000000	
Total YLD if die (from effect), Alternative scenario	0.000000	0.000000	
Total YLL if die (from effect), Reference scenario	0.000000	16.673234	
Total YLL if die (from effect), Alternative scenario	0.000000	14.900700	
TOTAL DALY if die (from effect), Reference scenario	0.000000	16.673234	
TOTAL DALY if die (from effect), Alternative scenario	0.000000	14.900700	

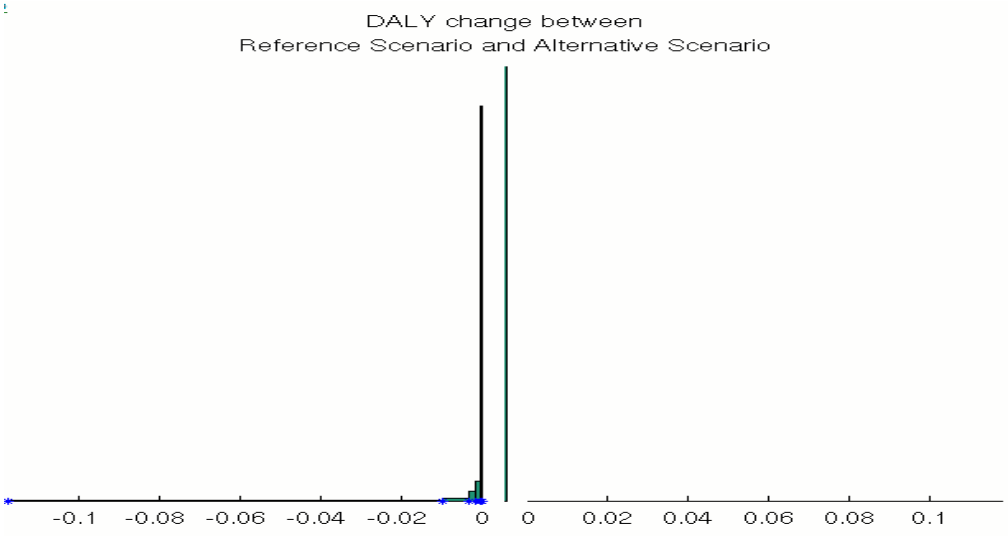

Figure 3. Histogram of individual contributions to the change in annual population DALYs between reference and alternative scenarios, for the assessment shown in Table 3. See text for details.

