MULTIPLE FRAMEWORK CONTRACT
WITH RE-OPENING OF COMPETITION
FOR SCIENTIFIC SERVICES FOR ECHA

REFERENCE: ECHA/2011/01

Service Request SR 19:

Quantification and valuation of the human health impacts of chemicals based on quality and disability-adjusted life-years

Final Report

19 August 2015

National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport

RPA
Risk & Policy Analysts

Quantification and valuation of the human health impacts of chemicals based on quality and disability-adjusted life-years

August 2015

Final Report

Quality Assurance	
Project reference / title	J857/SR19
Report status	Final Report
Author(s)	Abby Mahal Anna Heinevetter Phil Holmes Meg Postle Rob Jongeneel Wouter ter Burg Gerlienke Schuur
Approved for issue by	Meg Postle, Director
Date of issue	19/08/2015

Document Change Record			
Report	**Version**	**Date**	**Change details**
Draft Final Report	1.0	29/05	Inclusion of latest WHO and Dutch study findings, as well as SG comments
Final Report	2.0	19/08	

Table of Contents

Executive Summary

Socio-economic analysis (SEA) is integral to the REACH Authorisation process and can also play a useful part in the preparation of Annex XV restriction dossiers. ECHA provides guidance on the preparation of SEAs for use in these processes, suggesting different approaches for quantifying the human health impacts associated with exposure to substances of very high concern (SVHC). This quantification can be expanded to develop monetised estimates of the health impacts of restrictions and authorisation, facilitating a comparison with other economic impacts. To support the development of a methodology for quantifying and monetising human health impacts as part of SEAs, ECHA commissioned a study, conducted by the Charles University in Prague and the VU in Amsterdam, to develop willingness-to-pay (WTP) estimates for a set of health endpoints relevant to SVCHs.

This study was commissioned by ECHA to build on the above study. We explore the potential for quantifying and valuing human health impacts related to SVHCs using quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs). The aim of the study was to identify the key health endpoints associated with SVHCs identified under REACH to date, to screen the QALY/DALY literature for utility and disability weights relevant to these health endpoints, and to assess the transferability of those weights to the REACH SEA context; furthermore, it was hoped that the study would recommend a set of weights which could be used in the future to calculate QALYs or DALYs for inclusions in cost-benefit analyses (or cost-utility analyses) to support future Authorisation or Restriction SEAs.

The report starts by defining what QALYs and DALYs are, including how they are calculated, the methods through which the utility/disability weights are elicited, and some of the issues which arise in their derivation and use (further discussion of these issues is presented in Annex 1). A review of 85 substances identified as SVHC under REACH (that have been subject to Restrictions or Authorisation, or are on the Candidate List or the Registry of Intentions) was undertaken to identify the types of health effects that may need to be addressed in REACH SEAs. This led to a list of 41 health effects relevant to exposures to SVHCs, and which then acted as the focus for the collation of QALY utility weights and DALY disability weights. For 36 of these 41 health effects, it was possible to identify weights within the existing literature.

An overview is provided of the key sources of weights, with this including some discussion of the differences in studies and caveats associated with their outputs. The study has focused on the weights rather than on QALY or DALY values, as information on the weights together with the associated time profile for the disease are the key information needed to calculate the QALYs that would be gained or the DALYs that would be avoided through a policy intervention. The key sources of QALY utility weights include the Tufts University Medical Centre CEA Registry, UK NICE's guidelines (related to medical interventions) and a few key academic studies. The key sources of DALY disability weights are the Global Burden of Disease studies undertaken in 2013 (published in 2015) and 2010, as well as the European Burden of Disease study (also published earlier this year), and a series of national burden of disease studies.

Although not part of the study remit, we also briefly consider the relationship between QALYs/DALYs and WTP estimates as found by other researchers, a relationship also examined by the Charles University. The conclusion of this work is that there is no constant relationship between WTP and unit QALY gains. This may be due to either there being scope insensitivity in the WTP results, or to a decreasing marginal utility with increasing gains in quality of life. This suggests that analysts should take care in choice of any value that used to convert a QALY or DALY to a monetary value; it also suggests that sensitivity analysis of selected values may be important.

Regarding the transferability of the collated weights, the study concludes that there are many weights in the literature which correlate well with the health effects linked to SVHCs, which suggests the potential for the use of QALYs and DALYs in REACH SEAs. The available studies vary, however, in terms of the methodologies used and the health states (or endpoints) covered, as well as the differentiation for any given health state for variations in the severity of the conditions. As a result, it is not possible to recommend a single source of either utility weights or disability weights. Instead, analysts will need to consider the manner in which the health states of concern have been defined in the original studies to determine that which most closely reflect the types of effects that would stem from chemical exposures. We have highlighted the importance of ensuring that the weights chosen for SEAs reflect the type of health effects identified by the toxicological data, bearing in mind the uncertainties that exist in moving from the toxicological data to health impacts.

Several assumptions must be made when calculating QALYs/DALYs, including life expectancy, duration of disease, comorbidity, discount rate, etc. An example is provided to illustrate how the weights collated by this study could be used to calculate QALYs/DALYs and how these can then be converted to a monetary value for use in a SEA. It has been suggested that DALYs are more appropriate for use in REACH SEAs as they are aimed at quantifying the burden of disease, and the aim of REACH is to reduce the burden of disease in society relating to chemical exposures. However, there is a clear preference for the use of QALYs in some Member States and there is also a literature on WTP for a QALY that may be of relevance to some REACH SEAs. We note though that it is possible to calculate a QALY or a DALY from the original utility or disability weights (subject to some conditions).

Going forward, the study recommends that ECHA should consider either:

- Commissioning further research to develop a consistent set of weights for use in REACH SEAs, based on a more in-depth screening of the original literature than was possible here. This may entail a meta-analysis across the available sources in order to derive the mean weights for different health effects and their associated confidence intervals.

- Or preparing guidance on the use of QALYs and/or DALYs in SEAs, highlighting the key methodological and analytical issues which SEA practitioners will need to be aware of when applying these methods. A key feature of this guidance would be the need for sensitivity analysis and the identification of a set of appropriate WTP values which can act as anchors for quantifying different health effects.

1 Introduction

1.1 Background

Socio-economic analysis (SEA) plays an important role in the REACH Authorisation process and can also be important in the preparation of Annex XV restriction dossiers. ECHA's guidance on the preparation of SEAs to support these processes[1] provides an overview of the different approaches that can be taken to estimating the human health impacts associated with changes in exposures to substances of very high concern. In particular, the focus is on quantified and monetised estimates of the health impacts of restrictions and authorisation as this facilitates a comparison with other economic impacts. Unfortunately, the existing health economics literature specific to the valuation of hazardous chemicals is sparse.

In order to address this lack of supporting literature, ECHA commissioned a major study in 2012 to develop a range of chemicals-relevant health states using stated preference (SP) methods (Service contract for the European Chemicals Agency No. ECHA/2011/123). The study was undertaken by a consortium led by Charles University, Prague, and its results were published in 2014. The work undertaken by the Charles University team focussed on a limited range of illnesses/health states/conditions, covering skin sensitisation, liver-related dose toxicity, cancer-related morbidity and mortality, and reduced fertility and fertility-related outcomes.

The Charles University study also included a brief review of the use of the metrics referred to as the "Quality Adjusted Life Year" and the "Disability Adjusted Life Year" – QALYs and DALYs – for the same set of health states. These two metrics have not been widely used in relation to chemical risk management but are increasingly being used in other fields. QALYs are used in health care decision making to assess the value of different types of medical interventions, while DALYs are used by the World Health Organisation and others to measure the burden of ill health at a national level.

The review undertaken by the Charles University team had a number of objectives. First, it might provide information (e.g. on health state descriptions) which could feed into the SP questionnaires; second, it could help provide an external validity check on the values estimated via SP methods; third, it might provide directly-relevant utility weights which could be used to calculate monetary values for health impacts via the application of a 'cost per QALY'; and, fourth, comparing the QALY/DALY-based values and the SP-based values might give an insight into the transferability of both types of value.

In order to build on the findings of the review carried out by the Charles University, ECHA commissioned this study (Service Request 19 – SR19). The aim of this service request is to

[1] See for example ECHA's Guidance on the preparation of socio-economic analysis as part of an application for authorisation, Version 1, January 2011.

build on the initial review to provide a more comprehensive assessment of the usefulness of the QALY/DALY literature for providing quantitative health impacts for use in REACH SEAs.

1.2 Aims of the study

Thus, the focus of SR19 is the quantification and valuation of the human health impacts of hazardous chemicals, based on the use of the QALY and DALY concepts. Furthermore, the focus is on the human health impacts relevant to substances of very high concern which have been subject to REACH Annex XVII restrictions/restriction dossiers, are listed in REACH Annex XIV or are on the 'Candidate List', (Candidate List of Substances of Very High Concern for Authorisation), or are anticipated to be subject to restriction or prioritisation.

In commissioning the work, ECHA also noted that some of the chemicals within the above sets are likely to be associated with health impacts which have already been subject to measurement using QALYs and DALYs, or which give rise to health impacts which are similar to those for which QALYs and DALYs are available.

Given the above, the overall aims of this study are:

1. To review existing and planned regulatory measures under Annex XVII, Annex XIV and the Candidate List to establish a set of human health effects of potential interest for socioeconomic analysis

2. To review the existing literature to collect QALY/DALY weights which have been estimated for this set of human health effects, and

3. To consider the extent to which these QALY/DALY weights, and the underlying health state descriptions, are applicable or 'transferable' to the chemicals/REACH context and the sorts of health effects likely to be experienced from chemical exposures.

Where appropriate and possible, the work is also to include making simple adjustments to the available weights to make them more relevant to REACH SEAs as part of consideration of transferability. In addition, the outputs are to include identification of gaps in the existing literature in terms of the availability of QALY/DALYs for the types of human health effects of interest.

1.3 Study approach

The approach to the study comprised desk-based research by the core study team comprising RPA and RIVM, with the addition of some external peer review by experts outside the study team but within RIVM. Within the resources available for the study, it was not possible to undertake a full, systematic review of the QALY/DALY literature. Instead, the aim was to identify key issues which may affect the degree to which weights exist for REACH relevant human health effects, and the transferability of these.

Our approach to the study was divided into three separate work packages:

- Work Package 1 (WP1): Selection of human health states – the aim of this work package was to identify a set of health states which are relevant for quantification/valuation for use in socioeconomic analysis in REACH restriction and authorisation processes

- Work Package 2 (WP2): Literature review – the aim was to review the existing literature on QALYs and DALYs and to collate a set of utility weights linked to the health conditions identified in work package WP1

- Work Package 3 (WP3): Evaluation of transferability – the aim of WP3 was to evaluate the extent to which the QALY/DALY weights identified in Work Package 2 are applicable to the health states identified in Work Package 1, and 'transferable' for possible use in socioeconomic analysis under REACH. 'Transferability' here is defined primarily in terms of the comparability of the selected health states with those for which utility weights were obtained.

We would also note that after submission of the Draft Final Report, the study team identified that new and important studies providing up-dated DALY weights had been published by the World Health Organisation and other key academics researchers. As a result, the study was extended to enable inclusion of these latest DALY weights into the study findings.

1.4 Organisation of the report

The remainder of this report has been organised as follows:

- Section 2 provides an introduction to QALYs and DALYs and how these are calculated, as well as providing an overview of the methods that are used in their derivation

- Section 3 summarises the approach taken to identifying the key human health effects (endpoints) that may be relevant to REACH SEAs and which act as the focus for considering transferability (corresponding to the outputs of WP1)

- Section 4 presents the weights that were collated from the review of the literature and highlights some of the key issues arising from a comparative analysis of these

- Section 5 examines the transferability of existing weights to the REACH context for preparation of SEAs. This includes consideration of theoretical and methodological issues, as well as the degree to which the health effects for which the weights were derived match the types of effects of concern under REACH. It also briefly discusses the potential for moving from the weights to willingness to pay equivalents.

- Section 6 provides a summary of our key conclusions.

2 What are QALYs and DALYs?

2.1 Introduction to QALYs and DALYs

2.1.1 Overview

In public health and medicine, health and mortality effects are often measured using some form of health adjusted life year, or a HALY, to focus on the impact a certain disease has on the individual. The most common of these measures is the 'Quality Adjusted Life Year' (QALY) and the more recently introduced alternatives of the 'Disability Adjusted Life Year' (DALY) and 'Healthy Years Equivalent' (HYE).

Each of these concepts can be used to measure the utility of a specified "health profile" (i.e. a time path of health states ending in death) in terms of an equally valuable length of time lived in full health. This study is concerned with the potential use of QALYs and DALYs, as these are the two main measures that are currently being used to measure the value of health interventions. Although both of these concepts focus on the impact a certain disease has on the individual, they essentially have different objectives and their original purposes are somewhat at variance.

2.1.2 What are QALYs and DALYs?

The QALY was developed in the 1970s and has become an internationally recognised standard tool since the mid-1990s to assess the relative efficiency of investments in health care interventions. It was developed to provide a measure that integrates quantity of life with quality of life, i.e. a quality adjusted life year. It is a measure of an individual's preferences for his/her own health and longevity that can be added across people to measure the social value of health improvements. QALYs were originally developed to provide a basis for undertaking cost-utility analyses (a form of cost-effectiveness analysis) of health interventions, where the aim is to maximise the gains in QALYs per unit of health care expenditure.

Mathematically, one QALY is the arithmetic product of life expectancy combined with a measure of the quality of life in those years. They are relatively simple to calculate: the time a person is likely to spend in a particular state of health is weighted by a utility score from standard valuations. In such valuation systems, '1' equal perfect (full) or normal health and '0' equals death. Since certain health states that are characterised by severe disability and pain are regarded as worse than death, they are assigned negative values.

The DALY is an alternative tool which emerged in the early 1990s as a means of quantifying the burden of disease at a national (and indeed global) level. It was therefore created to serve a different purpose from QALYs. DALYs were developed to reflect the sum of years of life lost (YLL) due to premature mortality and years lived in disability/disease (YLD). YLLs are calculated as the number of deaths at each age multiplied by the standard life expectancy

for each age. YLDs represent the number of disease/disability cases in a period multiplied by the average duration of disease/disability and weighted by a disease/disability factor.

For DALYs, the scale used to measure health states is inverted to a 'severity scale', whereby '0' equals perfect health (i.e. zero disability) and '1' equals death. The weight factors may be age-adjusted to reflect social preference towards life years of a young adult (over life years of an older adult or young child). Furthermore, DALYs may be discounted over time, thus favouring immediate over future health benefits (EUFIC, nd).

Gold et al (2002) highlight the main differences between the two concepts as the aspect of health which is valued differently; while the QALY focuses on health profiles, the DALY has the disability profile as its focus point. QALYs could therefore be regarded as measuring the 'positive', the quality of life in certain health states, with the aim to maximise the quality of life; while DALYs are measuring the 'negative' impact of a disease or the years lived in poor health (Sassi, 2006). Another important difference is the populations from whom such values are taken (Gold et al, 2002). More generally, though, DALYs were created to provide a means of measuring the size of the health problem, i.e. the burden of disease or ill health within a country as a means of prioritising the health problems that require national action. In contrast, QALYs are used to measure the level of health improvement that can be gained from a specific intervention and to enable comparison of the cost-effectiveness of that intervention compared to others for budgetary allocation purposes within health care systems.

2.1.3 Overview of QALY and DALY calculations

QALYs are calculated by deriving a utility weight for a particular health state and then weighting the time spent in that state by that weight, and calculated using the following formula:

$$QALY = Utility\ Weight\ of\ Health\ State\ i \times Years\ Lived\ in\ Health\ State\ i$$

As noted above, a utility weight of 1 represents perfect health and implies that one year in that health state is equivalent to one year in perfect health, i.e. to 1 QALY; similarly, a utility score of 0.5 implies that two years in the associated health state are equivalent to one year in perfect health, or to 1 QALY. If an intervention provided perfect health for one additional year, it would therefore produce 1 QALY. Likewise, an intervention providing an extra twenty years of life at a health status of 0.5 would produce 10 QALYs. This result can then be related to the net cost to produce that QALY, i.e. the cost per QALY. For example, if a new treatment gives an additional 0.5 QALYs and the net cost of the new treatment per patient is €5,000, then the cost per QALY is €10,000, i.e. €5,000/0.5 (EUFIC, nd).

DALYs take into account the number of years of life lost due to either premature mortality or to living in a less than perfect health state, and are calculated as follows:

$$DALY = YLD + YLL$$

YLD, which stands for Years Lived with Disability, is calculated as follows:

YLD = Number of Cases x Average Disease Duration x Disability Weight

YLL, which stand for Years of Life Lost due to premature death, is calculated as:

YLL = Number of Deaths x Life Expectancy at Age of Death in Years[2]

For example, let us consider a woman with a standard life expectancy of 82.5 years who turns blind at age 45 and lives for 5 years in this state, dying at age 50. This woman spends 45 years without disability, followed by 5 years in a disability state with weight factor of 0.195[3] followed by premature death at age 50. To calculate the number of DALYs she suffers we simply add the number of years of life lost due to premature death (32.5 YLLs) and the number of years of life lived with disability (5 x 0.195 = 0.322 YLDs). This amounts to 32.822 DALYs lost[4].

The figure below, taken from Robberstad (2005), illustrates the difference between QALYs and DALYs (when DALYs are not age-weighted).

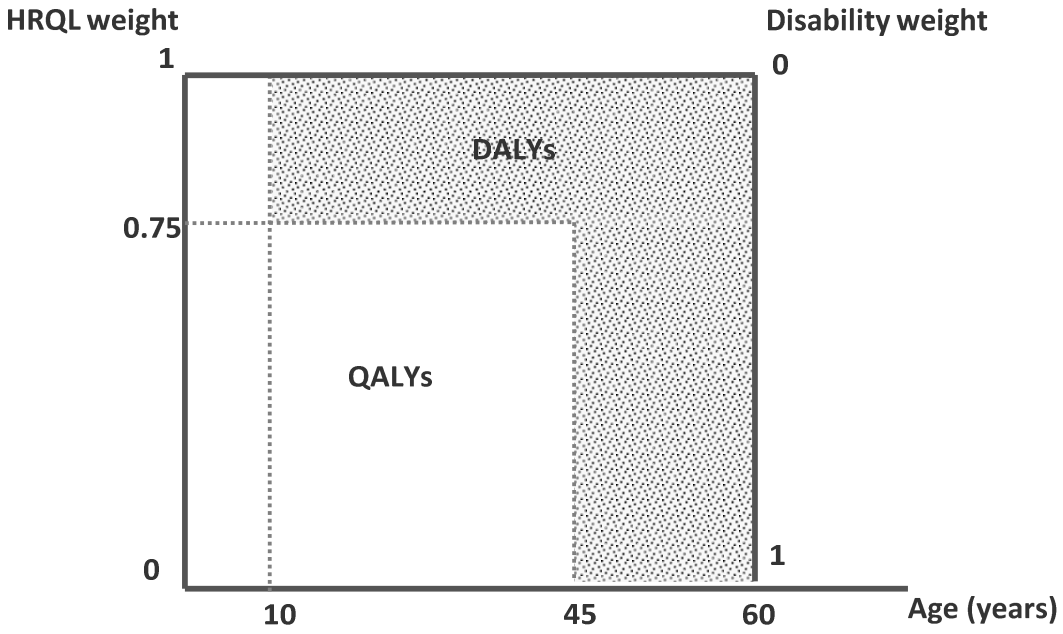


Figure 2-1: Differences in what is measured by QALYs and DALYs

It is based on the example of a person who gets a disability at the age of 10, lives with the disease for another 35 years and then dies prematurely at the age of 45. The Health Related

[2] This refers to life expectancy minus age of death i.e. 82.5 – 50 = 32.5

[3] The weight factor 0.195 is the disability weight associated with being blind, as cited by the WHO.

[4] In this example, the number of cases and the number of deaths is simply 1 as we are considering an individual. To calculate the burden of the disease, as is the case in the WHO Global Burden of Disease Study, one would take the average duration of the disease and multiply this by the number of incident cases.

Quality of Life (HRQL) associated with this disability is 0.75. Life expectancy is given as 60. In this case, the person's lifetime QALYs are given as (1 x 10) + (0.75 x 35) = 36.25 QALYs. The person's DALYs can be calculated as the number of life years lost, given expected life years: 60 – 36.25 = 23.75 (or conversely = (15 x 1) + (35 x 0.25). In this example, DALYs are not weighted for age or time and so they are the exact inverse of the QALYs. However, this may not necessarily be the case for all approaches to the calculation of a DALY, as discussed below.

Deriving the utility or disability weights for use in either of these measures is difficult: collection of accurate data from a representative sample of the population is not straightforward. Furthermore, there are problems with defining the health state which is being measured.

In addition, the calculation of a DALY may involve the weighting of social preferences in terms of age and time[5]. Both of these are additional weighting factors which were historically incorporated into DALY calculations, although the more recent literature suggests that age weighting is becoming more "discretionary". It is important to recognise that the implication of these weightings is that not all life years lost are equal; a higher weight is often applied to those between the ages of 9 and 54, the period of life thought to be more valuable to society (Murray, 1994).

These issues are discussed further below.

2.2 QALYs

As noted above, the QALY was originally developed as a measure of health effectiveness for use in cost-effectiveness analysis, with the intention of assisting decision-makers to allocate scarce resources across competing health-care programmes (Weinstein et al, 2009).

The conventional QALY is based on decision science and expected utility theory. Its basic construct is that individuals move through health states over time and that each health state has a value attached to it, while the aim is to maximise health. Here, health is defined in terms of the value-weighted time (i.e. life-years weighted by their quality) which is accumulated over the relevant time horizon. Health states are valued on a scale where dead = 0 because the absence of life is worth 0 QALYs. The upper end of the scale, or perfect health, has a value of 1. Health states worse than being dead can exist within the schema – they would have a negative value and subtract from the number of QALYs. Importantly, a QALY values health states and not *changes* in health states. The amount of time spent in a health state is weighted by the utility score given to that health state to yield a QALY.

[5] There is a literature which argues about whether or not social preferences should be taken into account when calculating QALYs. For example, Dolan et al. (2004) make the point that social value is not linear with respect to quality of life (Q) or length (T) and that it actually diminishes in marginal increments of both. Some argue the point that QALYs are strongly tied to economic theory which advocates discounting in most contexts.

Figure 2-2 builds on Figure 2-1 and illustrates how to calculate gains in QALYs from a medical intervention.

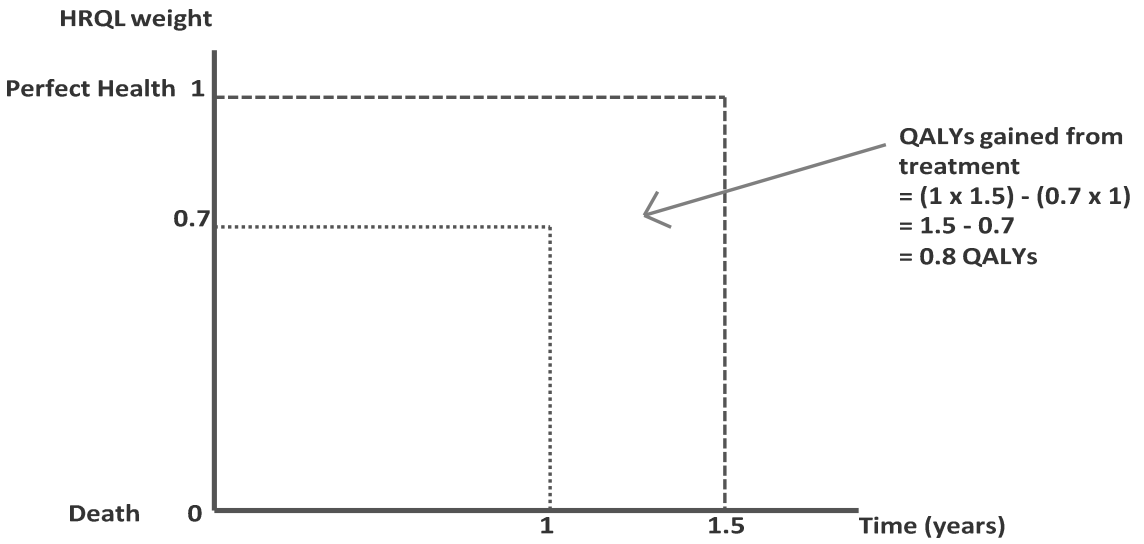


Figure 2-2: Calculation of gains in QALYs from an intervention

Because a health state that is more desirable is more valuable, for a QALY, value is equated to preferences or desirability. The preferences in conventional QALYs are usually elicited by those who do not have the disease rather than those who have it (or have had it).

QALYs, when combined with the costs of an intervention, can be used to calculate cost-utility ratios. A cost-utility ratio is simply the cost of the intervention divided by the number of QALYs gained as a result. When comparing potential interventions, one can compare their respective cost-utility ratios. This may be relevant to the assessment of Restriction proposals under REACH.

For example, if intervention 1 costs £3,000 and leads to a gain of 5 QALYs, the resulting cost-utility ratio would be (3000/5) =£600/QALY; if intervention 2 costs £4,000 and leads to a gain of 7 QALYs, the resulting cost-utility ratio would be (4000/7) = £571/QALY gained. The result is that intervention 2 has a lower cost-utility ratio than intervention 1. This exercise can be performed for interventions which are not similar or generally comparable by nature (Prieto and Sacristan, 2003). In this respect, a QALY provides a common currency for measuring the extent of health gain that results from an intervention and, when combined with the costs associated with the interventions, can be used to assess their relative worth from an economic perspective (Phillips and Thompson, 2001).

Within the context of Authorisation, however, QALYs are more important as a means of measuring health status and the potential gain in health from a change in status. This quantified change in health gain can be combined with monetary values (e.g. the value of a life year or a willingness to pay value) to provide the basis for a fuller cost-benefit analysis.

It is important to note that QALYs as quoted in the academic/health literature will reflect the gain from a treatment rather than the gain which would be obtained from the

prevention of an illness. However, the utility weight used in the calculation of the QALY gains stemming from an intervention will reflect the difference between full health and the current health state; it can therefore be used to calculate the gain in health that would be obtained if the illness was prevented in the first place.

2.3 DALYs

As indicated above, a disability adjusted life year (DALY) is an indicator of the burden of disease in a population. It takes into account not only premature mortality but also disability caused by disease or injury. It is a time-based measure that combines years of life lost due to premature mortality and years of life lost due to time lived in health states reflecting less than ideal health. One DALY can be thought of as one lost year of 'healthy life', and the burden of disease can be thought of as a measurement of the gap between current health status and an ideal situation where everyone lives into old age, free of disease and disability.

The concept gained prominence with the first Global Burden of Disease study in 1996, followed by the Dutch disability weight study from 1997. The disability weights that have been developed in both of these studies functioned as the basis of several of the later studies, such as the Victoria Burden of Disease study and several follow-up studies on the Global Burden of Disease (GBD) for the World Health Organisation (WHO).

The approach to estimating DALYs can vary across different applications depending on the choices of analysts. This includes choices across assumptions concerning lifespans, gender differences in lifespan (e.g. 80 years for men and 82.5 for women), age weighting (e.g. a higher 'value' of individual life at younger age based on higher economic productivity), as well as discounting of future years (Donev et al, 2010; Hammitt, 2002a; Anand & Hanson, 1997).

There is a large body of literature relating to the DALY concept itself, as well as its application in large-scale examinations of disease burden; the GBD studies, in particular, have been criticised for various reasons, as described below and in Section 4.

2.4 Methods used to derive utility and disability weights

2.4.1 Summary of the methods

Several methods have been developed to derive utility and disability weights and these can be categorised as either direct or indirect methods. The former includes the most widely-used techniques, the standard gamble and the time-trade off methods, which are grounded in economic utility theory. Other direct methods include the visual analogue scale, rating methods, person trade-off methods and pairwise comparisons, all of which are less grounded in utility theory but are easier to implement. The indirect methods rely on the use of multi-attribute utility instruments and health state valuation methods to obtain utility weights; these include for example the EuroQol-5D (EQ-5D) questionnaire or the McMaster's Health Utility Index (e.g. the HU12 or HU13). Alternatively, weights can be

calculated using disease-specific classifications or Quality of Life (QoL) instruments.

These methods are described below (with a more detailed discussion provided in Annex 1):

- Standard Gamble: the standard gamble requires the individual to make a choice between two alternatives: a health state which is certain [e.g. a certain frequency (once a month, once a week) of migraines of a certain severity] or a gamble between a better health state (e.g. no migraines; full health) and a worse health state (e.g. death).

- Time Trade-Off (TTO): the time trade-off (TTO) method for eliciting utility weights requires the individual to make a decision between living in an imperfect health state for t years or living for x years in perfect health, where x<t. The individual is required to indicate the value of x which would make them indifferent between t and x.

- Visual Analogue Scale (VAS): the VAS consists of a single line anchored by two health states, for example, "best imaginable health" and "worst imaginable health". The individual is then asked to indicate on this scale their preference for a particular health state.

- Person Trade-Off (PTO): the PTO method tends to be used to derive disability weights for use in DALYs and consists of asking people how many outcomes of one kind they consider equivalent in social value to X outcomes of another kind, with reference to the EuroQol or other such descriptive system. The disability adjustments are developed by assigning disability weightings to life years – diagnostic groups are chosen and defined, descriptions of these diagnostic groups are developed, and health states are assigned a disability weight to indicate the relative severity of each. Two approaches are frequently used: the PTO1 compares life extensions for disabled and healthy people; while the PTO2 compares cures for illness with extension of life.

- Pairwise comparisons: these are a simplified version of the PTO method in that pairwise comparison approaches consist of asking people which outcome is preferred in terms of X outcomes of one kind compared to Y outcomes of another kind. The disability weights are then derived from these choices with reference to descriptions of the health states that would correspond to the different outcomes[6].

- Multi-Attribute Utility Instrument (MAUI): a MAUI is an indirect method of measuring utility which involves two stages of assessment: description of a health state using a generic health-related quality-of-life questionnaire such as the EQ-5D, the SF-6D or the HU12 or HU13; followed by valuation. For example, the method

[6] It is not possible to tell from the literature examined for this study whether or not these exercises are undertaken in a manner that would be consistent with Saaty's priority theory, but from the descriptions provided it is assumed that they are not.

requires the individual to describe his/her current health state by completing the EQ-5D; subsequently, the health state as given by that person is translated into a value that has been generated by population-based valuation research. Here, large samples of the population have valued pre-defined health states using direct methods such as the SG or the TTO. Econometric modelling techniques are then used to infer valuations for all possible health states of that MAUI instrument.

- Discrete Choice Experiments: Discrete choice experiments involve presenting an individual with choices of various health scenarios described in terms of characteristics and associated levels. They must choose the scenario which they prefer. These choices are aligned with a rating system, the results of which are modelled using a regression function which generates information on the relative importance of characteristics, the rate at which an individual is willing to trade-off a characteristic for another, etc.

Empirical studies have shown that these different methods generate very different utility weights, with this being a key issue with the use of particular QALYs or DALYs, and their use more generally. This issue is discussed in greater detail below and in Section 4.

2.4.2 Comparative advantages and disadvantages

As indicated above, each of the methods has advantages and disadvantages over the other methods. Below is a table taken from Blinman et al. (2012) which summarises key characteristics of each of the methods used in eliciting health utility weights for QALYs. To the extent that these methods are also used to elicit disability weights, this summary equally applies in the context of DALYs.

The question of which method is best for eliciting utilities is not easily resolved; much depends on the intended purpose of these weights, the nature of the survey (e.g. sample size), etc. Blinman et al. (2012) suggest that the SG, TTO and discrete choice experiments are the optimal methods for clinical decision making, and that MAUIs are better suited to health policy decisions relating to the allocation of resources. Furthermore, research has found that the choice of method can impact on the utilities being generated, e.g. valuations with the VAS tend to be higher compared to equivalent ones with choice-based methods (Haagsma et al., 2014)[7].

A review of submissions to the Australian Pharmaceutical Benefits Scheme (Scuffham et al, 2008) assessed the methods that were used for estimating QALY weights included in submissions to the scheme. The academics undertaking the review rated approaches involving the use of MAUI administered to patients currently experiencing the health states as more appropriate, together with health state valuation experiments (i.e. stated

[7] It is important to note that Haagsma et al. conduct their evaluation in the context of DALYs therefore implying that "higher" has a different connotation in a DALY concept than in a QALY concept. It is generally agreed that in the QALY concept, VAS scores tend to be lower than scored generated by TTO and SG.

preference based elicitation) of either the general population (based on QoL data) or of a population of patients valuing their own health state. Interestingly all other approaches were considered less appropriate, including non-preference based approaches based on rating scales, mapping transformations and consensus opinions (note that the use of the VAS, as discussed in Section 2 is not considered valid in Australia).

This distinction is made because the latter set of techniques is considered by Scuffham et al. (2008) to be inconsistent with the QALY approach, as they do not explicitly trade of quality of life against time lived in a particular health state. Even then, of the approaches using more appropriate populations and techniques, 56% were rejected by the Advisory Committee compared to 66% using less appropriate methods.

Table 2-1: Summary of methods used to elicit preferences for different health states

Method	Description	Involves uncertainty	Involves trade-off between different health states	Direct or indirect comparison of health states	Suitability for clinical or policy decisions?
Standard gamble (SG)	Choose either a gamble between perfect health and death or a certain but intermediate health state	Yes	Yes	Direct	Both
Time trade-off (TTO)	Choose either an intermediate health state for time t or perfect health for time x < t	No	Yes	Direct	Both
Visual analogue scale (VAS)	Assign preference for a health state on a line anchored by perfect health and death	No	No	Direct	Both but best used in conjunction with SG or TTO
Multi-attribute utility instruments (MAUIs)	Complete a generic HRQL instrument which is then converted into a utility score using population values	No	No	Indirect	Policy decisions, e.g. cost-utility analyses
Discrete choice experiments (DCEs)	Choose between scenarios that describe a health state by different levels of attributes of that health state	No	Yes	Direct	Clinical decisions and policy decisions

Source: Blinman et al. (2012): Preferences for cancer treatments: an overview of methods and applications in oncology

The authors concluded that the variability in the methods and approaches used to derive QALY weights is a concern, but also noted the increasing level of guidance that now exists on this aspect. Scuffham et al (2008) further note that both NICE and the US Panel on Cost

Effectiveness in Health and Medicine have indicated that the QALY weights should be derived:

- through the use of choice based techniques aimed at eliciting preferences
- using generic MAUIs to describe and value health states
- and using a representative sample of the public as the most appropriate source of preferences.

Similarly, the disability weights used in the calculation of DALYs may be developed thorough a range of valuation methods, as highlighted by Haagsma et al (2014). This includes use of the VAS, the TTO and MAUI methods as described above for QALYs. In addition, various studies have used "interpolation", the person trade-off technique (PTO) and pairwise comparisons. The latter has been used on its own or in combination with the VAS for many of the key studies, including those generating disability weights for the WHO's Global Burden of Disease initiative.

Haagsma et al. (2014) reviewed twenty two studies that developed disability weights and critically assessed their methodological design choices (health state and time description, panel composition and valuation method), and the disability weights generated for eight specific conditions. The authors note that there has been considerable variation in the methods used to generate disability weights and that most of the studies have relied on non-preference based valuation methods for assessing the values for the majority of weights. Most of the weights have been derived using ranking, interpolation, pairwise comparisons or the VAS – all of which lack an explicit trade-off feature and hence may not fully assess and therefore reflect people's preferences. As a result, most of the studies provide information only on the relative desirability of one health state compared to another and do not take into account the full intangible impacts associated with living in a health state (e.g. in relation to impacts on relatives, etc.). The implication is that they provide weights that reflect health trade-offs rather than welfare trade-offs (see also further discussion on this issue in Sections 4 on the 2010 Global Burden of Disease Study and Section 5).

Research has also been carried out as part of the European Disability Weights Project (Schwarzinger et al., 2003) to assess the use of different methods for eliciting disability weights. This study used the VAS to measure the severity of health states, the TTO to measure health outcome trade-offs and the PTO to elicit health decision-makers' trade-offs. Two panels were used for the valuation of health states – health care professionals and members of the general public with an academic background. A key conclusion of the study is that there is a reasonably high level of agreement on the ranking of disability weights across countries. In particular, they found good correlation between countries using the VAS and TTO, although this did not hold when looking across different types of survey respondent. Correlation was not good across countries in the case of the PTO method which demonstrated systematic effects related to both health states and countries.

2.5 Key issues from underlying and methodological assumptions

2.5.1 Overview

Review of the literature highlights a number of issues arising from the assumptions underlying the different methods used to elicit weights, as well as other assumptions made within this process. The most important issues are associated with assumptions regarding:

- utility independence, risk neutrality and time preferences
- discounting
- whose weights are sought
- time profile assumptions for a disease, and
- Ability to consider co-morbidities and the basis for measuring disability.

2.5.2 Utility independence, risk neutrality and time preferences

The QALY assumes that a major objective of decision makers is to maximize health or health improvement across the population subject to resource constraints. As such, it also assumes that health or health improvement can be measured or valued based on amounts of time spent in various health states (Weinstein et al, 2009). Therefore, the conventional QALY is a valuation of health benefit, based on utility theory.

There is also an assumption of risk neutrality over life-years (i.e. each individual is risk neutral with respect to longevity). In addition, there is the assumption that the value of being in a health state depends neither on the length of time spent in that health state nor on the sequence of health states preceding or following it (this takes the time dimension out of the utility assessment process). Lastly, as the literature makes no mention of the extent to which health-related quality of life (HRQL) depends on wealth, income, or consumption, it is assumed that HRQL is derived with no consideration of income or wealth and that it is independent of wealth (Hammitt, 2002).

More specifically, the key assumptions are as follows:

- **Utility Independence:** if one of the factors of utility is held constant at a particular level, preferences for lotteries over the other factors are independent of the first factor, which is held constant. For example, if T is fixed at $T=t_0$, preferences for lotteries over Q are independent of t_0, i.e. preferences over Q do not depend on the fixed level of T. If the reverse is also true (i.e. T is independent of q_0) then we have mutual utility independence between the two factors.

- **Constant Proportional Trade-Offs** (CPTO): assuming utility independence, an individual is willing to give up the same constant proportion of time to move from health state q_3 to q_4 as to move from q_1 to q_2, independent of the number of life years left. This assumption implies that the individual's preferences are independent of the amount of life years remaining.

- **Risk Neutrality*:*** the individual is indifferent between a lottery over life years and the expected duration (T) of that lottery, with quality of life, Q, held constant. The assumption of risk neutrality is required to calculate the quality-adjusted life expectancy (QALE).

Linked to the above assumptions is the issue of whether or not QALYs should be discounted. Many economists argue that the QALY model ought to be discounted in order to properly reflect individual preferences. Shepard and Thompson (1979) argue that people have positive time preferences for health, i.e. they would prefer to experience health benefits sooner rather than later. As a result, discounting should be applied. Gafni and Torrance (1984) argue that the individual's time preferences are already captured without the application of a discount rate and that then applying discounting could lead to a "double counting" of the time preference effect (where time is a feature of the elicitation method).

However, it is more generally argued that, although an individual's time preference may be reflected in the weights, social time preferences are not. As a result, QALYs should be discounted at the social time preference rate and practitioners should just highlight that there could be double-discounting but that the effect of this is unknown. The social time preference rate (STPR) is defined as the value that society attaches to present as opposed to future consumption, and is based on comparisons of utility across different points in time or across different generations[8]. For the purposes of REACH SEAs, it may be appropriate to consider both a discount rate of 0% based on the above discussion, as well as applying the standard rate of 4% that is used in EU impact assessments more generally.

Further discussion on these issues is provided in Annex 1, including the consequences of choosing different discount rates, with examples. In addition, a brief discussion is provided on non-expected utility theory, which is a growing field of investigation in relation to decision making in the context of risk.

2.5.3 Whose weights?

Blinman et al. (2012), in their review of the different methods for measuring preferences for cancer treatments, suggest that the perspective taken (i.e. whose weights) depends on the intended outcome of the preference assessment: if one is using the results of the assessment for clinical decision-making, the perspective of the intended recipient (i.e. the patient) is most relevant. They argue that taking the viewpoint of medical professionals is helpful for better understanding how differences between their perspective and the patients' preferences can influence decisions made about cancer treatment. They also suggest that taking the perspective of the general public is the most suitable for questions about more general allocation of health-care expenditure, particularly as the main objective

[8] The STPR has two components: one which reflects the rate at which individuals discount future consumption over present consumption based individuals' pure time preferences; and a second component which reflects the product of the annual growth in per capita consumption and the elasticity of the marginal utility of consumption. See for example HM Treasury (2011): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf

of the economic evaluation is to maximise the allocative efficiency of limited resources in a way that maximises social welfare.

Nord (1994) comments that the PTO seems to be rarely used to derive weights for QALYs, as it has been argued that cardinal measures of utility are required for them to be valid. He argues that this should not be the case and, indeed, argues that individuals should not be asked to value health states for themselves, but instead that a procedure that reflects community views on trade-offs is required. In other words, individual valuations are considered to provide a poor proxy for social preferences and hence of the social value of achieving different health care outcomes, which will be a function of other factors such as the initial severity of the patient's health state.

Another argument is that although patients will have an intimate understanding of the reductions in function associated with their health state, their ability to make comparative judgments with other states is limited by information asymmetry; such individuals may also have been able to adapt to their circumstances leading to an underestimation of the initial health losses associated with a chronic health state. It can also be argued though that their assessment of the health loss is likely to be a more informed assessment than that of a person who has no experience with the health state and, thus not to take their assessment into account could lead to overestimation. As noted above, therefore, where resource allocation is the focus, then some analysts have argued that the general public is the most appropriate population on the basis that, in a democratic society, the views of the general public are relevant in comparative assessments that inform public policy (Salomon et al., 2012).

Many of the studies generating utility and disability weights have actually involved use of panels of medical experts. Fox-Rushby (2002) suggests that "without 'experts' the DALY would have crumbled" (Fox-Rushby J, 2002). However, others have criticised this type of approach for several reasons. Some argue that experts do not have a better judgement than the patients affected by the disease, while others (Brock, n.d.) indicate that medical experts should not be asked to develop the weights as "health professionals may have systematic biases that skew their value judgements about quality of life from those of ordinary persons". There are also concerns that experts do not reflect societal preferences or that, even if they do, this has not been validated (Fox-Rushby, 2002). The counter argument put forward by Fox-Rushby (2002) and others is that it is not clear whether the general population is able to understand the magnitude of a certain disease or even the impact of the disease itself. The response to such concerns is one of the reasons the more recent studies, such as that carried out by Salomon et al. (2012), utilised simple lay descriptions and asked large sample populations what their preferences were for one disease over another.

Haagsma et al. (2014) note that in studies which ask both experts and the public, significant differences in the resulting weights are found. They highlight that disability weights based on societal preferences have been recommended elsewhere, due to the fact that burden of disease studies are indeed used as tools for guiding decision making on resource allocation at the population level. An external peer reviewer of this report from RIVM also noted that health state valuation methods are increasingly being used to generate disability weights, so

there is no longer such a strict division between generating QALY weights and generating DALY weights.

Another issue that is debated within the literature relates to eliciting weights from people with disabilities. For example, the Burden of Disease studies exclude people with disabilities because they will hold a different view on quality of life compared to healthy people. It has been argued that individuals with disabilities "typically overstate their quality of life relative to how non-disabled people perceive it to be" (Grosse et al., 2009). In part, it is suggested that disabled individuals overstate their own quality of life due to the ability to adapt to certain conditions, leading them to rate their quality of life as improving. This issue has been described as a dilemma in "determining the appropriate evaluative standpoint, for ranking the importance of different disabilities, to avoid the potential for bias inherent in differing perspectives" (Brock, 1998; Fox-Rushby, 2002). As a worst case, failure to recognise the perspectives of the disabled has led to life extending programmes for disabled people having a lower value than for abled people when relying on DALYs (Arnesen T & Nord E, 1999).

That the quality of life of a disabled person is a matter of perception can be seen from an assessment undertaken by a NICE Citizens Council group on QALYS and the severity of illnesses. When the participants were asked to use the EQ-5D method which has one dimension focusing on the mobility of a patient a wheelchair, the group criticised the method because it would have led to the wheelchair user accruing a negative quality of life score (NICE, 2008). In this Citizen's Council meeting, the question whether the health related quality of life calculation places too much emphasis on mobility was discussed, raising valid questions about perspectives.

2.5.4 Time profile

One aspect of the DALY calculation which is debated in the literature is the so-called time profile of a health state. A disease usually develops in a certain way; in other words, in some cases it can get worse if it is left untreated, with further complications developing as part of the progression of some diseases while, in other cases the effects of the disease, such as congenital anomalies like a cleft lip, will remain the same over time.

The key issue here is that the potential for the development or the improvement of a disease such as coronary heart disease (CHD) is not necessarily taken into account in the calculation of QALYs or DALYs, unless this is addressed by the analyst by assuming different weights over time.

For example, if a patient is diagnosed with CHD, then he/she can reduce the progression of the disease through a better diet, smoking cessation and increased exercise. The disease does not follow a strict trajectory inevitably resulting in a heart attack once the arteries are entirely blocked, but can improve over time if the patient improves thereby reducing the risk factors. If this patient is being diagnosed with a moderate form of coronary heart disease at the age of 40, then the disability weight will be relatively high; but if the patient improves and reduces the likelihood of the disease progressing, the disability weight should

not remain the same for the rest of his/her life-time but should reduce. On the other hand, some diseases, such as a common cold or an asthma attack, do not persist for the entirety of someone's life-span. Similarly, some diseases are easily treated with a one-off treatment, such as a cleft lip. These possibilities also have to be considered with regards to the duration of a disease and the calculation of its impact.

In a very recent review of disability weight studies, Haagsma et al (2014) screened 22 studies on several parameters which included an assessment of the time profile assumed for diseases. This included identifying whether the time which a patient spends in a certain disease stage is valued using a 'period profile' or an 'annual profile' approach. The period profile reflects an episode of the disease and focuses on the acute phase of that disease and tends to assume a constant health state over the period. The annual profile reflects the course of a disease over the period of one year, including both the periods of illness and good health over the year. Of the 22 studies, most of them apply the period profile while one uses a mixed approach of period profile and annual profile; two studies apply only the annual profile approach, while the approach could not be determined for the third (Haagsma et al., 2014, p.5). The approach adopted for the original GBD study (based on the PTO) by Murray (1996) was based on the period profile approach (and was therefore more appropriate to chronic conditions).

Haagsma (2010) suggests that the period profile assumption is weak for complex diseases with changing severity over time. In order to overcome this issue, the annual profile approach (APA) is recommended (Haagsma, 2010), which was applied for the first time in the Dutch disability weight study (1997). Haagsma (2010, p.10) indicates that, with "the annual profile approach, the assumption of separated states of constant health is released and the health profile is valued as a whole, alleviating the contrived assumption of constancy of health". The Dutch disability weight study therefore adopted a combined approach and generated weights that are either annual or period in nature (allowing the latter to reflect episodes of disease).

However, Haagsma, et al (2014) also note that the annual profile approach has not been widely used when eliciting weights for multiple reasons including the fact that it may overvalue certain diseases, such as mild forms of a disease or diseases with a rapid progression (Haagsma et al., 2014). Another reason given by the same authors has been described as the inflexibility of the time dimension: "If health states with different durations than the duration included in the annual profile are needed, they cannot be derived by back-calculation" (Haagsma et al., 2014).

Similarly, the 'European Centre for Disease Control and Prevention' (ECDC) notes: "For short duration diseases, disability weights can be determined per episode by focusing only on the phase of acute disease (period profile) or by focusing on a year in which an episode of acute illness is experienced (annual profile). In contrast to the conclusions of Haagsma (2010), the ECDC (2011) concludes that "using the annual profile may overvalue disability weights" (ECDC, 2011, p.20). As a result, the ECDC recommends that if the patient is suffering from a disease which affects the individual for less than a year or only for a short period of time, such as an asthma attack, the annual profile should be utilised, while for

long-term disease the period profile should be used in multiplying the disability weights by the years the condition is present (ECDC, 2011, p.20).

2.5.5 Comorbidity

Comorbidity refers to the simultaneous occurrence of two or more diseases in an individual. The QALY model is argued as allowing for consideration of comorbidity through the aggregation of QALYs applying to different health states, where these QALYs have been derived in a consistent manner. The same is not the case with DALYs, and this issue has been examined using the example of elderly populations in developed nations where someone with a heart condition most likely has an accompanying condition such as diabetes or hypertension (Haagsma, 2010). The DALY concept does not readily permit the consideration of multiple diseases in one individual or a population, as well as another disease that arises as the sequela of an existing condition. Gold et al. explain this as follows (2002):

> "if treatment of arthritis with nonsteroidal anti-inflammatory medications resulted in peptic ulcers, there is no method within the DALY lexicon to describe the accompanying alteration in HRQL that accompanies abatement of arthritic symptoms and simultaneous onset of the symptoms that accompany peptic ulcer disease. Because the QALY family of measures is grounded in domains of health, rather than descriptions of specific diseases and disability, it is at least theoretically possible to describe, and therefore value, combinations of illness."

Haagsma et al. (2011) conclude that when comorbidity is accounted for it has a significant impact on the valuation of the disease. Within the context of disease burden, the authors indicate that by ignoring comorbidity, the concept of measuring the burden of disease becomes invalid for multi-morbid populations. This issue was addressed in some of the global burden of disease studies through simulation modelling, to develop probabilities that members within the population would have more than one disease based on prevalence data.

3 Key Health Effects

3.1 Introduction

A key element of the study was the identification of the types of health effects which are likely to be the main focus of SEAs prepared to support authorisation applications and restriction proposals within the context of REACH. It is for these effects that weights would be sought from the academic literature. Our approach to this task was as follows:

- We carried out a detailed review of the hazardous properties ascribed to currently identified SVHCs or substances of concern that are relevant to human health
- We then mapped these chemical properties against health effects, taking into account information from the toxicological and epidemiological literature, and
- We then selected a sub-set of relevant health effects for prioritisation in the literature review concerning the availability of QALYs and DALYs.

Further details are provided below on the outputs of each of these steps.

3.2 SVHCs and associated health effects

3.2.1 The substances considered

Our starting point for this work was the substances listed on ECHA's website as being Substances of Very High Concern (SVHCs). This included those substances listed as:

- Substances already subject to Authorisation
- Substances which have been recommended for inclusion in the Authorisation List
- Substances which are currently the subject of a restriction proposal, and
- Substances for which MS have indicated their "intentions" to produce Restriction dossiers.

In total, the study team examined 85 substances, with these falling under the groupings as indicated in Table 3-1.

Table 3-1: Substances being analysed for work package 1	
Substances subject to Authorisation	31 substances (29 human health related)
Recommendation for inclusion in the Authorisation List and public consultation	22 substances (20 human health related)
Substances which are the subject of a restriction proposal	27 substances
Current Restriction intentions	5 substances
Total	**85 substances**
Notes: Numbers as at 3 September, 2014, when this task was completed	

Of the 31 substances subject to Authorisation, two were prioritised due to environmental concerns, leaving 29 relevant to consideration of human health effects. Similarly, of the 22 recommended for inclusion in the List, two have been recommended because of their PBT/ vPvB properties.

For each of the substances, information available on ECHA's website, e.g. from relevant Annex XV dossiers or the restriction proposal or previous risk assessments, on their hazardous properties and the potential associated health effects were collated. Tables were developed to provide the basis for summarising this information, and these are provided in Annex 2 to this report. Note that separate tables are given for the substances subject to or proposed for Authorisation and those subject to Restriction.

In developing the tables, it became clear that it was possible to group families of substances, where the hazard endpoints and the associated potential health effects of concern were essentially the same across several substances. This enabled the number of table entries to be reduced to 15 groups for substances already subject to Authorisation, and to 11 grouped entries for those recommended for Authorisation (although there is also overlap between these two sets of entries with respect to phthalates).

In the tables summarising the main health effects associated with Restrictions, we distinguished between adopted opinions (information based on the final Background Document - BD) and proposed Restrictions (information based on Restriction report prepared by the Dossier Submitter - DS). For some intended Restrictions, no Background Document was publicly available at the time of undertaking this work; where this was the case, we shaded the entries in yellow in the table provided in Annex 2.

Note that no authoritative data sources were available on the risks of concern for 6 of the substances subject to, or with intentions for, Restriction. Of the remaining 19 substances[9], it was possible to group them in terms of effects into 11 groups. However, it should be noted that one of these groups – the phthalate grouping - overlaps with the Authorisation tables.

[9] 27 minus 2 there for environmental reasons minus a further 6 = 19

3.2.2 Linking health effect endpoints to human conditions

In completing the Authorisation tables, the focus was on those hazard classifications upon which SVHC status was based, rather than on all possible endpoints that may be associated with a particular substance. For example, some of the chromates are classified as trans-generational or heritable mutagens, but they were put forward for authorisation solely on the basis of their carcinogenicity and mutagenicity; as a result, effects related to their classification for reproductive toxicity do not feature in the tables. Similarly, in completing the Restriction tables either the initial proposal made by the Dossier Submitter, or the identified risk and subsequent health endpoints considered in the HIA as set out in the final Background Document, are used as the basis for identifying the relevant health endpoints. As an observation, the health endpoints considered in the risk characterization are not always fully reflected by the human conditions stated in the HIA. Furthermore, for some substances, no Background Document was publically available at the time of undertaking the work.

Table 3-2 is a summary table containing all of the health effects identified through the above work. Note that these have been organized by main classification, e.g. carcinogenicity and mutagenicity, reproductive toxicity, and other chronic and acute classifications. Impaired or reduced male fertility is the most common set of effects (linked to 32 of the substances), and there is also a high frequency (linked to 16 of the substances) of impaired female fertility as a key health effect; both male and female fertility are generally accompanied by offspring developmental effects (although the nature of these shows more variation). In particular, impacts on foetal growth (size for gestational period or birth weight) and increases in the rate of spontaneous abortions are both linked to 19 of the substances.

Different forms of cancer are the next most common type of health effects, if these are grouped together; lung is the most frequent of the cancers (linked to 21 of the substances), followed by kidney, urinary bladder, and liver cancer.

3.3 Screening and prioritisation of effects

In total, 41 different types of effect (some representing multiple types of impact) were identified. In order to prioritise them in terms of undertaking the literature search for utility or disability weights, a set of criteria were developed to act as the basis for establishing those health effects that are the most relevant from a regulatory perspective as well as a socio-economic one. A key boundary condition set by the team was that there should be an equal or balanced coverage of health endpoint categories, i.e. carcinogenic endpoints, reproduction toxic endpoints and other identified endpoints (e.g. chronic and acute effects). Furthermore, the frequency of a given endpoint is important as it provides an indication of the likelihood that this type of health effect will also have to be assessed in future socio-economic analyses under REACH.

As indicated by Table 3-2, across all of the health effects that were identified, impaired or reduced fertility in males, in particular, was linked to 32 substances; lung cancer was the

Table 3-2: Screening of health effects by frequency			
Health effects for HIA		**No. of chemicals linked to condition**	**Rank based on frequency**
Attributable to Carcinogenicity and Mutagenicity			
Cancer	Haematopoietic	1	10
	Breast	**2**	7
	Kidney	8	2
	Liver	6	4
	Lung	21	1
	Mesothelioma	1	10
	Lymphatic	1	10
	Muscle	3	6
	Pancreatic	1	10
	Prostate	1	10
	Skin	5	5
	Stomach	2	7
	Thyroid	2	7
	Urinary bladder	8	2
Attributable to Reproductive Toxicity classification			
Impaired or reduced fertility	Female	16	3
	Male	32	1
	Increase in Spontaneous abortion	19	2
	Testicular degenerative changes	3	4
	Ovarian follicular cysts	2	5
	Disruption of ovarian cycles	1	6
	Endometriosis	1	6
Reduced foetal growth	Small for gestational age; low birth weight	19	1
Impaired cognitive development and Developmental anomalies	IQ	9	2
	Developmental neuro-impairment / neuro-logical disorders	14	1
	High cholesterol (obesity and increased rate of heart disease, etc.)	1	9
	Cardiovascular	7	5
	Skeletal	3	7
	Cleft palate	4	6
	Cryptorchidism	9	2
	Hypospadias	9	2
	Abnormalities of limbs	1	9
	Osteosporosis (Bone fracture)	1	9
	Neural tube (spina bifida)	1	9
	Renal abnormalities	3	7
	Urinogenital abnormalities	1	9
Other health effects			
	Nephritic syndrome	5	1
	Respiratory (tract) irritation	2	2
	Allergic contact dermatitis	2	2
	Skin irritation	1	3

next most common effect linked to 21 substances, followed by being small for gestational age or a low birth weight being linked to 19.

3.4 Literature search for QALY/DALY weights

It was agreed with ECHA that we would screen the literature for weights for the 41 health endpoints, listed in Table 3-2. In some cases, the description of the health state differed in the medical literature from the descriptions used in the screening of effects. For example, 'spontaneous abortion' is often referred to as 'miscarriage' in the medical literature. Additionally, in some cases, the health effect identified in the screening was too general a term and so was divided into further categories. For example, the 'cardiovascular' sub-category was broken down into three separate diseases: stroke, coronary heart disease and acute myocardial infarction. Some endpoints identified in the screening of effects had to be dropped later as weights were not available; an example of this is hypospadias. After re-defining some of the endpoints and screening for weights for the 41 endpoints listed in Table 3-2, we were able to find weights for 36 of these endpoints, listed in Table 3-3 below. The endpoints for which weights were not available are highlighted in bold type. The endpoints which were redefined are indicated in brackets, next to the alternative health effect description.

Table 3-3: Disease/Health endpoints identified in QALY/DALY literature	
Cancer	
Lung	Stomach
Kidney	Leukaemia
Urinary bladder	Lymphoma
Liver	Pancreatic
Skin	Prostate
Breast	
Cardiovascular	
Coronary heart disease Acute Myocardial infarction	Stroke
Impaired fertility and impaired foetal growth	
Infertility (male infertility; female infertility)	
Disruption of ovarian cycle (low)	***Spontaneous abortion***
Low birth weight	Spina bifida
Idiopathic intellectual disability (IQ, developmental neuro-impairment/neuro-logical disorders)	Renal abnormalities
	Urogenital abnormalities
	Cryptorchidism
Skeletal abnormalities	***Hypospadias***
Cleft palate	
Other	
Nephritic and nephrotic syndrome	Liver cirrhosis (inflammation)
Chronic obstructive pulmonary disease	Stomach ulcer
Asthma (respiratory (tract) irritation)	Pancreatitis
Allergic contact dermatitis	Parkinson-like condition
Skin irritation	Alzheimer disease and other dementia
Osteoporosis	Hearing impairment
Anaemia (no cancer)	Obesity (high cholesterol)
Glaucoma	

4 Collation of the Available Weights

4.1 General approach to the collation of weights

Our approach to the identification and collection of utility and disability weights varied to some degree, depending on whether we were trying to collect QALY or DALY information. This is due to the differences in how these weights have generally been developed (in response to the differences in how they are used) and thus how they are reported on.

In general the work included:

1. Specification of search terms (including Boolean terms)
2. Screening of key search engines (Scopus, Science Direct, Web of Science, etc.) and databases
3. Search to identify other on-line authoritative reviews and databases
4. Assessment of identified studies / papers for relevance
5. Prioritisation of key papers/databases for review.

For QALYs and DALYs, as well as collecting information on the weights, we also tried to collate the information needed to make an assessment of the robustness or quality of the weights based on information from either the reporting study or, time allowing, the original study. This has included collation of the following information (based on the approaches of Scuffham et al., 2008; Tengs and Wallace, 2000; Haagsma et al., 2014):

- The reference sources for the utility weight (where a study adopts a weight originally reported in another study)
- Whether the source assesses only one disease condition or multiple conditions
- The methods used to generate the weight, with a focus on the valuation method but also taking into account the use of a common health state description
- The number of individuals / experts in the population used to generate the weights
- The region or countries that the weights are considered representative of
- Whether the sample population, if it is based on individuals, reflected the general population or those with the health effect, and
- Date of the original study.

Note that, where possible, we have tried to use the original references for the sources of the weights. This has proved difficult in some cases and has been a key reason for rejecting some studies (and hence weights).

We have also given priority to sources that cover more than one relevant disease; the higher the number of relevant diseases covered by a study, the more useful it was considered to be for the purposes of this work. This is due to the findings highlighted in Section 2 and further discussed below that differences in methodology can lead to significant variations in

utility/disability weights. As a result, it was judged that studies which generated weights for multiple diseases would provide more internally consistent sets of weights and therefore help ensure greater consistency than those that provided weights for only one or a few of the diseases of interest to this study. No real priority in data collection terms could be given to whether weights were derived from the general population, patients, medical experts or health professionals, as this aspect varied considerably across valuation methods and studies.

Similarly, our initial approach was to give priority to European-based weights, followed by North American weights, and, if no weights were identified from these jurisdictions, weights from other regions would be quoted. This is due to the finding reported by key studies (such as Haagsma et al., 2014 and Tengs and Wallace, 2000) that national and cultural contexts can be important to the value of the weights assigned to individual diseases. This view is challenged, however, by others (such as Salomon et al., 2012) who note that the construct of health loss associated with different health states may be more universal than is the construct of welfare loss. As a result, global weights are also considered.

Finally, as a starting point, we looked for the most recent references and then considered older sources of data. It became clear though when undertaking the research that many of the more recent studies, particularly in relation to QALYs, actually draw on older studies.

More generally, internet searches were undertaking of both the published academic and medical literature, as well as the grey literature.

4.2 Collated QALY weights

4.2.1 Sources of weights

Although some relevant studies were identified from the general literature searches, it also became clear that for the sake of efficiency a few key sources of information should provide the focus of the review (particularly in light of the need to prioritise research efforts). As a result, for QALYs, we focused on the following key sources:

1) The CEA Registry managed by Tufts University School of Medicine
2) The various UK National Institute for Health and Care Excellence (NICE) guidance documents (identified from internet searches) that provide QALY weights for a range of different health effects
3) Weights presented in Tengs and Wallace (2000) which provides a summary of key studies covering over 1000 original weights (stemming from work related to the CEA Registry and pre-dating its transfer to Tufts University's School of Medicine), and
4) And a national catalogue of utility weights for chronic conditions for the US and the UK, developed to meet NICE guideline requirements, and identified through the review of other literature.

These reference sources are discussed in more detail below. It is clear that in relation to QALY utility weights, few attempts have been made to develop a uniform set of weights which span across a range of health endpoints.

4.2.2 CEA Registry

It was originally believed that a screening of the CEA Registry would enable us to identify the most up-to-date reference sources for utility weights, whether linked to a specific medical intervention or not. The CEA Registry is a comprehensive database of over 4,000 cost-utility analyses on a wide variety of diseases and treatments, and which is administered by Tufts University. It reportedly holds information on over 11,000 cost-effectiveness ratios and more than 15,300 utility weights[10].

In using the Registry, the database of weights was screened (by disease/effect) to identify the source of each respective utility weight. In most cases, it became necessary to move from the reference cited in the Registry to a review of other studies, as the sources identified in the database were often not the original source of the weight. In some cases, it was not possible to identify the actual study that acted as the original basis for a particular weight.

When possible, the original academic references were examined to identify the source of the utility weights, and which parameters and methods were used to determine those weights. This means that, where possible, the context for the elicitation of the weights (how many participants, in which country and around what time (date) the study was conducted) has been identified. Unfortunately, this was not possible for most of the weights and it quickly became clear that it was not time efficient to continue with this process. Within the resources available for this study, it would not be possible to trace back and then check all of the original references for the full list of diseases.

4.2.3 NICE guidance documents

The 'National Institute for Health and Care Excellence' (NICE) has produced a series of guidance documents for assessing the cost-effectiveness of medical interventions and these were searched for relevant utility weights. These guidance documents are very detailed and have been used to pull out individual weights. Initially, it was felt that this would enable us to fill data (and quality) gaps remaining after examination of the CEA Registry. However, it was not possible to identify weights for all of the diseases through this approach and, once again, the information needed to assess the quality of the original studies was lacking in many cases.

Details of those weights that were collected using the various NICE guidance and reference sources are presented in Table A3-1 in Annex 3. It is of note that these utility weights were developed using a range of methods and reflect a range of different regions. Again, as with the CEA Registry, it was not possible to establish information on the methods and the

[10] https://research.tufts-nemc.org/cear4/AboutUs/WhatistheCEARegistry.aspx

population used to develop weights in all cases within the time and budget constraints for this study. Thus, the quality of many of these weights is uncertain.

4.2.4 UK catalogue of weights

Searching for weights used by NICE led to the identification of a study by Sullivan et al. (2011) which developed disutility weights for 135 diseases with the aim of creating a catalogue for future use by NICE. This study drew on work carried out by Sullivan to develop a catalogue of weights for use in the US. It drew on the already available EQ-5D Community-based index data from a survey of over 79,000 individuals in the US to derive a MAUI-based catalogue of EQ-5D scores and weights for the UK. The authors note that the use of US EQ-5D data is not ideal, but argue that the US data are already drawn upon in many of the models that are submitted to NICE. Importantly, the research drew upon a MAUI scoring algorithm based on UK Community preferences; this algorithm was used to calculate the UK equivalent of the EQ-5D scores elicited from the US population.

Sullivan et al. (2011) note that "an 'off-the-shelf' catalogue of EQ-5D scores provides a standardized and consistent source for use in cost-effectiveness analyses" and that their approach will promote a consistent source of utility scores. However, they also recognise the limitations of the approach, most notably that it is based on a modelling exercise and US EQ-5D data.

Interestingly, the outputs from the work provide not just utility weights for a series of different health conditions (15 of which are relevant to this study), but also the disutility decrements associated with moving from a standard health state for an individual of a certain age, gender, race, ethnicity, income and education. The advantage of this is that their approach enables calculation of the disutility associated with the development of one or more health conditions, i.e. it could be used to address the issue of comorbidity which may be relevant to exposures to SVHCs. An example is given to illustrate the calculation of the development of diabetes over time, with the additional development over time of hypertension and stroke, with progression to myocardial infarction and then congestive heart failure.

4.2.5 Tengs and Wallace's repository

The final source of utility weights is Tengs and Wallace (2000), who provide a comprehensive review of available quality of life weights so as to develop a "national" repository (although it is of note that they included a review of the then available UK studies held by the University of York). Their work involved identifying relevant original studies (154 in total), and then extracting the utility (QoL) weights from these, together with information on the assessment method, respondents and the QoL scale that was used when generating the weights. This work led to the identification of 1,000 original weights (and the authors note that they were not able to review all existing studies due to time and space limitations). The authors hoped that this review would lead to more consistent use of utility weights and hence more comparable cost-effectiveness analyses.

From this reference, it was possible to identify 20 relevant utility weights. In many cases, there was a choice as to what weight was selected. This was done through the following approach:

1. One of the study team's toxicologists reviewed the different disease outcomes and combined disease and treatment outcomes and selected those combinations considered most likely to be comparable to the effects of chemical exposures. So for example, out of around 100 health outcomes related to breast cancer, only a subset were considered relevant; these were outcomes prior to treatment and included a generic "cancer, breast" and "cancer, breast, metastatic".

2. The economics team members then examined the valuation method applied and the number of "subjects", as well as the nature of the lower and upper bound used in the valuation exercise (e.g. perfect health, good health, etc.).

3. Priority was then given to the selection of the weights derived using a trade-off method, and based on a "community" survey, over those derived from a small panel of experts or reflecting the "author's" own valuation. So, for example, in the case of "cancer, breast, metastatic" a weight derived using the TTO from a population of 54 community members was selected over one developed using the SG and a survey of 4 experts; this is despite no information be available on the boundaries for the utility scale for this study.

Tengs and Wallace (2000) purposefully did not subjectively rate the quality of the studies. They instead suggest that this is something that should be undertaken by "meta-analysts who wish to weight studies according to their quality". They note though the difficulties that exist in assessing the quality of studies based on published reports, as key details on methodology are often omitted in source documents, particularly when the weights were elicited as part of a larger cost-effectiveness study. They also note that studies may have flaws that are not apparent from a journal article.

4.2.6 The collated weights

The utility weights collated from the above sources are given in Table 4-1 below. This includes the weights derived to form the basis of the UK and US catalogues developed using MAUI methods (both are shown here to highlight the differences between weights). The selected weights taken from Tengs & Wallace are then given, highlighting any important qualifications regarding the disease outcome selected, and then the various weights identified through the review of the CEA and existing NICE guidelines are reported. The UK catalogue of weights data are reported before the weights from Tengs & Wallace, the CEA or NICE guidelines, as they should be internally consistent and are based on the recent EQ-5D index scores, albeit not from an European population. As can be seen from the table, the weights available from the different sources are highly consistent for some health outcomes but vary widely for others (e.g. lung cancer, allergic contact dermatitis, hearing disorders and pancreatitis). This may be due to methodological differences but also due to differences in the health states actually being assessed (e.g. lung cancer) by the different studies.

Table 4-1: Collated utility weights from the sources reviewed				
Disease/Health outcome	US "Catalogue"	UK "Catalogue"	Tengs & Wallace	CEA and various NICE guidelines*
Cancer				
Lung	n/a	0.56 [11]	0.15 [12] - 0.58	0.473 [13]
Kidney	n/a	n/a	0.7 **	0.69
Urinary bladder	n/a	0.711	0.7 **	0.9
Liver	n/a	n/a	0.49	0.73
Skin	n/a	0.787 [14]	0.7 **	0.65
Breast	0.81	0.756	0.63	0.77
Stomach	n/a	n/a	0.75	0.65
Leukaemia	n/a	0.684	0.5	0.6
Non-Hodgkin's Lymphoma	n/a	0.717	0.76	0.618
Pancreatic	n/a	n/a	0.62	0.72
Prostate	0.767	0.687	0.58	0.58
Cardiovascular				
Coronary heart disease[15]	0.724	0.623	0.8	0.718
Acute Myocardial infarction	0.704	0.605	0.87	0.8
Stroke	n/a	n/a	0.76	0.707
Impaired Fertility				
Infertility	n/a	n/a	n/a	0.82
Disruption of ovarian cycle (low)	0.872	0.835	n/a	0.81
IMPAIRED FOETAL GROWTH				
Low birth weight	n/a	n/a	0.75	0.94 [16]
Idiopathic intellectual disability	n/a	n/a	n/a	0.62 [17]
Cleft palate	n/a	n/a	n/a	0.78
Spina bifida	n/a	n/a	0.3[18]	0.454
Renal abnormalities	n/a	0.733 [19]	0.63	0.7
Urogenital abnormalities	n/a	0.739	n/a	0.895
Other				
Chronic obstructive pulmonary disease	0.802	0.732	n/a	0.82
Asthma	0.797	0.722	0.93	0.93
Allergic contact dermatitis[20]	0.819	0.77	0.95 [21]	0.93
Osteoporosis	0.753	0.67	n/a	0.72
Anaemia (no cancer) [22]	0.758	0.672	n/a	0.76

** Generic weight for "cancer"

[11] Cancer of bronchus (a form of lung cancer)

[12] Refers to small cell lung cancer which is significantly different from non-small cell lung cancer but no weights are reported for non-small cell. Other studies report a mean QoL weight for lung cancer of 0.58.

[13] Progressive non-small cell lung cancer

[14] Melanoma

[15] US catalogue and UK catalogue refer to "coronary atherosclerosis and other heart disease"

[16] Very low birth weight but no disability

[17] Mild retardation

[18] The same sources were used by Tengs and Wallace and in the CEA Registry, although different weights are reported where these were generated using different methods; note the difference in weights depending on the method.

[19] Other congenital anomalies

[20] US catalogue and UK catalogue refer to "inflammatory skin condition"

[21] Rash

Table 4-1: Collated utility weights from the sources reviewed				
Disease/Health outcome	US "Catalogue"	UK "Catalogue"	Tengs & Wallace	CEA and various NICE guidelines*
Glaucoma	0.782	0.715	n/a	0.92
Hearing impairment[23]	0.814	0.743	n/a	0.421
Liver cirrhosis	n/a	n/a	0.92	0.82
Stomach ulcer	0.734	0.635	n/a	0.881
Pancreatitis	n/a	0.487	n/a	0.89
Obesity	n/a	n/a	n/a	0.707
Parkinson-like condition	n/a	0.432	n/a	0.74
Alzheimer disease and other dementia[24]	0.563	0.442	n/a	0.68
* See Annex 3 for further details of the sources of these weights				

Further details on these studies and the weights are provided in a summary table given in Annex 3.

4.2.7 Overview of advantages and disadvantages of different sources

Table 4-1 presented four different sources of utility weights: the weights generated for the so-called "UK catalogue"; the weights summarised in Tengs and Wallace (2000); and the weights we were able to collate from the CEA Registry and various NICE guidelines. To summarise, the potential advantages and disadvantages of these different sets are as follows:

- The UK catalogue:
  - Advantages: internally consistent set of weights
  - Disadvantages: based on US health state data, calculated using a UK algorithm, only covers a sub-set of the diseases of interest
- Weights reported in Tengs and Wallace:
  - Advantages: data on original methods available, allowing selection of weights based on quality criteria
  - Disadvantages: older studies, not all weights needed are associated with studies that would be judged of high quality, studies do reflect a range of regions
- Weights extracted from the CEA Registry and NICE guidelines:
  - Advantages: newer studies and from the combination of the two sources weights can be identified for the full list of diseases

[22] US catalogue and UK catalogue refer to "deficiency and other anaemia"
[23] US catalogue and UK catalogue refer to "other ear and sense organ disorders"
[24] US catalogue and UK catalogue refer to "senility and organic mental disorder"

- Disadvantages: data on original methods not always available, some of the weights that are reported are not consistent with those given in the original studies; studies reflect a range of regions.

Which weights?

Note that although there was the potential for weights drawn from the list developed by Tengs and Wallace (2000) and those extracted from the CEA Registry and NICE guidelines to overlap, this was the case for only one disease outcome. Further checking of the original references found that the reporting studies had extracted different weights. As the CEA Registry sources was the newer reference it was examined to see why, but no justification could be identified other than the potential for either an alternative disease specification to have been adopted or for the authors to have selected weights derived using different methods being reported together.

This latter problem is highlighted by one of the studies reported in Tengs and Wallace. Even when all other factors in the assessment are kept constant (i.e. the number of participants, the upper and lower bounds, etc.), different methods will not necessarily elicit the same or similar weights. This is illustrated by the utility weights reported by Tengs and Wallace (2000) for spina bifida; these are taken from an original study by Kelly et al. (1996). In each case, the assessment method for deriving the weights used a panel of six experts or clinicians other than the author of the study. Furthermore, the upper and lower bound of the study is the same for each method, as can be seen from Table 4-2.

Table 4-2: Example of discrepancies in weights and method reporting

Health State	Utility Weight	Assessment Method (No. of subjects)	Lower/Upper Bound
Spina bifida, thoracic/higher lumbar lesion, adult	0.184	EQ	Death/Perfect Health
Spina bifida, thoracic/higher lumbar lesion, adult	0.30	HUI	Death/Perfect Health
Spina bifida, thoracic/higher lumbar lesion, adult	0.465	QWB[25]	Death/Perfect Health

Source: Kelly et al (1996) as reported in Tengs and Wallace (2000)

As can be seen from Table 4-2, the QOL weights resulting from use of the three different assessment methods are very different. The importance of these differences can be demonstrated using an theoretical example of a cost-effectiveness exercise for a measure that treats or prevents cases of spina bifida: if we assume that there is a treatment which can prevent spina bifida, that it costs €20,000, and that the patient has an added life expectancy of another 10 years as a result of this treatment, we have the following costs per QALY for these different weights (see Table 4-3).

[25] QWB is an abbreviation of "Quality of Well Being".

Table 4-3: Example of differences in impact associated with differences in weights		
Assessment Method	QOL Weight	Cost per QALY
EQ	0.184	€10,870
HUI	0.3	€6,670
QWB	0.465	€4,300

The cost per QALY is more than double using the EQ-5D based weight than it would be if we were to use the QWB-based weight. This highlights the potential implications of adopting one weight in preference to another, without sound justification for so doing.

Indeed, although Tengs and Wallace developed their list of weights to act as single repository for use by analysts, they also warn potential users. In particular, they highlight potential issues when trying to combine weights from multiple sources, e.g. as part of a meta-analysis. To quote:

> "… as with all meta-analytic studies, caution should be exercised. In particular, studies should be selected with care, and analysts should consider whether they want to incorporate all data or just data from the largest or best studies. It would also be foolhardy to combine weights elicited with different methods without giving careful thought to the implications of this strategy. One should not simply average weights elicited with the standard-gamble method with those derived through the EuroQol, for example. Even among those weights elicited with the same method, analysts will want to take into account heterogeneity in the upper and lower bounds of the elicitation scales and variability in subject populations. Meta-analysts will want to consider carefully the merits of combining data elicited from groups that vary by such measures as gender, age, race, severity of illness, and presence of comorbidities."

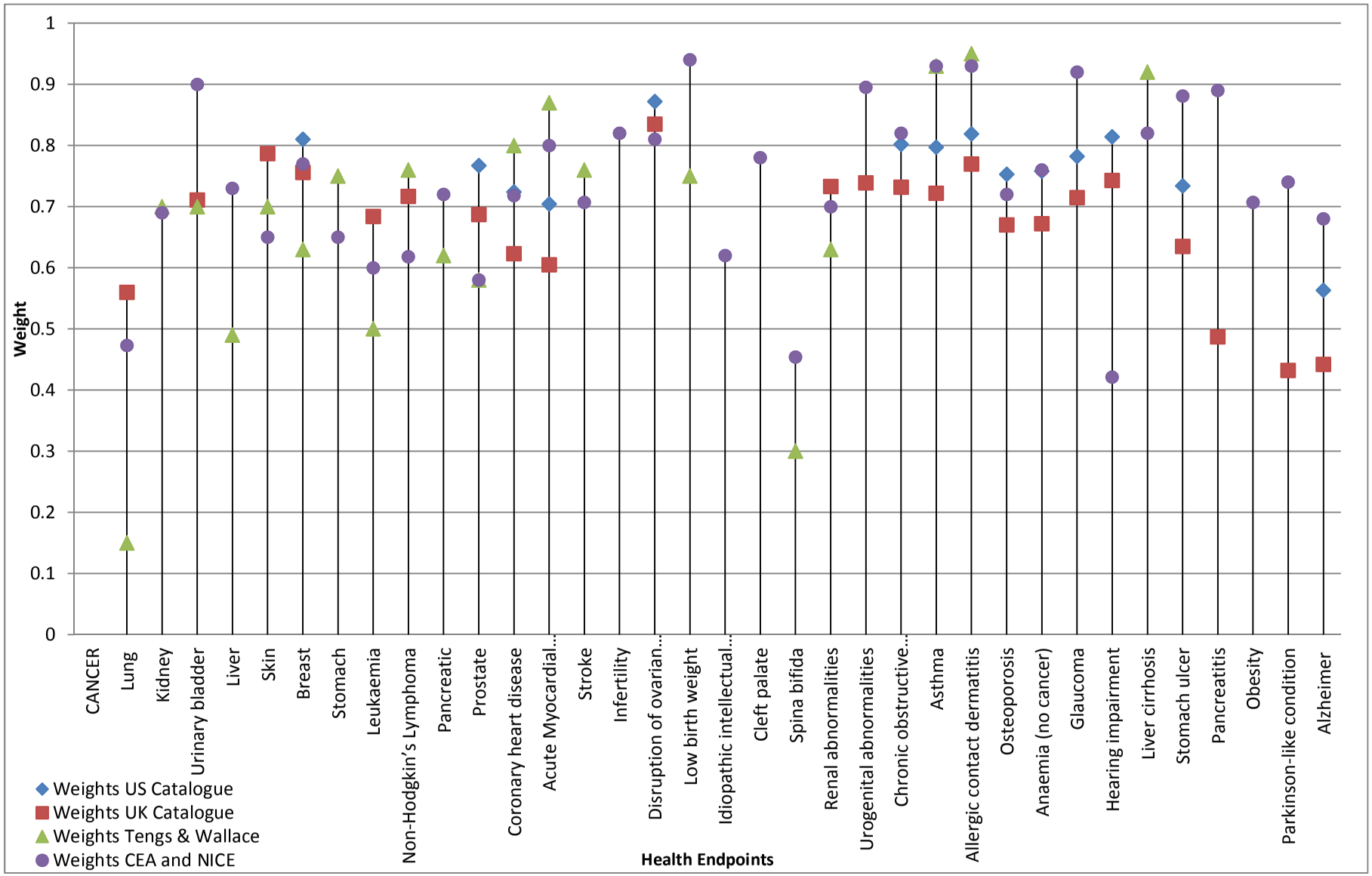
However, it is not recommended here that the weights from Tengs and Wallace are carried forward for use in REACH SEAs. This is due to the age of the studies reported in this repository. Developments in lifestyle, technology, medical understanding of diseases and other such factors mean that the most reliable weights will be those which were generated more recently in time, so that they are more accurate reflections of the preferences and utilities of today's population.

Consistency in weights across the different sources

In order to assess the consistency of the weights pulled out of the different sources, we have graphed them, with the results provided in Figure 4-1. This graph is interesting and highlights some of the disparities that exist in the weights that are reported in the literature. It also highlights some anomalies in what the different weights imply in terms of the relative level of disutility associated with different diseases.

Starting with a consideration of the patterns associated with the different sets of weights, it is interesting to note that, although the UK catalogue is based in part on US data, there are some differences. In particular, all of the UK utility weights are lower than those for the US suggesting that the conditions reflect a lower health quality than in the US. As would be expected though, the two sets move in a roughly similar pattern. In contrast, the weights

Figure 4-1: Comparative assessment of utility weights from different sources



taken from Tengs and Wallace show much more variability and there are several outliers in this set, in particular for lung cancer and spina bifida (although as shown above, selection of a different valuation method from the original study would have an impact on this value). More generally, there is relatively good agreement of these values with some of the weights related to cancer extracted from the CEA Registry and the various NICE guidelines. Finally, the utility weights reported in the CEA Registry database are significantly higher than those given by the other sources for many of the non-cancer outcomes, although they are also lower for others (i.e. hearing impairment).

We note though that the pattern of the weights drawn from the CEA Registry and NICE guidance documents overall is more as was anticipated – in other words, there is a lower quality of life associated with health outcomes such as cancer than living with health outcomes such as COPD, dermatitis, etc.

To further analyse these relationships, the correlation coefficients for the different sources of weights have been calculated, and are presented in Table 4-4. The high correlation between the CEA weights and those reported in Tengs and Wallace is not surprising, given our finding that they may be drawing on the same studies (even if they sometimes report different weights). Similarly, the correlation between the US and UK catalogues is not surprising given that they are based on the same underlying EQ-5D profiles (although their modelling algorithm will differ). The low correlation between Tengs and Wallace and the CEA and NICE weights with those of the UK catalogue, however, is more worrying – these different sources do not appear to provide weights that are highly correlated. This raises questions over the validity of the weights given in the UK catalogue in terms of their consistency with other studies more generally, or over the actual similarity in disease outcomes associated with the different sets of weights.

One reason for some of the potential differences between the older (Tengs and Wallace) and newer sets of weights may be that, over time, our understanding of and ability to treat different diseases has changed. As noted above, such changes may have impacted on the weights associated with different diseases over time, resulting in our recommendation not to draw from the Tengs and Wallace repository for REACH SEAs (unless part of a sensitivity analysis).

Table 4-4: Correlation coefficients for the utility weights

	Weights US Catalogue	Weights UK Catalogue	Weights Tengs & Wallace	Weights CEA and NICE
Weights US Catalogue	-	0.9952114	0.002892	0.153628
Weights UK Catalogue	-	-	0.450844	0.101096
Weights Tengs & Wallace	-	-	-	0.767608
Weights CEA and NICE	-	-	-	-

Conflicting weights

It is clear that decisions made by the study team may also affect the consistency found in the above analysis. Within the resources available for the study, it was not always possible to track through the full trail of source documents. This can be illustrated using lung cancer.

Initially, a utility weight for lung cancer of 0.8 was identified from the CEA registry, referring to a study which evaluates a smoking-cessation program, Txt2stop (Guerriero et al., 2013). See the Box below which provides a "snip" from the CEA Registry.

Note that the study by Guerriero et al., 2013 initially reports that it uses a health state value of 0.58 for lung cancer, taken from another study by Flack et al. (2007). This paper, in turn, refers to Tengs and Wallace (2000) who report QoL weights for 6 different health states for small cell lung cancer; in order to derive a weight for lung cancer, they simply take the average of these weights. This gives a mean QoL weight for lung cancer of 0.58.

Search Results (Back) Article/Ratios
Your search returned 300 results

Pick Columns to Display(Sort by)

Article ID	Health State	Weight
2013-01-11070	Disease progression of advanced non-small cell lung cancer	0.47
2013-01-11070	Progression-free survival (PFS) of advanced non-small cell lung cancer using carboplatin-gemcitabine	0.56
2013-01-11070	Progression-free survival (PFS) of advanced non-small cell lung cancer using erlotinib	0.65
2013-01-10819	Overall survival associated with non-small cell lung cancer	0.47
2013-01-10819	Progression-free survival associated with non-small cell lung cancer	0.65
2013-01-10769	Advanced nonsquamous non small-cell lung cancer (Progression survival)	0.47
2012-01-09731	Chronic obstructive pulmonary disease	0.73
2012-01-09731	Myocardial infarction	0.8
2012-01-09731	Lung cancer	0.8
2012-01-09731	Coronary heart disease	0.48
1 2 3 4 5 6 7 8 9 10		

* In 2013 US dollars

The only mention of the figure 0.8 with regard to lung cancer in the Txt2stop study is in the probabilistic sensitivity analysis. The CEA registry website does not explain why it has listed 0.8 rather than 0.58. Furthermore, when one searches the Registry for diseases such as myocardial infarction, COPD and coronary heart disease, the Txt2stop study (Article ID 2012-01-09731) was not always identified on the Registry as holding relevant weights, even though weights for these conditions are reported in the study.