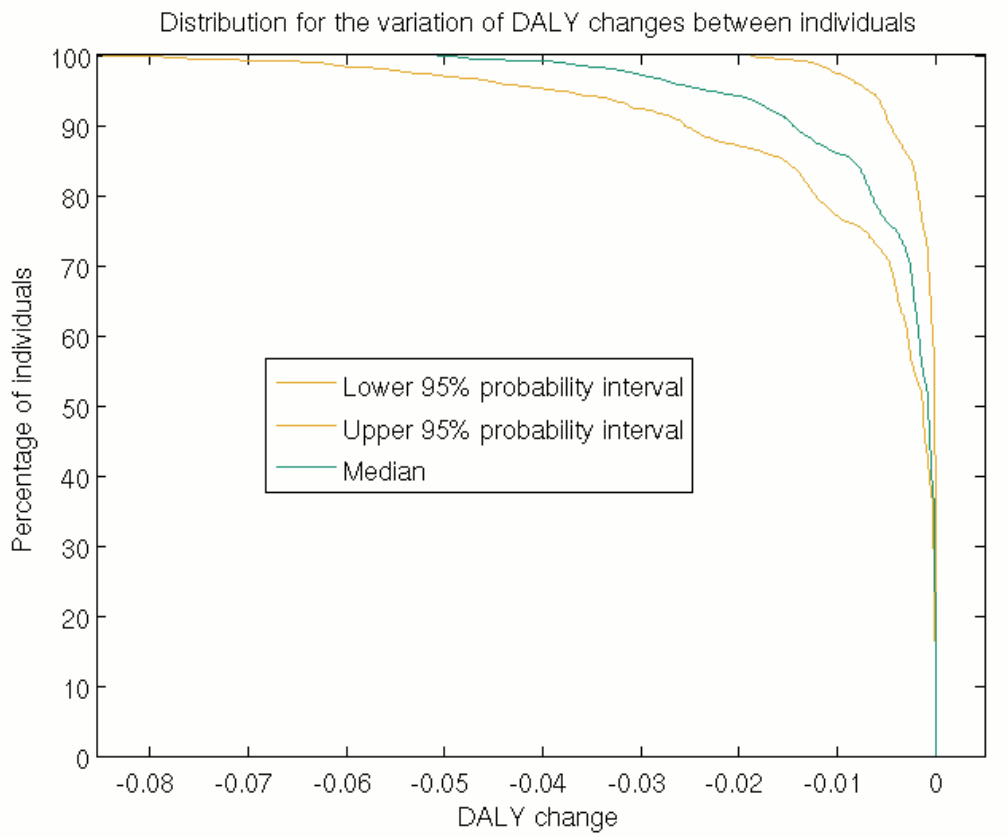

Figure 4. Complementary cumulative distribution of individual contribution to annual change in population DALYs between reference and alternative scenarios, for the assessment shown in Table 3. See text for details. The curve shows the percentage of the population (on the vertical axis) with annual DALY changes more positive than each point on the horizontal axis.

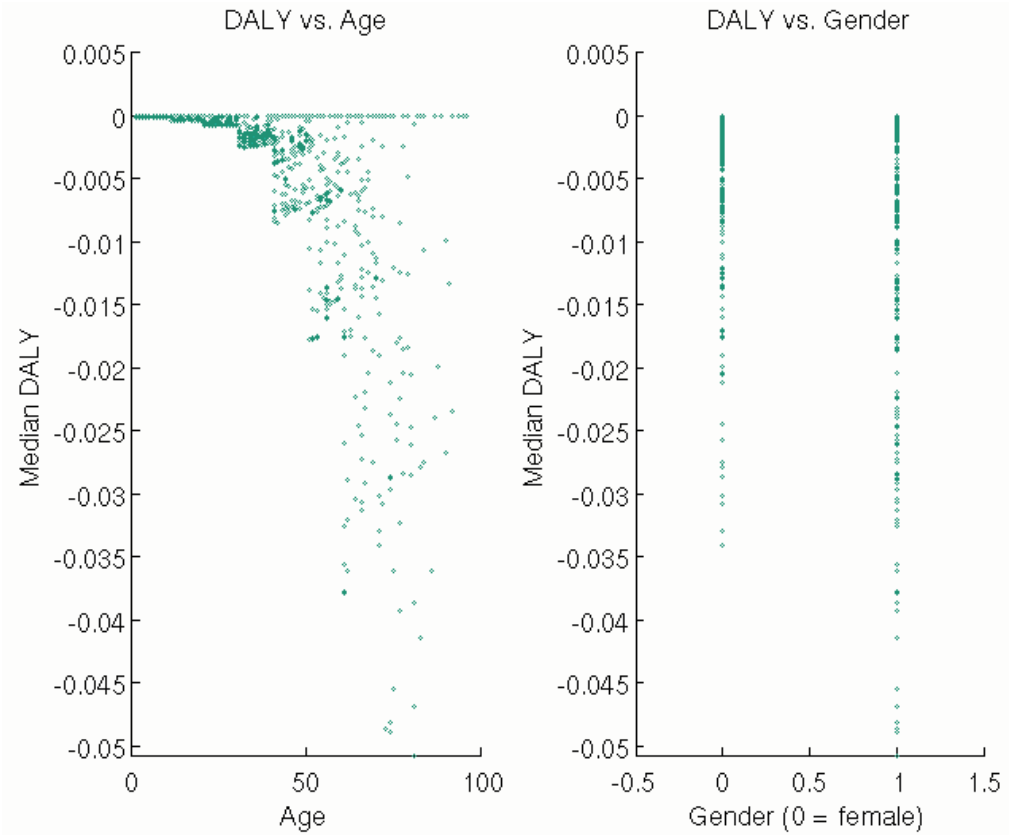

Figure 5. Individual contributions to annual change in population DALYs between reference and alternative scenarios shown in relation to age (left hand graph) and gender (right hand graph) for the assessment shown in Table 3. See text for details.