Similarly, the CEA Registry reports a utility value of 0.73 for hepatocellular carcinoma (HCC) which is the most common form of liver cell cancer. This value comes from a study by Hutton et al. (2010) which assesses cost-effectiveness of hepatitis B treatment in China. The utility value used is not derived by this study. Instead it is taken from three other sources: Kanwal et al. (2005), Chen et al. (2007) and Sheperd et al. (2006). We were only able to gain access to the latter, and the authors of this study do not elicit the weights themselves either. Instead, the weights are taken from another source, Wong et al. (1995), who used both the TTO and SG methods. Sheperd et al. report very different weights to those given by Hutton et al. Sheperd et al. differentiate between age groups and report weights for each age group. In this case, we adopted the weights provided by Sheperd et al.

This is a recurring problem when trying to identify the most appropriate weights for several of the diseases of most interest. It also highlights the importance of performing independent reviews of the original studies and not relying on secondary sources.

4.3 Collated DALY weights

4.3.1 Sources of weights

A slightly different approach was taken to the collation of the DALY disability weights, from that adopted for QALY utility weights. As part of the early literature searches, it became apparent that there are several larger studies which contain comprehensive lists containing disability weights. These acted as the starting point, although a broader literature search identified the review of 22 disability weight studies by Haagsma et al. (2014).

To summarise, identification of sources of DALY weights was undertaken as follows:

1) Work related to the WHO's Global Burden of Disease (GBD) initiative was reviewed to identify the availability of up-dates to the weights used in the calculation of DALYs since the development of the original set of weights in 1996 by Murray et al. The most recent update was published in 2015 using 2013 data; this is referred to as the Global Burden of Disease 2013 update.

2) At the national level, three major initiatives were identified:

    i. the Victorian Government's (*Australia*) Burden of Disease report (2001)

    ii. work undertaken in the Netherlands, referred to as the "Dutch Report" (Stouthard et al. (1997), and which acts as the basis for some of the weights adopted in the Australian and the (Australian State) Victorian Government's study, and

    iii. work undertaken to develop disability weights for Estonia in line with the GBD (Lai et al., 2009).

3) Work undertaken as part of the European Disability Weights Project was also reviewed, and

4) Towards the end of the study, the newly published European Burden of Disease (BoD) study (2015) became available, and was incorporated into the analysis.

Although the review by Haagsma et al. (2014) identifies other studies, many of them covered irrelevant health outcomes (injury) or were undertaken for countries with significantly different cultures than Europe (Korea, Iran, Zimbabwe, Burkina Faso). As a result, they were not examined further.

4.3.2 The Global Burden of Disease studies

The first GBD study undertaken by Murray et al. (1996) used two variants of the PTO to develop disability weights. In the first variant, referred to as PTO1, respondents were asked to decide for how many individuals (N, where N> 1000 persons) in health state X they would be willing to trade one year of life extension of 1000 healthy individuals for the extension of life by one year for the group in health state X. In the second variant (PTO2), the respondent is asked to estimate for how many individuals in health state X he would be prepared to surrender one year of extended life for 1000 individuals in perfect health in exchange for complete recovery followed by one year of perfect health for the group in the given health state. Thus, Murray incorporated a trade-off element into his operationalization of the PTO approach (see also Stouthard et al., 1997). Weights were elicited based on a period profile approach.

The 2012 up-date of the GBD disability weights, based on 2010 data and thus also referred to as the GBD 2010 study, did not use the PTO method as applied by Murray or in the intervening 2004 up-date. Instead, Salomon et al. (2012) used a combination of face to face and telephone surveys using paired comparison questions; in these paired comparisons, respondents were asked to consider two hypothetical individuals with different, randomly selected health states and to then identify which person they regarded as healthier. The web survey added questions about population health equivalence, which sought respondents' views on the overall health benefits of different life-saving or disease-prevention programmes. In this case, the weights are stated as referring to both short term and longer term health states, although it is not always clear from examination of the health state descriptions whether a specific state is considered to fall into one or the other category.

The face to face surveys were carried out in Bangladesh, Indonesia, Peru, and Tanzania, while the telephone interviews were carried out in the US, with a sample population of over 13,900. In addition, an open-access web-based survey was carried out, with this having over 16,300 respondents (for a total of over 32,000 respondents). Analysis of the paired comparison responses indicated a high degree of consistency across the surveys: correlations between individual survey results and results from analysis of the pooled dataset were 0.9 or higher in all surveys except in Bangladesh (r=0.75).

Additionally, Salomon et al. (2012) compared the weights generated through their research with those used in WHO's most recent update of the Global Burden of Disease Study for 2004. They found:

- There is a broad pattern of agreement between the old and new weights, particularly for health states outcomes in the moderate-to-severe range
- For health states in the mild range (with a disutility below 0.2), they found lower weights for many states than previously
- As a result, they conclude that contrary to the view that disability weights will vary across cultural environments, there may well be a high level of consistency.

Salomon et al. (2012) are explicit that in developing their pairwise comparison approach that their aim was to measure health loss rather than welfare loss; they argue that this is

appropriate as health loss is the focus of the GBD work. Note they also argue that the PTO methods used by Murray, because they included an element of trade-off, did not clearly distinguish between health and wellbeing (i.e. welfare). However, it is also clear that the method used for eliciting the disability weights had to be appropriate to the size and nature of the populations being surveyed.

There were a number of criticisms of the GBD 2010 study, mainly with respect to how health states were defined. As a result, an up-date was carried out in 2013, with the results of this published in 2015. The GBD 2013 study utilised the same weight elicitation methods as used in the GBD 2010 study with a few key refinements to take into account the criticisms made about the previous study. Firstly, the cause and sequelae list was expanded from 1160 to 2337 sequelae, to include asymptomatic sequelae; the nature-of-injury categories were expanded from 23 to 47; the cause list was expanded from 289 to 301 causes. Furthermore, the lay descriptions of some of the health states were revised so that the changes in severity of the health state are more consistent across different health states. Although there had also been criticisms of the use of paired comparisons, this aspect of the methodology was retained.

The authors of the GBD 2013 study (Haagsma et al., 2015) collaborated with the European Centre for Disease Prevention and Control, who were also conducting a European burden of disease study, using the same protocol as Salomon et al. (2012). As a result, the national surveys being carried out for the European BoD (see below) included 140 of 220 GBD 2010 health states for which the lay descriptions had not been revised, 32 health states with revised lay descriptions, and 42 new health states, 16 of which were included in GBD 2013. They then pooled the data from the European disability weights study and the GBD 2010 weights study into a single analysis, thus doubling the number of respondents to 60,890 across both studies. For states where the lay description was not previously included, was revised, or which were new, only the European disability weights measurement study data were used. This means that all disability weights in GBD 2013 differ from the GBD 2010 disability weights. The authors note that most of the disability weights generated in the 2013 study vary only slightly from the 2010 weights, though for some health states the change is more significant. A comparison of the weights of interest to this study, however, indicates only slight changes.

Weights elicited for the GBD 2013 study reflect either a long-term or short-term time profile, as was the case for the GBD 2010 study. The time profiles can be found in Appendices C, D and E for non-fatal cancer outcomes, injuries and impairments in a supplement available from The Lancet's website[26]. Similar supplements are available for the GBD 2010.

[26] http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2815%2960692-4/fulltext

4.3.3 The "Dutch" study

Although undertaken in 1997 as a follow-up to the 1996 WHO GBD study, the Dutch Disability Weights study (Stouthard et al., 1997) is important as it acted as the basis for the methods used by several of the subsequent studies estimating disease burdens.

The Dutch study adapted the GBD measurement approach applied by Murray et al. (1996), including making modifications to the descriptions of health states and adding new health states. All health states were characterised using the EQ-5D extended for cognitive function for 53 diseases and 175 associated disease stages which were then assessed directly, with weights elicited from a panel of 45 medical experts using the PTO method and the VAS. Weights for the remaining disease stages were derived through interpolation with reference to the calibrated VAS. As noted earlier, the study adopted a combination of annual and period profiles, with this being the first study to trial the annual profile approach. Note that the Dutch study also involved a pilot phase evaluating the different methods. From this, the authors note that the TTO valuation method seemed to offer no practical advantages over the PTO.

4.3.4 The Australian and Victoria State studies (1999)

The Australian burden of disease study (Mathers et al., 1999) used the methods developed for the GBD and for the Dutch study but again adapted them to the Australian context. The report covers 176 disease and injury categories and relied on the Dutch weights where possible as it was considered that they related to conditions of most relevance to the health of the Australian population. Where Dutch weights were not available, GBD weights were used where these were available. Where there were both Dutch and GBD weights, the researchers carried out a comparison and concluded that the two studies appeared to have been valuing the same conditions in a similar way and thus that it was valid to use both sets in the same study. In total they used the Dutch study for 370 weights and the GBD for 118 weights. A set of 46 further weights was developed using the EQ-5D extended for cognitive function plus a regression model. For six additional mental disorders, Australian experts were asked to assess weights using a value rating scale to compare them with Dutch weights for other mental disorders (Mathers et al., 1999). In addition, the study adjusts the final disability estimates for comorbidities between mental disorders and between physical disorders at older ages.

The Victorian Burden of Disease Study (Victorian Government Department of Human Services, 2005) reports on the burden of disease within this Australian State in 2001. Essentially, the study up-dates the national study and adapts it for the Victorian State context. As for the national study, the Dutch weights are adopted and complemented by the GBD weights where Dutch weights are not available.

It is also of note that neither of these two studies adopt age weighting in the derivation of DALYs. In addition, as the weights are based on a combination of the Dutch and the GBD weights, it is assumed that they reflect both an annual and a period profile approach.

4.3.5 European Burden of Disease Study

As noted above, the European Centre for Disease Prevention and Control conducted a burden of disease study specific to Europe, which was commissioned as part of a study looking at the burden of communicable diseases in the European Economic Area. Haagsma et al. (2015), having critically evaluated the Global Burden of Disease 2010 (GBD 2010) study, set out to develop a set of disability weights specific for Europe for 255 health states (including 43 additional health states as indicated above); they built on the GBD 2010 study, conducting web-based surveys which employed the same paired comparison technique used by Salomon et al. (2012), as well as a discrete choice questions in the form of population health equivalence (PHE) questions. The sample population of the study was made up of 30,660 respondents from four European countries (all of which are Member States): Hungary, Italy, the Netherlands and Sweden. The authors indicate that one of the reasons for adopting the approach taken is that it allows the incorporation of the view so the general public into resource allocation decision making, which is important in democratic societies.

Note that it is not clear for Haagsma et al. (2015) what time profile all of the weights relate to. It is assumed here that a combined approach was adopted as for the previous GBD studies. Within the time and budget available for this study, it was not possible to try and collate information on time profiles for all of the weights of interest.

4.3.6 Burden of disease in Estonia

The Estonian national burden of disease study (Lai et al., 2009) was based on the general approach used in 1996 GBD and Dutch studies, but importantly included the derivation of national disability weights. A panel of 25 medical experts was used to derive weights for 283 health conditions. As a first step, 26 indicator conditions were assessed using the PTO, and the results of these valuations were then plotted to a VAS which was used as the reference point for the valuation of the remaining conditions. The endpoints for which disability weights were derived are specific to Estonia, despite being based on previous studies. One advantage of the outputs of this study is the clear indication of the time period which each weight relates to.

As noted by Lai et al. (2009), "Duration index used in YLD calculation represented a fraction of a year needed for full recovery in case of mild conditions like ordinary influenza, other acute upper respiratory infections, superficial injuries, etc. The duration index had the value '1' in case of conditions which require longer than a year for full recovery or were full recovery is not expected."

4.3.7 The collated weights

Table 4-5 provides the final collated set of the most relevant disability weights for this study. The table presents weights taken from the recent GBD 2013 update, the European BoD study and the Estonian BoD study. All efforts have been made to extract weights related to the same health outcomes, although this has been difficult given the variation in endpoint descriptions that have been adopted by the different studies. In addition, for the Estonian

Table 4-5: Collated disability weights from different sources					
Disease/Health Endpoint*		Global Burden of Disease (2015)	European Burden of Disease (2015)	Estonia Burden of Diseases	
				Weight	Time
CANCER**					
Lung	Operable	0.288	0.265	0.507	1
	Inoperable	0.451	0.358		
	Disseminated	0.540	0.515		
Kidney	Primary	0.288	0.265	0.422	1
	Disseminated	0.451	0.358		
	Terminal	0.540	0.515		
Urinary bladder	Primary	0.288	0.265	0.448	1
	Disseminated	0.451	0.358		
	Terminal	0.540	0.515		
Liver	Primary	0.288	0.265	0.677	1
	Disseminated	0.451	0.358		
	Terminal	0.540	0.515		
Skin	Primary	0.288	0.265	0.362	1
	Disseminated	0.451	0.358		
	Terminal	0.540	0.515		
Breast	Primary	0.288	0.265	0.178	1
	Disseminated	0.451	0.358		
	Terminal	0.540	0.515		
Stomach	Primary	0.288	0.265	0.420	1
	Terminal	0.5401	0.515		
Leukaemia	Primary	0.288	0.265	0.182	1
	Pre-terminal	0.451	0.358		
	Terminal	0.540	0.515		
Non-Hodgkin's Lymphoma	Primary	0.288	0.265	0.368	1
	Disseminated	0.451	0.358		
	Terminal	0.540	0.515		
Pancreatic	Primary	0.288	0.265	0.587	1
	Disseminated	0.451	0.358		
	Terminal	0.540	0.515		
Prostate	Primary	0.288	0.265	0.239	1
	Hormone refractory cancer	0.451	0.358		
	Terminal	0.540	0.515		
CARDIOVASCULAR					
Heart Disease***	Mild	0.033	n/a	*0.239 or 0.383*	*1 and 0.5*
	Moderate	0.08	0.103		
	Severe	0.167	n/a		
Acute myocardial infarction		0.074	0.098 (days 3-28)		
Stroke	Mild (Effects)	0.019	n/a	0.547	1
	Moderate (Effects)	0.07	0.075		
	Severe (Effects)	0.552	0.580		
IMPAIRED FERTILITY					
Infertility		0.008	*0.008*		
Disruption of ovarian cycle		n/a	*n/a*	0.227	0.4

Table 4-5: Collated disability weights from different sources					
Disease/Health Endpoint*		Global Burden of Disease (2015)	European Burden of Disease (2015)	Estonia Burden of Diseases	
				Weight	Time
IMPAIRED FOETAL GROWTH					
Low birth weight		n/a	n/a	0.442	1
Idiopathic intellectual disability	Mild	0.043	0.053	0.242	1
Cleft palate		n/a	n/a	0.379	1
Spina bifida	Mild	n/a	n/a		
	Moderate			0.412	1
	Severe				
Renal abnormalities	Unilateral renal agenesis/dysgenesis	n/a	n/a	n/a	
	Bilateral renal agenesis/dysgenesis			0.832	0.2
	Chronic kidney	0.104	0.108	0.300 ('severe')	1
	End-stage renal failure (on dialysis)	0.571 (0.024 with transplant)	0.487 (0.030 with transplant)		
Urogenital abnormalities		n/a	n/a	0.140 (M) 0.169 (F)	0.04 0.04
OTHER					
Chronic obstructive pulmonary disease	Mild	0.019	0.025	0.299	1
	Moderate	0.225	0.284		
	Severe		0.418		
Asthma	Mild		0.020	0.093	1
	Severe		0.045		
Allergic contact dermatitis		0.408	n/a	0.183	0.1
Anaemia	Mild	0.133	0.004	0.168	0.5
	Moderate	n/a	0.045		
	Severe	n/a	0.118		
Glaucoma / Distance vision	Mild	0.004	0.004	0.168	1
	Moderate	0.052	0.034		
	Severe	0.149	0.158		
Hearing impairment	Mild	0.003	0.011	0.254	1
	Moderate	0.031	0.037		
	Severe	0.184	0.152		
Liver cirrhosis		0.01	0.163	0.475	1
Stomach ulcer		0.027	n/a	0.104	0.25
Pancreatitis		0.158	n/a	0.499	0.25
Obesity		0.178	n/a	0.168	1
Parkinson-like condition	Initial/Mild	n/a	0.016	0.244	1
	Intermediate	n/a	0.239		
	End-stage/Severe	n/a	0.530		
Alzheimer disease and other dementia	Mild	0.01	0.059	0.261	1
	Moderate	0.267	0.434		
	Severe	0.575	n/a		

*Notes: * Some values may not be directly comparable as the definitions used may vary from source to source*

*** The Estonian cancer weights relate to 'malignant neoplasms' and, as such, have been equated to 'primary' site-specific cancers used by other authors.*

**** The GBD and the European study refer to angina pectoris whereas the Estonian study refers to ischemic heart disease or inflammatory diseases of the heart*

study, we give the fraction of year used for calculating YLDs. These time periods could also be applied to the other weights, where there appears to be a good correlation between the health states.

We do not report the weights listed in the Victoria study or the original Dutch study, in part because of the age of these studies. However, disability weights reported from the Dutch study are reported in Annex 3 due to the fact that this study provides a more detailed differentiation in disease stages and severities compared to the GBD or the Estonian BoD. In addition, it covers some health effects that are not covered GBD studies (such as low birth weight); this may be important for sensitivity purposes or for gap filling.

4.3.8 Overview of advantages and disadvantages of different sources

The weights presented in Table 4-5 for the GBD 2013 study and the European BoD study are very similar, as expected, given that the former uses the results of the latter in developing new weights. One might anticipate greater parity across the European study and the Estonian study, given Estonia's status as an EU Member State. A key issue in comparing the different sets of weights though is that (as previously mentioned) the health states adopted for the different studies differ. For example, where the GBD and European study provide values for moderate angina pectoris, the Estonian study gives weights for ischemic heart disease and inflammatory diseases of the heart.

In addition, the Estonian study does not cover different severity stages for the diseases, with the result that fewer of the disease outcomes of interest are covered by the study. The European study also does not differentiate across severity levels for many of the diseases, despite being based on the GBD 2010 study; it also does not cover all of the endpoints of interest, as can be seen by a comparison of the availability of weights from this study compared to the 2013 study.

As noted earlier, the study by Haagsma et al. (2014) reviewed twenty two studies that developed disability weights and also found that the different studies have defined different health stages of a specific disease differently (which will obviously affect the associated weights); however, they also concluded that the weights for similar health states can vary significantly, particularly for mild diseases; they note differences of a factor of two or more.

Haagsma et al. (2014) further found that 10 out of 17 studies that they reviewed used medical experts or health professional to value health states; three relied on a population panel; and two on panels involving both medical experts and the more general populations. The two studies that used medical and general population panels found differences in the disability weights derived from the groups. One of the studies found a correlation of only 0.32, while the other found an average correlation of 0.76 but noted that the medical experts valued five of the nine health states significantly lower than the general population.

The approaches used for eliciting the weights reported in Table 4-5 also varied, and this is likely to be responsible for some of the differences. For the Estonian study, the weights were elicited using a panel of medical experts, following the procedure employed in the previous 1996 and 2004 GBD studies and the Dutch study. This contrasts to the more recent

GBD 2010 and 2013 studies and the European study, which respond to the academic literature recommending that instead of relying on experts disability weights should be elicited using a sample from the general population; and that, where necessary, attention should be given to educating the general population on the consequences and implications of the different disease states. Although medical experts are in a good position to judge a disease from their perspective as doctors, their ability to properly assess a disease's impact on behalf of someone who is actually suffering the disease has been questioned. This type of criticism (which arises frequently throughout the literature) was one of the reasons that the 2010 GBD collected data through the large population-based household surveys (Salomon et al., 2012), as also emulated in the most recent European study.

The review by Haagsma et al. (2014) also considered the rank order correlation between the disability weights generated by the GBD 2010 study and those generated by the 1996 GBD study, by the Dutch Disability Weights study and by the Estonian study (with the GBD 2013 and the European study published after this review). They found that there was a lack of consistency in the rankings between the GBD 2010 study and these other studies. They also note that studies that relied on rankings and VAS provided slightly worse disability weights when only a short disease-specific health state description when used, compared to these that also presented generic information on health conditions.

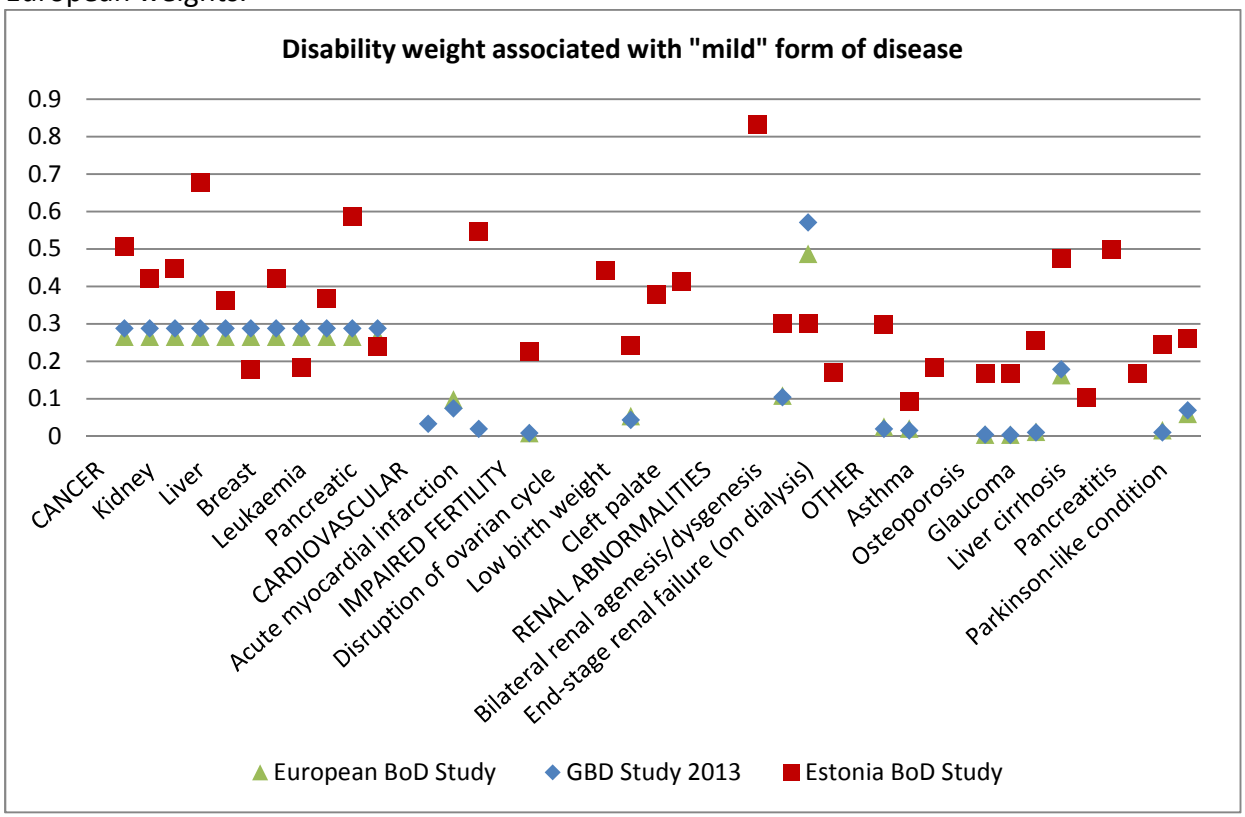
Another issue, discussed in further detail in section 5, is that the GBD studies set out to measure health loss, rather than welfare loss and so the weights derived through these studies do not necessarily reflect the welfare losses suffered through illness. This may have consequences for their use in a REACH SEA, as they may underestimate the true welfare losses from an illness for an individual.

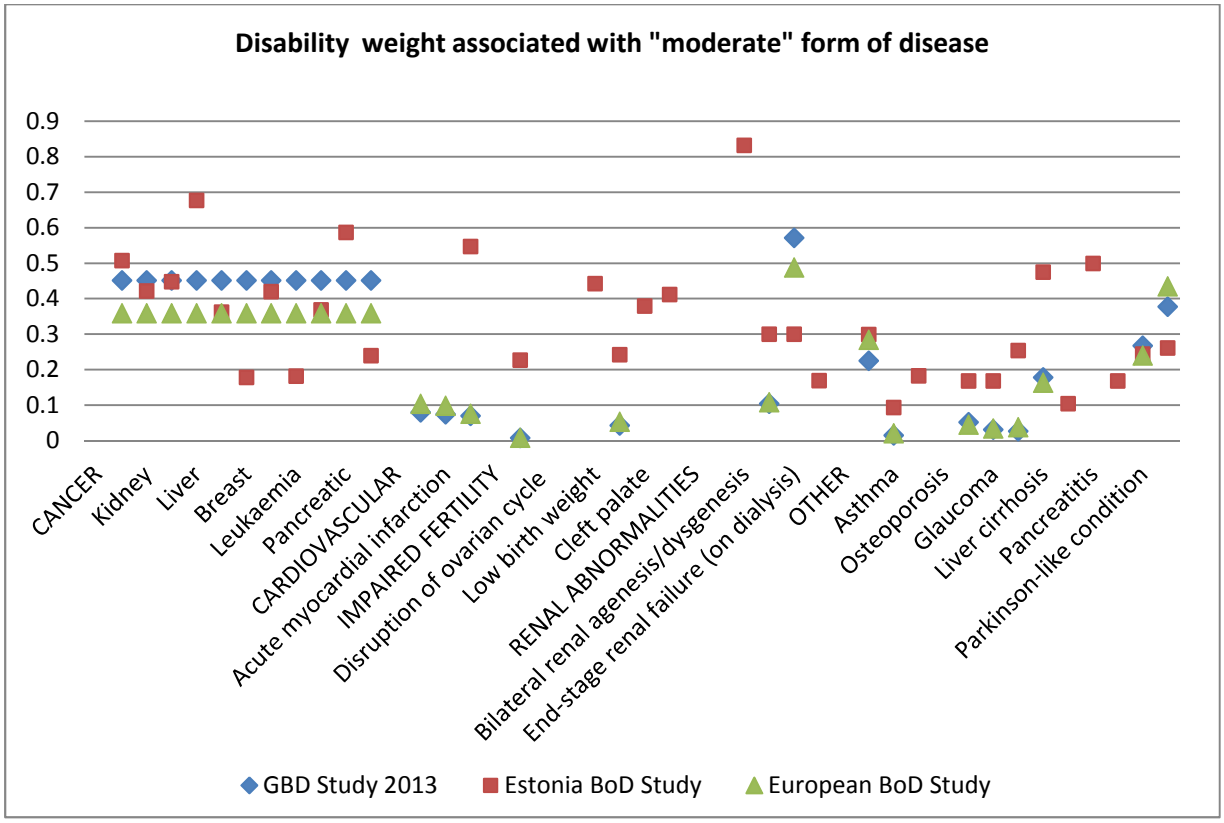
Haagsma et al. (2014) also note that valuations can vary significantly across countries, due to clear contextual differences in the ways people perceive health problems and how they affect their lives. This may be one reason the Estonian study weights differ from the GBD 2013 and European study weights, although the authors of the European study (Haagsma et al., 2015) found that the estimated weights were highly correlated across the four European countries studied. In this respect, it is also of note that the GBD 2010 study also showed consistency across the countries for which weights were derived, with the exception of Bangladesh.

Consistency in weights across the different sources

As for the utility weights, we have graphed the disability weights presented in Table 4-5 to examine their consistency across the 36 diseases for which weights could be found. Because it is important to respect the fact that weights differ for different severities of disease, we have developed three graphs, given as Figure 4-2 below. These depict weights for the mild health outcomes from all three sources where available, the same then for the moderate and most severe outcomes. Where only one weight is available, as for the Estonian study, then this weight is used in all three graphs for the disease outcome regardless of severity.

At risk of over-simplifying the potential differences that exist between the different studies in terms of what was actually valued, some patterns can be identified from these graphs. Across all three severities, the weights from the European study and the weights from the GBD 2013 study are very similar across all endpoints; however, it must be remembered that this is due to in part to the fact that GBD 2013 draws directly on the European study weights. Because, in many cases, the Estonian study only provided a single value, this affects the relative position of its weights compared to the other sets of weights. Even so, there does not appear to be any strong consistency between the GBD 2013 weights and the Estonian weights across any severity levels. The same is true of the former and the European weights.





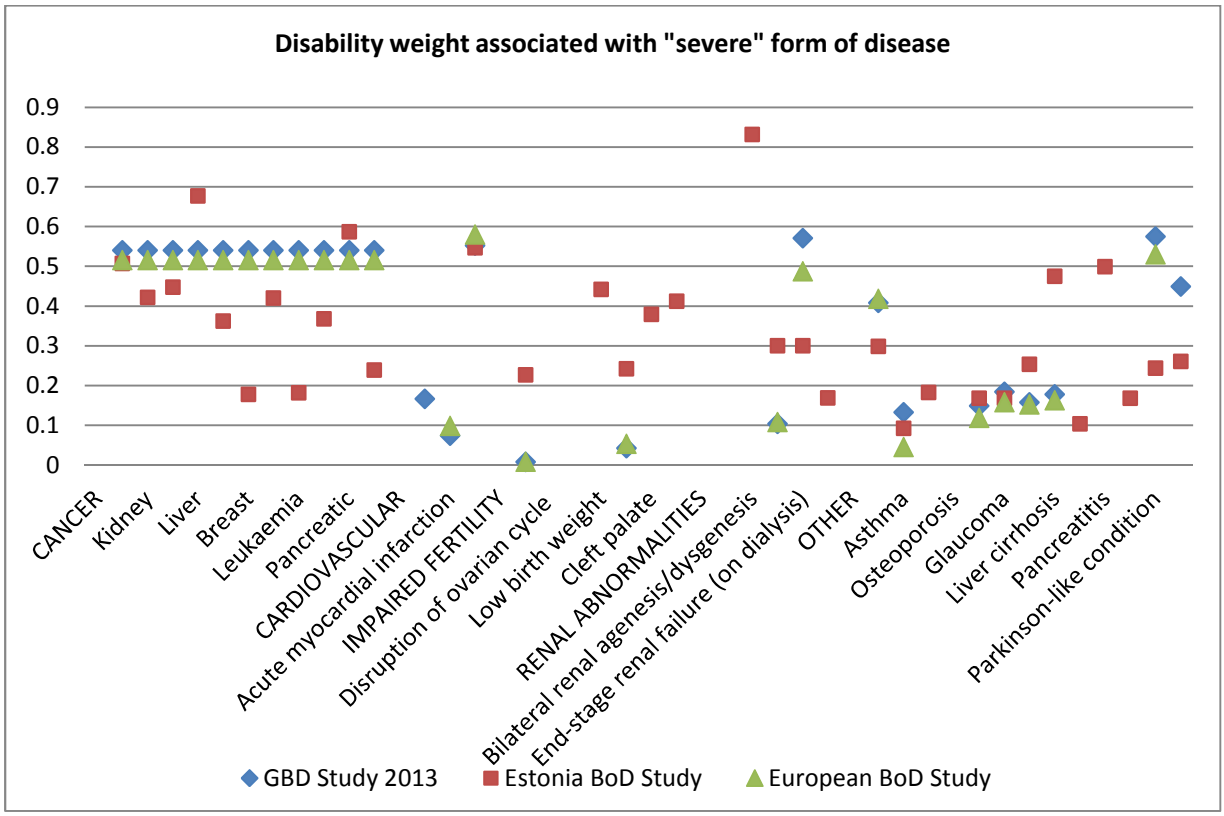


Figure 4-2: Comparative assessment of disability weights

The above findings are also illustrated by the correlation coefficients between the different sets of weights for each severity level, as presented in Table 4-6.

Table 4-6: Correlation coefficients across the three sources of disability weights			
	Mild	Moderate	Severe
Correlation GBD and Estonian	0.3825	0.3991	0.456
Correlation GBD and European	0.9979	0.9779	0.9921
Correlation Estonian and European	0.4976	0.3648	0.5067

4.3.9 Summary

QALY utility weights

The above issues highlight the uncertainty that surrounds any potential set of utility weights for potential use in REACH Authorisation or Restriction SEAs. There are no large studies which cover the full range, or even most, of the disease outcomes for utility weights (as available for the disability weights) other than the work by Sullivan et al. in developing the UK "catalogue of weights"; this is, however, based on US ratings of health states. Furthermore, as indicated by Tengs and Wallace (2000), weights derived using different methods should not be combined to develop mean values. They could be used individually or separately side by side, although this might limit the potential for considering comorbidities as there are only a small number of sources that consider a range of health outcomes.

In this respect the UK "catalogue of weights" is interesting in that it provides figures for the marginal disutility associated with the onset of a disease, taking into account different starting health states. This approach may be highly attractive in the context of REACH, as it allows consideration of comorbidities (which can be relevant to chemical exposure related disease cases – something that cannot be done with disability weights) and calculation over the analyst's choice of time periods.

DALY disability weights

Haagsma et al. (2014) note that, in the absence of a standard for estimating disability weights, it is hard to evaluate the validity of different sets of weights. They conclude that "disability weights from studies with different designs cannot be used interchangeably". This suggests that, if internal consistency is important, then this study should recommend the set of weights that covers the largest number of disease endpoints, which, in this case, would be the GBD 2013update; alternatively, one would recommend the European BoD study as it covers almost as many endpoints but is specific to a set of EU Member States and is consistent with the WHO's broader GBD studies. However, it is of note that the original Dutch study (Stouthard et al., 1997) provided a greater differentiation in health states and covers diseases not addressed by the GBD studies or the European Bod.

A comparison of the three sources of disability weights indicates that the Estonian study performs comparatively worse, particularly in the context of this report: the weights are

elicited from medical experts; there are fewer endpoints than the other studies and these endpoints and weights are specific to Estonia.

In terms of time profiles, the Estonian study provides an indication of the time profile to be assumed for each weight in a summary table. For the GBD 2013 study and the European BoD, analysts will have to examine the detailed health state descriptions set out in the detailed study annexes to establish the time profiles used when eliciting weights. These are available on-line from The Lancet[27].

The GBD 2013 update provides weights for more endpoints, including more severity levels for particular diseases. However, these weights are not European specific, as opposed to those generated by the European BoD study. Unfortunately the European study does not cover as many endpoints. As noted above, it should also be remembered that the weights derived by both of these studies were generated for use in measuring health loss, rather than welfare loss[28].

Table 4-7: Advantages and disadvantages of different sets of DALY weights

Study	Advantages	Disadvantages
GBD 2013	• Consistent with European weights • Greater no. of endpoints • Weights consistent across most of the countries studied	• Weights elicited from general population (including those pooled from the European BoD) • Weights derived with a view to measure health loss and not welfare loss
European BoD	• Consistent with GBD study • Specific to Europe • Weights taken from general public • Weights consistent across the four countries studied	• Fewer endpoints than GBD • Weights derived with a view to measure health loss and not welfare
Estonian BoD	• Provides clear time durations to be used when calculating YLDs*	• Weights elicited from medical experts • Endpoints and weights are specific to Estonia • Fewer endpoints than GBD
*Time durations are likely to be available from the other studies, but are not reported alongside the weights		

[27] For example, for the 2013 GBD see: http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2815%2960692-4/fulltext

[28] The difference between health loss and welfare loss is that the former does not take into account the intangible effects which arise from illness such as the economic burden for an individual. Welfare is defined here in the economic theory context, specifically welfare economics. Please see section 5 for further discussion on this topic.

5 Transferability of the Available Utility Weights

5.1 Introduction

The QALY model has been a popular tool in cost-utility analyses because it is a simple, intuitive concept which allows two outcomes of health-care (quality and quantity of life) to be measured using one index. Similarly, DALYs have been adopted by the World Health Organisation and others as providing a simple metric for measuring the burden of disease within a population. An advantage of both approaches is that they allow a variety of diseases and health endpoints to be compared using a common unit of measurement (i.e. a QALY or a DALY).

The previous sections have reviewed the methods used to elicit the utility and disability weights that form the starting point for QALY and DALY calculations, and presented information on the availability and pedigree of the weights available for the types of health effects likely to be of most interest in forthcoming REACH SEAs.

Table 5-1 summarises the key characteristics of the DALY and QALY approaches, highlighting their similarities and differences (adapted from ECDC, 2011).

Table 5-1: Characteristics of the QALY and DALY methods

Factor	DALY	QALY
Summary measure of population health type	Health loss	Health gain
Unit of measurement	Years of life lost in a population due to premature death and disability, referring to an arbitrarily predefined health goal Sum of Standard Expected Years of Life Lost (SEYLL) and Years Lost due to Disability (YLD)	Is a product of survival time and quality of time QALYs are generally expressed as QALYs obtained by comparing two populations
Health dimension	Disability – i.e. loss of functional capacity	Health – related quality of life associated with certain health states
Information on mortality	Yes – SEYLL	Yes – standardised life expectancy of the population under study (i.e. general population; drug-users, etc.)
Information on non-fatal outcomes	Yes – YLD	Yes – through quality of life associated with non-fatal health outcomes
Disease specific approach	Yes – e.g. conditions can be linked to an ICD classification	Yes and No – approach can be disease specific or can be based on the use of generic health state descriptions which can then be linked to a disease or ICD classification

Table 5-1: Characteristics of the QALY and DALY methods		
Factor	**DALY**	**QALY**
Characteristics of measurement	DALYs developed for assessments of the Global Burden of Disease Latest sets measure health losses and rather than welfare losses Comparability of DALY estimates between population and disease Like events are treated equally – i.e. like events contribute to the same numbers of DALYs, irrespective of the individual's environment Individual characteristics are restricted to age and sex	Developed by economists, decision scientists and psychologists Preference-based measure Intended use for evaluation and intervention planning Comparison for interventions Primary use for cost-utility analysis Applies cost-utility ratio which describes the incremental price of obtaining a unit of health effect Utilities may differ from country to country/region to region Utilities may be derived directly or indirectly
Data required for REACH	Life expectancy data, duration of disease, disability weights, number of cases (incidence)	Life expectancy data, expected duration of disease, time-to-event or event probabilities, utility weights, number of cases (incidence)
Value choices	Age weighting Sex specific weights Time discounting Disability weights	"If…then" applied, sex-specific weights Time discounting Utility weights
Limitations	No co-morbidity Different sources have adopted varying approaches so not easily compared Weights may be affected by national/cultural perspectives, but latest studies are large pooled populations and researchers indicate that national variances are not significant	Weights derived from different countries may not be comparable so may not be transferrable Resource and time intensive to assess the appropriateness of differently derived utility weights

Although there are some important differences in the way that QALY and DALYs are used, these are not necessarily relevant for this study. The key issues for this study relate to some of the underlying assumptions, how the weights were derived, how each measure is calculated and what this means for their use directly in REACH SEAs, including through conversion to monetary equivalents. In addition, a further important consideration is the degree to which the weights that are available reflect the types of health outcomes of concern within REACH assessments.

The remainder of this section considers the following issues in terms of their implications for the transferability of utility or disability weights to the REACH context:

- Theoretical considerations
- Methodological considerations
- The relevance of the assessed health effects to REACH SEAs
- The correspondence of weights with willingness to pay estimates, and
- What would be required in order to apply the weights.

This discussion draws on the more detailed discussions provided in Sections 2-4 and the Annexes. The aim here has not been to rehearse the same arguments but to summarise the implications of those arguments in relation to REACH.

5.2 Theoretical considerations

Section 2.5 provided a summary of the key assumptions underlying QALYs and the methods used to derive the associated utility weights. These include: accounting for the individual's time preferences and their attitude towards risk. Further discussion on these is also provided in Annex 1 to this report.

Key conclusions from this discussion are as follows:

- The assumption of risk neutrality that is assumed as part of the derivation of utility weights has been found not to hold by a number of studies, with individuals exhibiting either risk aversion or risk seeking behaviour across diseases with different durations. For example, risk seeking attitude has been linked to the increasing duration of a disease (e.g. to cancer). Figures given in Annex 1 illustrate how an incorrect interpretation of an individual's risk attitude can affect QALY calculations, potentially leading to significant over or underestimates of the loss of utility. Combining these findings with the assumptions underlying the different methods indicates that if the individuals are risk averse, the utility elicited using the TTO method will be lower than the utility elicited using the standard gamble technique. Unfortunately, there is unlikely to be any information on whether or not the individuals' from who the weights were generated are risk averse or not.

- Several authors have proposed that the QALY model is still valid even if the risk neutrality assumption is violated, so long as the assumptions of utility independence (the utility derived from a particular health state does not depend on the utility attached to the health state which precedes or follows) and Constant Proportional Trade-Off (CPTO) hold. The discussion in Section 2.4 and Annex 1 examines these issues and highlights that research has found that CPTO in particular may not hold and that this has implications for the validity of weights generated using the time trade-off method, for example.

- Discounting of QALY values is undertaken to reflect the fact that individuals have positive time preferences for health. Although it has been suggested that the assumption of risk neutrality implies that there is no need to apply a discount rate in the QALY calculation, for example, discounting is nevertheless standard practice. The implications of discounting will be different, however, depending on the risk attitude of the individuals' whose weights are assumed. Nevertheless, most authors accept that discounting will be undertaken, although they note that care should be taken to ensure that QALYs are not discounted twice when being combined with monetary estimates.

- Discounting is also raised as an issue with DALYs, however, the literature tends to focus more on the equity of discounting future years of disability rather than on the theoretical validity of discounting.

REACH SEA Implications

What are the implications of the above points for the transfer of weights to a REACH SEA? Essentially, these are all issues that cannot readily be addressed as part of the transfer of weights to REACH, accept through the careful choice of studies. The fact that such concerns are raised within the academic literature does, however, indicate that uncertainties will arise from the selection of studies acting as the source for the weights and the analytical choices that were made within the original study. In this respect, it may be important for analysts to consider multiple weights derived through different elicitation methods, as part of a sensitivity analysis.

With respect to discounting, it is important to note that this is carried out as part of the calculation of the QALY or DALY (rather than in the elicitation of the weights) and whether or not to discount, and at what rate, is therefore the analyst's decision. However, it is also important to recognise that existing estimates of disease burdens, for example, are based on varying discounting assumptions. The original WHO study on the global burden of disease, as well as others such as the VoBD, applied discounting in the calculation of DALYs. In contrast, the Estonian study and the recent up-dated to the global burden of disease by Salomon et al. (2012) did not apply discounting. As a result, care should be taken in using any of the actual DALY values or in using the estimates of disease burdens as a context within a REACH SEA, for example, to inform on the potential total number of DALYs occurring due to chemicals exposure; in particular, the analyst should consider whether the application of discounting or lack thereof may affect the consistency between such sources and own estimates.

5.3 Methodological considerations

5.3.1 Introduction

As the theoretical basis and methodological decisions made when eliciting weights varies across different studies, so may the resulting weights, as highlighted by the discussion presented in Sections 2 and 4, and Annex 1. This was illustrated in Section 4 through a comparison of the weights generated by the different identified studies. In particular, the following methodological choices are important:

- Choice of valuation method
- Whose weights are elicited
- What time profile is assumed for a disease
- Potential to account for co-morbidity
- Assumptions regarding life expectancy and the use of age weighting, and
- Whether weights reflect a health or welfare loss.

It is also important to recognise a key difference between the studies that have elicited utility weights and those that have elicited disability weights. Because QALYs have been developed to assist in medical intervention decision making, there is not the same pressure for the underlying utility weights to be consistent across different diseases. Consistency across diseases is important to the DALY model, as it is aimed at providing an overall comparative summary of the burden of disease within a population and of the relative importance of different diseases within this overall burden.