19. Interpretation of results

It will be clear from the preceding sections that interpreting the results of a quantitative risk-benefit assessment requires great care, full consideration of the associated assumptions and uncertainties, and significant expertise in the various scientific disciplines involved (toxicology, epidemiology, nutrition, intake modelling, etc.).

Due to all the uncertainties involved, the results should not be presented as estimates of "real" impacts on real individuals. Instead, they should be communicated as providing *an indication of the potential average annual health impact of the dietary change for the population* as a whole. Furthermore, it should be emphasised that the estimate represents the steady-state outcome over long time periods. This is because the directly attributable health loss approach implicitly considers a steady-state scenario and does not consider transition effects.

Among other considerations, when using the directly attributable health loss approach, the assessor will need to evaluate how the overall health impact might be

affected by the way the individual effects combine. The calculation simply sums the DALYs or QALY losses for all the effects that onset in a given year. This means that:

1. The combination of two effects is estimated as the sum of their individual effects. In reality, the combined effect could be larger (i.e. having both diseases has a greater impact on quality of health than the sum of their individual impacts) or smaller (e.g. if both diseases affect the same bodily function, and the additional impact of having both diseases is marginal). If there are many concurrent effects, or if there are a small number of major impacts, simply taking the sum of the DALY weights or QALY losses might exceed the maximum value of 1 (= death).
2. The DALYs or QALY loss caused by an effect will be overestimated to some degree because Equations (1)-(4) imply an assumption that individuals not affected by the dietary change will remain in full health until their normal life expectancy whereas, in reality, health generally declines in later years due to background diseases. However, this only affects the contribution from years lived with the disease, because the years of life lost to background diseases should already be accounted for in the normal life expectancy.

The assessor will need to consider how complications (1) and (2) are affecting their assessment, and take this into account together with other uncertainties when drawing conclusions (general approaches for evaluating uncertainties are discussed in section 15.1). The combined effect of (1) and (2) across all the effects may often cause over-estimation of combined health impact, unless some of the effects show substantially more than additive impact. If effects are overestimated, this implies the change in health impact between the two scenarios is likely to be overestimated also (subject to the effect of other uncertainties). However, the overestimation of health impacts is likely to cancel out to some extent when calculating the difference between the two scenarios. The assessor should consider carefully these possibilities and take them into account as part of the qualitative evaluation of unquantified uncertainties (i.e. as part of Table 2, section 15.1).

If the uncertainties affecting the assessment results are too great for the assessor and risk manager to form a sufficiently firm conclusion about the impact of the dietary change, two options are open: seek more data to reduce one or more of the uncertainties and repeat the assessment, or attempt to improve the clarity of the outcome by quantifying more of the key uncertainties probabilistically.

20. Risk management considerations

Based on the outcome of the risk-benefit assessment and consideration of the associated uncertainty, policy-makers may consider which scenario they prefer. This will depend on how opportunistic or risk averse they are, and on other information that is available to them including issues that were not incorporated in the health risk-benefit analysis (e.g. ‘other legitimate considerations’ in the sense of the EU Food Regulation, such as legal, economic, social and ethical considerations).

21. Final remarks

Risk-benefit assessment is inherently complex. This complexity is not introduced by the approaches used in the Qalibra framework, it is an inevitable consequence of the multidimensional nature of dietary change, of its effects on health, and of major limitations in the amount, quality and relevance of available data.

As already stated, risk-benefit assessment requires a high level of expertise in the relevant fields of science, and quantitative risk-benefit assessment additionally requires significant expertise in modelling and statistics. It requires substantial data or assumptions, is affected by many uncertainties, and the results require very careful interpretation and communication.

In this context it is hoped that the Qalibra framework and software will help by providing a common conceptual framework, assist users to identify important issues and data gaps, and provide a user-friendly software environment within which users can start with a simple deterministic assessment and progressively refine it by treating key elements probabilistically (when needed). The software also helps the user to organise the large number of datasets and model runs that may be needed, and to share them with colleagues of their choosing.

Acknowledgements

The authors gratefully acknowledge the contributions of everyone involved in the Qalibra project, its Science Advisory Panel, and the participants of Qalibra workshops. The project was funded by the European Commission's 6th Framework Programme, contract number FOOD-CT-2006-022957. The participation of Fera was partly supported by funding from the UK Food Standards Agency (project T01042).

References

Anand, S. and Hanson, K. (1998) DALYs: Efficiency versus equity. *World Development* **26**, 307-310.

Arnesen, T. and Kapiriri, L. (2004) Can the value choices in DALYs influence global priority-setting? *Health Policy* **70**, 137-49.

Cullen AC, Frey HC (1999) The use of probabilistic techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs. New York, NY, Plenum Press.

De Wit, G.A., Busschbach, J.J. and De Charro, F.T. (2000) Sensitivity and perspective in the valuation of health status: whose values count? *Health Econ* **9**, 109-26.

EFSA (2006a) Summary report EFSA scientific colloquium 6 - Risk-benefit analysis of foods: methods and approaches, 13-14 July 2006, Tabiano, Italy. *Scientific colloquium series of the European Food safety authority* **6**,

EFSA 2006b. Guidance of the Scientific Committee on a request from EFSA related to Uncertainties in Dietary Exposure Assessment. The EFSA Journal, 438, 1-54.

EFSA 2007. Opinion of the Scientific Panel on Plant protection products and their Residues on a request from the Commission on acute dietary intake assessment of pesticide residues in fruit and vegetables. The EFSA Journal, 538, 1-88.

Essink-Bot, M.L., Stuifbergen, M.C., Meerding, W.J., Looman, C.W. and Bonsel, G.J. (2007) Individual differences in the use of the response scale determine valuations of hypothetical health states: an empirical study. *BMC Health Serv Res* **7**, 62

Ezzati, M., Hoorn, S.V., Rodgers, A., Lopez, A.D., Mathers, C.D. and Murray, C.J. (2003) Estimates of global and regional potential health gains from reducing multiple major risk factors. *Lancet* **362**, 271-80.

FAO/WHO Expert Consultation FAO/WHO Expert Consultation, (Ed.) (1995) Geneva: WHO.

Fryback, D.G., Dunham, N.C., Palta, M., Hanmer, J., Buechner, J., Cherepanov, D., Herrington, S.A., Hays, R.D., Kaplan, R.M., Ganiats, T.G., Feeny, D. and Kind, P. (2007) US Norms for Six Generic Health-Related Quality-of-Life Indexes From the National Health Measurement Study. *Med Care* **45**, 1162-1170.

Gold, M.R., Stevenson, D. and Fryback, D.G. (2002) HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annu Rev Public Health* **23**, 115-34.

Havelaar, A.H., De Hollander, A.E.M., Teunis, P.F.M., Evers, E.G., Van Kranen, H.J., Versteegh, J.F.M., Van Koten, J.E.M. and Slob, W. (2000) Balancing the risks and benefits of drinking water disinfection: disability adjusted life-years on the scale. *Environmental Health Perspectives* **108**, 315-321.

Havranek, E.P. and Steiner, J.F. (2005) Valuation of health states in the US versus the UK: two measures divided by a common language? *Med Care* **43**, 201-2.