5.3.2 Choice of valuation method

Several valuation methods have been described in Section 2 and the question of which method is best for eliciting utilities is not easily resolved. As noted in Section 2.4, it has been suggested that the SG, TTO and discrete choice experiments are the optimal methods for clinical decision making, and that MAUIs are better suited to health policy decisions relating to the allocation of resources. Similarly, Scuffham et al (2008) note that both NICE and the US Panel on Cost Effectiveness in Health and Medicine have indicated that the QALY weights should be derived using generic MAUIs to describe and value health states. The review of submissions to the Australian Pharmaceutical Benefits Scheme (Scuffham et al, 2008) also rated approaches involving the use of MAUI administered to patients currently experiencing the health states as more appropriate, together with health state valuation experiments (i.e. stated preference based elicitation) of either the general population (based on QOL data) or of a population of patients valuing their own health state.

Haagsma et al. (2014) have observed that historically most of the disability weights have been derived using ranking, interpolation, pairwise comparisons or the VAS – all of which lack a proper trade-off feature and hence do not fully assess people's preferences. As a result, most of the studies provide information only on the relative desirability of one health state compared to another. As discussed in Section 4, the implication is that they provide weights that reflect health trade-offs rather than welfare trade-offs.

What perspective and whose weights?

The question of whose weights should be sought is widely debated within the literature as discussed in Section 2. Historically, there has been a difference in the populations whose weights were sought when eliciting QALY versus DALY weights, with QALYs more often being sought from patient groups or the general population (although not always) and DALYs being derived through the use of expert panels. Panel composition is an important factor when weights are derived. The cultural background and level of expertise (patients, general public or medical experts) influences the outcome of the weights whereas the size determines the accuracy.

There is also an ongoing discussion as to whether weights derived from one country could be applied in a different country. As noted in Section 4, there is a lack of country-specific data or data generated in a manner that would allow generalizable results (e.g. through multi-national trials). In this respect, it is of note that UK and US preferences elicited using the EQ-5D have been found to differ substantially, as have UK and Japanese QALY weights. This difference raises uncertainties around the applicability of QALY weights to populations other than those from whom they were drawn. Similarly for DALYs, several key studies

(such as Haagsma et al., 2014 and Tengs and Wallace, 2000) report that national and cultural contexts can be important to the value of the weights assigned to individual diseases. This view is challenged, however, by others (such as Salomon et al., 2012) who note that the construct of health loss associated with different health states may be more universal than is the construct of welfare loss.

The level of medical expertise or familiarity with the disease also influences the outcome. Patients suffering from the selected health state tend to have a limited ability to make comparative judgments with other states due to information asymmetry. However, their assessment of the health loss is likely to be a more informed assessment than that of a person who has no experience with the health state. Medical experts or public health professionals are also used in panels on the basis that they have knowledge of a diverse set of health states and are therefore able to make comparative judgments. Where resource allocation is the focus, though, some analysts have argued that the general public is the most appropriate population on the basis that, in a democratic society, the views of the general public are relevant in comparative assessments that inform public policy (Salomon et al., 2012; Haagsma et al., 2014). Contextual differences may be stronger amongst lay people compared to health professionals according to Jelsma et al (2000). As noted earlier, though, studies which ask experts and the public have found significant differences in the resulting weights (Haagsma et al., 2014).

Most recently, there is no longer such a strict division between the panels/populations and methods used to generate QALY weights and DALY weights. The more recent DALY studies have elicited weights through surveys of the general population, in line with recent recommendations that disability weights based on societal preferences be developed since burden of disease studies are indeed used as tools for guiding decision making on resource allocation at the population level (Haagsma et al, 2014).

However, it is also important to note that:

- Most studies do not account for the way patients might adapt to a particular health state, a point made by Salmon et al. (2012). For example, when a person is first diagnosed with diabetes, they may value the severity of the condition highly yet, with time, they may become accustomed to the condition and thus value its severity less. This argument follows a similar line to the concepts of the time profile of disease and how time preferences and risk attitude might change with time.

- Most studies distinguish between those who are born with a particular condition and those who get a condition later in life, for example, blindness. Similar to the above point, those born with a condition will be used to the condition and not necessarily consider it a disability; they are therefore likely to value its severity lower than a person who becomes blind later in life.

More generally, there should probably be a preference for choosing weights that stem from more recent studies than from older studies. This is because of changes over time in both the population's and medical practitioners understanding of different diseases and how they can be treated. As a result, newer studies are more likely to reflect current

understanding of the disability associated with different health states and hence the impacts of this on a person's health. This is the reason that the 1997 Dutch study, on which many burden of disease studies are based, is not considered in detail in this report.

REACH Implications

What is the importance of the above arguments within the context of this study? The fact that, historically, different studies have drawn their weights from medics, patients, and more general populations is one of the reasons that there is such variability across studies in the resulting weights. With respect to their use in REACH SEAs, weights elicited from the general population or patients should be given priority. The general population is relevant to establishing social preferences and hence resource allocation. Patients are relevant as they have the detailed knowledge of the disease and its impacts on their life needed to indicate what impact it has on their welfare.

In practice, if SEA analysts are to try and choose weights from the available sources, they may be forced to draw on weights developed through panels due to a lack of alternatives. In such cases, it may be important that the characteristics of the panel that generated the weights is made clear, to highlight any uncertainties that this could introduce.

5.3.3 Time profile

When eliciting weights or generating them within a MAUI system, the health state of interest can be described in general terms or disease specific terms. For example, generic descriptions based on health-related quality of life descriptions, such as the EQ-5D or the Health Utilities Indexes (HUI), can be used to depict the functional health independent of the actual underlying condition. However, they may not be sensitive to specific disease contexts and it can be difficult to use generic descriptions for measuring the impacts of an acute disease rather than the chronic condition.

A disease specific description depicts the clinical description of the disease with the specific health effects of the condition. These disease specific descriptions are more sensitive for the detection and quantification of small changes. However, the amount of information provided by disease specific descriptions might lead to cognitive overload for the population from whom the weights are to be elicited. In addition, when using disease specific description, the manner of depicting the health state affects preferences for them and therefore can result in information bias.

As discussed in Section 2 and 4, most DALY studies historically have applied a period profile approach to the health state valuation, with the underlying assumption being that that the value of the health state is not affected by the duration of the health state. This time-profile is weak for complex diseases with changing severity over time and conditions with a short duration or periodic health effects (such as asthma attacks). Instead, an annual time-profile is suggested in which "the course of the health state – the disability profile which may reflect a few weeks of illness followed by the remainder of the year in good health – is described over a period of 1 year time".

The more recent GBD studies and the European study appear to have adopted a combined approach. The Estonian study is helpful in this respect as it provides the time duration assumed for each weight within the same summary table.

REACH Implications

Analysts will need to examine the health state descriptions associated with the disability weights that they propose to adopt to ensure that it is appropriate to the disease in question before adopting it for use in a REACH SEA; this will require accessing the full annexes available on The Lancet's website[29] (Vos et al., 2015).

This same issue may arise with QALYs, although analysts should be able to find information on the disease description and the disease duration/profile used when eliciting the utility weights from the original academic papers.

5.3.4 Co-morbidity

Co-morbidity is a phenomenon that is growing in the developed world in particular and has been defined as the presence of 'a distinct additional clinical entity' (Jones, 2010). There is an increasing recognition that patients suffer from multiple diseases, with estimates that, for example, in the Dutch population comorbidity is present in over 25% of the population with more than four conditions in about 55% of the elderly patients (Jones, 2010).

Co-morbidity is not always directly related to the condition that the patient is primarily suffering from but can occur under different circumstances. Three types of co-morbidity have been identified: "unrelated (the most common case of two conditions happening by chance on the same individual), related through common risk factors (pathophysiology is grossly unrelated, in particular the pathway to symptoms), and directly related where one condition can be regarded as natural consequences or parallel manifestation of the other condition" (Haagsma, 2010). An example of a typical condition that usually comes together with other diseases is Chronic Obstructive Pulmonary Disease (COPD), which is linked to coronary heart disease, diabetes mellitus or osteoporosis (Chatila et al., 2008).

As discussed in Section 2, the DALY concept does not always include consideration of multiple diseases in one individual or a population, or consideration of a disease that arises as the sequelae of an existing condition (Gold MR et al, 2002). The criticisms linked to the non-consideration of co-morbidity have been described as resulting in DALYs overestimating the impact that treatment can have because it assumes that once the cause of a disease is eliminated all successfully treated patients return to perfect health (Fox-Rushby, 2002).

In order to overcome this issue, some studies have adjusted the disability weights assumed for a disease using one of three approaches. The first approach has been described as the 'maximum limit (maxlimit) approach' which only counts the condition which has the highest overall disability weight. This "approach assumes that a comorbid disease does not affect

[29] http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2815%2960692-4/fulltext

the disability of a patient with a primary disease, unless the comorbid disease – in general terms - exceeds the disability of the former" (Haagsma, 2010). The second approach is the additive approach and, as the name indicates, adds up the additional or comorbid implications on the patient. The third approach is the multiplicative approach which "assumes that a comorbid disease does increase the utility loss of a patient, though it is less than the sum of the utility loss of both diseases independently" (Haagsma, 2010). The more recent GBD studies addressed this issue through the adoption of modelling methods to predict the number of people within the population who would have co-morbidities based on prevalence data.

REACH Implications

In contrast to the DALY model, co-morbidities can be taken into account when calculating QALYs by adopting an additive approach as described above. However, if consideration of co-morbidities is important then, as discussed in Section 4, the utility weights underlying the calculation of the QALYs should be taken from a single study or at least a consistent set of studies in terms of the methodologies that were used to elicit the weights. In this respect the UK "catalogue of weights" is interesting in that it provides figures for the marginal disutility associated with the onset of a disease, taking into account different starting typical health states. This approach may be highly attractive in the context of REACH, as it allows consideration of comorbidities (which can be relevant to chemical exposure related disease cases – something that cannot be done with disability weights) and calculation over the analyst's choice of time periods.

There is no reason that an additive type of approach could not also be adopted with the use of disability weights and the calculations of DALYs for a REACH SEA, although it would be important for the analysts to note that their approach varied from the standard methods used in the calculation of national disease burdens.

However, it is not clear how significant an issue co-morbidity is likely to be in the context of REACH SEAs. In most restriction assessments to date, co-morbidities have not been taken into account, with the focus being on the main health endpoint identified as giving rise to risks at the EU-wide level. Similarly, SEAs prepared to support Authorisations will focus on the risks identified as the reason for prioritisation.

5.3.5 Life expectancy and age weighting

Calculation of both QALYs and DALYs requires making assumptions on the standard life expectancy of an individual. Within the DALY literature this is identified as a key issue, mainly because the larger studies have assumed a standard life expectancy for men and another for women when developing estimates of disease burden at the global level, regardless of whether the country is developed or developing. When undertaking studies of national disease burdens, analysts have generally adopted life expectancies that are age and gender specific as appropriate to the national population.

Related to this is the potential for age weighting to be incorporated into the calculation of DALYs. It is argued by several authors that the process of age-weighting within the calculation of DALYs means that although they may be appropriate for measuring the

burden of disease, they are less appropriate for allocating resources by DALY minimisation (see for example Anand and Hanson, 1996). This may, in part, be one of the reasons for the inclusion of age-weighting in DALYs to increasingly become a more discretionary feature. It has not been part of the methodology applied in either of the latest updates of the GBD weights (Salomon et al, 2012), or in the European BoD study. Furthermore, age weighting is not inherent to the elicitation of weights and thus does not affect their use in other contexts.

REACH Implications

Within the context of REACH, analysts will need to define the appropriate life expectancy for the population at risk, taking into account gender-specific differences as appropriate. Perhaps a key issue will be the differences in life expectancy that exist across EU countries, as illustrated by Table 5-2, which gives average life expectancy figures. However, it is expected that most SEAs would adopt an EU average, unless there is concern for particular regions.

Table 5-2: Life Expectancy of some European countries

Country	Life Expectancy
Sweden	82
Germany	81
Italy	83
Greece	81
Bulgaria	75
Romania	74
United Kingdom	81
Estonia	77
Source: WHO (2015)	

5.3.6 Measurement of health versus welfare loss

An issue raised by practitioners is that some of the studies are focused on health loss rather than welfare loss; in other words, their weights were elicited so as to reflect the health ipact alone and to include no measure of the broader intangible effects of the disease on an individual. This is the case for the most recent burden of disease studies; Salomon et al. (2012) note that, in eliciting disability weights for the 2010 Global Burden of Disease update, they specifically focused on health loss as opposed to welfare loss (p.2139).

This failure to capture welfare loss may be important, as it may significantly underestimate the impact of a disease on the quality of life for the individual affected. This is illustrated nicely in an example from Brock (1998): "in a setting in which most labour is manual, limitations in physical functioning will have greater importance than they would in a setting in which most individuals are engaged in non-physical, knowledge-based occupations, where certain cognitive disabilities are of great importance" (p.73). Fox-Rushby further supports this argument, noting that "the burden of disease can only be judged in the context of the culture of the individual" (2002, p.85). She refers in her analysis to Sayers and Fliedner who pointed out the impact on the environment of the individual as contributing to the real burden of disease: "Thus the impact on disability depends on the

economic, family and social circumstances of the individual; this must influence the true "burden" that the disability represents for the individual, the family, and the community and the appropriate level of additional support that should be provided" (Sayers and Fliedner, 1997).

REACH Implications

This failure to consider welfare loss is in part due to the manner in which the weights are elicited, for example, through the use of paired comparisons rather than the more sophisticated utility based elicitation methods. It suggests that the use of weights elicited to reflect only health loss is likely to underestimate the true economic impacts of a health effect. It also suggests that there may be a mismatch between such weights and economic measures of health impact based on willingness to pay, as individuals should implicitly take into account their economic, family and social circumstances when responding to a willingness to pay question.

This potential for underestimating impact should be noted in any REACH SEA.

5.4 Linking QALYs and DALYs to economic measures of impact

5.4.1 Cost-utility analysis and cost-benefit analysis

The aim of the health impact assessment (HIA) under REACH is to assess the impacts on human health including morbidity and mortality effects from exposure to substances and the impacts of this on social welfare. Such impacts cover health related welfare effects (intangible costs), lost production due to workers' sickness (indirect costs) and health care costs (direct costs). The morbidity and mortality costs are estimated for the various identified relevant health effects. Within this, there is the potential that the QALY and DALY concepts may be relevant as another method to estimate the health related welfare loss.

As noted above, usually only one health endpoint is accounted for in a quantitative assessment of impacts in REACH SEAs. In such assessments, a QALY/DALY based measure of health effects could be used to provide a measure of the intangible component of the cost of illness. For example, the number of QALYs gained or of DALYs avoided could be divided by the costs of a proposed Restriction to provide a cost-utility assessment. Ratios across different CUAs could be compared to assess whether a particular Restriction proposal appear to reflect good value for money (note that the original purposes of the CEA Registry now held by Tufts University was developed for this purpose – to provide ratios of different interventions that could be used to assess whether or not a particular action appeared proportionate) [30].

The second approach would be to combine information on the QALYs gained or DALYs avoided with other components of the cost of illness, such as direct health care costs and

[30] It is important to note that the QALY/DALY does not necessarily need to be converted into a monetary value; it can still be used in a CUA or CEA in its regular unit.

indirect costs due to the lost production of sick workers, as part of a cost-benefit analysis. In this case, for the QALY/DALY measure to be of additional value, it could be converted into a monetary value. Such a conversion has been carried out for other decision making purposes and there is a growing literature on the willingness to pay for a QALY.

Ryen and Svensson (2014) recently reviewed the empirical literature on the willingness to pay for a QALY. They note that there are two decision rules when using QALYs as an outcome measure in cost-effectiveness analyses (CEAs) (Weinstein and Zeckhauser, 1973): (i) choose interventions in ascending order of cost per QALY until the budget is exhausted and (ii) select interventions with a cost per QALY less than or equal to a specified threshold value. These thresholds can be determined on different theoretical perspectives; the opportunity cost approach or the willingness to pay (WTP) approach. The latter is linked to welfare economics where the cost of an intervention is compared with a societal WTP for a QALY (sometimes referred to as WTP-Q).

5.4.2 Placing a monetary value on a QALY or DALY

While the QALY concept is a well-established approach for evaluating the cost-effectiveness of a health intervention, it is only used by a sub-set of EU countries to assist in the analysis of whether medical products or interventions offer value for money. Within the UK, the National Institute of Health and Care Excellence (NICE) advises the National Health Service on appropriate treatments. NICE has set a threshold value of how much a treatment can cost, and which is calculated using the QALY concept. For example, if a treatment costs more than £20,000-30,000 per QALY, then it would not be considered cost effective and hence is not advised as a treatment for the patient (NICE, 2010). There have been several cases in the media where a treatment option has not been advised because it was considered to be too expensive (PharmaTimes, 2010; The Guardian, 2014; PMLiVE, 2013). This has opened a debate over whether or not the current threshold is too low.

In order to address the debate about the empirical basis for the cost-per-QALY threshold and whether QALYs gained by different beneficiaries of health care should be weighted equally, the Social Value of a QALY (SVQ) project was commissioned (Donaldson et al, 2011). It addressed the issue of the threshold in two ways: 1) by combining, via models, the current UK Value of a Prevented Fatality (used in transport policy) with data on fatality age, life expectancy and age-related quality of life; and 2) via a survey designed to test the feasibility of combining respondents' answers to WTP and health state utility questions to arrive at values of a QALY. The results yielded values of £10,000 - £70,000 per QALY. Via survey research, most methods of aggregating the data resulted in values of a QALY of between £18,000 - £40,000 (Donaldson et al, 2011).

The second set of figures falls within the range used since 2004 by NICE (2004), which was based on expert judgement and is a reflection of past recommendations. Overall, though, the study concludes that it is not clear whether it is feasible to estimate a constant monetary value of a QALY. In addition, they conclude that based on the population average values derived from the survey research, there is no compelling evidence for moving the current NICE threshold either up or down.

In another study, Ryen and Svensson (2014) also examine the economic value of a QALY. They note that two empirical approaches have been used to estimate the WTP-Q. A first and direct approach is to ask respondents about their WTP for small health increases or QALYs using stated preference (SP) techniques such as discrete choice experiments or contingent valuation. The WTP estimates can subsequently be used to estimate the WTP for a gain in a full QALY. As with the derivation of a utility weight, WTP-Q studies can be substantially different methodologically even when using the same SP technique. These differences include whether an individual or societal perspective is given; a general or specific population is used; the type of good (i.e. whether quality of life or life extension is being valued); and the subject of valuation (general health or specific conditions).

The second approach is to use the monetary value of preventing fatalities (the value of a statistical life or the value of a life year lost (which is often calculated starting from the value of a statistical life)), on which there is a substantial empirical literature, in order to implicitly derive the WTP-Q assuming a certain life expectancy (LE) and discount rate for the sample from which the value of life is derived.

The analysis by Ryen and Svensson considers 24 papers with 383 WTP-Q estimates, based on empirical papers where the aim is to estimate the WTP-Q and original estimates are used. Most studies are European (14) or from the USA (5). The overall mean and median WTP-Q estimates are €118,839 and €24,226, respectively (2010 Euros). The interval of estimates ranges from less than €1000 to €4,800,000. However, 80% of all estimates are below €75,000. Estimates based on direct stated preferences approaches are generally lower compared with estimates from value of statistical life (VSL)-conversion studies. By definition, VSL conversions value changes in the length of life, but there is also a difference in SP studies based on whether quality of life (QoL) changes are being valued solely, or if length of life is also part of the valuation exercise. Regression analysis indicates that WTP-Q tends to be higher if a risk of premature death is included in the valuation scenario than if QoL changes are being valued.

Ryen and Svensson (2014) also tested if the WTP is proportional to the QALY change, which is required for a constant WTP-Q. A regression analysis was conducted for those estimates where the size of the QALY gain is explicitly stated in the article and based on pure QoL changes (161 estimates from nine papers). The authors state that if the WTP is proportional to the QALY change, which is required for a constant WTP-Q, the coefficient of ΔQALY should be equal to zero. They found a negative and statistically significant coefficient estimate (-1.028). Based on the results, they reject the null hypothesis that the WTP-Q is constant across different QALY changes. A negative coefficient estimate implies that larger QALY gains give lower WTP-Q estimates (proportionally). The magnitude of this difference was found to have a substantial economic effect (for each increase in the unit of QALY change, the WTP-Q is 64% smaller). This finding could suggest either scope insensitivity in the WTP estimates, or diminishing marginal returns for each additional improvement in quality of life.

One of the studies included in the review by Ryen and Svensson that may be worth further attention is the EuroVAQ project (2010)[31], a large pan-European project with the aim of developing robust methods to estimate WTP-Q in 10 European countries. The report describes three studies; the first estimates a WTP-based monetary value of a QALY from existing contingent valuation studies of the value of prevented fatalities and serious injuries; the second estimates a WTP-based value of a QALY through survey-based research. The third study reports views on health care priority setting amongst the public and decision makers across 10 European countries.

Of particular interest in this report is the second study on the estimation of a WTP-based value of a QALY through survey research. In this study, a WTP-based value of a QALY was derived through a "chained" and "direct" approach. The chained approach is the most common approach adopted by other previous research and is described as follows:

> *´The chained approach set out to build upon previous research that had attempted to estimate WTP per QALY by breaking the exercise down into two distinct components. First, respondents would be asked to complete a utility assessment exercise in order that their utility value (between 0 and 1) for a given health state could be ascertained. Next they would be asked their WTP to avoid a given duration/risk of that health state. Combining the respondent's answers to both components then allows that respondent's WTP per QALY gained to be estimated (essentially by "multiplying up" their WTP for a known fraction of a QALY into one whole QALY). For example, if we know that a respondent is willing to pay £1000 to avoid one year with certainty in a health state with utility value of 0.95 (i.e. a loss of 0.05 or 1/20th of a QALY), we can estimate their WTP per QALY to be 20 * £1000 =£20,000 per QALY gained (assuming linearity). The basic principle behind the chained approach is that the "health losses" being considered in the WTP component of the exercise are not too large such that stated WTP values would likely be subject to "budget constraints".'*

The direct approach is described as:

> *The direct questionnaire tested the notion of presenting health gains "directly" using a simple graphical and textual description. Most of the health gains presented were of one QALY, avoiding the need to "multiply up" WTP values for smaller health gains to generate a value for one QALY. Hence all of the gains were for durations of at least one year and most involved no risk or probabilities. However, we sought to build in an "overlap" with the chained approach and included questions offering smaller health gains of similar magnitude (0.05 to 0.25 QALYs) to those in the chained questionnaire. Three of these questions presented the health gain using a risk format where respondents paid to avoid a fixed probability (5 or 10%) of a loss of one QALY.*

In addition, two different types of WTP question were used: "risk variant" and "time variant" questions. Risk variant WTP questions asked respondents about their WTP to avoid some risk of a health state, with the risk allowed to vary across respondents in order to keep the QALY gain constant across respondents. In the time variant WTP questions,

[31] http://research.ncl.ac.uk/eurovaq/EuroVaQ_Final_Publishable_Report_and_Appendices.pdf.

respondents were asked about their WTP to avoid some duration of a given health state and the duration varied in order to keep the QALY gain constant across respondents.

The survey was conducted in 10 countries (Netherlands, UK, France, Spain, Sweden, Norway, Denmark, Poland, Palestine and Hungary) and in total 39,922 people completed the survey (overall response rate of about 60%). Responses were converted to international dollars at Purchasing Power Parity rates to allow aggregation across all ten countries prior to calculating the mean and median. With the chained approach, the overall range of mean WTP per QALY is $18,247 to $77,323. The overall range of median WTP per QALY is $3,723 to $8,211. The authors suggest that many readers will consider the median WTP per QALY estimates to be conservative and they are certainly lower than the commonly used "thresholds". The direct approach yielded means for the questions offering one QALY range from $4,854 to $20,719 across all 10 countries. Smaller health gains generated higher means per QALY ranging from $27,977 to $82,347. The raw mean values are heavily influenced by a few respondents giving very high values. Trimming generally had a larger impact on the questions offering smaller health gains, perhaps because erroneous responses are scaled up to generate values per QALY for these questions.

The authors indicate that there were significant differences in mean WTP values across the 10 countries, with respondents in the UK, France and the Netherlands consistently giving lower mean WTP values. The Danes generally gave the highest values closely followed by the Spanish.

The study by Ryen and Svensson (2014) shows that WTP-Q estimates vary widely and are dependent on several methodological factors. The EuroVAQ report gives a range for WTP-Q estimates based on a large European sample and might be useful in the context of REACH SEAs, although the range of the estimates is wide. Overall, a "common" societal value for one QALY may not be appropriate, as the regression analysis by Ryen and Svensson suggests that the hypothesis of a constant WTP-Q may not hold. Instead, the WTP-Q studies have found that the magnitude of the QALY gain seems to determine the height of the WTP-Q, with small changes leading to a higher WTP-Q. As noted earlier, this may be due to either diminishing marginal returns to improvements in health or due to scope insensitivity within the WTP studies themselves.

These findings may also be relevant to DALYs. They indicate the potential importance of sensitivity analysis in any assessment that moves from use of just the calculated QALY or DALY change to the monetary valuation of that change. It may be important to consider more than one valuation for that change, where this is backed by explanation as to the choice of values for use in the analysis, given the lack of an empirical basis for assuming a constant WTP for all levels of change in QALYs (or presumably DALYs).

Charles University Study

As noted in Section 1, the background context to this study is work carried out by the Environment Centre at Charles University in Prague and VU University Amsterdam. Researchers from these organisations conducted a stated-preference study to estimate values for the WTP to avoid selected adverse health outcomes. The intention of the study is

that these values could be used by ECHA and other bodies in conducting REACH SEAs and other HIAs.

The study elicited WTP values for a range of health outcomes and included within these were values for chronic severe dermatitis, chronic kidney disease, acute kidney disease and acute dermatitis. The study also derived QALY weights and losses for the four illnesses using a standardised VAS method. The authors indicate that there are some issues with the results from this exercise as some respondents unexpectedly reported preferences for the illnesses compared to their current level of health. They therefore calculated annual QALY losses for different groupings of respondents to account for this to derive estimates of mean annual QALY losses. They then recalculated WTP per QALY for the four illnesses; the results of this exercise are presented in Table 5-3.

The authors also acknowledge that "Such a spread of WTP per QALY values given the varying size of health gain in terms of quality of life and duration is not uncommon in existing studies". As noted earlier this may be due to either scope insensitivity within the results or due to decreasing marginal returns associated with increases in health state.

Table 5-3: WTP values taken from Charles University Study

Health Endpoint	Mean WTP to avoid a case of the health outcome (euros)	QALY loss	WTP-Q (euros)
Chronic severe dermatitis	982	0.381	4,016
Chronic kidney disease	2,568	0.558	4,656
Acute kidney disease	503	0.028	17,500
Acute dermatitis	225	0.008	25,028

Source: Charles University in Prague and VU University in Amsterdam (2014): Stated preference study to examine the economic value of benefits of avoiding selected adverse human health outcomes due to exposure to chemicals in the European Union, Part I: sensitization & dose toxicity

5.5 Relevance of health effects to REACH SEAs

As described earlier, the QALY and DALY concepts can be used to measure the impact on an individual's quality of life of a specified "health profile" (i.e. a time path of health states ending in death) in terms of an equally valuable length of time lived in full health. They are calculated using the weights that have been the focus of in this study (utility or disability weight) for the particular health state of interest, using additional information on age of onset of the health state and expected age of death.

In the literature QALY and DALY weights are only derived for noticeable adverse health effects. Therefore, observed adverse effects in animal toxicological data[32] need to be

[32] Adverse effect: Any change in toxicological endpoint (often observed in animal studies) that is considered by toxicologists and/or risk assessors to be undesirable for human health. According to the World Health Organization (WHO): "Adverse Effect: Change in the morphology, physiology, growth, development, reproduction, or life span of an organism, system, or (sub) population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences"

translated into matching clinical effects in humans. This extrapolation from the risk assessment to noticeable human health effects in the health impact assessment is critical for the use of QALYs and DALYs. The extrapolation into assumed clinical human health effects is usually done qualitatively in the health impact assessment (HIA) and most of the time results in a worst-case estimation of the health impact due to the exposure.

Section 3 of this report set out the 41 health effects which were identified as being relevant to chemicals identified as being SVHCs. Table 3-2 in Section 3 (see also Annex 2) identifies the health effects relevant to SVHC for REACH health impact assessments. These are compared in Table 5-4 to the health effect descriptions used in the elicitation of the QALY and DALY weights for the 36 health effects for which we were able to find weights.

Note that due to time limitations, we were constrained as to the number of original sources that could be reviewed concerning the availability of utility and disability weights, and from which specific health effect descriptions could be extracted. The table should thus be considered to give some general insight on the various health effect descriptions that have been assessed and are available rather than providing a complete overview (although the key sources of disability weights in particular have been used).

Carcinogenicity and Mutagenicity

For the health effects attributable to carcinogenicity and mutagenicity, it appears that weights are available that discriminate between the various forms of cancer, as well as the various stages of the specific cancers. DALY disability weights are available that differentiate generally between primary, disseminated and terminal stages of the specific cancers. For lung cancer an extra distinction is made between operable or not and for prostate cancer hormone refractory cancer is included. Skin cancer is subdivided into melanomas and non-melanoma skin cancers (basal cell carcinoma). For hematopoietic cancer, weights are found for various forms of leukemia. In the case of leukemia, a differentiation is made between acute and chronic leukemia. In the case of lymphatic cancer, the health effect description only covers non-Hodgkin's lymphoma.

The differentiation into various stages of cancer is less profound with the QALY utility weights. For lung cancer, weights were found for cancer of bronchus, small cell lung cancer and non-small cell lung cancer. With other cancer types such as liver, kidney, urinary bladder, breast and prostate cancer no further description on the stage or severity of the cancer could be easily retrieved. Skin cancer is described as melanomas and subdivided into stage I or stage IV (both stable). Also for stomach cancer a division is made between stage I&II and stage IV. Hematopoietic cancer is described as leukemia cancer and a utility weight was also derived for a progressed health state for chronic lymphocytic leukemia. The lymphatic cancer is described as non-Hodgkin's lymphoma. A distinction is made between progression free follicular non-Hodgkin's lymphoma and a progressive form of follicular non-Hodgkin's lymphoma. For pancreatic cancer a weight was also found for metastatic pancreatic cancer.

Thus, it appears that for all cancers, disability weights are available relating to the various stages of cancer, while for some cancers utility weights relating to several stages of cancers are also available. Furthermore, with some cancers (lung, leukemia) weights are derived for

different forms of the cancer. The health effects endpoints attributable to carcinogenicity and mutagenicity identified in our initial work state only the general description of the cancer and do not give any information on the severity (stage) or specific type of cancer. It is questionable whether the underlying toxicological data will give any information on the specific type of cancer or the stage at which they will be discovered in humans. The health effect descriptions from the weight derivation studies are thus probably more detailed than the information available from the toxicological data.

In the HIA, assumptions have to be made regarding the specific disease profile of the cancer to link it to a specific weight. The human disease conditions identified for the SVHCs assessed by this study (see Section 3 and Annex 2) do not directly match the various descriptions of the identified endpoints in the derivation of QALY and DALY weights. A better match may be possible by looking in more detail at the toxicological data or making assumptions to match the most appropriate endpoints for which weights are derived. Overall, it therefore seems that the health effects attributed to carcinogenicity match the health states underlying the weights available in the literature (NB. this does not indicate that these weights are accurately derived and transferable into the REACH context).

Reproductive toxicity

The health effects from chemical exposures associated with reproductive toxicity are subdivided into impaired or reduced fertility, reduced fetal growth and developmental anomalies/abnormalities. For impaired or reduced fertility, only the QALY utility weights make a distinction between male and female infertility. Furthermore, the weights assigned to infertility reflect full infertility whereas the toxicological data usually only indicate an impaired or reduced fertility. The health effect description used in the weight derivation for infertility therefore overestimates the health effects related to the chemicals. This overestimation could be seen as a worst case estimate as the relationship between the effects observed in the toxicological studies and the human health correlate is always uncertain. It is also questionable whether the disruption of ovarian cycle is matched by the descriptions used in the weight derivation for menorrhagia (heavy menstrual bleeding) and menstrual disorders. The term menstrual disorder can relate to irregularity of menstruation or excess bleeding. Menorrhagia relates to the amount of blood loss or the length of the menstrual period and does not have a clear relationship with a disruption of the ovarian cycle as the menstruation is at regular intervals. Thus, when using QALY or DALY weights for the disruption of the ovarian cycle, the full study details of the weight derivation should be examined to determine whether the health effect description in the weight derivation matches a disruption of the ovarian cycle instead of other menstrual disorders.

The reduced fetal growth health effect descriptions seems to correlate quite well with the health effect description in the QALY utility weight study where parents of very low birth weight infants were interviewed immediately after birth. However, one has to make assumptions on the perspective to take regarding reduced fetal growth health. The perspective from the QALY utility weight reflects the impact on the infant as perceived by the parents. The DALY disability weight is taken from the infants' perspective for mild disabilities in later life due to the low birth weight and reflects a permanent disability weight for mild to moderate early acquired hearing loss. In the HIA, assumptions have to be made

regarding the consequences of a low birth weight and whether the impact of low birth weight on the infant, as perceived by its parents, is truly a validation of a health state.

Developmental abnormalities

The health effects concerning developmental anomalies/abnormalities vary widely depending on the underlying specific effects and whether the health effects descriptions match well. When there are clear toxicological effects that directly link to the clinical effects that can be seen with newborns, such as the spina bifida and cleft palate, the descriptions match well. For skeletal abnormalities (including abnormalities of limbs), the health effects description linked to SVHCs is too general and can include various different clinical outcomes with subsequent different impacts (such as a club foot or hip dislocation). No weights were identified for limb abnormalities. A disability weight for urinogenital abnormalities is taken from the weight for renal failure and therefore does not reliably represent urinogenital abnormalities. For QALYs, the utility weight is derived for abnormal aspect scrotum congenital UDT (undescendent testis) and is more relevant to cryptorchidism than for urinogenital abnormalities. This illustrates the difficulty of finding weights for endpoints specific to REACH SEAs.

There are different DALY disability weights found for various stages and of renal abnormalities/failure. Only for the end-stage renal disease was a QALY utility weight found. The interpretation of the renal abnormalities seen in toxicological studies to a disease profile for the health impact assessment is challenging and various assumptions probably need to be made.

For developmental neuro-impairment/neuro-logical disorders (IQ) weights are available for various gradations of retardation or intellectual disability. Comparing the health effect descriptions, it seems that they match fairly well for the general description of the health effect. As for the renal abnormalities, the challenge will be in the translation between the observed effects in the toxicological studies to the degree of clinical health effect used in the HIA. Alternatively, one can also argue that these health effect descriptions do not match well at all; the developmental neuro-impairment/ neurological disorders (IQ) as health effects endpoints state only the general description of the potential retardation and do not give any information on the severity (as is available for the different stages of cancers). It is questionable whether the underlying toxicological data will give any information on the severity that will be introduced in humans. The health effect descriptions from the weight derivation studies are thus probably more detailed than the information available from the toxicological data. In the HIA, assumptions have to be made regarding the specific disease profile of the neurological disorder to link it to a specific weight. A better match could be possible by looking in more detail to the toxicological data or make assumptions to match the most appropriate diseases/endpoints that is used in the derivation of QALY and/or DALY weights.

We would note though that the weights available from the literature imply a significant level of IQ loss and do not correspond to the low (i.e. fractional) levels of IQ loss that have been the focus of recent Restriction SEAs. In addition, there was debate (and dissent) over the appropriateness of including intellectual disability (together with vision loss and a few other endpoints) within the GBD studies and it appears it was done so in part due to the fact that

impacts on IQ may result from common diseases in some countries such as malaria, tetanus, etc. (Solomon et al., 2012).

As for the renal abnormalities, then, the challenge will be in the translation between the observed effects in the toxicological studies to the degree of clinical health effect used in the health impact assessment.

For the cardiovascular reproductive health effects, the corresponding health descriptions of the various weights reflect coronary heart diseases, acute myocardial infarction and stroke varying in severity in adults, rather than specifically relating to congenital heart defects. It is unlikely that the impact of congenital heart defects in newborns are fully reflected by these three health effects and potentially other effects may occur as well depending on the type of cardiovascular abnormality. Therefore, the weight derived for these three health effects may not fully cover developmental cardiovascular abnormalities. The health endpoint description of developmental cardiovascular anomalies/abnormalities from our review of REACH dossiers and the health effects descriptions as found in the collation of weights can both be regarded as too generic for a match. More detailed analysis of the specific type of cardiovascular abnormality observed in the toxicological study may be needed to link the particular effect to a health effect description used in a utility or disability weight study which is more specific to congenital heart defects in newborns.

Other effects

No specific weight was found for high cholesterol level alone. If obesity would be the health effects used in the health impact assessment, then obesity utility and disability weights are available but it is not clear if these are relevant in terms of the health effect descriptions. The link between high cholesterol level and obesity is one that would have to be made within the health impact assessment. Yet one might argue that cholesterol is not suitably assessed using the QALY or DALY model as the symptoms specific to high cholesterol do not, themselves, impair the individual's everyday quality of life[33]. A similar argument can be put forward for a loss of IQ: the effects of the loss of IQ associated with chemical exposure may not have a real impact on one's health or quality of life and so very small changes in IQ could not be assessed using this method. This is discussed further below.

The other effects identified as relevant to SVHCs are nephritic and nephrotic syndrome; respiratory (tract) irritation; allergic contact dermatitis and osteoporosis (bone fracture). For the nephritic and nephrotic syndrome only DALY disability weights could be found for end stage renal failure (with various treatment forms). It is questionable that exposure to chemicals will always lead to end stage renal failure. However, this overestimation can be seen as a worst case estimate as the relationship between the effects observed in the toxicological studies and the human health correlate is always uncertain. Respiratory (tract) irritation as such is not used as a health effect description in the QALY and DALY weight

[33] For example, the scales used in the derivation of weights using the EQ-5D or the HUI2 have no basis for assigning utility decrements linked to cholesterol alone, as they relate to mobility, self-care, ability to undertake usual activities (or senses/cognitive ability), pain/discomfort, and anxiety or depression (emotion) and fertility in the case of the HUI2.

derivation studies. Instead, various degrees of COPD or asthma are used. Depending on the specific description for the various form of COPD and asthma, the weight might be used for respiratory (tract) irritation. Specific attention should be paid to the probable reversible and episodic character of the respiratory (tract) irritation in the case of exposure to chemicals.

Allergic contact dermatitis is for the QALY utility weight described as atopic eczema or severe inflammatory skin condition. For the DALY disability weight, no further details are available. As with respiratory (tract) irritation, caution should be taken when correlating the different health descriptions with regard to the episodic character, as the rash is of reversible character and is only caused when in contact with the chemical. The frequency of the rash occurring is therefore correlated with the frequency of exposure to the chemical, and it is expected that the frequency of the rash will affect the value of the associated utility or disability weights. Whether the health descriptions match is dependent on the expected frequency of effects in the HIA and the frequency of effects used in the weight derivation studies. An additional complicating factor is that with the chemicals that can cause allergic contact dermatitis, the severity of the rash can differ depending on the chemical.

Various QALY utility weights are derived for osteoporosis (bone fracture), depending on the location of the fracture (wrist, clinical vertebral or other) all for women. A disability weight is available for diagnosed cases of osteoporosis. It seems the health effect descriptions match quite well.

General conclusions

Overall, the match between the identified health effects relevant to SVHCs and the health effect descriptions of the utility and disability weights varies considerably across health effects. For cancers, some of the more specific reproductive abnormalities and some of the other health effects the match might be quite suitable. Other health effects, for instance on disruption of ovarian cycles, need more detailed information from the health effect description in the weight derivation study to conclude whether the described health effects match the relevant health effects found in the toxicological data. In all situations, a case by case approach is warranted to obtain both the full substance specific toxicological data as the full health effect description from the weight derivation to assess whether the heath effect descriptions truly match.

There are numerous challenges to translate the effects observed in toxicological studies into a disease profile suitable for the use of DALYs or QALYS in the HIA. The disease profile chosen in the HIA will dictate the type of specific health effect for which a utility or disability weight is sought. However, it might not always be appropriate to directly assign disease profiles to the observed toxicological effects (such as linking high cholesterol directly to obesities), as one would understand that such a small change in one parameter does not directly relate to the clinical disease and would highly overestimate the health impact.

In a more realistic approach towards translating the observed animal effects into the human correlate at the same exposure levels, it is more likely that subclinical effects will occur instead of clinical effects. The difference would be that a subclinical effect is defined as: *"An adverse effect that has not (yet) progressed to the level of the clinical health effect"* and a clinical effect as: "*An adverse effect that is manifested clinically (i.e., recognized and*

considered as a disability, disease, or illness by a medical professional)"(ter Burg et al,. 2015). However, the QALY and DALY weights are only derived for the clinical health effects and can only be used if clinical effects are anticipated from the exposure.

The HIA will, therefore, be predominantly focused on clinical diseases, while the actual human exposure levels mostly lie below the levels where these clinical effects are anticipated. Nonetheless, at those levels, toxicologically relevant adverse effects might still have an impact on health. The etiology of such an effect is deemed as an adverse effect of which an individual is not noticeably affected, but the effect may reduce overall health status or resistance. The impact of subclinical effects at the population level may be substantial when a high number of subjects are affected. Highly prevalent subclinical effects could thus strongly contribute to the ultimate health status of the population.

De Hollander, Melse et al. (1999) argued that standard subclinical effect variables cannot be properly translated to a burden of disease estimate and also not specific for disease genesis. It may be taken into consideration that one should not attempt to characterize the impact of subclinical effects on a disability (clinical) level but the health impact on the level of subclinical effects. ter Burg, Bokkers et al. (2015) explored this possibility of estimating the health impact on the level of subclinical effects and introduced a severity weight for subclinical effects in analogue with a utility and disability weight. Although the approach looked feasible, several difficulties were identified such as the identification of the level where the subclinical effects may be considered to progress to a clinical disease. As a result, it was concluded that more experience is required, even more so when a clinical level in the test animals has not been reached. Furthermore, such an approach would only be feasible in a quantitative SEA if valuation surveys would be performed to estimate the societal willingness to pay for avoidance of subclinical effects. Currently, we are not aware of such appraisal studies and the same difficulties encountered with the non-existence of a constant value for a QALY or DALY would apply with the valuation of one unit of avoided subclinical effects.

More generally, it should be recognized that moving from the toxicology to the prediction of health impacts is a highly uncertain task and can introduce significant uncertainty into the assessment; indeed, the uncertainty associated with this step may be orders of magnitude greater than the uncertainty associated with any assumptions made as part of the use of QALYs or DALYs within an economic analysis.

Table 5-4: Mapping of health effects against weights							
Health effects for HIA from WP1		**QALYs general**	**Specific description based available from Table A2-1**		**DALYs general**	**Specific description based available from Table A2-2**	
Attributable to Carcinogenicity and Mutagenicity							
Cancer	Lung	Lung	Cancer of bronchus		Lung	Operable	Diagnosis and primary therapy
			small cell lung cancer			Inoperable	
			Non small cell lung cancer			Disseminated	Relapse/terminal stage small cell cancer
	Kidney	Kidney	-		Kidney	Primary	
						Disseminated	
						Terminal	
	Urinary bladder	Urinary bladder	-		Urinary bladder	Primary	
						Disseminated	
						Terminal	
	Liver	Liver	-		Liver	Primary	
						Disseminated	
						Terminal	
	Skin	Skin	Melanoma—stage I—stable		Skin	Primary	Melanoma primary treatment Non-melanoma skin cancers Basal cell carcinoma
			Melanoma—stage IV—stable disease			Disseminated	
						Terminal	Melanoma primary treatment Non-melanoma skin cancers Basal cell carcinoma
	Breast	Breast	-		Breast	Primary	non-invasive tumour

Table 5-4: Mapping of health effects against weights							
Health effects for HIA from WP1		**QALYs general**	**Specific description based available from Table A2-1**		**DALYs general**	**Specific description based available from Table A2-2**	
Cancer							<2 cm
						Disseminated	
						Terminal	
	Stomach	Stomach	Gastric cancer stage I and II		Stomach	Primary	
			Gastric cancer stage IV			Terminal	
	Haematopoietic	Leukaemia	Progressed health state for chronic lymphocytic leukaemia		Leukaemia	Primary	Acute myeloid leukaemia Chronic myeloid leukaemia
						Pre-terminal	
						Terminal	Acute myeloid leukaemia Chronic myeloid leukaemia
	Lymphatic	Non-Hodgkin's Lymphoma	Progressive disease: follicular non-Hodgkin's lymphoma		Non-Hodgkin's Lymphoma	Primary	
			Progression free follicular non-Hodgkin's lymphoma			Disseminated	
						Terminal	
	Pancreatic	Pancreatic	metastatic pancreatic cancer		Pancreatic	Primary	
						Disseminated	
						Terminal	
	Prostate	Prostate	-		Prostate	Primary	
						Hormone refractory cancer	
						Terminal	
Attributable to Reproductive Toxicity classification							
Impaired or reduced fertility	Female	Infertility			Infertility		Infertility: primary
	Male						
	Increase in Spontaneous abortion	Not available			Not available		

Table 5-4: Mapping of health effects against weights							
Health effects for HIA from WP1		**QALYs general**	**Specific description based available from Table A2-1**		**DALYs general**	**Specific description based available from Table A2-2**	
	Disruption of ovarian cycles	Disruption of ovarian cycle (low)	Heavy menstrual bleeding		Disruption of ovarian cycle	Menstrual disorders	
Reduced foetal growth	Small for gestational age; low birth weight	Low birth weight	Very low birth weight infants immediately after birth		Low birth weight	Mild permanent disability	
Developmental anomalies/abnormalities	IQ	Idiopathic intellectual disability	No disability		Idiopathic intellectual disability	Mild	Mild intellectual disability
	Developmental neuro-impairment / neuro-logical disorders		Mild mental retardation				Moderate intellectual disability
							Severe intellectual disability
							Profound intellectual disability
	Cleft palate	Cleft palate	Infants		Cleft palate		Cleft palate—treated Cleft lip—treated
			Children				
			Adolescents:				
			Adults				
	Neural tube (spina bifida)	Spina bifida	-		Spina bifida	Mild	
						Moderate	
						Severe	
	Renal abnormalities	Renal abnormalities	End-stage renal disease		Renal abnormalities	Unilateral renal agenesis/ dysgenesis	
						Bilateral renal agenesis/ dysgenesis	
						Chronic kidney disease (stage IV):	
						End-stage renal failure (on dialysis)	End-stage renal disease: with kidney

Table 5-4: Mapping of health effects against weights							
Health effects for HIA from WP1		**QALYs general**	**Specific description based available from Table A2-1**		**DALYs general**	**Specific description based available from Table A2-2**	
							transplant:
	Urinogenital abnormalities	Urogenital abnormalities	Abnormal aspect scrotum congenital UDT Unilateral		Urogenital abnormalities	Other urinary tract malformations	
			Abnormal aspect scrotum congenital UDT Bilateral				
	High cholesterol (obesity and increased rate of heart disease, etc.)	Obesity			Obesity		
	Cardiovascular	Coronary heart disease	coronary atherosclerosis and other heart disease		Coronary heart disease (Angina pectoris)	Mild	
						Moderate	
						Severe	
		Acute Myocardial infarction	-		Acute myocardial infarction	Acute myocardial infarction: days 1-2:	
						Acute myocardial infarction: days 3-28:	
		Stroke	-		Stroke	Mild (Effects)	Mild permanent impairments
						Moderate (Effects)	Moderate permanent impairments
						Severe (Effects)	Severe permanent impairments
	Skeletal	Not available			Not available		
	Cryptorchidism	Not available			Not available		
	Hypospadias	Not available			Not available		
	Abnormalities of limbs	Not available			Not available		
Other							
	Nephritic syndrome	na	na			End-stage renal failure with dialysis	
						End-stage renal failure with transplant	
						Transplanted patient	
						Untreated end-stage renal failure	
	Respiratory (tract)	Chronic	Mild COPD		Chronic	Mild	

Table 5-4: Mapping of health effects against weights							
Health effects for HIA from WP1		**QALYs general**	**Specific description based available from Table A2-1**		**DALYs general**	**Specific description based available from Table A2-2**	
	irritation	obstructive pulmonary disease	Moderate COPD		obstructive pulmonary disease	Moderate	
			Severe COPD			Severe	
			Very severe chronic obstructive pulmonary disease (COPD)				
		Asthma	Good asthma control		Asthma	Mild	Asthma: controlled:
			Mildly reduced asthma control				Asthma: partially controlled
			Moderately reduced asthma control			Severe	Asthma uncontrolled
			Poor asthma control				
	Allergic contact dermatitis	Allergic contact dermatitis	Atopic eczema: Severe		Allergic contact dermatitis		
			inflammatory skin condition				
	Osteosporosis (Bone fracture)	Osteoporosis	Clinical vertebral fracture for women with osteoporosis, first year		Osteoporosis	Diagnosed cases	
			Clinical vertebral fracture for women with osteoporosis, subsequent years				
			Wrist fracture for women with osteoporosis, first year				
			Wrist fracture for women with osteoporosis, subsequent years				
		Anaemia (no cancer)	Other fracture for women with osteoporosis, first year		Anaemia	Mild	
			Deficiency and other anaemia			Moderate	
						Severe	
		Glaucoma	Mild		Glaucoma	Mild	
			Moderate			Moderate	

Table 5-4: Mapping of health effects against weights							
Health effects for HIA from WP1		**QALYs general**	**Specific description based available from Table A2-1**		**DALYs general**	**Specific description based available from Table A2-2**	
			Severe			Severe	
		Hearing impairment	other ear and sense organ disorders		Hearing impairment	Mild	
			Profound deafness and no cochlear implant			Moderate	
						Severe	
		Liver cirrhosis	Compensated cirrhosis		Liver cirrhosis		
			Decompensated cirrhosis				
		Stomach ulcer	30-39 years		Stomach ulcer	Peptic ulcer disease	
			40-49 years				
			50-59 years				
		Pancreatitis	Mild		Pancreatitis		
			Severe				
		Parkinson-like condition	0–25 % OFF time		Parkinson-like condition	Initial/Mild	
			26–50 % OFF time			Intermediate/ Moderate	
			51–75 % OFF time			End-stage/Severe	
			76–100 % OFF time				
		Alzheimer disease and other dementia	senility and organic mental disorder		Alzheimer disease and other dementia	Mild	
			Mild, community			Moderate	
			Mild nursing			Severe	
			Moderate, community				
			Moderate, nursing home				
			Severe, community				
			Severe, nursing home:				
	Skin irritation	Not available			Not available		

6 Conclusions

6.1 Overview

Socio-economic analysis (SEA) plays an important role in the REACH Authorisation process and can also be important in the preparation of Annex XV restriction dossiers. ECHA's guidance on the preparation of SEAs to support these processes[34] provides an overview of the different approaches that can be taken to estimating the human health impacts associated with changes in exposures to substances of very high concern (SVHC). In particular, the focus is on quantified and monetised estimates of the health impacts of restrictions and authorisation as this facilitates a comparison with other economic impacts.

In order to support the preparation of SEAs, ECHA commissioned a willingness to pay study by the Charles University, aimed at developing new estimates for health endpoints relevant to SVHCs. As part of their work, the Charles University study team also examined the use of QALYs and DALYs, with some of the results presented in Section 5 of this report.

This study was commissioned with the aim of building on the work carried out by the Charles University, to provide a more comprehensive assessment of the usefulness of the QALY/DALY literature for providing quantitative health impacts for use in REACH SEAs. The focus has been on the quantification and valuation of the human health impacts associated with SVHCs on the Authorisation Candidate List or Registry of Intentions for Restriction. In addition, the focus was on the weights that are used in the calculation of QALYs and DALYs rather than on the end QALY or DALY figures themselves. This is because information on the weights together with data on the duration of a disease are the key pieces of information to enable analysts to develop own estimates the QALY gains or DALYs avoided that would result from reduced exposures to a SVHC.

More specifically, the aims of the study can be summarised as follows:

1. To review existing and planned regulatory measures under REACH Annex XVII, Annex XIV and the Candidate List to establish a set of human health effects potential interest for socioeconomic analysis

2. To review the existing literature to collect QALY/DALY weights which have been estimated for this set of human health effects, and

3. To consider the extent to which these QALY/DALY weights and the underlying health state descriptions are applicable or 'transferable' to the chemicals/REACH context and the sorts of health effects likely to be experienced.

[34] See for example ECHA's Guidance on the preparation of socio-economic analysis as part of an application for authorisation, Version 1, January 2011.

In addition, the outputs are to include identification of gaps in the existing literature in terms of the availability of QALY/DALYs for the types of human health effects of interest. We were also tasked with making any simple adjustments to weights that could be made to enable their transfer, if this was considered possible and appropriate.

Section 3 provides details of the 41 health effects that were identified from the review of existing and planned regulatory measures and that have acted as the focus for our work. Section 4 sets out the collated sets of utility and disability weights that were identified from the literature for these health effects, and gives an overview of the studies from which they were drawn. This section builds on Section 2, which presents an introduction to how these weights are generated methodologically and some of the issues that arise in their derivation and use.

All three of these sections then fed into our discussion on the degree to which the weights identified from the literature are transferrable to the REACH context. In this respect, it is of note that as we were able to identify weights for 36 of the endpoints (or potential proxies for them based on linked or related health effects), no steps were taken by the study team to make any simple adjustments to the weights to enable their use in REACH SEAs. Instead, we have highlighted the importance of ensuring that the weights chosen reflect the type of health effects identified by the toxicological data, bearing in mind that moving from the toxicological data to health impacts is itself characterised by uncertainty.

6.2 Key Findings

Our key findings are as follows:

- There are utility and/or disability weights for almost all of the types of health effects identified as being linked to SVHC which have been: subject to REACH Annex XVII restrictions/restriction dossiers; are listed in REACH Annex XIV or are on the 'Candidate List' (Candidate List of SVHC for Authorisation); or are anticipated to be subject to restriction or prioritisation.

- The availability of utility and/or disability weights in the literature that correlate well to many of the health effects identified for SVHC suggests the potential for the use of QALYs and DALYs in REACH SEAs. This includes their use in both cost-utility analyses, for example to support Restriction proposals, or in cost-benefit analyses to support applications for authorisation.

- It was not possible, however, to identify a single consistent reference set of utility weights which cover the exact health endpoints identified as being associated with exposure to SVHCs. The available DALY studies cover a fuller set of the health effects, especially when taken together (although there are still gaps). Furthermore, the more recent studies (GBD 2013 and the European Burden of Disease) have applied the same approach to the derivation of weights across different health states, they have the attraction of providing a more consistent basis for valuing different health effects within SEAs. However, it should be noted that the Dutch study (Stouthard et al., 1997) provides greater differentiation between disease

stages and severities, and covers some endpoints of concern not addressed by the GBD or the European BoD.

- It has been argued that DALYs are more appropriate in the context of REACH, as the aim is reduce the burden of disease within society. We note though that there is a clear preference for the use of QALYs in some Member States and that there is also a literature on willingness to pay for a QALY that may be of relevance to some REACH SEAs.

- This study focuses on utility and disability weights and they can be more flexibly used than the actual QALY or DALY estimates provided for particular medical interventions or a given disease. For example, QALYs as quoted in the academic/health literature will reflect the gain from a treatment rather than the gain which would be obtained from the prevention of an illness. However, the utility weight used in the calculation of the QALY gains stemming from an intervention will reflect the difference between full health and the current health state; it can therefore be used to calculate the gain in health that would be obtained if the illness was prevented in the first place. Similarly, QALY and DALY estimates will include assumptions on average remaining life expectancy and disease duration that may not be appropriate for a REACH SEA.

- Given the varying availability of weights across the different health endpoints, it should be recognised that the weights can be used interchangeably. In other words, the inverse of a utility weight is the corresponding disability weight (see also Section 2).

- Assumptions will be required of analysts when translating the effects observed in toxicological studies (or extrapolated from them) into disease profiles suitable for the use of QALYs or DALYs (for example, the exact severity or stage of cancer which may be caused). For some health effects, the lack of relevant weights is due to problems in applying the QALY and DALY concepts to diseases which do not impair the individual's everyday quality of life (e.g. very low levels of impact on IQ and high cholesterol levels on their own rather than linked to a secondary effect in the future).

- Where there is a good match in health effects, the transfer of the weights to the chemicals context requires careful consideration of the choices made in the original studies. Potential methodological issues are detailed in Section 2 and which may affect the reliability of the utility or disability weights: the size and nature of the sample population, the age of the study, the elicitation method used, the degree to which the method used reflects a welfare loss or just a health loss, etc. There are clear examples of how variations in these factors can influence the weights elicited; for example, the utility weights for spina bifida produced by researchers in a single study using the same sample population but different elicitation methods vary significantly due to differences in the elicitation methods.

- In general, there should be a preference for choosing weights that stem from more recent studies than from the older studies. This is because of changes over time in both the population's and medical practitioners' understanding of different diseases, how they can be treated, and hence how they affect quality of life.

- When using disability weights, analysts will need to try and establish the time period associated with a disability weight; similarly, disease duration as assumed in deriving utility weights will also be important. Failure to do so may result in a significant over or underestimate of impacts. Within the context of a REACH SEA, this may be a particular important issue and consideration may need to be given as to whether a case of the disease due to chemical exposures would correspond better to a weight that reflects a short term or a longer term and more constant condition. The Estonian study presents information on the time period associated with individual weights in a summary table. For the GBD studies, one needs to examine the detailed health state descriptions used in the surveys to establish the appropriate time profile. This requires accessing the detailed annexes to the main reports (via The Lancet[35]).

- If co-morbidities are important to the health impact assessment, then it may be important to either adopt an additive or a maximum approach (as discussed in Section 5) to assessing the total level of impact. If using QALYs, the UK "catalogue of weights" may be worth considering as it provides figures for the marginal disutility associated with the onset of a disease, taking into account different starting typical health states.

- Within the context of REACH, analysts will need to define the appropriate life expectancy for the population at risk, taking into account gender-specific differences as appropriate.

- In any assessment using the existing weights, some level of uncertainty will arise due to a combination of methodological assumptions and the asymmetry of information between the outputs of a REACH risk assessment and the descriptions used when eliciting weights. It will therefore be important for analysts to undertake sensitivity analysis, for example, by undertaking calculations using weights derived by different studies and reflecting either different definitions of health states or different populations and elicitation approaches. Otherwise, the justification for choosing only a single weight should be clearly explained, together with the potential uncertainty that this may lead to in the SEA's conclusions. The magnitude of variation between the different existing weights for one endpoint and the uncertainty in methodological assumptions, however, seem to fit within the overall range of variation and uncertainties commonly observed in REACH SEAs.

[35] http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2815%2960692-4/fulltext

- Research carried out as part of other EU projects has found that there is no empirical relationship between QALYs and WTP values. In particular the study by Ryen and Svensson (2014) shows that WTP-Q estimates vary widely and are dependent on several methodological factors. They conclude that a "common" societal value for one QALY may not be appropriate as a constant WTP-Q does not hold. This finding appears to be confirmed by the work carried out by the Charles University. This does not exclude the possibility to monetise QALY /DALY estimates, however, special attention should be paid to the justification for the choice of monetary value used in the transfer. Again, instead of assuming a single monetary value per QALY gained/DALY avoided, it may be appropriate to consider more than one value for sensitivity purposes.

6.3 Example calculations

To illustrate how one might use either QALYs or DALYs, a set of example calculations are provided here for one of the health states identified in Section 5 as showing reasonable correlation with the type of effects that may stem from chemical exposures. For this example it is assumed that chemical exposures for the mother have led to a very low birth weight infant. As discussed in Section 5, both a QALY utility weight and a DALY disability weight are available for this endpoint, although they reflect different perspectives. The utility weight reflects the parent's view of the impacts of reduced growth impacts on the child. Two disability weights are available. One from the Estonian BoD study, which provides no additional detail on the assumed health state, and a second from the original Dutch study which is a from the perspective of the child in later life due to low birth weight resulting in mild disabilities in later life, e.g. permanent hearing loss).

The weights are as follows (see also the tables in Annex 3):

- Utility weight for very low birth weight in infants (Korvenranta et al, 2010):
  0.940

- Disability weights for low birth weight:
  - Mild permanent disability (Stouthard et al, 1997): 0.110
  - Low birth weight (no further details) (Lei et al, 2009):
    0.442

The utility weight given by Korvenranta et al (2010) can be inverted to provide the weight for calculation of the potential gains in QALYs that would result from an intervention; inverting the weight also translates this into a weight that could be used in DALY calculations of 0.06. Comparing this weight to the two disability weights shows reasonable correlation with the weight elicited by Stouthard et al (1997) but the weight from the Estonian study is significantly higher.

The Estonian study provides information on the duration (time period) to be assumed for the health effect, with this given as 1 year – in other words, the impacts should be assumed to reflect a chronic condition.

For the purposes of this example calculation, it is assumed that the average life expectancy for a newborn child is 80 years and that the value of a life year is €75,000 (based on a VoSL of around €1.5 million for a 40 year old person). Based on these assumptions, the undiscounted impacts are as follows per infant:

- QALY gains = 4.8 QALYs over the 80 years
- DALYs avoided = 8.8 DALYs over the 80 years based on Stouthard et al (1997)
  = 35.36 DALYs over the 80 years based on Lei et al (2009)

Discounted at 4% and valued at €75,000 per QALY or DALY, the present value impacts are valued as follows per infant:

- QALY gains = €107,619
- DALYs avoided= €197,302 based on Stouthard et al (1997)
- DALYs avoided = €792,796 based on Lei et al (2009)

Estimates could also be prepared assuming a discount rate of zero (i.e. by multiplying the above total QALY gains or DALYs avoided by €75,000); this results in economic impacts valued at between €360,000 to €2.65 million per infant. These end estimates would then be multiplied by the population of children at risk (e.g. an estimate based on the estimated number of female workers of childbearing age exposed to the chemical multiplied by the average number of children per woman for the EU).

6.4 Recommendations for ECHA

The above conclusions suggest that there may be two approaches going forward:

1. ECHA commissions further research aimed at developing a consistent set of weights for use in REACH SEA based on a more detailed exploration of the original literature than was possible here. This could include a meta-analysis across the available sources to derive mean values for different health effects and associated confidence intervals.

2. ECHA prepares some advice or guidance on the use of either QALYs or DALYs highlighting the key methodological and analytical issues that SEA practitioners need to take into account in using these approaches, with a key feature of this being the need for consideration of multiple weights through sensitivity analysis. This guidance would include a series of examples of how weights could be used within cost-utility analysis and cost-benefit analysis.

As part of either of the above approach, ECHA may also want to suggest appropriate WTP values to act as anchors for quantifying different health effects. We do not believe that these can be derived analytically given the lack of a consistent relationship between WTP and QALYs. Expert judgement could be used to set these anchors, although they would clearly be open to challenge on the basis of the findings of the academic literature.

7 Bibliography

Abellan-Perpiñan JM, Bleichrodt H & Pinto-Prades JL (2009): The predictive validity of prospect theory versus expected utility in health utility measurement. *Journal of Health Economics,* Vol. 16. 28: 1039-1047.

Anand S & Hanson K (1997): Disability-adjusted life years: a critical review. *Journal of Heath Economics,* Vol. 16.

Arnesen T & Nord E (1999): The value of a DALY life: problems with ethics and validity of disability adjusted life years. *BMJ*, Vol. 319.

Arnesen T & Norheim O (n.d.): Disability Adjusted Life Years: Possibilities and Problems, presentation downloaded from http://www.pitt.edu/~super7/2011-3001/2911.ppt

Attema AE & Brouwer WBF (2014): Constant Proportional Trade-Offs and Health State Evaluations. *Encyclopedia of Quality of Life and Well-Being Research*, 1207-1212.

Bleichrodt H & Johannesson M (1997): The Validity of QALYs: An Experimental Test of Constant Proportional Tradeoff and Utility Independence. *Med Decis Making*, 17(1): 21-32.

Bleichrodt H & Gafni A (1996): Time preference, the discounted utility model and health. *Journal of Health Economics*, 15: 49-66.

Blinman P, King M, Norman R, Viney R, & Stockler, M (2012): Preferences for cancer treatments: an overview of methods and applications in oncology. *Annals of Oncology*, 1104-1110.

Bognar G (2008): Age-weighting. *Economics and Philosophy*, Vol.24, 167–189.

Brock DW (1998): Ethical issues in the development of summary measures of population health status. In Field MJ, Gold MR (eds.) (1998): Summarizing Population Health: Directions for the Development of Population Metrics, Division of Health Care Services, Institute of Medicine, National Academy Press, Washington.

Brock DW (n.d.): Ethical Issues in the Use of Cost Effectiveness Analysis for the Prioritization of Health Care Resources, information downloaded from http://www.nyu.edu/gsas/dept/philo/courses/bioethics/Papers/EthicalIssues.PDF

Cairns J (1994): Valuing future benefits. *Health Economics*, 2: 161-165.

Chatila WM, Thomashow BM & Minai OA et al. (2008): Comorbidities in Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society,* Vol. 5.

Charles University in Prague and VU University in Amsterdam (2014): Stated preference study to examine the economic value of benefits of avoiding selected adverse human health outcomes due to exposure to chemicals in the European Union, Part I: sensitization & dose toxicity, Service contract for the European Chemicals Agency No ECHA/2011/123.

Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC & Sculpher M (2013): Methods for Estimation of the NICE Cost Effectiveness Threshold, Final Report, CHE Research Paper 81, University of York.

Department of Health (2014): Comorbidities: A framework of principles for system-wide action, information downloaded from https://www.gov.uk/government/publications/better-care-for-people-with-2-or-more-long-term-conditions

Department of Human Services (2005): Victorian Burden of Disease Study: Mortality and morbidity in 2001, information downloaded from http://docs.health.vic.gov.au/docs/doc/6AEAFAB1BAE696B9CA257886000158A0/$FILE/bod_2001.pdf

Doctor JN, Bleichrodt H, Miyamoto J, Temkin NR & Dikmen S (2004): A new and more robust test of QALYs. *Journal of Health Economics*, 23:353-367.

Dolan P & Stalmeier P (2003): The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. Journal of Health Economics, 22: 445-458.

Dolan P (2000): A note on QALYs versus HYEs: health states versus health profiles. *International Journal of Technology Assessment in Health Care*, 16(4): 1220-1224.

Drummond M, Stoddart G & Torrance GW (1987): Methods for the Economic Evaluation of Health Care Programmes, Oxford, England: Oxford Medical Publications.

European Centre for Disease Prevention and Control (ECDC) (2011): Methodology protocol for estimating burden of communicable diseases. Stockholm: ECDC.

Fox-Rushby J (2002): Disability adjusted life years (DALYs) for decision-making?: an overview of the literature. Office of Health Economics, London. Working Paper No. 596.

Gold MR, Stevenson D & Fryback DG (2002): HALYs and QALYs and DALYs, Oh My: Similarities and Differences in Summary Measures of Population Health. *Annu. Rev. Public. Health,* Vol.23.

Grosse SD et al. (2009): Disability and Disability-Adjusted Life Years: Not the same. *Public Health Reports*, Vol., 124, 197-202.

Haagsma JA et al. (2014): Review of disability weight studies: comparison of methodological choices and values. *Population Health Metrics*, Vol. 12.

Haagsma JA (2010): Disability Adjusted Life Years and Acute Onset Disorders: Improving estimates of the non-fatal burden of injuries and infectious intestinal disease, information downloaded from https://www.bmg.eur.nl/fileadmin/ASSETS/bmg/Onderzoek/Promoties/Promoties_2010/iBMG_proefschrift_juanita_haagsma_disability_adjusted_life_years_and_acute_onset_disorders_improving_estimates_of_the_non-fatal_burden_of_injuries_and_infectious_intestinal_disease.pdf

Hammond P (1976): Changing tastes and coherent dynamic choice. *Review of Economic Studies*, 43: 159-173.

Harris J (1985): The Value of Life. London: Routledge & Kegan Paul.

Highfill T & Bernstein E (n.d.): Using Disability Adjusted Life Years to Value the Treatment of Thirty Chronic Conditions in the U.S. from 1987-2010, information downloaded from http://www.bea.gov/papers/pdf/highfill_bernstein_2014_dalysall.pdf

Jakovljević M & Ostojić L (2013): Comorbidity and multimorbidity in medicine today: challenges and opportunities for brining separated branches of medicine closer to each other. *Medicina Academica Mostariensia,* Vol. 1, No.1.

Jones R (2010): Chronic Disease and Comorbidity, information downloaded from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2880735/

Kahneman D & Tversky A (1979): Prospect theory: an analysis of decision under risk. *Econometrica*, 47:263-291.

King KM, Norris CM, Knudston ML & Ghali WA (2009): Risk-taking attitudes and their association with process and outcomes of cardiac care: a cohort study. *BMC Cardiovascular Disorders*, 9(36).

Lai T, Habicht J, Kiivet RA (2009): Measuring burden of disease in Estonia to support public health policy, information downloaded from http://eurpub.oxfordjournals.org/content/19/5/541

Lajoie J (2014): Understanding the Measurement of Global Burden of Disease, information downloaded from https://cdn.metricmarketing.ca/www.nccid.ca/files/Influenza/Global_Burden_Disease_Influenza_ENG.pdf

Lipscomb J (1989): The Preference for Health in Cost-Effectiveness Analysis. Medical Care, 3 (supplement): S233-53.

Mackiegan LD, Lawson LN, Drangalis JR, Bootman JL & Burns LR (1993): Time preference for health gains versus health losses. *Pharmaco Economics*, 3: 374-386.

McNeil BJ, Weichselbaum R & Pauker SG (1978): Fallacy of the five-year survival in lung cancer. *New England Journal of Medicine*, 299:1397:1401.

Mehrez A & Gafni A (1989a): Quality-adjusted life years, utility theory and health-years equivalents. *Med Decis Making*, 9: 142-149.

Mehrez A & Gafni A (1989b): An empirical evaluation of two assessment methods for utility measurement for life years. *Socio-Economic Planning Sciences*, 21: 371-375.

Ministry of Health (2012): Ways and Means: A report on methodology from the New Zealand Burden of Diseases, Injuries and Risk Factors Study, 2006–2016. Wellington: Ministry of Health.

Miyamoto JM (1999): Quality-Adjusted Life Years (QALY) Utility Models under Expected Utility and Rank Dependent Utility Assumptions. *Journal of Mathematical Psychology*, 43: 201-237.

Murray CJL et al. (2012): GBD 2010: design, definitions, and metrics. *The Lancet,* Vol.380.

National Institute for Health and Care Excellence (2008): Report on NICE Citizens Council meeting: Qality Adjusted Life Years (QALYs) and the severity of illness, information downloaded from https://www.nice.org.uk/Media/Default/Get-involved/Citizens-Council/Reports/CCReport10QALYSeverity.pdf

National Institute of Health and Care Excellence (NICE) (2010): Measuring effectiveness and cost effectiveness: the QALY, information downloaded from https://www.nice.org.uk/proxy/?sourceurl=http://www.nice.org.uk/newsroom/features/measuringeffectivenessandcosteffectivenessTheQALY.jsp

Olsen JA (1993a): Time preference for health gains: An empirical investigation. *Health Economics*, 2: 257-265.

PharmaTimes (2010): NICE advises against breast cancer drug Tyverb once more, information downloaded from http://www.pharmatimes.com/Article/10-06-10/NICE_advises_against_breast_cancer_drug_Tyverb_once_more.aspx?rl=1&rlurl=/12-02-14/NICE_sticks_with_no_for_GSK_s_Tyverb_Roche_s_Herceptin.aspx

Pliskin J, Shepard DS & Weinstein MC (1980): Utility Functions for Life Years and Health Status. *Operations Research*, 28(1): 206-224.

PMLiVE (2013): NICE: Pfizer's lung cancer drug Xalkori too expensive for NHS, information downloaded from http://www.pmlive.com/pharma_news/nice_pfizers_lung_cancer_drug_xalkori_too_expensive_for_nhs_469241

Polinder S et al. (2012): Systematic review of general burden of disease studies using disability-adjusted life years. *Population Health Metrics,* Vol.10.

Redelmeier DA & Heller DN (1993): Time preference in medical decision making and cost effectiveness analysis. *Medical Decision Making*, 13: 212-217.

Richardson J (1999): Age Weighting and Discounting: What are the Ethical Issues?, information downloaded from http://www.buseco.monash.edu.au/centres/che/pubs/wp109.pdf

Roberts E (2000): Age Discrimination in Health and Social Care, information downloaded from http://www.kingsfund.org.uk/publications/age-discrimination-health-and-social-care

Robinson A, Loomes G & Jones-Lee M (2001): Visual analog scales, standard gambles and relative risk aversion. *Med Decis Making*, 21(1): 17-27.

Ryen L & Svensson M (2014): The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. *Health Economics*.

Salomon JA et al. (2012): Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010, information downloaded from http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2812%2961680-8/abstract

Sayers BMA & Fliedner TM (1997): The critique of DALYs: a counter-reply. *Bulletin of the World Health Organization*, Vol. 75 (4).

Shepard DS & Thompson MS (1979): First Principles of Cost- Effectiveness Analysis in Health. *Public Health Reports*, 94: 535-543.

Spencer A & Robinson A (2007). Tests of Utility Independence When Health Varies over Time. *Queen Mary University of London, Department of Economics*, Working Paper No. 596.

Stiggelbout AM, Kiebert GM, Kievit J, Leer JWH, Stoter G & De Haes JCJM (1994): Utility Assessment in Cancer Patients: Adjustment of Time Tradeoff Scores for the Utility of Life Years and Comparison with Standard Gamble Scores. *Med Decis Making* 14(1): 91-97.

Stouthard MEA et al. (1997): Disability weights for diseases in the Netherlands, information downloaded from http://dare.uva.nl/document/2/3276

Strotz RH (1956): Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23: 165-180.

Sullivan PW, Slejko JF, Schulpher MJ & Ghushchyan V (2011): Catalogue of EQ-5D Scores for the United Kingdom. *Medical Decision Making*, 31: 800-804.

The Guardian (2014): Breast cancer drug Kadcyla is too expensive for the NHS – not to pay for it is the correct decision, information downloaded from http://www.theguardian.com/commentisfree/2014/apr/25/kadcyla-cancer-drug-too-expensive-nhs

Tversky A & Kahneman D (1992): Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5: 297-323.

Verhoef LCG, de Haan AFJ & van Daal WAJ (1994): Risk attitude in gambles with years of life: Empirical support for prospect theory. *Med Decis Making*, 14(2): 194-200.

Victorian Government Department of Human Services (2005): Victorian Burden of Disease Study: Mortality and morbidity in 2001, information downloaded from http://docs.health.vic.gov.au/docs/doc/6AEAFAB1BAE696B9CA257886000158A0/$FILE/bod_2001.pdf

Voigt K & King NB (2014): Disability weights in the global burden of disease 2010 study: two steps forward, one step back? *Bulletin of the World Health Organization*.

Vos T et al. (2015): Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013, information downloaded from http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2815%2960692-4/fulltext

Wagland R (2012): A fair innings or a complete life: another attempt at an egalitarian justification of ageism. In, Lesser, H (ed.) *Justice for Older People.* Amsterdam.

Weinstein M & Zeckhauser RJ (1973): Critical ratios and efficient allocation. *Journal of Public Economics*, 2: 147-157.

World Health Organisation (2015): Life expectancy at birth, information downloaded from http://gamapserver.who.int/gho/interactive_charts/mbd/life_expectancy/atlas.html

World Health Organisation (2001): National Burden of Disease Studies: A Practical Guide, information downloaded from http://www.who.int/healthinfo/nationalburdenofdiseasemanual.pdf

World Health Organisation (n.d.): The Global Burden of Disease concept, information downloaded from http://www.who.int/quantifying_ehimpacts/publications/en/9241546204chap3.pdf

Annex 1:

The Methods Used to Elicit Health Utilities and Disability Weights and Key Underlying Assumptions

1.1 The Methods

1.1.1 Standard gamble

The standard gamble (SG) method is used for the measurement of individuals' preferences under uncertainty, where these preferences indicate the utility that an individual gains from a certain set of choices. Within the context of QALYs, the utility weights derived using the standard gamble are cardinal values that represent the strength of an individual's preferences for specific health-related outcomes or conditions.

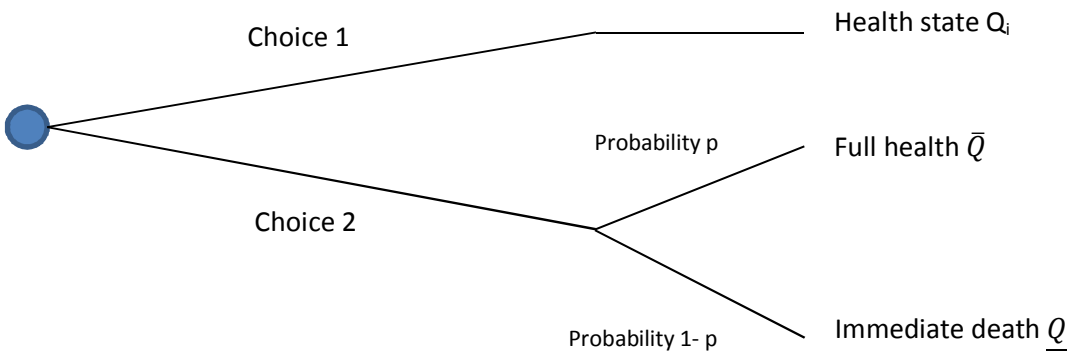
More specifically, the SG approach is based on the von Neumann-Morgenstern (vNM) utility theory, and it is fully consistent with the axioms of rational individual behaviour which are defined in this theory (known as the axioms of preference). It is important to note that the SG is only a valid measure of vNM utility if these axioms hold.

To derive health utility weights, the SG requires the individual to make a choice between two alternatives: a health state which is certain [e.g.. a certain frequency (once a month, once a week) of migraines of a certain severity] or a gamble between a better health state (e.g. no migraines; full health) and a worse health state (e.g. death).

Consider the following notation. Let Q represent the health state of the individual: represents the better health state and represents the worse health state. Qi refers to the certain health state. T denotes remaining life-years. A von Neumann-Morgenstern (vNM) utility function, which describes being in health state Q, starting now for a period of T years (followed by death), is given as U(Q,T). The individual is offered two alternatives:

1. The individual remains in the certain health state Q_i for T years (i.e. for the rest of their life).
2. A treatment with two possible outcomes: either the patient returns to full health, $\bar{Q}$, and lives for another T years (this occurs with probability p) or the patient dies immediately, $\underline{Q}$ (probability 1-p).

The probability p is varied until the individual is indifferent between the two alternatives. This probability is denoted as p* and is the preference value for living in health state Qi for T years, i.e. $U(Q_i,T) = p^*$. The utilities generated are dependent on the risk behaviour of the individual: a more risk averse individual will have a higher utility than a risk-seeker. This can be illustrated using the following diagram:



QALYs are calculated using the Standard Gamble approach as the summation of the years of life adjusted for quality:

$$QALY = \sum_{t=1}^{T} p^{*}T$$

where $p^*=U(Q_i,T_i)$ and T refers to years of life left to live. For example, if an individual has a preference value $p^*=0.7$ and they are expected to live in health state Q_i for 5 years, this is equivalent to 3.5 QALYs (0.7 x 5).

As noted above, the SG only provides a valid measure of an individual's vNM utility if the axioms of preference hold. According to Gafni (1994), the main reasons that interpretation of SG utility weights may be difficult are: the inappropriate treatment of the time-dimension (either it is ignored completely or it is counted twice); the addition of other assumptions, on top of the axioms set out in vNM utility theory (as stated above); and problems arising from the use of the SG technique in measuring temporary health states (as opposed to chronic health states, an assumption made in the majority of SG weight elicitations).

The standard gamble method for eliciting health utilities is preferred to other methods, such as the Time Trade-Off (TTO) and Visual Analogue Scale (VAS), because it incorporates a risk element which is very much key to medical decision-making (Bleichrodt, van Rijn and Johannesson, 1999). However, Richardson (1990) argues that the preference value, p, generated using the standard gamble approach does not accurately reflect real world uncertainty, given that the individual is unaware of the nature and magnitude of this risk.

Weaknesses of the standard gamble approach

As stated previously, the standard gamble approach elicits true utilities only if the individual behaves in agreement with expected utility theory. Schoemaker (1982), Llewellyn-Thomas et al. (1982) and Stiggelbout et al. (1994), amongst others, find that there are deep flaws in the axioms outlined in von Neumann-Morgenstern utility theory and that these axioms do not hold empirically. Though there is little in medical literature to suggest this, several studies have reported violations of the axioms of rational individual behaviour in practical applications. They therefore recommend that they should not be assumed in any scenario.

Another issue is the assumption of risk neutrality, as defined by Pliskin, Shepard and Weinstein (1980). Bleichrodt, Wakker and Johannesson (1997) detail a methodology for deriving QALYs by only imposing the risk neutrality assumption. This assumption enables easy manipulation which is useful when using this methodology for non-chronic, or temporary, health states. However, empirical findings suggest that the risk neutrality assumption does not hold (Stiggelbout et al., 1994). Additionally, Bleichrodt and his colleagues concede that QALYs generated using vNM utility theory "can at best be used as an approximation" in the cases where the risk neutrality assumption is violated excessively.

Richardson (1990) also reports that there is empirical evidence that the standard gamble has been found to be difficult to interpret by individuals; people often have difficulties in understanding the extreme values and their implications. Another point of contention with

this method is that the choices offered to the individual are unrealistic; instant death is an unlikely outcome for most medical procedures and so the preference values elicited may not truly reflect the individual's preferences.

In addition, an individual's cognitive abilities may have to be relative high in order to answer the SG questions, i.e. the method includes "risks", reflected in probabilities, yet many people have difficulties in understanding "risks" and probabilities. As a result, less literate or less educated individuals may face problems in responding to the questions they are being asked, even when devices such as probability wheels are used to help make the trade-offs clearer.

The method is also based on the use of face to face interviews, which increases the costs and administrative requirements of applying it to large sample populations. Finally, in a clinical context, repeated measurements to monitor changes in quality of life over time are difficult to organise, time consuming and, as a consequence, expensive.

Summary of standard gamble

The standard gamble is one of the most appropriate methods for eliciting health utility weights, and is used on its own and as a component of other methods for eliciting health utility weights such as part of the multi-attribute utility function based approaches. However, due to the practical problems associated with its use (e.g. the need to undertake face to face surveys, limiting the number of individuals surveyed) and issues concerning the information requirements of its use and whether individuals fully understanding what they are being asked to do (particularly those that are less educated or have a reduced cognitive ability), other methods appear to be preferred by practitioners.

1.1.2 Time trade-off

Developed by Torrance et al. (1972), the time trade-off (TTO) method for eliciting utility weights requires the individual to make a decision between living in an imperfect health state for t years or living for x years in perfect health, where x<t. The individual is required to indicate the value of x which would make them indifferent between t and x. Formally, the utility value is derived using the following equation:

utility value = $\frac{x}{t}$

For example, if the individual is expected to live 20 years (t) in imperfect health state i but is told there is a treatment which guarantees they will have perfect health but which will also decrease their life expectancy, they might be willing to give up 4 years of life in order to have perfect health. Then the utility value of perfect health is given as:

utility value = $\frac{20-4}{20} = \frac{16}{20} =$ ***0.8***

The TTO method is easier to implement in that, unlike the standard gamble, the individual is not presented with probabilities which may be difficult to comprehend. Although the TTO is grounded in utility theory, it does not reflect decisions made under uncertainty which means it is not consistent with vNM expected utility theory. However, an advantage of the

TTO method is that the way of framing the questions is more consistent with the QALY concept of valuing life years in less than perfect health.

Future discounting and Constant Proportional Time Trade-Off

TTO is subject to discounting but in order for QALYs to be calculated from the utility values elicited through the TTO method, they must satisfy the constant proportional time trade-off assumption (CPTTO). The CPTTO assumption implies that if the individual equates 16 years in perfect health to 20 years in imperfect health state i then they equate 12 years in perfect health to 15 years in imperfect health state i. However, as detailed later in this Annex (Section A1.2), there is empirical evidence which suggests that the CPTTO assumption does not hold. This is likely due to time discounting, whereby future gains in utility are undervalued by the individual in the present.

If, on average, CPTTO does hold, it would only be applicable to those states for which more time is always preferred to less, and there is some evidence that this may not always be the case. For example, Sutherland et al. (1982), having found from a sample of 20 colleagues that the proportion preferring death to varying durations in each of five health states increased as the duration of the states increased, postulate that for some states there exists a maximum endurable time (MET) beyond which people do not wish to live. In other words, the value of those states becomes negative after some threshold. This concept has been reinforced by the results from a much larger general population study (Dolan, 1996). These results suggest that it might be inappropriate to calculate QALYs using TTO values of states for which there might come a point at which death would be preferred to any more time in those states, despite (positive) TTO values that suggest otherwise (Dolan, 2002).

Risk neutrality

Utilities generated by the TTO method are based on value theory rather than expected utility theory (like the standard gamble technique). Therefore the TTO assumes that the individual makes their choice under certainty, i.e. there are no risks. The implication of this is that if the individuals are risk averse, the utility elicited using the TTO method will be lower than the utility elicited using the standard gamble technique. One ought to keep this in mind when comparing results from the TTO and standard gamble methods, as we have considered previously the implications of assuming risk neutrality.

See also the discussion provided in A1.2 below.

The TTO has been used on its own and as a component of other methods for eliciting health utility weights, such as the generation of multi-attribute utility functions. Because it is easier to implement, it has been the method of choice for many of the studies identified in the literature.

An external peer reviewer of this report from RIVM that is an expert in the derivation of utility weights noted that have noted that there is also a further issue with the use of the method, in that a non-insignificant part of the population is not willing to trade any life years, no matter how awful the health state presented, which may be due to ethical reasons.

1.1.3 Visual analogue scale

A range of different rating methods have been used to derive weights. The most commonly applied of these methods would appear to be the visual analogue scale (VAS). The VAS consists of a single line anchored by two health states, for example, "best imaginable health" and "worst imaginable health". The individual is then asked to indicate on this scale their rating of a particular health state.