Hoekstra, J, Hart A, Boobis A, Claupein E, Cockburn A, Hunt A, Knudsen I, Richardson D, Schilter B, Schütte K, Torgerson P, Verhagen H, Watzl B, Chiodini A. BRAFO Tiered approach for benefit-risk assessment of foods. Submitted to *Food Chem Toxicol.*

Hoekstra J., Verkaik-Kloosterman J., Rompelberg C., van Kranen H., Zeilmaker M., Verhagen H. & de Jong N. (2008) Integrated risk-benefit analyses: Method development with folic acid as example. *Food Chem Toxicol* 46, 893-909

Hoogenveen R.T., van Baal P.H. & Boshuizen H.C. (2009) Chronic disease projections in heterogeneous ageing populations: approximating multi-state models of joint distributions by modelling marginal distributions. *Math Med Biol*

Kopec, J.A. and Willison, K.D. (2003) A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol* **56**, 317-25.

Madelin R, 2004. The importance of scientific advice in the Community decision making process. Opening address to the Inaugural Joint Meeting of the members of the Non-Food Scientific Committees. Directorate General for Health and Consumer Protection, European Commission, Brussels.

Murray, C.J. (1994) Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bull World Health Organ* **72**, 429-45.

Murray, C.J. and Acharya, A.K. (1997) Understanding DALYs (disability-adjusted life years). *J Health Econ* **16**, 703-30.

Murray, C.J. and Lopez, A.D. (1996) *The global burden of disease: a comparative assessment of mortality and disability from disease, injuries, and risk factors in 1990 and projected to 2020.*, Cambridge, Mass: Harvard School of Public Health, on behalf of the WHO and the World Bank.

Murray, C.J. and Lopez, A.D. (2000) Progress and directions in refining the global burden of disease approach: a response to Williams. *Health Econ* **9**, 69-82.

Nord, E. (2005) Concerns for the worse off: fair innings versus severity. *Soc Sci Med* **60**, 257-63.

Paalman, M., Bekedam, H., Hawken, L. and Nyheim, D. (1998) A critical review of priority setting in the health sector: the methodology of the 1993 World Development Report. *Health Policy Plan* **13**, 13-31.

Ponce, R.A., Bartell, S.M., Wong, E.Y., LaFlamme, D., Carrington, C., Lee, R.C., Patrick, D.L., Faustman, E.M. and Bolger, M. (2000) Use of quality-adjusted life year weights with dose-response models for public health decisions: a case study of the risks and benefits of fish consumption. *Risk Analysis* **20**, 529-542.

Ponce, R.A., Wong, E.Y. and Faustman, E.M. (2001) Quality adjusted life years (QALYs) and dose-response models in environmental health policy analysis -- methodological considerations. *Sci Total Environ* **274**, 79-91.

Sassi, F. (2006) Calculating QALYs, comparing QALY and DALY calculations. *Health Policy Plan* **21**, 402-8.

Schwarzinger, M., Stouthard, M.E., Burstrom, K. and Nord, E. (2003) Cross-national agreement on disability weights: the European Disability Weights Project. *Popul Health Metr* **1**, 9

U.S. EPA. (1997). Guiding Principles for Monte Carlo Analysis. EPA/630/R-97/001. US EPA Risk Assessment Forum. Available at http://www.epa.gov/raf/publications/pdfs/montecar.pdf

Van der Voet H., Gerie WAM van der Heijden, Peter MJ Bos, Sieto Bosgra, Polly E. Boon, Stefan D. Murie and Beat J. Brüschweiler (2009). A model for probabilistic health impact assessment of exposure to food chemicals. Food and Chemical Toxicology (47) 2926-2940.

Van der Voet, H. and Slob, W. (2007). Integration of probabilistic exposure assessment and probabilistic hazard characterization. Risk Anal. 27: 351-371.

van Kreijl, C.F., Knaap, A.G.A.C., and van Raaij, J.M.A. (2006). Our food, our health. Health diet and safe food in the Netherlands. RIVM, National Institute for Public Health and the Environment, The Netherlands.

Wong, E.Y., Ponce, R.A., Farrow, S., Bartell, S.M., Lee, R.C. and Faustman, E.M. (2003) Comparative risk and policy analysis in environmental health. *Risk Anal* **23**, 1337-49.