The main advantage of the VAS method for eliciting health state preferences is that it is very quick and very easy to implement. However, this method has several disadvantages which nullify its singular use in eliciting weights. Blinman et al. (2012) identify several shortcomings of the VAS technique: the ratings are given under certainty, rather than uncertainty; the anchors are poorly described which means they limit the ability to make comparisons across different individuals as different people have different perceptions of "best possible health" and "worst possible health"; there is no element of a "trade-off" and therefore the VAS elicits values rather than utilities. Parkin and Devlin (2006) review the commonly asserted criticisms against the VAS method and propose arguments against these criticisms: firstly, they address the notion that VAS does not have a strong theoretical foundation (Johannesson et al., 1996). They disagree as they state that VAS has a foundation based on psychometric research and other psychological theories relating to responses to stimuli. Furthermore, they argue that a VAS is just as inconsistent with QALY theory as any other measure (other than SG): if we are considering QALYs based on utility, all other tools apart from SG fail to derive QALYs based on utility and so in this sense, VAS is just as inadequate as any of the other measures such as TTO. However, if QALYs are calculated based on quality of life weights, for example, then the VAS is a valid measure of the individual's valuation of a particular health state.

Another common complaint against VAS is that it does not require the individual to make a choice or trade-off. The authors argue that, although the individual is not forced to make a choice between two alternatives (as is the case with SG and TTO), the individual must still choose a point on the scale which reflects their preference – this is still a choice. It could be argued that the "choices" presented to the individual under the VAS are more realistic than those offered under TTO and SG experiments. Parkin and Devlin also find contention with the view held by many health economists that "choiceless" methods like the VAS are not based on economic theory and so they are not relevant. They argue that utility theory is not based on choice either and so the argument that "choiceless" methods are not appropriate for eliciting health preference weights for use in QALY calculation is not valid.

The VAS method is widely used because of its simplicity, in terms of the participant's understanding and its execution. Yet opinions are divided with respect to the VAS's credibility as a method for eliciting individual's health preferences. Many authors, such as Tolley (2009), advocate the use of VAS in conjunction with other techniques like SG or MAUIs, as a "warm-up" exercise to allow the participants to familiarise themselves with the concept of valuing health. Few support the VAS as an outright method to be used on its own.

1.1.4 Person trade-off method

The PTO method consists of asking people how many outcomes of one kind they consider equivalent in social value to X outcomes of another kind, with reference to the EuroQol or other such descriptive system. The disability adjustments are developed by assigning disability weightings to life years – diagnostic groups are chosen and defined, descriptions of these diagnostic groups are developed, and health states are assigned a disability weight to indicate the relative severity of each. Disability weights are obtained by posing two different Person Trade-off (PTO) questions to expert panels. PTO1 compares life extensions for disabled and healthy people. PTO2 compares cures for illness with extension of life. In addition to adjusting the value of life years with disability weights and choosing a particular life expectancy, the value of a life year is modified by discounting and age weighting. In discounting, the value of a life year in the present is set higher than the value of a life year in the future. In age weighting, life years of children and old people are counted less (Arnesen and Norheim, nd).

Nord (1994) suggests that while the technique is theoretically appealing, it is in practice quite demanding as it needs to be applied in fairly large groups of subjects to keep random measurement error at an acceptable level. Possible framing effects include the effects of argument presentation and the choice of start points in numerical exercises. Nord (1994) also notes that to control for these effects, it is important to take subjects through a multi-step procedure, in which they are induced to carefully consider the various arguments that might be relevant in each exercise and to reconsider initial responses in the light of their implications. The investigator must also think through which decision context he/she wishes to study and make his/her choice of context very clear when reporting the results.

Advantages and limitations of the PTO

Although we did not find a conclusive statement which indicates why pairwise comparisons, the PTO and panel based approaches seem to be used in the context of disability weights, there is a literature which discusses their proposed advantages over the preference based methods used to derive weights for QALYs. Interpretation of this literature suggests that, because DALYs are concerned with the societal burden of disease, several analysts have favoured adopting a non-individual preference based measure of value.

Nord (1994) comments that the PTO seems to be rarely used to derive weights for QALYs, as it has been argued that cardinal measures of utility are required for them to be valid. He argues that this should not be the case and indeed argues that individuals should not be asked to value health states for themselves, but instead that a procedure that reflects community views on trade-offs is required. In other words, individual valuations are considered to provide a poor proxy for social preferences and hence of the social value of achieving different health care outcomes, which will be a function of other factors such as the initial severity of the patient's health state.

Prades (1997) explores this argument and assesses (using a limited experiment) how the PTO performs against the standard gamble and the visual analogue methods to establish priorities amongst patients waiting for treatment. He finds that PTO may perform better,

although his main conclusion is that there may be reasons for this in terms of how probabilities are presented and people's abilities to process information.

Dolan and Green (1998) also explore this argument and run a PTO exercise, again on a small sample of respondents. They conclude that considerations about individual health gains may not fully reflect social value; instead respondents seemed to be concerned more about the health gain associated with different treatments and perhaps also the health states that people end up in after treatment, than they are about the severity of the pre-treatment health states.

There are several other criticisms of the PTO protocol which can be regarded as validity problems. These include lack of simplicity, forced consistency between essentially different questions, inability to consider all individuals as equally valuable, and inability of the expert panel to represent the values of other people.

1.1.5 Multi-attribute utility instruments (MAUIs)

A multi-attribute utility instrument (MAUI) is an indirect method of measuring utility which involves two stages of assessment: description of a health state using a generic health-related quality-of-life questionnaire such as the EQ-5D, the SF-6D or the HUI I, HUI II or HUI III, followed by valuation.

For example, the method requires the individual to describe his/her current health state by completing the EQ-5D; subsequently, the health state as given by that person is translated into a value that has been generated by population-based valuation research. Here, large samples of the population have valued pre-defined health states using direct methods such as the SG or the TTO. Econometric modelling techniques are then used to infer valuations for all possible health states of that MAUI instrument.

The benefits of using indirect methods (like MAUIs) include that they are quick and easy to implement. As a result, they are now more widely used in health policy decision making and are recommended for use in health policy decisions relating to the allocation of resources (Blinman et al., 2012). However, several issues arise with the use of MAUIs: firstly, as Tolley (2009) suggests, because the MAUIs are so generic, they may not be sensitive to specific disease contexts. Furthermore, it is more difficult to use a MAUI for measuring the impacts of an acute disease, such as an acute asthma attack, rather than the chronic condition (i.e. atopic to respiratory sensitizers). Another issue is that there is evidence of ceiling effects with the use of EQ-5D and floor effects with the use of the SF-6D. For the EQ-5D, the ceiling effect means that for the patients who give the highest score possible for a health dimension, if their health improves, this improvement will not be accounted for by this scale. For the SF-6D, the implication of the floor effect is that patients in a severe health state will report the lowest possible score for some health dimensions which means that if their situation deteriorates further, the SF-6D will not be able to account for this. This means there is a compromise in the validity of the scores obtained. To overcome such sensitivity problems, EQ-5D has now been changed from a 3-level to a 5-level descriptive system, implying that the instrument can describe 3125 different health states instead of 243.

Disease-specific utility instruments

An area of health utility which has been emerging in recent years is the development of preference-based, disease-specific measure whereby utilities are attached to disease-specific HRQL instruments and are generated in the same way that they are generated with a generic instrument. For example, the International Prostate Symptom Score (IPSS) was valued using the TTO method. The advantage of these instruments is that they help to generate disease-specific utilities and that the instrument closely matches the problems a specific patient group may encounter. However, it is difficult to compare disease-specific utilities across different diseases. Another disadvantage is the sheer amount of research that would be required to elicit such utilities across all diseases of interest. It remains questionable whether such disease-specific utilities are appropriate for health care resource allocation.

1.1.6 Discrete choice and stated preference experiments

Discrete choice experiments (DCE) and other related stated preferences approaches are an attribute-based measure of an individual's preferences. They are an effective way of illustrating the importance of different health states, treatments and aspects of treatment (i.e. effectiveness, tolerability) to the individual. As they appear to be a common approach to preference measurement, discrete choice experiments are considered further here.

The underlying notion of discrete choice experiments is that the value of the health-state or treatment is determined by the values of its attributes (Lancaster, 1966). Discrete choice experiments involve presenting an individual with choices of various health scenarios described in terms of characteristics and their relevant levels. They must choose the scenario which they prefer. These choices are aligned with a rating system, the results of which are modelled using a regression function which generates information on the relative importance of characteristics, the rate at which an individual is willing to trade-off a characteristic for another, etc.

Designing, implementing and analysing a discrete choice experiment is not a straight-forward task. First, qualitative analysis is required to determine the key questions, target demographic and health-state and treatment attributes. Then the experiment must be designed, defining appropriate "choice sets". The DCE must be piloted in order to fine tune the design and make data collection as efficient as possible. The third step is to conduct the experiment, collecting the data. Finally, statistical and econometric analysis (e.g. using a random effects model) is conducted to quantify the participants' health-state preferences.

Benefits

Discrete choice experiments allow combination of an individual's values towards several health aspects into one measure of preference, reflecting the utility of a certain scenario, i.e. a health state. They provide a direct indication of the different trade-offs an individual is willing to make amongst the different possible health states. This is one of the positive aspects of discrete choice experiments - the ability to conduct a simultaneous assessment of several attributes of the health-state or treatment being assessed.

In addition, researchers such as Ryan and Gerard (2003), Viney, Lancsar and Louviere (2002) and Ryan, Watson and Amaya-Amaya (2003) have found that individuals' responses in discrete choice experiments are internally valid and consistent. This is due to the advantage within DCE that respondents don't need to keep a golden standard in mind, just repeated choices between health states. Note that it is also argued that DCE discriminates better between health states than other health state valuation methods.

Limitations

A major limitation of using a discrete choice experiment to derive preferences is the complexity that the study may suffer. For example, there may be several characteristics of which could be included in the experiment (such as the other risks which might be involved, other information which the test could give, etc.), and analysts have to decide which characteristics to include in the DCE and how many can be included to provide detailed information on individual preferences, while also ensuring that the questionnaire can be completed by the participant. However, pilot testing should enable analysts to identify the most important attributes and careful design of the survey instrument should mean that any errors due to missing variables are minimised.

1.2 Further Discussion on Key Assumptions

1.2.1 Utility independence, constant-proportional trade-offs, risk attitude and time preferences

Utility independence

Utility independence implies that if one of the factors of utility is held constant at a particular level, preferences for lotteries over the other factor are independent of the other factor, which is held constant. For example, if T is fixed at $T=t_0$, preferences for lotteries over Q are independent of t_0, i.e. preferences over Q do not depend on the fixed level of T. If the reverse is also true (i.e. T is independent of q_0) then we have mutual utility independence between the two factors.

Spencer and Robinson (2007) find somewhat mixed empirical evidence but generally support utility independence as a reasonable assumption about individuals' preferences, on the condition that the model specified is suitable. They refer to previous studies which support utility independence for chronic health states (as cited in Spencer and Robinson, 2007: Miyamoto and Eraker 1988; Bleichrodt and Johannesson, 1997; Doctor et al., 2004; Bleichrodt and Pinto, 2005).

Many authors argue that the assumption of utility independence is restrictive and that it does not accurately reflect individuals' preferences. This is because it is implausible to suggest that the individual's decision-making process does not depend on either length of time left to live or the individual's time preferences. Furthermore, although there is evidence to suggest utility independence might hold for chronic health states, many studies find that this is not the case for health states which vary over time.

Constant proportional trade-offs

Constant proportional trade-offs (CPTO) indicate that, assuming utility independence, an individual is willing to give up the same constant proportion of time to move from health state q_3 to q_4 as to move from q_1 to q_2, independent of the number of life years left. It implies that the individual's preferences are independent of the amount of life years remaining, thus nullifying the concept of the time effect described by Gafni and Torrance (1984). In this sense, the utility which the individual associates with being in health state Q is "timeless" (Gafni, 1994). However, as previously stated, it is unrealistic to assume that the individual's preferences are not affected by considerations of time.

The benefit of adding the assumption of CPTO is that it allows one to construct a general utility scale whereby the preference value generated is simply equivalent to the value of the health state Q_i *and* time (T_i) (Gafni, 1994). A key advantage of employing a timeless measure of utility is that it allows for easier evaluation of healthcare programs using multi-period utility models (Gafni, 1994). A study by Bleichrodt and Johannesson (1997) finds support for the CPTO assumption: they find that the CPTO holds approximately. However, there has been criticism of this CPTO assumption: the assumption is argued as being restrictive and misrepresentative of an individual's behaviour (Pliskin et al, 1980). Bleichrodt and Johannesson (1997) also concede in the same paper that the evidence supporting the CPTO is not very strong in the case of utility independence. Additionally, Loomes and McKenzie (1989) find in their review of the relevant literature that the CPTO assumption is not supported by empirical evidence. Attema and Brouwer (2010) conclude that health state valuations depend on life expectancy and that one cannot assume CPTO. Bala et al. (1999), in their study, find that for most of the participants, preference scores are not independent of time.

Gafni, despite outlining the benefits of a timeless utility scale, also argues that one must not simply ignore the effect that the time spent in a health state might have on an individual's decision, as doing so may compromise the exactitude of the preference value derived; it may even reverse the individual's preference: Sutherland et al. (1982) looked at individual preferences over several different health states under certainty and for varying time periods. Their results showed that for a short survival period (three months) most respondents preferred all other health states to death. However, when the survival period was longer (eight years), most of the respondents preferred death to all other health states.

Risk attitude

The assumption of risk neutrality states that the individual is indifferent between a lottery over life years and the expected duration (T) of that lottery, with quality of life, Q, held constant. This assumption is required to calculate the quality-adjusted life expectancy (QALE). However, much of the evidence does not favour this assumption: it is widely accepted that individuals are risk averse with respect to the time which they will spend in a particular health state. Yet, there is an argument that individuals may adapt to their circumstances and so they may overvalue the severity of a health state (leading to a higher QALY weight).

Gafni and Torrance (1984) attribute the individual's risk attitude to three components: the gambling effect, the quantity effect and the time effect. The implications of the time effect on QALY calculations, relating to the individual's time preferences, are discussed in greater detail below. The risk attitude as a whole is considered here.

The assumption of risk neutrality (that the individual is risk neutral with respect to longevity) is argued by many authors as being restrictive and an inaccurate representation of an individual's preferences or behaviour as most people exhibit either risk averse attitudes or risk seeking attitudes with respect to their health. In their study of 14 patients undergoing treatment for bronchogenic carcinoma, McNeil et al. (1978) find that most patients exhibit risk aversion with respect to additional time spent in poor health; Stiggelbout et al. (1994) find risk aversion amongst men with testicular cancer; Verhoef et al. (1994) find evidence of risk aversion amongst a sample of healthy women. Yet there is also an equal amount of evidence supporting risk prone behaviour amongst patients: in a study of cardiac patients, King et al. (2009) find that 45% of their 6294 patients were risk prone rather than non-risk-prone in their choice to undergo invasive cardiac surgery (CABG surgery); Verhoef et al. (1994) also find evidence of risk-seeking behaviour involving shorter time frames; Mehrez and Gafni (1989) find risk-seeking attitudes to gambles are more likely as duration increases. Figures 1 and 2 illustrate how an incorrect interpretation of the individual's risk attitude can affect QALY calculations[1].

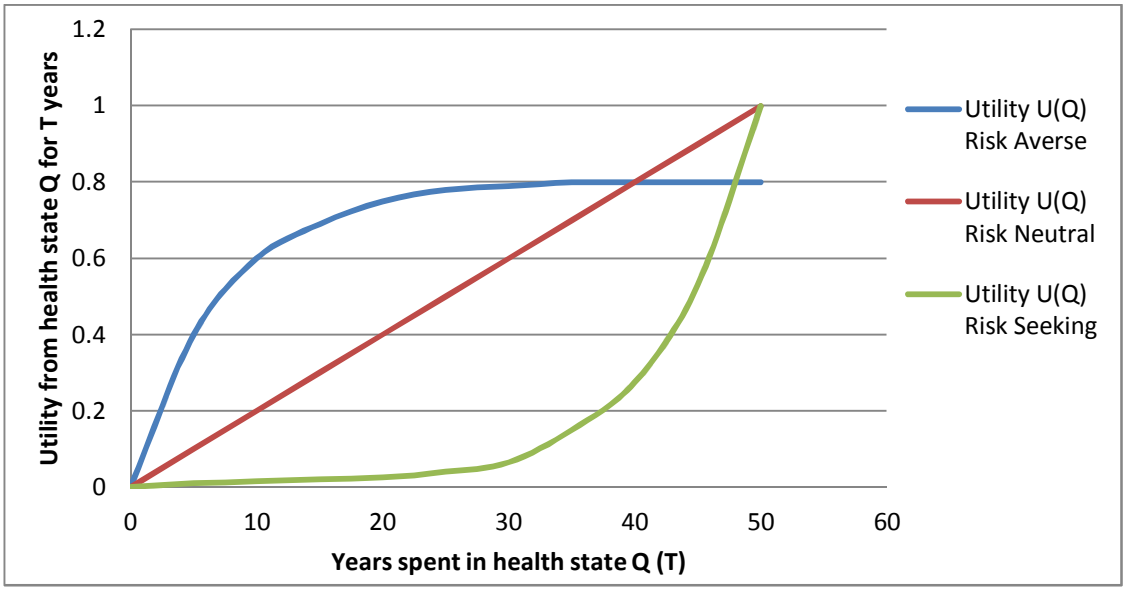


Figure 1: Utilities derived from health state Q for different risk attitudes

[1] The numbers used to create the figures above are only an example and are not necessarily an accurate estimation of real world preferences and utilities.

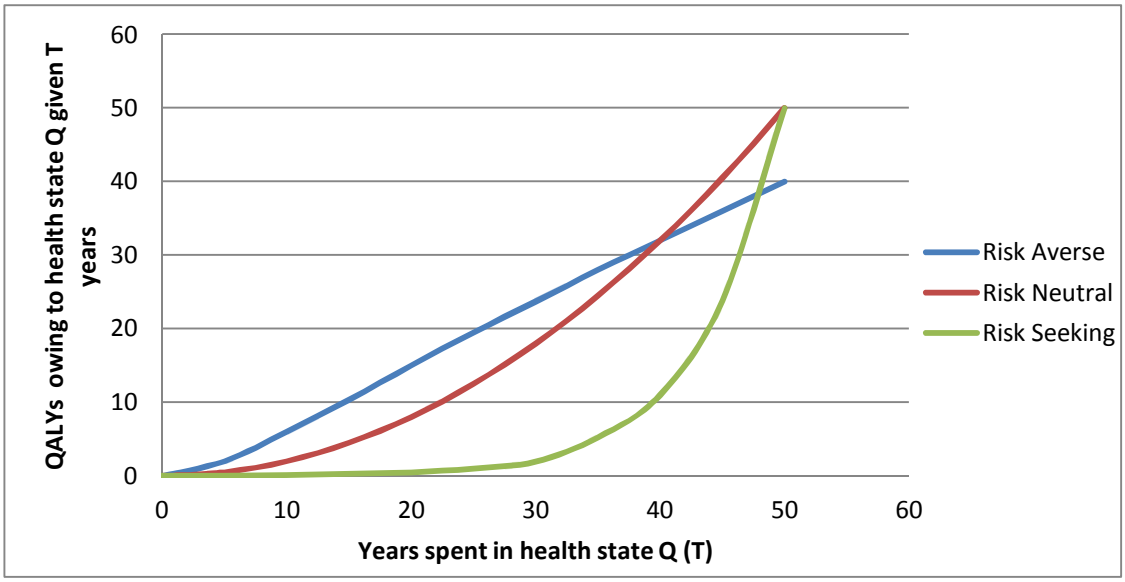


Figure 2: Impact of risk attitude on QALY calculation

It is clear that risk attitude potentially has a significant effect on the number of QALYs calculated. The assumption of risk neutrality underestimates the utility the individual derives from spending T years in health state Q which leads to an underestimation of the number of QALYs. This occurs only up to a certain number of years, after which the risk averse individual becomes indifferent to the number of years remaining and so the assumption of risk neutrality overestimates the number of QALYs the individual is subject to. Conversely, the assumption of risk neutrality overestimates the individual's utility and so over-estimates the number of QALYs which the individual is subject to given health state Q. This can greatly affect the calculations employed in an SEA or CUA thus care must be taken when considering risk attitude.

Several authors have proposed that the QALY model is still valid even if the risk neutrality assumption is violated, so long as the assumptions of utility independence and CPTO still hold. We have examined above the ways in which the CPTO assumption is often violated in reality and how this affects the QALY model's validity in measuring health utility. The assumption of utility independence implies that as the individual's health varies over time, the utility they derive from a particular health state does not depend on the utility attached to the health state which precedes or follows. Bleichrodt and Johannesson (1997) find weak evidence to support the assumption of utility independence, though they find strong evidence of the validity of the CPTO assumption. Yet Spencer and Robinson (2007) find empirical support for utility independence in health states which vary over time.

Constant proportional trade-offs and discounting

Constant proportional trade off (CPTO) alludes to the idea that the individual is willing to give up the same proportion of remaining life years for the same improved health state, regardless of the number of life years remaining.

Bleichrodt and Johannesson (1997) find evidence to support that CPTO holds, though their study suggests it holds only approximately. Attema and Brouwer (2010), in their literature review, find many violations of the CPTO assumption. Yet the impact of these violations varies: in some cases, the violation leads to higher valuations for a health state which could be attributed to a failure to correct for utility of life durations. In other instances, the opposite is true. This leads to an inaccurate calculation of QALYs. More worryingly for proponents of the QALY model is that the authors, even having corrected for utility of life duration curvature, find cause to reject CPTO. The authors conclude that the conventional QALY model is too simple for its purpose and that health state preferences depend on the number of life years spent in that health state; preferences are not constantly proportional.

Dolan and Stalmeier (2003) assess the CPTO[2] assumption as enforced under the time trade-off method for calculating QALY weights for health states associated with maximum endurable time (MET) preferences. Their study examines the responses of 91 participants to determine the extent to which they satisfy the CPTTO assumption or whether the proportional heuristic is a significant determinant of the responses. They conclude that most responses are generated by a proportional heuristic rather than the CPTTO and thus, one cannot readily accept that the QALY weights generated using the CPTTO are valid.

Many economists argue that the QALY model ought to be discounted in order to properly reflect individual preferences. Shepard and Thompson (1979) argue that people have positive time preferences for health i.e. they would prefer to experience health benefits sooner rather than later. This is an assumption supported by general utility theory. Drummond, Stoddart and Torrance (1987) propose the following discounted QALY model:

$$D_T = \sum_{i=1}^{T} \frac{1}{(1+r)^i}$$

where D_T is the appropriate discount factor, r is the discount rate and the number of QALYs (living T years in health state Q) is calculated by:

$$QALY = U(Q,T) \cdot T \cdot D_T$$

We use this formula to calculate the impact of different discount rates on QALY calculations, as illustrated in Figures 3 to 5 below[3]. Note that it is suggested that the assumption of risk neutrality implies that an individual does not exhibit time preferences and so there is no need to apply a discount rate if risk neutrality is assumed in QALY calculation. Nonetheless, we have calculated discounted QALYs for risk neutral individuals, as well as for risk averse and risk seeking individuals. As expected, for shorter durations, the choice of discount rate has little effect on the differences in QALYs calculated. Yet for longer durations, we see greater disparity amongst the discount rates chosen and the number of QALYs calculated.

[2] Dolan and Stalmeier refer to this as the constant proportional ***time*** trade-off (CPTTO) rather than CPTO.

[3] These calculations make use of the same numbers used to produce Figure 1.

This serves to illustrate that several factors can affect the outputs of the QALY model including risk attitude, duration and choice of discount rate.

Not all authors champion the use of a discounted QALY model. Bleichrodt and Gafni (1996) explore literature relating to constant rate discounted utility models and conclude that the constant discounted utility model is not empirically robust in its representation of individuals' preferences. They make reference to work by Lipscomb (1989) which looks at a discounted utility model and compares this to a "scenario strategy" which imposes less restrictions; the findings of this study are that the two different models generate reversed preferences (under the discounted model, B is preferred to A; under the scenario strategy, A is preferred to B). The authors suggest that the scenario strategy is the best indicator of the individual's true preferences as it imposes fewer restrictions than the discounted model, which reports the wrong prediction. Other studies which try to elicit a constant discount rate for health utilities have also rejected the constant rate discounted utility model (Redelmeier and Heller, 1993; Olsen, 1993a; Mackeigan et al., 1993; Cairns, 1994).

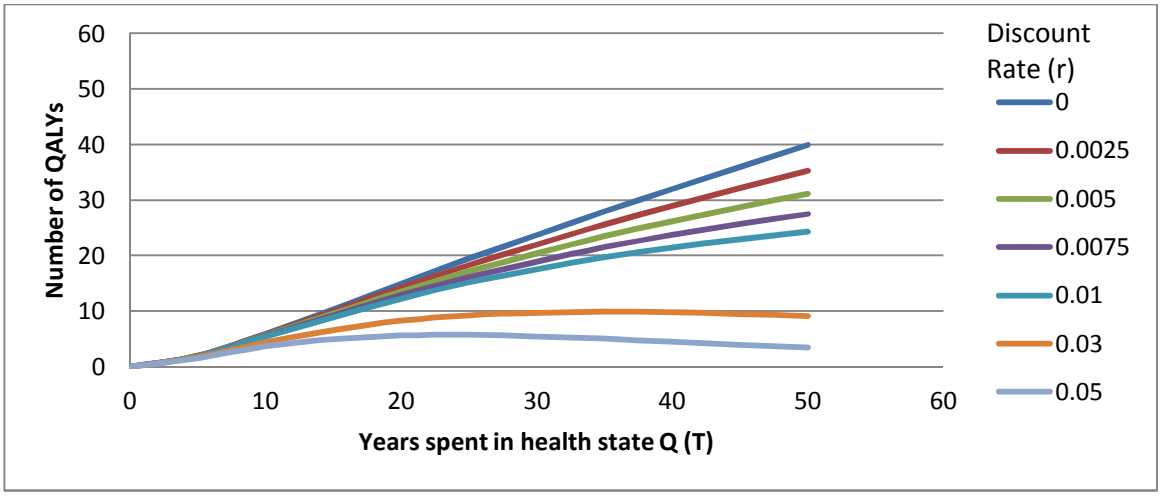


Figure 3: Impact of discount rate on QALY calculation assuming risk aversion

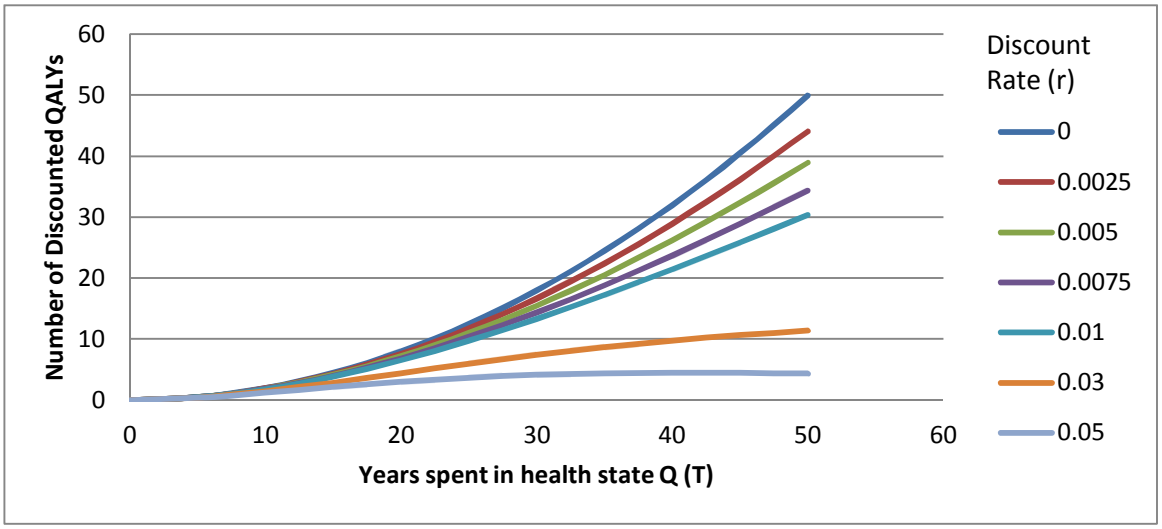


Figure 4: Impact of discount rate on QALY calculation assuming risk neutrality

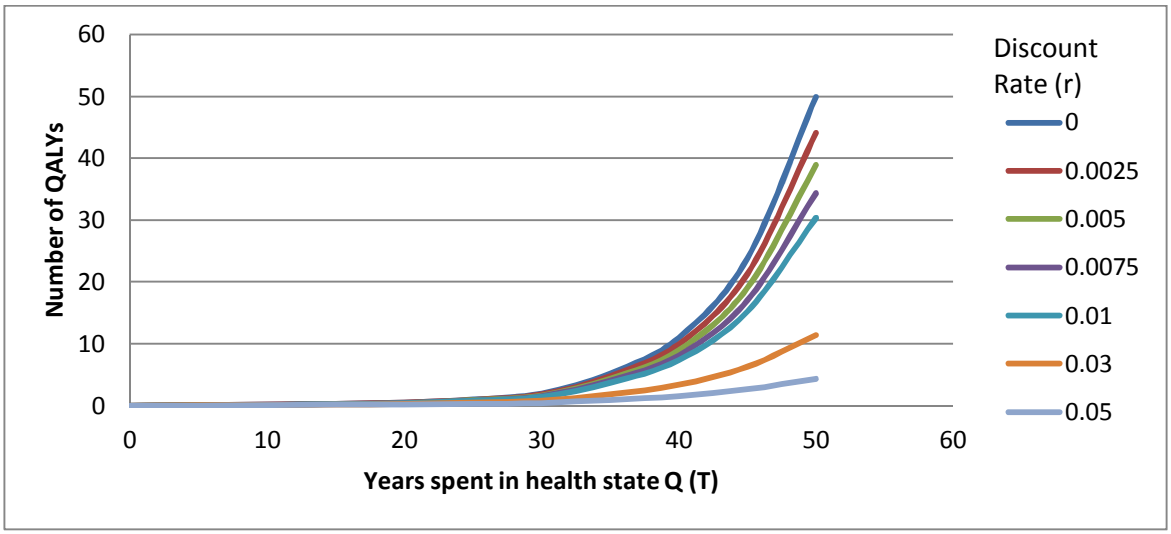


Figure 5: Impact of discount rate on QALY calculation assuming risk seeking behaviour

If used, one must proceed with caution when using discounted QALYs in cost-effectiveness analysis: often in CEAs a discount rate is applied to the entire equation, including both QALYs and monetary costs; one must take care not to apply a discounting factor twice to the QALY. There is the potential, also, to account for time preferences twice when QALY weights are generated using the TTO method, which, inherently, requires the individual to demonstrate time preferences.

1.2.2 Expected vs. Non-Expected Utility Theory

Failings of Expected Utility Theory in Health Economics

More recent work in utility theory looks at alternatives to expected utility theory (EUT) such as prospect theory and rank-dependent expected utility theory. Many authors propose that non-expected utility theory (NEUT) is able to overcome the challenges faced by utility theory relating to the assumptions on which the theory is built (i.e. CPTO, risk neutrality, etc.). The main proponents of NEUT for use in health utility assessments are, most notably, Bleichrodt, Doctor and Johannesson, amongst others. They build on work by Tversky and Kahneman (1979, 1992) and adapt it for application in health utility theory.

Introduction to Non-Expected Utility Theory

NEUT models include the rank-dependent model and the prospect theory model put forward by psychologists Tversky and Kahneman (1979, 1992) to better reflect human behaviour. Rank-dependent utility generalises EUT by allowing for probability weighting and prospect theory allows for both probability weighting and loss aversion. Loss aversion is an important element in an individual's decision-making process, particularly under the health context, and so it is optimal to use a model which allows for loss aversion to be taken into account, i.e. the model described under prospect theory.

Below is a table summarising the key differences between EUT model and the two key NEUT models:

Table 4.2: Overview of different utility models

Characteristic	Expected Utility model	Rank-Dependent Expected Utility Model	Prospect Theory Model
Overview of model	When faced with risky prospects, individuals make choices under uncertainty which maximise their expected utility	Generalised version of EUT model which allows for probability weighting in order to overcome Allais paradox	Two-stage decision-making process: first, individual studies lotteries (or "prospects") and eliminates all but two using decision heuristic. Second stage similar to other models as individual calculates utility based on possible outcomes and their respective probabilities

Probability Weighting	Linearity in probability assumed; no probability weighting	Function is inverse S-shaped: small probabilities over-weighted; large probabilities under-weighted	Separate probability weighting function for gains and for losses
Risk Attitude and Time Preferences	Utility assumed to be linear with respect to duration i.e. risk neutrality assumed	Individuals can be both risk averse with respect to duration and have linear utility for duration	Model assumes sign-dependence which allows loss aversion to be accounted for

Critique of Non-Expected Utility Theory

Miyamoto (1999) reiterates that EUT is systematically violated by individual preferences and considers the use of rank-dependent utility theory (RDUT) as a possible alternative for use in health utility assessments.

Doctor et al. (2004) perform tests of QALYs which are valid under both EUT and NEUT; specifically, the NEUT models they look at are the rank-dependent utility model and the prospect theory model. They conclude that QALYs are a valid method if the utility weights are derived on the basis of NEUT rather than EUT.

Abellan-Perpiñan et al. (2009) compare prospect theory to EUT and conclude that for decisions which involve risk, prospect theory offers a much more theoretically viable framework for preference decisions made under risk than EUT. However, they find that the utilities derived under one context are not necessarily transferable to a different context; they considered the transfer of utilities elicited under risk conditions to an inter-temporal context and find that the method which elicits the most consistent weights in one context doesn't do so for the alternative context. They suggest taking caution when assuming that utilities are transferrable to different contexts, i.e. situations other than risk.

The scope for non-expected utility theory in place of expected utility in health economics is promising. As of yet, however, the theory is still relatively new and there areas within the theory which are not fully understood and so it may be premature at this stage for it to be used practically in a crucial sector such as health.

1.2.3 Age weighting within the calculation of DALYs

Another question relating to weighting is the discounting and age weighting that may be employed for the calculation of disability adjusted life years and has been described as "probably the most controversial societal value choice to be built into the DALY estimates" (Lajoie, 2014, p.7). The underlying principle of age weighting is that individuals in the age group of 20-40 are valued higher than children or elderly because of their increased productivity and also because of their role as care takers for children or disabled or elderly relatives which has been referred to as 'productivity ageism' (Bognar, 2008, p.169).

This concept serves as a justification on the grounds of efficiency when it comes to rationing health care because if the age of the recipient of the treatment is taken into account then the benefits in terms of productivity are likely to be higher with a younger person (Bognar, 2008, p. 169). This mechanism allows DALYs to be seen as a measurement of social value of health which means that the input to society is the focus but it has been argued that from this viewpoint "many other variables besides age are relevant for the evaluation of health: income and level of education" (Bognar, 2008, p.174). Another concern with regards to age weighting is the possibility of double counting, as the factor of a reduced quality of life in old age might have already been taken into account. In other words this double counting will occur "if the age weight represents, not only a 'pure age effect', but also reflects the changing quality of life (QoL) and the contribution to society of individuals at different ages. In this case there would be double counting if age weights were included in an analysis which also includes QoL weights and indirect production benefits" (Richardson, 1999, p.14).

The concept of 'productivity ageism' has been argued as placing too much weight on the economic value of individual's productivity capacity and as a consequence age weighting has not been carried out in the more recent studies, such as GBD 2010, the GBD 2013 and the European Burden of Disease study. The Dutch disability weights appear not to have been age weighted.

A second justification for the use of age weighting is one based on equity arguments; in order to allocate scarce health resources, age is taken into account for fairness reasons, with this known as the 'fair innings ageism' (Bognar, 2008, p.169). The concept of fair innings has the underlying assumption "that people who had achieved old age or who were closely approaching it would not have their lives further prolonged when this could only be achieved at the cost of the lives of those who were not nearing old age" (Wagland, 2012, p.1). The fair innings argument was introduced in the mid-1980s by John Harris who explains in his book that "the fair innings argument takes the view that there is some span of years that we consider a reasonable life, a fair innings" (Harris, 1985, p.367). The fair innings argument is however only valid if all other factors besides age are equal, which is rarely the case and has been acknowledged by John Harris himself (Harris, 1985; Rivlin, 2000).

Age weighting is not relevant to the use of DALYs within the context of REACH SEAs. Within a cost-benefit analysis framework productivity effects can be taken into account directly when estimating health impacts. In this respect, it is important that analysts ensure that any age weighting effects are excluded from the DALYs that are used in assessments.

Annex 2:

Mapping Substance Properties against Health Effects

Table A2-1: Substances currently subject to Authorisation						
Substance	**EC Number**	**CAS Number**	**Classification on which Authorisation based**	**Principle endpoint(s) of concern**	**Potential health correlates**	**Potential major human condition/ disease warranting Impact assessment[1]**
Ammonium dichromate Potassium dichromate	232-143-1 231-906-6	7789-09-5 7778-50-9	Carcinogenic	Lung[A, b, d] & other cancers[d]	Lung cancer & potentially other cancers Infertility in males & females Developmental delay	Lung cancer Reduced male fertility Reduced female fertility Increase in spontaneous abortion Low birth weight
			Mutagenic	In vivo somatic cell mutagen[A, c, d]		
			Toxic for reproduction	Reduced fertility (e.g. reduction in implantation, increase in resorption)[A, d]; Developmental toxicity (e.g. reduced fetal weight, & crown-rump length; delay in cranial ossification)[A, d]		
Sodium dichromate Sodium chromate	234-190-3 231-889-5	7789-12-0; 10588-01-9 7775-11-3	Carcinogenic	Lung[A, b, d] and other cancers[d]	Lung cancer & potentially other cancers Infertility in males & females Developmental delay	Lung cancer Male infertility Increase in spontaneous abortion Low birth weight
			Mutagenic	In vivo somatic cell mutagen[A, c, d]		
			Toxic for reproduction	Reduced fertility (e.g. reduction in implantation, increase in resorption)[A, d]; Developmental toxicity (e.g. reduced fetal weight, & crown-rump length; delay in cranial ossification)[A, d]		
Potassium chromate Chromium trioxide	232-140-5 215-607-8	7789-00-6 1333-82-0	Carcinogenic	Lung[A, b, d] and other cancers[d]	Lung cancer & potentially other cancers	Lung cancer
			Mutagenic	In vivo somatic cell mutagen[A, c, d]		
Acids generated from chromium trioxide & their oligomers Group containing: Chromic acid, Dichromic acid, Oligomers of chromic acid & dichromic acid	231-801-5, 236-881-5	7738-94-5, 13530-68-2	Carcinogenic	Lung cancer[d]	Lung cancer	Lung cancer

Table A2-1: Substances currently subject to Authorisation						
Substance	**EC Number**	**CAS Number**	**Classification on which Authorisation based**	**Principle endpoint(s) of concern**	**Potential health correlates**	**Potential major human condition/ disease warranting Impact assessment[1]**
Lead chromate	231-846-0	7758-97-6	Carcinogenic	Lung cancer[A, b, c, d] Urinary bladder cancer[A, b, c, d] Kidney cancer [A, b, c, d]	Lung cancer & potentially other cancers	Lung cancer Urinary bladder cancer Kidney cancer
			Toxic for reproduction	Impairment of male and female fertility[A, d] Fetotoxicity (e.g. reduced fetal weight & survival; developmental neuro-impairment)[A, b, d]	Male & female infertility Fetal growth impairment; Neurodevelopmental deficiency	Male infertility Female infertility Spontaneous abortion Low birth weight Impaired cognitive development (e.g. based on IQ impairment)
Trichloroethylene	201-167-4	79-01-6	Carcinogenic	Lung cancer[A] Liver cancer[A] Renal adenoma & carcinoma[A, b]	Cancer of various tissues	Lung cancer Liver cancer Kidney cancer
2,4 – Dinitrotoluene (2,4-DNT)	204-450-0	121-14-2	Carcinogenic	Hepatobiliary cancer[A, b, c, d] Urothelial[b, c] & renal cell cancer[A, b, c] Other cancers[A, c]	Cancer of various tissues	Lung cancer Liver cancer Kidney cancer Bladder cancer
Tris (2-chloroethyl) phosphate (TCEP)	204-118-5	115-96-8	Toxic for reproduction	Impaired male & female fertility[A]	Male & female infertility	Male infertility Female infertility
Diarsenic pentaoxide Diarsenic trioxide	215-116-9 215-481-4	1303-28-2 1327-53-3	Carcinogenic	Lung tumours [A, B, c, d] Benign hepatic tumours[A, b, c, d] Urinary bladder tumours[A, B, c, d] Renal tumours[A, b, c, d] Skin cancer (mainly squamous cell carcinoma)[B, c, d] Prostate[b, c, d] Other cancers [A, b, c, d] Trans-generational or hereditary types of cancer[A, c, d]	Cancer of various tissues	Lung cancer Skin cancer Urinary bladder cancer Kidney cancer

Table A2-1: Substances currently subject to Authorisation						
Substance	**EC Number**	**CAS Number**	**Classification on which Authorisation based**	**Principle endpoint(s) of concern**	**Potential health correlates**	**Potential major human condition/ disease warranting Impact assessment[1]**
Lead sulfochromate yellow (C.I. Pigment Yellow 34)	215-693-7	1344-37-2	Carcinogenic	Lung tumours [A, b, c, d] Renal tumours[A, b, c, d] Skin cancer A, B, [c, d] Muscle cancer[A, c, d]	Cancer of various tissues Infertility	Lung cancer Skin cancer Urinary bladder cancer Kidney cancer Muscle cancer Male infertility Increase in spontaneous abortions in partners of exposed males
Lead chromate molybdate sulphate red (C.I. Pigment Red 104)	235-759-9	12656-85-8	Toxic for reproduction	Impaired male fertility & testicular development[A, b, c, d]		
Benzyl butyl phthalate (BBP)	201-622-7	85-68-7	Toxic for reproduction	Testicular atrophy[A, c, d] Reduced sperm production[A, c, d] Male infertility[A, b, c, d] Female infertility[a] Developmental anomalies/abnormalities (e.g. hypospadias; cryptorchidism; reduced A:G ratio) in males[A, b, c, d] Altered development of male reproductive organs[A, c, d] Altered development of female organs[a]	Reduced male fertility Reduced female fertility Developmental abnormalities in males (e.g. hypospadias; cryptorchidism) Minor developmental anomalies in males (e.g. reduction in A:G ratio)	Male infertility Female infertility Hypospadias Cryptorchidism
Bis(2-ethylhexyl) phthalate (DEHP)	204-211-0	117-81-7	Toxic for reproduction	Testicular atrophy[A, c, d] Reduced sperm production[A, c, d] Male infertility [A, b, c, d] Female infertility[a] Developmental anomalies (e.g. reduced A:G ratio in males; nipple retention in males) Altered development of male reproductive organs[A, c, d]	Reduced fertility, particularly in males Developmental abnormalities in males (e.g. hypospadias; cryptorchidism) Minor developmental anomalies in males (e.g. reduction in A:G ratio)	Male infertility female infertility Hypospadias Cryptorchidism
Dibutyl phthalate	201-557-4	84-74-2	Toxic for	Testicular atrophy[A, c, d]	Reduced fertility, particularly in	Male infertility

Table A2-1: Substances currently subject to Authorisation						
Substance	**EC Number**	**CAS Number**	**Classification on which Authorisation based**	**Principle endpoint(s) of concern**	**Potential health correlates**	**Potential major human condition/ disease warranting Impact assessment[1]**
(DBP)			reproduction	Reduced sperm production[A, c, d] Male infertility [A, b, c, d] Female infertility[A] Developmental anomalies/abnormalities (e.g. cleft pallet; skeletal changes; hypospadias; cryptorchidism; mammary gland changes in males; reduced A:G ratio in males; nipple retention in males) & delays in development (e.g. time of preputial separation) [A, b, c, d]	males Developmental abnormalities in males (e.g. cleft palate, hypospadias; cryptorchidism) Minor developmental anomalies in males (e.g. reduction in A:G ratio)	Female infertility Hypospadias Cryptorchidism Cleft pallet
Diisobutyl phthalate (DIBP)	201-553-2	84-69-5	Toxic for reproduction	Testicular atrophy[A, c, d] Reduced sperm production[A, c, d] Male infertility [A, b, c, d] Developmental anomalies/abnormalities (e.g. cleft pallet; hypospadias; cryptorchidism; mammary gland changes in males; reduced A:G ratio in males; nipple retention in males) & delays in development (e.g. time of preputial separation)[A, b, c, d]	Impaired male fertility Developmental abnormalities in males (e.g. cleft palate, hypospadias; cryptorchidism) Minor developmental anomalies in males (e.g. reduction in A:G ratio	Male infertility Hypospadias Cryptorchidism Cleft pallet
4,4'-Diaminodiphenyl methane (MDA)	202-974-4	101-77-9	Carcinogenic	Benign & malignant hepatic cancer[A, c] Benign & malignant thyroid cancer[A, c] Urinary bladder cancer[b]	Cancer of various tissues	Liver cancer Thyroid cancer Urinary bladder cancer
Pentazinc chromate octahydroxide	256-418-0	49663-84-5	Carcinogenic	Lung[Abd] and other cancers[d]	Lung cancer & potentially other cancers	Lung cancer
Potassium hydroxyoctaoxodi zincatedichromate	234-329-8	11103-86-9				
Dichromium	246-356-2	24613-89-6				

Table A2-1: Substances currently subject to Authorisation						
Substance	**EC Number**	**CAS Number**	**Classification on which Authorisation based**	**Principle endpoint(s) of concern**	**Potential health correlates**	**Potential major human condition/ disease warranting Impact assessment[1]**
tris(chromate)						
Strontium chromate	232-142-6	7789-06-2				
2,2'-dichloro-4,4'-methylenedianiline (MOCA)	202-918-9	101-14-4	Carcinogenic	Lung cancer[A, c, d] Liver tumours[A, c, d] Heamangiosarcoma[A, c, d] Mammary tumours[A, c, d] Bladder cancer [A, b, c, d]	Cancer of various tissues	Bladder cancer
1,2-Dichloroethane (EDC)	203-458-1	107-06-2	Carcinogenic	Lung cancer[A, c, d] Lymphatic cancer[A, b, c, d] Liver cancer[A, c, d] Mammary cancer[A, c, d] Uterine cancer[A, c, d] Haematopoietic cancer[A, b, c, d] Stomach cancer[b, c, d] Pancreatic cancer[b, c, d]	Cancer of various tissues	Lymphatic cancer Haematopoietic cancer Stomach cancer Pancreatic cancer
Arsenic acid	231-901-9	7778-39-4	Carcinogenic	Urinary bladder cancer[B c, d] Lung cancer[A, B, c, d] Skin cancer[B, c, d] Liver cancer [A, b, c d] Kidney cancer[b, c, d] Prostate cancer[b, c, d]	Cancer of various tissues	Urinary bladder cancer Lung cancer Skin cancer Liver cancer Kidney cancer Prostate cancer
Formaldehyde, oligomeric reaction products with aniline (technical MDA)	500-036-1	25214-70-4	Carcinogenic	Benign & malignant hepatic cancer[A, c, d] Benign & malignant thyroid cancer[A, c, d] Urinary bladder cancer[b, d]	Cancer of various tissues	Liver cancer Thyroid cancer Bladder cancer
Bis(2-methoxyethyl) ether (Diglyme)	203-924-4	111-96-6	Toxic for reproduction	Testicular atrophy[A] Reduced sperm production[A] Male infertility[A]	Impaired male fertility	Male infertility Increase in spontaneous abortion Renal and urinogenital

Table A2-1: Substances currently subject to Authorisation						
Substance	**EC Number**	**CAS Number**	**Classification on which Authorisation based**	**Principle endpoint(s) of concern**	**Potential health correlates**	**Potential major human condition/ disease warranting Impact assessment[1]**
				Fetal resorption[A] Major developmental abnormalities (e.g. abnormalities of urinogenital and kidney, cardiovascular, neural tube, limb structures)[A]	Increase in spontaneous abortion Major abnormalities of organs and limbs	abnormalities Cardiovascular abnormalities Abnormalities of limbs Neural tube defects (e.g. spina bifida)

- Key: 1 –Major human health impact that may warrant consideration as part of a comprehensive socioeconomic assessment.
  - A - Animal evidence identifies effect
  - a – Animal evidence suggestive of effect
  - B – Human evidence identifies effect
  - B – Human evidence suggestive of effect
  - c – Other experimental evidence indicative of effect
  - d – Mechanistic evidence or read-across supportive of effect

Table A1-2: Substances recommended for Authorisation List					
Substance	**EC Number**	**CAS Number**	**Classification on which SVHC status based**	**Principle hazard endpoint(s) of concern**	**Potential health conditions / disease for impact assessment**
1,2-Benzenedicarboxylic acid, di-C6-8-branched alkyl esters, C7-rich	276-158-1	71888-89-6	Toxic for reproduction	Male and female infertility[A, d] Embryolethality[A, d] Reduced foetal weight[A, d] Developmental abnormalities (e.g. anasarca, exencephaly, cranioschisis, anophthalmia, micophthalmia, cleft palate, cardiovascular abnormalities, ectopic gonads, skeletal abnormalities)[A, d] Developmental anomalies (e.g. reduced AG distance, developmental delays, retention of thoracic nipples)[A, d]	Reduced fertility in males & females Increase in spontaneous abortion Low birth weight Developmental abnormalities and anomalies (e.g. cranial and other skeletal abnormalities, abnormalities of eye, cardiovascular abnormalities, cleft palate; cleft palate)
Disodium tetraborate, anhydrous	215-540-4	1330-43-4, 12179-04-3, 1303-96-4	Toxic for reproduction	Testicular degenerative changes[A, d] Reduced sperm development[A, d] Reduced foetal growth[A, d] Minor developmental anomalies (rib anomolies)[A, d] Developmental abnormalities (cardiovascular)[A, d]	Male infertility Low birth weight Cardiovascular developmental abnormalities
Acetic acid, lead salt, basic	257-175-3	51404-69-4	Toxic for reproduction	Impairment of male and female fertility[A, d] Fetotoxicity (e.g. reduced fetal size & survival; developmental neuro-impairment)[A, d]	Reduced male fertility Reduced female fertility Spontaneous abortion Low birth weight Impaired cognitive development (e.g. based on IQ impairment)
Lead monoxide (lead oxide)	215-267-0	1317-36-8			
Pentalead tetraoxide sulphate	235-067-7	12065-90-6			
Orange lead (lead tetroxide)	215-235-6	1314-41-6			
Pyrochlore, antimony lead yellow	232-382-1	8012-00-8			

Table A1-2: Substances recommended for Authorisation List					
Substance	**EC Number**	**CAS Number**	**Classification on which SVHC status based**	**Principle hazard endpoint(s) of concern**	**Potential health conditions / disease for impact assessment**
Silicic acid, lead salt	234-363-3	11120-22-2			
Tetralead trioxide sulphate	235-380-9	12202-17-4			
1-bromopropane (n-propyl bromide)	203-445-0	106-94-5	Toxic for reproduction	Male & female infertility[A] Reduced sperm quantity and quality[A] Elongation of oestrus cycle[A] Ovarian follicular cysts[A] Reduced foetal growth [A] Developmental anomalies (skeletal)[A]	Reduced male fertility Reduced female fertility Ovarian cysts Low birth weight Skeletal developmental abnormalities
Bis(2-methoxyethyl) phthalate	204-212-6	117-82-8	Toxic for reproduction	Testicular atrophy[A, d] Male infertility [A, d] Developmental abnormalities (hydronephrosis, cardiovascular, skeletal)[A] Developmental delays (e.g. time of preputial separation)[d] Developmental abnormalities (e.g. hypospadias; cryptorchidism; mammary gland changes in males; reduced A:G ratio in males; nipple retention in males)[d]	Reduced male fertility Hypospadias Cryptorchidism; Cleft pallet Developmental abnormalities of renal and cardiovascular systems
1,2-Benzenedicarboxylic acid, di-C7-11-branched and linear alkyl esters	271-084-6	68515-42-4	Toxic for reproduction	Testicular atrophy[A] Foetal loss[A] Developmental abnormalities (e.g CNS, skeleton, urogenital tract, renal agenesis)[A]	Reduced male fertility Spontaneous abortion CNS diseases
Dipentyl phthalate (DPP)	205-017-9	131-18-0	Toxic for reproduction	Testicular damage[A, d] Reduced sperm cell development[A,d] Impaired male fertility[A,d] Impaired female fertility[A,d] Developmental abnormalities & anomalies (e.g. abnormalities of male reproductive organs,	Reduced male fertility Reduced female fertility Hypospadias Cryptorchidism

Table A1-2: Substances recommended for Authorisation List					
Substance	**EC Number**	**CAS Number**	**Classification on which SVHC status based**	**Principle hazard endpoint(s) of concern**	**Potential health conditions / disease for impact assessment**
				reduced AG distance, developmental delays, retention of thoracic nipples) [d]	
Diboron trioxide	215-125-8	1303-86-2	Toxic for reproduction	Testicular degenerative changes [A, d] Reduced sperm development [A, d] Reduced foetal growth [A, d] Developmental abnormalities (cardiovascular) [A, d] Minor developmental anomalies (rib anomolies) [A, d]	Reduced male fertility Low birth weight Developmental abnormalities of cardiovascular system
Boric acid	233-139-2, 234-343-4	10043-35-3, 11113-50-1			
Tetraboron disodium heptaoxide, hydrate	235-541-3	12267-73-1			
N-pentyl-isopentylphthalate	-	776297-69-9	Toxic for reproduction	Testicular damage [d] Reduced sperm cell development [d] Impaired male fertility [d] Impaired female fertility [d] Developmental abnormalities & anomalies (e.g. abnormalities of male reproductive organs, reduced AG distance, developmental delays, retention of thoracic nipples) [d]	Reduced male fertility Reduced female fertility Hypospadias Cryptorchidism
Diisopentylphthalate	210-088-4	605-50-5	Toxic for reproduction	Testicular damage [A, d] Reduced sperm quality and quantity [A, d] Foetal resorption [A] Developmental abnormalities & anomalies (e.g. abnormalities of male reproductive organs, reduced AG distance, developmental delays, retention of thoracic nipples) [d]	Reduced male fertility Spontaneous abortion Hypospadias Cryptorchidism
1,2-Benzenedicarboxylic acid, dipentylester,	284-032-2	84777-06-0	Toxic for reproduction	Testicular damage [d] Reduced sperm quality and quantity [d] Developmental abnormalities & anomalies (e.g.	Reduced male fertility Hypospadias Cryptorchidism

Table A1-2: Substances recommended for Authorisation List						
Substance	**EC Number**	**CAS Number**	**Classification on which SVHC status based**	**Principle hazard endpoint(s) of concern**	**Potential health conditions / disease for impact assessment**	
branched and linear				abnormalities of male reproductive organs, reduced AG distance, developmental delays, retention of thoracic nipples)[d]		
Pitch, coal tar, high temp.	266-028-2	65996-93-2	Carcinogenic	Lung cancer[A, B, c] Urinary bladder cancer[B, c]	Lung and bladder cancer	Lung cancer Urinary bladder cancer
Key: 1 –Major human health impact that may warraent consideration as part of a comprehensive socioeconomic assessment. A - Animal evidence identifies effect a – Animal evidence suggestive of effect B – Human evidence identifies effect B – Human evidence suggestive of effect c – Other experimental evidence indicative of effect d – Mechanistic evidence or read-across supportive of effect						

Table A2-3: Substances which are subject to or proposed for Restriction					
Substance	**EC Number**	**CAS Number**	**Risk on which Restriction is based**	**Principle hazard endpoint(s) associated with risk of concern**	**Potential major human condition/ disease warranting impact assessment** • *(entries in italics are the potential human conditions as assessed by the consultants as no effects were identified* • *y the DS or in the opinion)*
Substances which are the subject of a restriction proposal					
Methanol	200-659-6	67-56-1			
Cadmium and its compounds (in artist paints)	231-152-8	7440-43-9	Carcinogenic Repeated dose toxicity	Postmenopausal breast cancer[b,d] Organ toxicity (osteoporosis)[B]	Postmenopausal breast cancer Osteoporosis leading to bone fractures
Chrysotile	---	12001-29-5, 132207-32-0	Carcinogenic	Lung cancer and mesothelioma[A,B]	Lung cancer and mesothelioma
bisphenol A; 4,4'-isopropylidened	201-245-8	80-05-7	Toxic for reproduction	Developmental toxicity (alteration of memory and learning functions concurrent with a decrease in the expression of NMDA receptors;	Alterations in memory and learning function (IQ function) Endometriosis; ovarian cysts; disruption of ovarian

Table A2-3: Substances which are subject to or proposed for Restriction					
Substance	**EC Number**	**CAS Number**	**Risk on which Restriction is based**	**Principle hazard endpoint(s) associated with risk of concern**	**Potential major human condition/ disease warranting impact assessment** • *(entries in italics are the potential human conditions as assessed by the consultants as no effects were identified* • *y the DS or in the opinion)*
iphenol				disruption of ovarian cycles; increase in body weight and cholesterol; effect on the buds and terminal breast ducts)[A]	cycles Raised cholesterol Breast cancer
Ammonium salts	---	---	Irritation	Irritation (respiratory and ocular irritation)[B]	Respiratory irritation
1-methyl-2-pyrrolidone (NMP)	212-828-1	872-50-4	Toxic for reproduction Repeated dose toxicity	Developmental toxicity (reduced fetal body weight; reduced live foetuses; delayed ossification; increased malformations)[A] General toxicity (decrease in body weight and body weight gain; mortality)[A]	Intra Uterine Growth Retardation (IUGR) Spontaneous abortions/stillbirths Decrease in body weight Decrease in body weight gain Decrease in food consumption Respiratory tract irritation
Lead and its compounds	231-100-4	7439-92-1	Toxic for reproduction	Developmental toxicity (developmental neurotoxicity[A,B])	Impaired cognitive development (IQ reduction)
1,4-Dichlorobenzene (p-dichlorobenzene)	203-400-5	106-46-7	Carcinogenic	Liver cancer[B]	Liver cancer
Chromium VI	---	18540-29-9	Sensitization	Skin sensitization (contact dermatitis)[A,B]	Allergic contact dermatitis
Benzyl butyl phthalate	607-430-00-3	85-68-7	Toxic for reproduction	Developmental toxicity (small male reproductive organ, minimal testis atrophy, reduced spermatocyte development, mammary gland changes, degeneration of seminiferous tubules and oligo-/azospermia in epididymides, reduced anogenital distance)[A]	Reduced male fertility
Diisobutyl phthalate	201-553-2	84-69-5			
Bis(2-ethylhexyl) phthalate	204-211-0	117-81-7			

Table A2-3: Substances which are subject to or proposed for Restriction					
Substance	**EC Number**	**CAS Number**	**Risk on which Restriction is based**	**Principle hazard endpoint(s) associated with risk of concern**	**Potential major human condition/ disease warranting impact assessment** • *(entries in italics are the potential human conditions as assessed by the consultants as no effects were identified* • *y the DS or in the opinion)*
(DEHP)					
Dibutyl phthalate (DBP)	201-557-4	84-74-2			
Mercury	231-106-7	7439-97-6	Repated dose toxicity Toxic for reproduction	neurotoxic and neurodevelopmental effects (not further specified)	*Neurological and behavioral disorders* *Impaired cognitive development (IQ reduction)*
Phenylmercuric octanoate	---	13864-38-5	Repeated dose toxicity	Organ toxicity (kidney damage e.g., tubular dilation, atrophy, granularity, fibrosis[A]; effects on central nervous system[B])	*Nephritic syndrome/decrease in renal function* *Neurological and behavioral disorders*
Phenylmercury 2-ethylhexanoate	236-326-7	13302-00-6			
Phenylmercury neodecanoate	247-783-7	26545-49-3			
Phenylmercury acetate	200-532-5	62-38-4			
Phenylmercury propionate	203-094-3	103-27-5			
Dimethylfumarate	210-849-0	624-49-7	Sensitization Irritation	Skin sensitization (contact dermatitis)[A,B] Skin irritation[A,B]	Contact dermatitis Skin irritation
Current Restriction Intentions – no Dossiers / Documents available					
Octamethylcyclo-tetrasiloxane (D4);	209-136-7	556-67-2			
Decamethylcyclopentasiloxane (D5)	208-764-9	541-02-6			
Grill lighters fluids and fuels	---	---			

Table A2-3: Substances which are subject to or proposed for Restriction					
Substance	**EC Number**	**CAS Number**	**Risk on which Restriction is based**	**Principle hazard endpoint(s) associated with risk of concern**	**Potential major human condition/ disease warranting impact assessment** • *(entries in italics are the potential human conditions as assessed by the consultants as no effects were identified* • *y the DS or in the opinion)*
for decorative lamps labelled R65 or H304					
Perfluorooctanoic acid (PFOA, , EC), and any other linear or branched perfluoroheptyl derivative as described	206-397-9	335-67-1			
N,N-dimethylformamide; dimethyl formamide	200-679-5	68-12-2			
Yellow highlighting indicates that the required data are not available from the Background Documents or currently available Dossiers.					

- Key: A - Animal evidence identifies effect
  - a – Animal evidence suggestive of effect
  - B – Human evidence identifies effect
  - b – Human evidence suggestive of effect
  - c – Other experimental evidence indicative of effect

  d – Mechanistic evidence or read-across supportive of effect

Data sources

Agency for Toxic Substances and Disease Registry (2014): Home, information downloaded from http://www.atsdr.cdc.gov/

Agency for Toxic Substances and Disease Registry (undated a): Chemical and physical information: arsenic, information downloaded from http://www.atsdr.cdc.gov/toxprofiles/tp2-c4.pdf

Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) (2014): CLH report: Proposal for Harmonised Classification and Labelling: Diisobutylphthalate (DIBP), information downloaded from http://echa.europa.eu/documents/10162/59594fc5-519a-4e97-b0f8-7a45b5db04ce

California Environmental Protection Agency (2004): Chemical Meeting the criteria for listing via the authoritative bodies mechanism, information downloaded from http://oehha.ca.gov/prop65/CRNR_notices/admin_listing/intent_to_list/pdf_zip/NOIL_ABpkg21a.pdf

Chemicalbook (2010a): Pigment Yellow 34, information downloaded from http://www.chemicalbook.com/ChemicalProductProperty_EN_CB8512280.htm

Chemicalbook (2010b): Arsenic pentoxide, information downloaded from http://www.chemicalbook.com/ChemicalProductProperty_EN_CB6369156.htm

Chemical substances (2010): C.I. Pigment Yellow 34, information downloaded from http://www.chemicalsubstanceschimiques.gc.ca/challenge-defi/summary-sommaire/batch-lot-2/1344-37-2-eng.php

Commission of the European Communities (2005): Commission Regulation (EC) No 1048/2005 of 13 June 2005 amending Regulation (EC) No 2032/2003 on the second phase of the 10-year work programme referred to in Article 16(2) of Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market, information downloaded from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32005R1048&from=EN

Danish Competent Authority for REACH (2011): Annex XV Restriction Report: Proposal for a Restriction; Substance Name: BIS(2-EHTYLHEXYL)PHTHLATE (DEHP), BENZYL BUTYL PHTHALATE (BBP), DIBUTYL PHTHALATE (DBP), DIISOBUTYL PHTHALATE (DIBP), information downloaded from http://echa.europa.eu/documents/10162/c6781e1e-1128-45c2-bf48-8890876fa719

Dent MP (2007): Strengths and limitations of using repeat-dose toxicity studies to predict effects on fertility, information downloaded from http://www.sciencedirect.com/science/article/pii/S0273230007000438

ECBI/56/03 Revision 2. Ispra, May 24, 2004 SUMMARY RECORD Meeting of the Commission Working Group on the Classification and Labelling of Dangerous Substances Ispra, 14-16 May 2003.

ECBI/42/02 Rev. 2 Ispra, 25 November 2002 SUMMARY RECORD Meeting of the Commission Working Group on the Classification and Labelling of Dangerous Substances Ljubljana, 15-17 May.

E.C. (2005). European Union Risk Assessment Report - Chromium trioxide (CAS-No: 1333-82-0), sodium chromate (CAS-No:7775-11-3), sodium dichromate (CAS-No: 10588-01-9), ammonium dichromate (CAS-No: 7789-09-5) and potassium dichromate (CAS-No: 7778-50-9) Risk Assessment. 415 p. (EUR 21508 EN - Volume: 53).

ECHA (2014): Draft background document for n-Pentyl-isopentyl phthalate, information downloaded from http://echa.europa.eu/documents/10162/ae7f5e42-8f94-4175-bc09-bba0656ef7b8

ECHA (2013): Committee for Risk Assessment (RAC); Committee for Socio-economic Analysis (SEAC): Opinion on an Application for Authorisation for Bis(2-ethylhexyl) phthalate (DEHP) used in the processing of a stop-off formulation during the diffusion bonding and manufacture of aero engine fan blades, information downloaded from http://echa.europa.eu/documents/10162/54858a96-fcc1-4260-9b3c-631c40cc4234

ECHA (2012): Responses to Comments Document (RCOM) on ECHA's Draft 4th Recommendation for Arsenic Acid (EC number: 231-901-9), information downloaded from http://echa.europa.eu/documents/10162/0363346d-0e3d-4c7d-b52f-c030362bda1d

ECHA (2011a): Recommendation of the European Chemicals Agency (ECHA) of 20 December 2011 for the inclusion of substances in Annex XIV, information downloaded from http://echa.europa.eu/documents/10162/13640/3rd_a_xiv_recommendation_20dec2011_en.pdf

ECHA (2011b): Responses to Comments Document (RCOM) on ECHA's Draft 3rd Recommendation for the Group of recommended Chromium (VI) Substances, information downloaded from http://echa.europa.eu/documents/10162/4b73a481-7530-4f1f-9403-901058e7ff4b

ECHA (2011c): Background document for Chromium trioxide, information downloaded from http://echa.europa.eu/documents/10162/ca724e1d-6d7c-4c61-9473-17dd92ababb0

ECHA (2011d): Background document for ammonium dichromate, information downloaded from http://echa.europa.eu/documents/10162/ee3db677-c1a2-497c-9ae5-f27dd024001c

ECHA (2011e): Member State Committee support document for identification of Pentazinc Chromate Octahydroxide as a substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/suppdoc_pentazinc_chromate_octahydroxide_20111129_en.pdf

ECHA (2011f): Member State Committee support document for identification of Potassium hydroxyoctaoxodizincatedichromate(1-) as a substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/suppdoc_potassium_+hydroxyoctaoxodizincatedichromate_1_20111129_en.pdf

ECHA (2011g): Member State Committee Support Document for Identification of Dichromium Tris(chromate) as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/suppdoc_dichromium+tris_20111129_en.pdf

ECHA (2011h): Member State Committee Support Document for Identification of Strontium Chromate as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_msca_strontiumchromate_ec_232-142-6_en.pdf

ECHA (2011i): Member State Committee Support Document for Identification of 2,2'-dichloro-4,4'-methylenedianiline as a substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/suppdoc_moca_20111124_en.pdf

ECHA (2011j): Member State Committee Support Document for identification of 1,2-dichloroethane as a Substance of Very High Concern because of its CMR properties, information downloaded from

http://echa.europa.eu/documents/10162/13638/suppdoc_1_2_dichloroethane_20111128_en.pdf

ECHA (2011k): Annex XV dossier: proposal for identification of a substance as a category 1A or 1B CMR, PBT, vPvB or a substance of an equivalent concern: 1,2-dichloroethane, information downloaded from http://echa.europa.eu/documents/10162/57c002b0-db53-4f7c-ac5e-7c8dd46a53bf

ECHA (2011l): Member State Committee Support Document for identification of arsenic acid as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/suppdoc_arsenic+acid_20111124_en.pdf

ECHA (2011m): Support Document for identification of Formaldehyde, oligomeric reaction products with Aniline as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_technical_mda_en.pdf

ECHA (2011n): Annex XV dossier: Proposal for identification of a substance as a category 1A or 1B CMR, PBT, vPvB or a substance of an equivalent level of concern: 1,2-Benzenedicarboxylic acid, di-C6-8-branched alkyl esters, C7-rich (DIHP), information downloaded from http://echa.europa.eu/documents/10162/eec0b364-e29e-48f8-970c-a4cdb78465b8

ECHA (2010a): Background document for Diisobutyl phthalate (DIBP), information downloaded from http://echa.europa.eu/documents/10162/3e43a3a5-ad26-427a-873b-6ff9028c2d42

ECHA (2010b): Member State Committee support document for identification of trichloroethylene as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_trichloroethylene_publication_en.pdf

ECHA (2010c): Annex XV report proposal for identification of a substance as a CMR cat 1 or 2, PBT, vPvB or as a substance of an equivalent level of concern, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_germany_cmr_acids_cr-trioxide_en.pdf

ECHA (2010d): Member State Committee support document for identification of Acids generated from Chromium Trioxide and their Oligomers as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/supdoc_acids_of_chromiumtrioxide_en.pdf

ECHA (2010e): Annex XV report: Proposal for identification of a substance as a CMR cat 1 or 2, PBT, vPvB or a substance of an equivalent level of concern, information downloaded from http://echa.europa.eu/documents/10162/20ee121d-0db9-4c97-ae32-d18d1f4b3ff4

ECHA (2010f): Annex XV dossier: Proposal for identification of a substance as a CMR cat 1 or 2, PBT, vPvB or a substance of an equivalent level of concern, information downloaded from

http://echa.europa.eu/documents/10162/13640/svhc_axvrep_france_cmr_potassium_chromate_en.pdf

ECHA (2010g): Member States Committee support document for identification of potassium chromate as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_potassiumchromate_publication_en.pdf

ECHA (2010h): Member State Committee support document for identification of potassium chromate as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_potassiumchromate_publication_en.pdf

ECHA (2010i): Member State Committee support document for identification of ammonium dichromate as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_ammonium_dichromate_publication_en.pdf

ECHA (2010j): Member State Committee support document for identification of sodium chromate as a Substance of Very High Concern because of its CMR properties, information downloaded from

http://echa.europa.eu/documents/10162/13638/svhc_supdoc_sodium_chromate_publication_en.pdf

ECHA (2010k): Annex XV dossier: Proposal for identification of a substance as a CMR cat 1 or 2, PBT, vPvB or a substance of an equivalent level of concern, information downloaded from http://echa.europa.eu/documents/10162/a43277c9-70d1-4914-a429-8dc42b9ddd09

ECHA (2010l): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: Disodium tetraborate, anhydrous, information downloaded from http://echa.europa.eu/documents/10162/a43277c9-70d1-4914-a429-8dc42b9ddd09

ECHA (2010m): Annex XV dossier: proposal for identification of a substance as Substance of Very High Concern, information downloaded from http://echa.europa.eu/documents/10162/9289e7af-16aa-47c8-8e3a-20179670803d

ECHA (2009a): Background document for benzyl butyl phthalate (BBP), information downloaded from http://echa.europa.eu/documents/10162/b2d894cb-dd0f-4f18-94f8-a165cb3fa75f

ECHA (2009b): Member State Committee Support Document for the Identification of Lead chromate molybdate sulphate red (C.I. Pigment Red 104) as a Substance of Very High Concern, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_lead_chromate_molybdate_publication_en.pdf

ECHA (2009c): Annex XV dossier Proposal for identification of a substance as a CMR cat 1 or 2, PBT, vPvB or a substance of an equivalent level of concern: Lead sulfochromate yellow (C.I. Pigment Yellow 34), information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_france_cmr_lead_sulfochromate_yellow_20090831_en.pdf

ECHA (2009d): Member State Committee Support Document for the Identification of Lead Sulfochromate Yellow (C.I. Pigment Yellow 34) as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_lead_sulfochromate_publication_en.pdf

ECHA (2009e): Annex XV Report Proposal for identification of a substance as a CMR cat 1 or 2, PBT, vPvB or a substance of an equivalent level of concern: lead chromate, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_france_cmr_lead_chromate_20090831_en.pdf

ECHA (2009f): Background document for diarsenic pentaoxide Document developed in the context of ECHA's first Recommendation for the inclusion of substances in Annex XIV, information downloaded from http://echa.europa.eu/documents/10162/13640/diarsenic_pentaoxide_en.pdf

ECHA (2009g): Member State Committee Support Document for the Identification of Tris (2-chloroethyl) phosphate as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/d0f5c171-5086-49c3-a6a3-3a31cb4e08eb

ECHA (2009h): Member State Committee Support Document for the Identification of 2,4 – dinitrotoluene as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_dinitrotoluene_publication_en.pdf

ECHA (2009i): Support document for identification of lead chromate as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_lead_chromate_publication_en.pdf

ECHA (2009j): Annex XV Report Proposal for identification of a substance as a CMR cat 1 or 2, PBT, vPvB or a substance of an equivalent level of concern, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_france_cmr_lead_chromate_20090831_en.pdf

ECHA (2009k): Support document for identification of lead chromate as a Substance of Very High Concern because of its CMR properties, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_lead_chromate_publication_en.pdf

ECHA (2009l): Annex XV Transitional report: Documentation of the work done under the Existing Substance Regulation (EEC) No 793/93 and submitted to the European Chemicals Agency according to Article 136(3) of Regulation (EC) No 1907/2006, information downloaded from https://echa.europa.eu/documents/10162/13630/trd_cover_page_pitch_coal_tar_high_temp_en.pdf

ECHA (2009m): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: Anthracene oil, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_germany_pbt_anthracene_oil_20090831_en.pdf

ECHA (2008a): Member State Committee Support Document for Identification of 4,4'-diaminodiphenylmethane (MDA) as a Substance of Very High Concern, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_mda_publication_en.pdf

ECHA (2008b): Member State Committee Support Document for the Identification of Dibutyl Phthalate (DBP) as a Substance of Very High Concern, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_dibutylphthalate_publication_en.pdf

ECHA (2008c): Member State Committee Support Document for the Identification of Bis(2-ethylhexyl)phthalate (DEHP) as a Substance of Very High Concern, information downloaded from https://noharm-uscanada.org/sites/default/files/documents-files/953/SVHC_DEHP_Support_Doc.pdf

ECHA (2008d): Member State Committee Support Document for the Identification of Benzyl butyl phthalate (BBP) as a Substance of Very High Concern, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_bbp_publication_en.pdf

ECHA (2008e): Member State Committee Support Document for the Identification of Diarsenic Trioxide as a Substance of Very High Concern, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_diarsenic_trioxide_publication_en.pdf

ECHA (2008f): Member State Committee Support Document for Identification of Diarsenic Pentaoxide as a Substance of Very High Concern, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_supdoc_diarsenic_pentaoxide_publication_en.pdf

ECHA (2008g): TRIS (2-CHLOROETHYL) PHOSPHATE, TCEP, Summary Risk Assessment Report, information downloaded from http://echa.europa.eu/documents/10162/f42be21b-33a3-4063-ad4d-2b0f937e41b4

ECHA (2008h): European Union Risk Assessment Report: 2,4-Dinitrotoluene, information downloaded from http://echa.europa.eu/documents/10162/b1176fd0-799d-4c08-a908-755a1c82181f

ECHA (2007): European Union Risk Assessment Report: Disodium tetraborate, anhydrous boric acid, boric acid crude natural, information downloaded from http://echa.europa.eu/documents/10162/ea3533df-1457-4664-98d6-51b2f904af36

ECHA (undated a): Authorisation List: Substance Details Diisobutyl phthalate (DIBP), information downloaded from http://echa.europa.eu/addressing-chemicals-of-concern/authorisation/recommendation-for-inclusion-in-the-authorisation-list/authorisation-list/-/substance/545/search/dibp/term

ECHA (undated b): Adopted opinions and previous consultations on applications for authorisation: Bis(2-ethylhexyl) phthalate (DEHP), information downloaded from http://echa.europa.eu/addressing-chemicals-of-concern/authorisation/applications-for-authorisation-previous-consultations/-/substance/367/search/+/term

ECHA (undated c): Adopted opinions and previous consultations on applications for authorisation, information downloaded from http://echa.europa.eu/addressing-chemicals-of-concern/authorisation/applications-for-authorisation-previous-consultations

ECHA (undated d): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: Calcium arsenate, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_norway_cmr_calcium_arsenate_20110829_en.pdf

ECHA (undated e): Information from the Existing Substances Regulation (ESR), information downloaded from http://echa.europa.eu/information-on-chemicals/information-from-existing-substances-regulation?p_p_id=substancetypelist_WAR_substanceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_pos=1&p_p_col_count=2&_substancetypelist_WAR_substanceportlet_delta=20&_substancetypelist_WAR_substanceportlet_keywords=&_substancetypelist_WAR_substanceportlet_advancedSearch=false&_substancetypelist_WAR_substanceportlet_andOperator=true&_substancetypelist_WAR_substanceportlet_orderByCol=ECNUMBER&_substancetypelist_WAR_substanceportlet_orderByType=asc&cur=1

ECHA (undated f): Summary of Classification and Labelling: chromic acid, information downloaded from http://echa.europa.eu/information-on-chemicals/cl-inventory-database/-/cl-inventory/view-notification-summary/86823

ECHA (undated g): Annex XV dossier Proposal for identification of a substance as a category 1A or 1B CMR, vPvB or a substance of an equivalent level of concern, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_france_cmr_dichromium_tris_chromate_20110829_en.pdf

ECHA (undated h): Annex XV dossier: Proposal for identification of a substance as a category 1A or 1B CMR, PBT, vPvB or a substance of an equivalent level of concern, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_france_cmr_pentazinc_chromate_octahydroxide_20110829_en.pdf

ECHA (undated i): Annex XV dossier: Proposal for identification of a substance as a category 1A or 1B CMR, PBT, vPvB or a substance of an equivalent level of concern: Potassium hydroxyoctaoxodizincatedichromate(1-), information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_france_cmr_potassium_hydroxyoctaoxodizinccatedichromate_20110829_en.pdf

ECHA (undated j): Annex XV dossier: proposal for identification of a substance as a category 1A or 1B CMR, PBT, vPvB or a substance of an equivalent level of concern: Dichromium tris(chromate), information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_france_cmr_dichromium_tris_chromate_20110829_en.pdf

ECHA (undated k): Annex XV dossier: proposal for identification of a substance as a category 1A or 1B CMR, PBT, vPvB or a substance of an equivalent level of concern: Strontium chromate, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_fr_cmr_srcro4_publ_en.pdf

ECHA (undated l): Annex XV dossier: proposal for identification of a substance as a category 1A or 1B CMR, PBT, vPvB or a substance of an equivalent concern: Formaldehyde, oligomeric reaction products with aniline, information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_germany_cmr_techmda_20110829_en.pdf

ECHA (undated m): Annex XV dossier: proposal for identification of a substance as a category 1A or 1B CMR, PBT, vPvB or a substance of an equivalent concern: bis(2-methoxyethyl) ether (Diglyme), information downloaded from http://echa.europa.eu/documents/10162/13640/svhc_axvrep_austria_belgium_poland_cmr_diglyme_20110829_en.pdf

ECHA (undated n): Recommendation for inclusion in the Authorisation List and public consultation, information downloaded from

http://echa.europa.eu/addressing-chemicals-of-concern/authorisation/recommendation-for-inclusion-in-the-authorisation-list?p_p_id=substancetypelist_WAR_substanceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_pos=1&p_p_col_count=3&_substancetypelist_WAR_substanceportlet_delta=20&_substancetypelist_WAR_substanceportlet_keywords=&_substancetypelist_WAR_substanceportlet_advancedSearch=false&_substancetypelist_WAR_substanceportlet_andOperator=true&_substancetypelist_WAR_substanceportlet_orderByCol=extraColumn2504&_substancetypelist_WAR_substanceportlet_orderByType=asc&cur=2

ECHA (undated o): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: Acetic acid, lead salt, basic, information downloaded from http://www.echa.europa.eu/documents/10162/9db5ed03-30fa-44c4-b477-7b4f939e9540

ECHA (undated p): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: 1-bromopropane; n-propyl bromide, information downloaded from http://echa.europa.eu/documents/10162/d7d30eaa-86a3-4fd1-a85e-122a96c29c3f

ECHA (undated q): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: Bis(2-methoxyethyl)phthalate, information downloaded from http://echa.europa.eu/documents/10162/38458518-7e1d-49ff-b53d-d07963c1bceb

ECHA (unknown r): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: 1,2-Benzenedicarboxylic acid dipentylester, branched and linear, information downloaded from http://echa.europa.eu/documents/10162/71dbce1d-0448-4b9f-8b2a-558072f18faf

ECHA (undated s): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: 1,2-Benzenedicarboxylic acid, di-C7-11-branched and linear alkyl esters (DHNUP), information downloaded from http://echa.europa.eu/documents/10162/248dd51f-b218-455f-bb00-fbe34d500a5c

ECHA (undated t): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: Lead monoxide [Lead oxide], information downloaded from http://www.echa.europa.eu/documents/10162/646ccf4f-bf51-44c2-b499-b3960271be87

ECHA (undated u): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: dipentyl phthalate (DPP), information downloaded from http://echa.europa.eu/documents/10162/d55c182b-f063-4955-969d-5684584d17b2

ECHA (undated v): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, vPvB or a substance of an equivalent level of concern: Pentalead tetraoxide sulphate, information downloaded from http://www.echa.europa.eu/documents/10162/0e4130c9-c101-4452-b7af-685a522b9374

ECHA (undated w): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: Orange lead [Lead tetroxide], information downloaded from http://www.echa.europa.eu/documents/10162/c703bcfc-6a0a-40f8-a37a-372fa54bccf8

ECHA (undated x): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: Pyrochlore, antimony lead yellow, information downloaded from http://www.echa.europa.eu/documents/10162/81fef8ab-25f2-4ab9-8e5d-74737495ce7a

ECHA (undated y): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: Diboron trioxide, information downloaded from http://echa.europa.eu/documents/10162/13638/svhc_axvrep_diboron_trioxide_en.pdf

ECHA (undated z): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: n-Pentyl-isopentyl phthalate, information downloaded from http://echa.europa.eu/documents/10162/48f63323-2ed7-453b-b1ca-c42987d0453f

ECHA (unknown ab): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: Diisopentylphthalate (DIPP), information downloaded from http://www.echa.europa.eu/documents/10162/b33d9431-e823-43fd-bf4b-d554b6aef968

ECHA (unknown cd): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: Silicic acid, lead salt, information downloaded from http://www.echa.europa.eu/documents/10162/7c92fbfb-7cb7-4264-bdce-acea6a9fb5f9

ECHA (unknown ef): Annex XV dossier: proposal for identification of a substance as a CMR 1A or 1B, PBT, vPvB or a substance of an equivalent level of concern: Tetralead trioxide sulphate, information downloaded from http://www.echa.europa.eu/documents/10162/d682bce5-13a3-4b4c-aecb-80d40a890586

Environment Canada (2008): Screening Assessment for the Challenge: C.I. Pigment Yellow 34, information downloaded from http://www.ec.gc.ca/ese-ees/default.asp?lang=En&n=A9AB1DAD-1

European Chemicals Bureau (2008): European Union Risk Assessment Report: bis(2-ethylhexyl)phthalate (DEHP), information downloaded from http://www.dehp-facts.com/upload/documents/webpage/DEHP%20RA%20report%20full.pdf

European Chemicals Bureau (2007a): European Union Risk Assessment Report: benzyl butyl phthalate (BBP), information downloaded from http://echa.europa.eu/documents/10162/bad5c928-93a5-4592-a4f6-e02c5e89c299

European Chemicals Bureau (2007b): European Union Risk Assessment Report: anthracene – part II human health, information downloaded from http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/5403/1/rar316_anthracenehh.pdf

European Chemicals Bureau (2005): European Union Risk Assessment Report: chromium trioxide, sodium chromate, sodium dichromate, ammonium dichromate and potassium dichromate, information downloaded from http://echa.europa.eu/documents/10162/3be377f2-cb05-455f-b620-af3cbe2d570b

European Chemicals Bureau (2004a): European Union Risk Assessment Report: dibutyl phthalate, information downloaded from http://echa.europa.eu/documents/10162/ba7f7c39-dab6-4dca-bc8e-dfab7ac53e37

European Chemicals Bureau (2004b): European Union Risk Assessment Report: Trichloroethylene, information downloaded from http://echa.europa.eu/documents/10162/83f0c99f-f687-4cdf-a64b-514f1e26fdc0

European Chemicals Bureau (2004c): European Union Risk Assessment Report: Edetic Acid (EDTA). Information downloaded from http://echa.europa.eu/documents/10162/65615721-ab6d-4f28-b48f-73cf9d8cc529

European Chemicals Bureau (2001): European Union Risk Assessment Report: 4,4'-methylenedianiline, information downloaded from http://www.baua.de/en/Chemicals-Act-biocide-procedure/Documents/RAR_008.pdf?__blob=publicationFile

Furst A, Schlauder M & Sasmore DP (1976): Tumorigenic activity of lead chromate, information downloaded from http://www.ncbi.nlm.nih.gov/pubmed/1268834

Golub M (2006): Metals, Fertility, and Reproductive Toxicity, information downloaded

http://books.google.co.uk/books?id=Qt8LEB7_HyQC&pg=PA100&lpg=PA100&dq=potassium+chromate+reproductive&source=bl&ots=sPNIGd4z3q&sig=wXk5FkYGzuA-60063xR-u1Kyuug&hl=en&sa=X&ei=lRoIVN_EJ8_naI-GgpAH&ved=0CFsQ6AEwCQ#v=onepage&q=potassium%20chromate%20reproductive&f=false

Instituto Nacional de Seguridad e Higiene en el Trabajo (INSHT) (n.d.): Occupational Chemical Hazards and Reproductive health, information downloaded from http://www.insht.es/InshtWeb/Contenidos/Instituto/Noticias/Noticias_INSHT/2010/Ficheros/METROnet%20White%20Book%20Occupational%20Chemical%20Hazard%20&%20Reproductive%20Health.pdf

International Agency for Research on Cancer (2014): IARC Monographs Volume 100C Arsenic and Arsenic Compounds, information downloaded from http://monographs.iarc.fr/ENG/Monographs/vol100C/100C-01-Index-tables.php

International Agency for Research on Cancer (2006): IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Volume 87: Inorganic and Organic Lead Compounds, information downloaded from http://monographs.iarc.fr/ENG/Monographs/vol87/mono87.pdf

International Agency for Research on Cancer (IARC) (1998): Arsenic and arsenic compounds, information downloaded from http://www.inchem.org/documents/iarc/suppl7/arsenic.html

International Agency for Research on Cancer (undated a): Chromium and Chromium Compounds, information downloaded from http://monographs.iarc.fr/ENG/Monographs/vol49/mono49-6.pdf

International Agency for Research on Cancer (undated b): Summary of epidemiological studies of arsenic in drinking-water and bladder and kidney cancers, information downloaded from http://monographs.iarc.fr/ENG/Monographs/vol100C/100C-01-Table2.7.pdf

International Agency for Research on Cancer (undated c): Arsenic and arsenic compounds, information downloaded from http://monographs.iarc.fr/ENG/Monographs/vol100C/mono100C-6.pdf

International Agency for Research on Cancer (undated d): IARC Monographs Supplement 7 Arsenic and arsenic compounds, information downloaded from http://monographs.iarc.fr/ENG/Monographs/suppl7/Suppl7-19.pdf

International Agency for Research on Cancer (undated e): Chromium (VI) Compounds, information downloaded from http://monographs.iarc.fr/ENG/Monographs/vol100C/mono100C-9.pdf

International Agency for Research on Cancer (undated f): 4,4-methylenebis(2-chlorobenzenamine), information downloaded from http://monographs.iarc.fr/ENG/Monographs/vol100F/mono100F-9.pdf

International Agency for Research on Cancer (undated g): 1,2-dichloroethane, information downloaded from http://monographs.iarc.fr/ENG/Monographs/vol71/mono71-21.pdf

International Programme on Chemical Safety (IPCS) (undated): IPCS International Programme on Chemical Safety, Inorganic Arsenic compounds other than arsine, information downloaded from http://www.inchem.org/documents/hsg/hsg/hsg070.htm#SectionNumber:2.5

National Industrial Chemicals Notification and Assessment Scheme (NICNAS) (2013a): Inventory Multi-tiered assessment and prioritisation (IMAP): Human health tier II assessment for lead chromates, information downloaded from

http://www.nicnas.gov.au/chemical-information/imap-assessments/imap-group-assessment-report?assessment_id=843

National Industrial Chemicals Notification and Assessment Scheme (NICNAS) (2013b): Inventory Multi-tiered Assessment and Prioritisation (IMAP), information downloaded from http://www.nicnas.gov.au/chemical-information/imap-assessments/imap-group-assessment-report?assessment_id=843

National Industrial Chemicals Notification and Assessment Scheme (NICNAS) (2008): A summary of physicochemical and human health hazard data for 24 ortho-phthalate chemicals, information downloaded from http://www.nicnas.gov.au/_data/assets/pdf_file/0008/4958/Phthalate-Hazard-Compendium.pdf

National Toxicology Program (2014): 13th Report on Carcinogens (RoC), information downloaded from http://ntp.niehs.nih.gov/ntp/roc/twelfth/profiles/chromiumhexavalentcompounds.pdf

National Toxicology Program (1985): Di-N-Pentylphthalate, information downloaded from http://ntp.niehs.nih.gov/testing/types/repro/abstracts/racb/index-25.html

Nestmann ER & Zhang B (2007): Chromosome aberration test of Pigment Yellow 34 (lead chromate) in Chinese hamster ovary cells, information downloaded from http://www.sciencedirect.com/science/article/pii/S1383571807001751

New Jersey Department of Health (2011): Hazardous Substance Fact Sheet, information downloaded from http://nj.gov/health/eoh/rtkweb/documents/fs/0097.pdf

OECD (2009): Initial Targeted Assessment Profile: 1,2-Benzenedicarboxylic acid, di-C7-11-branched and linear alkyl esters (Di(heptyl, nonyl, undecyl) phthalate), information downloaded from http://webnet.oecd.org/Hpv/ui/handler.axd?id=d9d0c5dd-43bb-4604-a5ce-129c70fb1ca5

OECD (2005): SIDS Initial Assessment Profile: 1,2-Benzenedicarboxylic acid, di-C6-8-branched alkyl esters, C7-rich (diisoheptyl phthalate ester), information downloaded from http://webnet.oecd.org/Hpv/UI/handler.axd?id=b74ff4f8-0ae0-4fe7-a133-27ad22bf4797

OECD (1995): Robust Study Summaries: Critical Studies Identified in Screening Assessment Targeted for Human Health, information downloaded from

http://webnet.oecd.org/Hpv/ui/handler.axd?id=9554f035-4a9a-4198-9499-225557cb6070

Oliveira H, Spanò M, Guevara MA, Santos TM, Santos C & Pereira Mde L (2010): Evaluation of in vivo reproductive toxicity of potassium chromate in male mice, information downloaded from http://www.ncbi.nlm.nih.gov/pubmed/19577443

Risctox (undated): dichromic acid, information downloaded from http://www.istas.net/risctox/en/iframe/dn_risctox_ficha_sustancia.asp?id_sustancia=954723

Rolls Royce Plc (2013): Chemical Safety Report, information downloaded from http://echa.europa.eu/documents/10162/0991d6a1-3a58-4b12-80bc-51f0b9269351

R S Clare & Co Ltd (2011): Safety Data Sheet Lead Chromate Pigments, information downloaded from http://www.rsclare.com/sites/default/files/Lead%20Chromate%20Pigments%20v3%20170611.pdf

Scientific Committee on Health and Environmental Risks (2009): Opinion on Voluntary Risk Assessment Report on lead and lead compounds Human Health Part, information downloaded from http://ec.europa.eu/health/ph_risk/committees/04_scher/docs/scher_o_114.pdf

Scientific Committee on Health and Environmental Risks (2008): Coal tar pitch, high temperature, information downloaded from http://ec.europa.eu/health/ph_risk/committees/04_scher/docs/scher_o_083.pdf

Scorecard (2011a): Arsenic Pentoxide, information downloaded from http://scorecard.goodguide.com/chemical-profiles/edf-risk-characterization.tcl?edf_substance_id=1303-28-2

Scorecard (2011b): Lead chromate, information downloaded from http://scorecard.goodguide.com/chemical-profiles/national-risk-characterization.tcl?edf_substance_id=7758-97-6

Scorecard (2011c): Lead chromate: safety assessment, information downloaded from http://scorecard.goodguide.com/chemical-profiles/safety-detail.tcl?edf_substance_id=7758-97-6

Sentury Reagents, Inc. (2014): Safety Data Sheet: Lead (II) chromate, information downloaded from http://www.senturyreagents.com/msds/lead_chromate.pdf

SSSEurope (2013): REACH Candidate List of SVHC (updated as on 16/12/2013), information downloaded from http://ssseurope.org/REACH%20Candidate%20List%20of%20SVHC%2016.12.2013.pdf

Toxnet (2008): Diamyl phthalate, information downloaded from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+131-18-0

Toxnet (2005a): Hazardous Substances Data Bank, information downloaded from http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB

Toxnet (2005b): Strontium Chromate, information downloaded from http://toxnet.nlm.nih.gov/cgi-bin/sis/search/a?dbs+hsdb:@term+@DOCNO+2546

Toxnet (2001): Arsenic trioxide, information downloaded from http://toxnet.nlm.nih.gov/cgi-bin/sis/search/a?dbs+hsdb:@term+@DOCNO+419

Toxnet (undated a): TOXNET Databases, information downloaded from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~CXxruc:2

Toxnet (undated b): TOXNET Databases, information downloaded from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~V0p0vN:2

Toxnet (undated c): TOXNET Databases, information downloaded from http://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~hKcuNy:2

United States Environmental Protection Agency (2014): Brief Records: EPA National Library Catalog, information downloaded from http://cfpub.epa.gov/ols/catalog/advanced_bibliography.cfm?&FIELD1=SUBJECT&INPUT1=Dogs&TYPE1=EXACT&LOGIC1=AND&COLL=&SORT_TYPE=MTIC&start_row=26

United States Environmental Protection Agency (2006): IUCLID Dataset: Diisoheptyl phthalate, information downloaded from http://www.epa.gov/hpv/pubs/summaries/benzene/c13467rr3s.pdf

Vermont Safety Information Resources, Inc. (1990): CABIN YELLOW CORDURITE, information downloaded from http://www.hazard.com/msds/f2/bkr/bkrpr.html

Xilinx, Inc. (2014): REACH Position Statement: EU Regulation (EC) Number 1907/2006, information downloaded from http://origin.xilinx.co.jp/system_resources/lead_free/reach_position_statement.pdf

Annex 3:

Collated QALY and DALY weights by Health Effect

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
Cancer							
Lung (non-small cell lung cancer)	Stable disease with no toxicity: 0.653 **Progressive disease: 0.473**	CEA	SG and VAS	United Kingdom 100 participants. Average age: 40.51 38% Female; 74% white; 14% black; 9% Asian	The authors complain that the existing utility weights reported for non-small cell lung cancer vary widely and thus are not suitable for clinical or analytical purposes.	2012	1
Kidney	Men: 0.76 Women: 0.62	CEA Registry (2005)	PTO and VAS	The Netherlands	Smoking cessation interventions. Time horizon is 75 years.	2005	2
Urinary Bladder	Men: 0.91 Women: 0.89	CEA Registry	PTO and VAS	The Netherlands	Same source than for kidney cancer.	2005	2
Liver	0.73	CEA (2010)	TTO and SG	China, Children and adolescents (age 1-19)	Time horizon is a life. HBV vaccination program for children aged 1 to 19 years in China	2010	3
Skin	Melanoma—stage I—stable disease: 0.96 Melanoma—stage IV—stable disease: 0.65	CEA Registry (2012)	TTO and SG	Australia	Time horizon is a life-time and the impact of active promotion of routine sunscreen use	2012	4
Breast	0.77	CEA Registry (2012)	Unclear	Canada	Time horizon is a life-time and the context were prophylactic surgical procedures	2012	5
Stomach	**Gastric cancer stage I and II: 0.65** Gastric cancer stage IV: 0.4	CEA Registry (2003)	Quality of life utilities scores were taken from Quality-of-Life Repository for gastric cancer	Singapore	Time horizon is a life-time. Comparison between 2-yearly endoscopic mass screening programme versus no screening.	2003	6
Leukaemia	Progressed health state: 0.6	NICE (2010)	Utility values from a manufacturer's submission. The	United Kingdom	Treatment for chronic lymphocytic leukaemia. Lifetime horizon	2010	7

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
			estimates of utility were not preference based, and were estimated by the authors of the report from condition-specific health-related quality-of-life data.				
Lymphoma	**Progressive disease: follicular non-Hodgkin's lymphoma: 0.618** Progression free follicular non-Hodgkin's lymphoma: 0.805	CEA Registry (2010)	Finish utility data were not available, applied utility weights based on a British study with 222 FL patients [FL = follicular non-Hodgkin's lymphoma)	United Kingdom 222 FL patients	Modelling of three health states "[progression free (PF), progressive disease (PD), and death]; all patients entered the model in the PF state. PF and PD states were defined separately to account for the higher quality of life (QoL) and lower treatment costs for PFS. […] A lifetime horizon was used	2010	8
Pancreatic/Pancreas	Stable disease - metastatic pancreatic cancer 0.72	CEA Registry (2010)	EQ-5D, Utilities were obtained by surveying medical oncologists across Canada using the EQ-5D.	Canada	Time frame is two years. Treatment of pancreatic cancer.	2010	9
Prostatic	0.58	NICE	"For health-related quality of life, investigators collected data during AFFIRM using the EQ-5D and Functional Assessment of Cancer Therapy-Prostate (FACT-P) questionnaires."	United Kingdom	"The manufacturer conducted the analysis from the perspective of the NHS and personal social services and chose a time horizon of 10 years."	2014	10
Cardiovascular							
Coronary heart	0.718	CEA	Unclear	United Kingdom	The time frame is a lifetime.	2010	11

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
disease					The objective of this study was "to determine the effectiveness and cost effectiveness of using information from circulating biomarkers to inform the prioritisation process of patients with stable angina awaiting coronary artery bypass graft surgery."		
Acute myocardial infarction	0.8	CEA Registry (2012)	Unclear	5.800 smokers (male 55%, female 45%) Region: United Kingdom, Europe	Time horizon is life and it relates to a smoking cessation program which is delivered by text message	2012	1
Stroke	0.707	CEA (2011)	Unclear	United States of America, post-menopausal women	A detailed, continuous time, mathematical model of breast cancer was used and health care processes was used to simulate a postmenopausal population aged < 55 years in a virtual trial comparing tamoxifen treatment with no treatment for lifetime follow-up.	2011	12
Impaired fertility							
Infertility	0.82	CEA Registry	HUI II	Canada	A committee of 14 experts derive preferences from an analysis of the responses taken from an experiment conducted by Torrance (based on a sample of Canadian parents answering the HUI II.)	1995	13
Disruption of ovarian cycle or menstrual disorder	Minimum utility value: 0.19 **Maximum utility value: 0.81**	NICE	HRQoL and EQ-5D but no further details	United Kingdom	Heavy menstrual bleeding A 5 year time horizon was chosen as this is the maximum time for which one of the treatments considered is licensed.	2007	14

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
Impaired foetal growth							
Low birth weight	Very low birth weight infants immediately after birth: 0.94	CEA but sourced from another article which is not available (Rautava et al)	(HRQoL) according to a 17-dimension parental questionnaire (17D), as described in detail in an article by	Finland	"2064 very preterm children (gestational age <32 weeks or birth weight <1501 g) and all 200 609 full-term control individuals (mean [SD] gestational age, 37 [0] to 41 [6] weeks) born from January 1, 2000, through December 31, 2003" participated.	2010	15
Idiopathic intellectual disability	No disability: 0.94 **Mild mental retardation: 0.62**	CEA	The authors use the EQ-5D and the HUI 2 to derive preference scores for 8 permanent sequelae after bacterial meningitis. The sequelae chosen are also potential sequelae which could result from chemical exposure. 36 paediatricians completed an EQ-5D and HUI-15Q survey. The mean weights were then computed and MANOVA was used to perform profile analysis. The scores stated are the mean preference scores.		The aim of this paper by Oostenbrinka et al. (2002) is to establish the best method for deriving quality weights. They conclude that the best method depends on the classification system used (i.e. EQ-5D) and the type of health states under evaluation. They also add that estimates of quality weights are sensitive to difference and that this must be taken into account in cost-effectiveness studies which use these values.	2002	16
Cleft palate (CLP = isolated cleft lip and palate)	**Infants: 0.78** Children: 0.82 Adolescents: 0.85 Adults: 0.89	Wehby GL, Ohsfeldt, RL & Murray JC (2006)	Health professionals, based on HRQL values, obtained through a VAS evaluation	United States of America	Representative group of health professionals working on craniofacial and/or cleft palate Values (between 0 and 1) representing the HRQL associated with isolated and non-	2006	17

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
					isolated oral clefting for infants, children, adolescents, and adults		
Spina bifida	0.454	CEA	Based on the Health Utility Index (HUI) 2	Australia	The time-frame is a lifetime. Mandatory folic acid to prevent birth defects is being assessed	2013	18
Renal abnormalities	End-stage renal disease: 0.7	CEA	Unclear.	United States of America	Time horizon is a lifetime.	2013	19
Urogenital abnormalities (e.g. cryptorchidism)	**Abnormal aspect scrotum congenital UDT** **Unilateral: 0.895** Bilateral: 0.757	Van den Akker-van Marle, ME et al. (2010)	VAS (the valuations indicated on the VAS scale are transformed to approximate time tradeoff (TTO) scores using the power transformation 1 - (1 - VAS/100)	The Netherlands, Dutch general population with 41 soundly completed questionnaires	A decision analysis is performed to determine the best age at which to perform orchidopexy.	2013	20
Other							
Chronic Obstructive Pulmonary Disease (COPD)	Very severe chronic obstructive pulmonary disease (COPD): 0.74 Severe COPD: 0.77 Moderate COPD: 0.80 **Mild COPD: 0.82**	CEA	HRQoL (?) But values taken from literature	United Kingdom	cost-effectiveness profile of indacaterol, the first once-daily long-acting beta2-agonist (LABA), compared with tiotropium and salmeterol, in patients with moderate to severe COPD	2013	21
Asthma	Poor asthma control: 0.52 Moderately reduced asthma control: 0.65 Mildly reduced asthma control: 0.76	CEA	Euro-Qol EQ-5D	United Kingdom	Time horizon is 1 year. Asthma diagnosis and management	2009	22

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
	Good asthma control: 0.93						
Allergic contact dermatitis	Atopic eczema: Severe: 0.73 (VAS) **0.93 (TTO)** 0.98 (SG)	Lundberg (1999) as quoted in NICE (2005)	Various: VAS, TTO and SG	United Kingdom	Cost-effectiveness of pimecrolimus for mild to moderate atopic eczema and tacrolimus for moderate to severe atopic eczema compared with current standard treatment in adults and children	2005	23
Osteoporosis	**Clinical vertebral fracture for women with osteoporosis, first year: 0.72** Clinical vertebral fracture for women with osteoporosis, subsequent years: 0.93 Wrist fracture for women with osteoporosis, first year: 0.94 Wrist fracture for women with osteoporosis, subsequent years: 1 Other fracture for women with osteoporosis, first year: 0.91	CEA (2013)	Unclear – extensive literature review	Belgium, the study population included women aged over 60 years with osteoporosis	These values have been derived from a study on the cost-effectiveness of Bazedoxifene in the treatment of postmenopausal osteoporotic Women.	2013	24

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
Anaemia	For short term QALY gains 0.76	NICE (11) from Brazier et al. (2004).	Standard gamble technique to transform them to SF-6D values. The Assessment Group then expressed the SF-6D values as EQ-5D values using regression analyses	United Kingdom	Lifetime time horizon. Treating anaemia in people with cancer having chemotherapy from an NHS and personal social services perspective	2014	25
Glaucoma	**Mild: 0.92** Moderate: 0.89 Severe: 0.86	CEA	SG	United States of America	Cost effectiveness of medication compared with laser treatment in newly diagnosed patients with open angle glaucoma has been assessed.	2012	26
Hearing impairment	Profound deafness and no cochlear implant: 0.421	NICE	HUI3 (health utility index 3)	United Kingdom	Currently available devices for cochlear implantation.	2009	27
Liver cirrhosis	**Compensated cirrhosis: 0.82** Decompensated cirrhosis: 0.54	CEA (2013)	Unclear	United States of America	Assessed vaccination for hepatitis B and outcomes for HBV infection to reflect the impact of hepatitis B in adults with diagnosed diabetes	2013	28
Stomach ulcer	Ulcer: **30-39 years: 0.881** 40-49 years: 0.889 50-59 years: 0.806	Song HJ, Kwon JW, Kim N & Park YS (2013)	"Specific utility weights for applicable ages were obtained using the 2007 to 2009 Korean National Health and Nutritional Examination Survey Data, which was a national survey representative of the South Korean population.	South Korea	Impact of a H. pylori screening/eradication strategy compared to a no-screening strategy among patients who required treatment with NSAIDs or aspirin	2013	29
Pancreatitis	**Mild: 0.89** Severe: 0.11	CEA (2007)	Calculation of QALYs are based on health utilities	United States of America	The intended population for this analysis was a typical female patient	2007	30

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
			determined by standard methodology assigning diagnoses and procedures a value ranging from 0 (utility of death) to 1 (utility of perfect health)."		18 years of age or older with symptomatic cholelithiasis and incidental CDL discovered at the time of LC/IOC." Time horizon is one year.		
Parkinson –like condition	**Utilities in H&Y 1: 0–25 % OFF time 0.74,** 26–50 % OFF time 0.68, 51–75 % OFF time 0.64 76–100 % OFF time 0.52; […] Utilities in H&Y 5 0–25 % OFF time 0.131 26–50 % OFF time 0.043 51–75 % OFF time 0.045 76–100 % OFF time 0.133	CEA (2013)	Although PDQ-39 data were collected at both visits within the key clinical study, there is currently no adequate tool for mapping these data to a generic measure of QoL suitable for use in decision-analytic modelling e.g., EQ-5D."	Unclear but UK NHS perspective is adopted by authors of study.	The study evaluated changes in quality of life via the Parkinson's disease questionnaire (PDQ-39), and changes in symptom severity using Part III of the UPDRS."	2013	31
Alzheimer disease and other dementia	**Mild, community: 0.680** Mild nursing home: 0.710 Moderate, community: 0.540	CEA (2000)	HUI2 and HUI3 "Quality-of-life weights for patients with Alzheimer disease were based on Health Utility Index Mark 2 (HUI 2)	United States of America, but data based on a nursing home community of Canadians	18 month time horizon.	2000	32

Table A3-1: Collation of available QALY utility weights							
Disease	**QALY utility weight**	**Original Date of study**	**Method**	**Sample size Healthy or sick Region**	**Context**	**Date of CEA/ NICE**	**Ref**
	Moderate, nursing home: 0.480 Severe, community: 0.370 Severe, nursing home: 0.310		scores that were published previously (8, 28, 29)" A QALY weight was derived for the age cohort 70-79 (excluding patients with a diagnosis of Alzheimer disease) from a study in which investigators administered the Health Utility Index Mark 3 (HUI 3) to a large sample of community-dwelling Canadians.				

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
Cancer							
Lung	Diagnosis and primary therapy: […] 0.440 Relapse/terminal stage small cell cancer: 0.930	Dutch disability weights (B), 1997	During the panel sessions, the indicator conditions were valued according to two valuation methods: person trade-off (PTO) [PTO1 + PTO2] and visual analogue scaling (VAS). Compared these results to EQ-5D values to makes sure that these values are sound. ("The health states were assigned standardized descriptions with the help of the standard EuroQol 5D classification system […] based on the	38 physicians (28 men, 10 women) from the Netherlands have been divided up into three panels. In addition a panel of layperson has been established asking them to rate the same conditions (The lay panel was composed of 7 members, 4 men and 3 women) (p.55).	"The outcome of the Dutch project on 'Disability weights for diseases' is a coherent set of disease-specific disability weights for 175 disease stages, derived from the 52 diseases."	2005 (1997)	A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
			descriptive (functional) health-state data available" (Stouthard et al, 1997, p.12).				
Kidney	Diagnosis and primary therapy: 0.270 […] Terminal stage: 0.930	Provisional based on Dutch weights/ Dutch weight for end-stage disease, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)
	Malignant neoplasms of liver: 0.422	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Urinary Bladder	Diagnosis and primary therapy: 0.270 […] Terminal stage: 0.930	Provisional weight based on Dutch weights/ Dutch weight for end-stage disease, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Malignant neoplasms of bladder: 0.448	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Liver	Diagnosis and initial treatment: 0.430 […] Terminal phase: 0.930	Dutch weight for colorectal cancer/ Dutch weight for end-stage disease, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)
	Malignant neoplasms of liver: 0.677	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Skin	Melanoma Primary treatment: no evidence of dissemination 0.190 […] Terminal phase 0.930 Non-melanoma skin cancers Basal cell carcinoma:	Dutch weight/ Dutch weight for end-stage disease Dutch weight/ Dutch weight for end-stage disease, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	0.050 […] Terminal phase 0.930						
	Malignant neoplasms of skin: 0.362	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Breast	Diagnosis and primary therapy: non-invasive tumour <2 cm 0.260 […] Terminal phase 0.930	Dutch weight/ Dutch weight for end-stage disease, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)
	Malignant neoplasms of breast: 0.178	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Stomach	Diagnosis and primary therapy 0.530 […] Pre-terminal and terminal stages 0.930	Dutch weight/ Dutch weight for end-stage disease, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
Haematopoietic Subcategory: Leukaemia	Acute myeloid leukaemia Diagnosis and primary therapy: 0.550 […] Terminal stage: 0.930	Provisional weight based on Dutch weights/ Dutch weight for end-stage disease, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)
	Chronic myeloid leukaemia Diagnosis and primary therapy: 0.550 […] Terminal stage 0.930						
	Leukaemia: 0.182	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Lymphoma (Non-Hodgkin's lymphoma)	Non-Hodgkin's lymphoma Low grade: dissemination stage I and II 0.190 […] Terminal phase 0.930	Dutch weight/ Dutch weight for end-stage disease, 1997	PTO + VAS	45, the Netherlands		2005 (1997)	A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Lymphoma: 0.368	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Pancreatic/Pancreas	Diagnosis and initial treatment 0.430 […] Terminal phase 0.930	Dutch weight for colorectal cancer, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)
	Malignant neoplasms of pancreas: 0.587	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Prostatic	Diagnosis and primary therapy: localised cancer 0.270 […] Terminal stage 0.930	Dutch weight/ Dutch weight end-stage disease, 1997	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)
	Malignant neoplasms of prostate: 0.239	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Impaired fertility							
Infertility	Infertility: primary 0.008	European weights	Paired comparison	Europe		2015	C
	Infertility: 0.180	GBD weights (2013)	Paired comparison	Global		2015	D

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
Disruption of ovarian cycle or menstrual disorder	Menstrual disorders 0.033	Estimated from EQ-5D+ regression mode	EQ-5D	Australia		2005	A
Reduced foetal growth							
Small for gestational age; low birth weight	Mild permanent disability 0.110	Dutch weight for mild to moderate early acquired hearing loss	PTO + VAS	45, the Netherlands		2005 (1997)	A (B)
	Low birth weight: 0.442	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Developmental anomalies/abnormalities							
IQ and other developmental neuro-impairment / neurological and behaviour disorders ('idiopathic intellectual disability')	Intellectual disability: Mild: 0.053 Moderate: 0.123 Severe: 0.141 Profound: 0.213	European weights	Paired comparison	Europe		2015	C
	Mental retardation: 0.242	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
	Intellectual disability/mental retardation: Mild: 0.043	GBD weights (2013)	Paired comparison	Global		2015	D

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Moderate: 0.1 Severe: 0.16 Profound: 0.2						
Cardiovascular							
Myocardial infarction	Acute myocardial infarction (days 3-28): 0.098	European weights	Paired comparison	Europe		2015	C
	Acute myocardial infarction: Days 1-2: 0.432 Days 3-28: 0.074	GBD weights (2013)	Paired comparison	Global		2015	D
Stroke	Mild permanent impairments 0.360 Moderate permanent impairments 0.630 Severe permanent impairments 0.920	Dutch weight	PTO + VAS	45, the Netherlands		2005 (1997)	A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Stroke: 0.547	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
	Stroke: long-term consequences, mild: 0.019 Stroke: long-term consequences, moderate: 0.07 Stroke: long-term consequences, moderate plus cognitive problems: 0.316 Stroke: long-term consequences, severe: 0.552 Stroke: long-term consequences, severe plus cognitive	GBD weights (2013)	Paired comparison	Global		2015	D

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	problems: 0.588						
	Stroke: long-term consequences, moderate: 0.075 Long-term consequences, severe plus cognitive problems: 0.580	European weights	Paired comparison	Europe		2015	C
Skeletal (includes also abnormalities of limbs)							
Cleft palate (CLP = isolated cleft lip and palate)	Cleft palate—treated 0.015 Cleft lip—treated 0.016	GBD weights (1996)	<1% PTO, 99% VAS	Global, panel of medical experts (10 people)		2005 (1996)	A
	Clefts of palate and lip: 0.379	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Spina bifida	Low-level spina bifida aperta 0.160 Medium-level spina bifida aperta 0.500 High-level spina bifida aperta	Dutch weight	PTO + VAS	45, the Netherlands		2005 (1997)	A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	0.680						
	Spina bifida: 0.473	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Renal abnormalities	Chronic kidney disease (stage IV): 0.108 End-stage renal disease: with kidney transplant: 0.030 End-stage renal disease: on dialysis: 0.487	European weights	Paired comparison	Europe		2015	C
	Chronic kidney disease (stage IV): 0.104 End-stage renal disease: with kidney transplant: 0.024 End-stage renal disease: on dialysis: 0.571	GBD weights (2013)	Paired comparison	Global		2015	D

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Acute conditions in kidney: 0.340 Severe chronic kidney disease: 0.300	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Urogenital abnormalities (e.g. cryptorchidism)	Other urinary tract malformations 0.290	Dutch weight for renal failure	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)
Other							
Nephritic and nephrotic syndrome	End-stage renal failure with dialysis 0.290 End-stage renal failure with transplant 0.290 Transplanted patient 0.110 Untreated end-stage renal failure 0.104	Dutch weight for diabetic nephropathy/ Dutch weight for diabetic nephropathy/ 1996 GBD weight for treated renal failure: Dutch weight for uncertain prognosis/ GBD weight	PTO + VAS; PTO1 + PTO2	The Netherlands, Global			A (B)
	Nephritis and nephrosis: 0.570	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
Respiratory (tract) irritation							
Chronic Obstructive Pulmonary Disease (COPD)	COPD and other chronic respiratory diseases: mild: 0.019 Moderate: 0.225 Severe: 0.408	GBD weights (2013)	Paired comparison	Global		2015	D
	COPD and other chronic respiratory diseases: mild: 0.025 Moderate: 0.284 Severe: 0.418	European weights	Paired comparison	Europe		2015	C
	COPD mild to moderate: 0.170 Severe: 0.530	Dutch weights	PTO + VAS	45 people, the Netherlands		2005 (1997)	A (B)
	COPD: 0.299	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Asthma	Asthma: controlled: 0.015 Asthma: partially controlled: 0.036	GBD weights (2013)	Paired comparison	Global		2015	D

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Asthma uncontrolled: 0.133						
	Asthma, controlled: 0.020 Asthma, partially controlled: 0.045	European weights	Paired comparison	Europe		2015	C
	Asthma: 0.093	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
	Mild asthma: 0.030 Severe: 0.230	Dutch weights for mild asthma/ Severe asthma has been estimated using EQ-5d+ regression model and Australian data on severity distribution of disability	PTO + VAS/ EQ-5D	45 people, the Netherlands/Australia		2005 (1997)	A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
Other							
Allergic contact dermatitis	Eczema 0.056	Estimated from EQ-5D+ regression model/ Estimated from EQ-5D+ regression model	EQ-5D	Australia			A
	Dermatitis: 0.183	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
Anaemia	Mild anaemia: 0.005 […] Very severe anaemia 0.250	GBD weight: Estimated using EQ-5D+ regression model, 1996	EQ-5D	Australia		2005	A
	Anaemia: mild: 0.004 Moderate: 0.052 Severe: 0.149	GBD weights (2013)	Paired comparison	Global		2015	D
	Anaemia: Mild: 0.004 Moderate: 0.045 Severe: 0.118	European weights	Paired comparison	Europe		2015	C
	Anaemia: 0.168	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Osteoporosis	Diagnosed cases 0.009	Estimated using EQ-5D+ regression model	EQ-5D	Australia			A

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
Glaucoma	Glaucoma—mild vision loss 0.020 Glaucoma—moderate vision loss 0.170 Glaucoma—severe vision loss 0.430	Dutch weight	PTO + VAS	45 people, the Netherlands			A (B)
	Near vision impairment: 0.012 Distance vision, mild impairment: 0.004 Moderate impairment: 0.034 Severe impairment: 0.158 Distance vision blindness: 0.173	European weights	Paired comparison	Europe		2015	C

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Distance vision, mild impairment: 0.003 Moderate impairment: 0.031 Severe impairment: 0.184 Distance vision blindness: 0.187	GBD weights (2013)	Paired comparison	Global		2015	D
Hearing impairment	Mild hearing loss (25–34 dBHTL) 0.020 Mild hearing loss (35–44 dBHTL) 0.040 Moderate hearing loss 0.120 Severe hearing loss 0.370	One half of Dutch weight for mild hearing loss/ Dutch weight/ Dutch weight	PTO + VAS	45 people, the Netherlands			A (B)

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Hearing loss: Mild: 0.01 Moderate: 0.027 Severe: 0.158 Profound: 0.204 Complete: 0.215	GBD weights (2013)	Paired comparison	Global		2015	D
	Hearing loss: Mild: 0.011 Moderate: 0.037 Severe: 0.152 Profound: 0.235	European weights	Paired comparison	Europe		2015	C
	Hearing decrease: 0.254	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Liver inflammation/degenerative changes (of non-infective origin)/cirrhosis	Decompensated cirrhosis of the liver: 0.163	European weights	Paired comparison	Europe		2015	C
	Liver cirrhosis: 0.475	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Stomach ulcer	Peptic ulcer disease 0.066	Dutch weight	PTO + VAS	45 people, the Netherlands			A (B)
	Gastrointestinal ulcers: 0.104	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
Pancreatitis	Cases 0.349	Estimated using EQ-5D+ regression model	EQ-5D	Australia		2005	A
	Acute Pancreatitis: 0.499	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Parkinson –like condition	Mild: 0.016 Moderate: 0.239 Severe:0.530	European weights	Paired comparison	Europe		2015	C
	Mild: 0.01 Moderate: 0.267 Severe:0.575	GBD weights (2013)	Paired comparison	Global		2015	D
	Parkinson's disease: Initial stage: 0.480 Intermediate stage: 0.790 End stage: 0.920	Dutch weight	PTO + VAS	45 people, the Netherlands			A (B)
	Parkinson's disease: 0.244	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E
Alzheimer disease and other dementia	Dementia: Mild 0.069 Moderate 0.377	GBD weights (2013)	Paired comparison	Global		2015	D

Table A3-2: Collation of available disability weights from the Global BoD 2013 update (2015), European BoD study (2015) and Estonian BoD study (2015)							
Disease	**DALY utility weight**	**Source and the date of the source**	**Method**	**Sample size Number of people Experience: healthy or sick Region (USA, Asia, Europe…)**	**What it is measuring**	**Date**	**Ref**
	Severe 0.449						
	Dementia: Mild 0.059 Moderate 0.434	European weights	Paired comparison	Europe		2015	C
	Alzheimer and other dementias: Mild: 0.270 Moderate: 0.630 Severe: 0.940	Dutch weight	PTO + VAS	45 people, the Netherlands			A (B)
	Dementia: 0.261	Estonian weights	9% PTO, 91% VAS	Panel, medical experts, Estonian population		2009	E

Data sources for QALY utility weights

(1) = Guerriero C et al. (2012): The cost-effectiveness of smoking cessation support delivered by mobile phone text messaging: Txt2stop, information downloaded from http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC3751449&blobtype=pdf

(2) = Feenstra TL (2005): Cost-Effectiveness of Face-to-Face Smoking Cessation Interventions: A Dynamic Modeling Study, information downloaded from http://onlinelibrary.wiley.com/doi/10.1111/j.1524-4733.2005.04008.x/epdf

(3) = Hutton DW et al. (2010): Cost-Effectiveness of Nationwide Hepatitis B Catch-up Vaccination among Children and Adolescents in China, information downloaded from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245734/pdf/nihms150156.pdf

(4) = Hirst et al. (2012): Lifetime Cost-Effectiveness of Skin Cancer Prevention through Promotion of Daily Sunscreen Use, information downloaded from http://ac.els-cdn.com/S1098301511035303/1-s2.0-S1098301511035303-main.pdf?_tid=f3aeb39c-b765-11e4-a391-00000aacb360&acdnat=1424261103_523237c6acd4d6ca3b3f8836d130428e

(5) = Kwon JS et al. (2013): Prophylactic Salpingectomy and Delayed Oophorectomy as an Alternative for BRCA Mutation Carriers, information downloaded from http://journals.lww.com/greenjournal/Fulltext/2013/01000/Prophylactic_Salpingectomy_and_Delayed.5.aspx

(6) = Dan YY et al. (2006): Endoscopic Screening for Gastric Cancer, information downloaded from http://www.cghjournal.org/article/S1542-3565%2806%2900328-4/abstract

(7) = National Institute for Health and Clinical Excellence (2010): Rituximab for the treatment of relapsed or refractory chronic lymphocytic leukaemia, information downloaded from https://www.nice.org.uk/guidance/ta193/resources/guidance-rituximab-for-the-treatment-of-relapsed-or-refractory-chronic-lymphocytic-leukaemia-pdf

(8) = Soini EJO, Martikainen JA & Nousiainen T (2010): Treatment of follicular non-Hodgkin's lymphoma with or without rituximab: cost-effectiveness and value of information based on a 5-year follow-up, information downloaded from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3082160/pdf/mdq582.pdf

(9) = Tam VC et al. (2013): Cost-effectiveness of systemic therapies for metastatic pancreatic cancer, information downloaded from http://www.current-oncology.com/index.php/oncology/article/view/1223/1162

(10) = National Institute for Health and Clinical Excellence (NICE) (2014b): Enzalutamide for metastatic hormone-relapsed prostate cancer previously treated with a docetaxel-containing regimen, information downloaded from https://www.nice.org.uk/guidance/ta316/resources/guidance-enzalutamide-for-metastatic-hormonerelapsed-prostate-cancer-previously-treated-with-a-docetaxelcontaining-regimen-pdf

(11) = Henriksson M et al. (2010): Assessing the cost effectiveness of using prognostic biomarkers with decision models: case study in prioritising patients waiting for coronary artery surgery, information downloaded from http://www.bmj.com/content/bmj/340/bmj.b5606.full.pdf(12) =

Noah-Vanhoucke J, Green LE, Dinh TA, Alperin P & Smith RA (2011): Cost-Effectiveness of Chemoprevention of Breast Cancer Using Tamoxifen in a Postmenopausal US Population, information downloaded from http://onlinelibrary.wiley.com/doi/10.1002/cncr.25926/epdf

(13) = National Institute for Health and Clinical Excellence (NICE) (2013): Fertility: assessment and treatment for people with fertility problems, information downloaded from http://www.nice.org.uk/guidance/cg156/resources/cg156-fertility-full-guideline3

(14) = National Institute for Health and Clinical Excellence (NICE) (2007): Heavy menstrual bleeding, information downloaded from http://www.nice.org.uk/guidance/cg44/evidence/cg44-heavy-menstrual-bleeding-full-guideline2

(15) = Korvenranta E et al. (2010): Hospital costs and quality of life during 4 years after very preterm birth, information downloaded from http://archpedi.jamanetwork.com/article.aspx?articleid=383428

(16) = Prosser LA et al. (2012): Decision analysis, economic evaluation, and new-born screening: challenges and opportunities, information downloaded from http://www.nature.com/gim/journal/v14/n8/pdf/gim201224a.pdf

(17) = Wehby GL, Ohsfeldt, RL & Murray JC (2006): Health Professionals' Assessment of Health-Related Quality of Life Values for Oral Clefting by Age Using a Visual Analogue Scale Method, information downloaded from http://www.cpcjournal.org/doi/pdf/10.1597/05-066.1

(18) = Rabovskaja V, Parkinson B & Goodall S (2013): The Cost-Effectiveness of Mandatory Folic Acid Fortification in Australia, information downloaded from http://jn.nutrition.org/content/143/1/59.full.pdf+html

(19) = Erickson KF et al. (2013): Cost-Effectiveness of Statins for Primary Cardiovascular Prevention in Chronic Kidney Disease, information downloaded from http://ac.els-cdn.com/S0735109713002623/1-s2.0-S0735109713002623-main.pdf?_tid=2430af42-b906-11e4-9a54-00000aacb360&acdnat=1424439855_964e268a32bf885b897e481f4a94daf7

(20) = Van den Akker-van Marle, ME et al. (2013): Management of Undescended Testis : A Decision Analysis, information downloaded from https://d1l066c6yi5btx.cloudfront.net/ncj/ncj/docs/publicatie_akker-vanmarle_et_al_mdm2013.pdf

(21) = Price D et al. (2013): A UK-Based Cost-Utility Analysis of Indacaterol, A Once-Daily Maintenance Bronchodilator for Patients with COPD, Using Real World Evidence on Resource Use, information downloaded from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3663982/pdf/40258_2013_Article_21.pdf

(22) = Price D, Berg J & Lindgren P (2009): An economic evaluation of NIOX MINO airway inflammation monitor in the United Kingdom, information downloaded from http://onlinelibrary.wiley.com/doi/10.1111/j.1398-9995.2008.01855.x/epdf

(23) = National Institute for Health and Clinical Excellence (NICE) (2005): The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation, information downloaded from http://www.journalslibrary.nihr.ac.uk/__data/assets/pdf_file/0018/65142/FullReport-hta9290.pdf

(24) = Hiligsmann M, Sedrine WB & Reginster JY (2013): Cost-Effectiveness of Bazedoxifene Compared With Raloxifene in the Treatment of Postmenopausal Osteoporotic Women, information downloaded from http://onlinelibrary.wiley.com/doi/10.1002/jbmr.1819/epdf

(25) = National Institute for Health and Clinical Excellence (NICE) (2014a): Erythropoiesis-stimulating agents (epoetin and darbepoetin) for treating anaemia in people with cancer having chemotherapy (including review of TA142), information downloaded from http://www.nice.org.uk/guidance/ta323/resources/guidance-erythropoiesisstimulating-agents-epoetin-and-darbepoetin-for-treating-anaemia-in-people-with-cancer-having-chemotherapy-including-review-of-ta142-pdf

(26) = Stein JD et al. (2012): Cost Effectiveness of Medications Compared with Laser trabeculoplasty in Patients with Newly-Diagnosed Open-Angle Glaucoma, information downloaded from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3324653/pdf/nihms-365501.pdf

(27) = National Institute for Health and Clinical Excellence (NICE) (2009): Cochlear implants for children and adults with severe to profound deafness, information downloaded from https://www.nice.org.uk/guidance/ta166/resources/guidance-cochlear-implants-for-children-and-adults-with-severe-to-profound-deafness-pdf

(28) = Hoerger et al. (2013): Cost-Effectiveness of Hepatitis B Vaccination in Adults with Diagnosed Diabetes, information downloaded from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3526214/pdf/63.pdf

(29) = Song HJ, Kwon JW, Kim N & Park YS (2013): Cost Effectiveness Associated with Helicobacter pylori Screening and Eradication in Patients Taking Nonsteroidal Anti-Inflammatory Drugs and/or Aspirin, information downloaded from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3607772/pdf/gnl-7-182.pdf

(30) = Poulose BK, Speroff T & Holzman MD (2007): Optimizing choledocholithiasis management: a cost-effectiveness analysis, information downloaded from http://archsurg.jamanetwork.com/article.aspx?articleid=399595

(31) = Eggington S et al. (2013): The cost-effectiveness of deep brain stimulation in combination with best medical therapy, versus best medical therapy alone, in advanced Parkinson's disease, information downloaded from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3895185/pdf/415_2013_Article_7148.pdf

(32) = McMahon PM et al. (2000): Cost-effectiveness of functional imaging tests in the diagnosis of Alzheimer disease, information downloaded from http://pubs.rsna.org/doi/pdf/10.1148/radiology.217.1.r00se1358

Data sources for DALY disability weights

A = Victorian Government Department of Human Services (2005): Victorian Burden of Disease Study: Mortality and morbidity in 2001, information downloaded from http://docs.health.vic.gov.au/docs/doc/6AEAFAB1BAE696B9CA257886000158A0/$FILE/bod_2001.pdf

B = Stouthard MEA et al. (1997): Disability weights for diseases in the Netherlands, information downloaded from http://dare.uva.nl/document/2/3276

C= Haagsma JA et al. (2015): Assessing disability weights based on the responses of 30,660 people from four European countries, information downloaded from http://www.pophealthmetrics.com/content/13/1/10

D = Vos T et al. (2015): Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013, information downloaded from http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2815%2960692-4/fulltext

(E) = Lai T, Habicht J, Kiivet RA (2009): Measuring burden of disease in Estonia to support public health policy, information downloaded from http://eurpub.oxfordjournals.org/content/19/5/541