Guidance on Derivation of DNEL/DMEL from Human Data

DRAFT

(Rev.:2.0)

CONTENTS

Note for the Guidance consultation process

The sections above are proposed amendments to the Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.8 "Characterisation of dose (concentration) – response for human health".

Chapter R.8 was initially designed to cover the derivation of DNEL/DMEL both from human and from animal data. However, it did not cover in detail the specificities of human data when used as the starting point for the derivation of DNEL /DMEL. The present Guidance, conceived as an addition to the actual Chapter R.8, deals exclusively with this topic. It aims to explain in detail how to evaluate the data, extract the dose descriptors and derive the DNEL/DMEL where relevant human studies are used. It also presents the criteria to be applied in case human and animal data are available for the same endpoint/exposure pattern.

Where possible, the new sections refer to tables and sections of Chapter R.8 to avoid duplication of material. References to Chapter R.4 are also included, where applicable, to connect the guidelines on human data evaluation presented here with the more general advice already given.

The Guidance on DNEL/DMEL derivation from human data has been developed as a relatively independent/stand alone document to be used in the chemical safety assessment of substances for which relevant human data are available. The major contents of the Guidance are included in Appendix R.8-15 which covers the different phases of the DNEL/DMEL derivation process from human data. The decision to include the Guidance mainly as an Appendix to Chapter R.8 was taken in order to avoid any major changes in the main bodies of the already available guidance documents at this stage of the registration process.

It is ECHA's intention to restructure Chapter R.8 and Chapter R.4 once the first registration deadline is over. This planned restructuring will integrate the derivation of DNEL/DMEL process from human data in the main body of Chapter R.8. It will also take out from Chapter R.8 those sections that deal with data evaluation and will incorporate them to Chapter R.4, where they properly belong.

Chapter R.8 planned restructuring will also offer an opportunity to add more practical guidance on the use of human data in the form of examples. Additionally, specific guidelines on the use of human data other than epidemiological studies will be further developed. This may become increasingly relevant for future registration dossiers of lower tonnage levels.

R.8.1.2.8 Human data as source for derivation of DNEL and/or DMEL

Since DNELs and DMELs are used in the assessment of risks to humans, human data are an appropriate basis also for the derivation of DNEL/DMEL. Human data are valuable as a source of hazard information because they apply directly to the human species, and the mode of action (MoA) is usually relevant. As a consequence, no inter-species assessment factor is needed when human data are used for derivation of DNEL/DMEL. Furthermore, human data have in most cases been obtained from relevant exposure conditions and are based on an adequate route of exposure. In addition, human data most often come from studies covering a more heterogeneous sample of the population than animal studies carried out on inbred strains. Nevertheless, the quality of the human data needs to be ensured. Under REACH human data, when available and relevant, are used in the human hazard assessment and as part of the Chemical Safety Assessment as described in Annex I of the REACH Regulation. More specifically, according to Annex I the human health hazard assessment comprises four steps:

Step 1: Evaluation of non-human information.

Step 2: Evaluation of human information.

Step 3: Classification and labelling.

Step 4: Derivation of DNELs.

Furthermore, according to the provisions of Annex XI of the REACH Regulation "Historical Human Data"[1] can be used to adapt the standard testing requirements of Annexes VII to X provided that the quality of the data is properly assessed and found to be adequate.

Human data can come from analytical epidemiology studies, descriptive or correlation epidemiology studies, case reports, clinical studies, poison centre information, occupational disease registries or other occupational surveillance systems. When they are already available, well-conducted controlled human exposure studies in volunteers, including low exposure toxicokinetics studies, can also be used in risk assessment. However, few human experimental toxicity studies are available due to the practical and ethical considerations involved in deliberate exposure of individuals. Such studies, e.g. studies carried out for the authorisation of a medical product have to be conducted in line with the World Medical Association Declaration of Helsinki, which describes the general ethical principles for medical research involving human subjects (World Medical Association, 2000). It is emphasised that testing with human volunteers is strongly discouraged, but when there are good quality data already available they should be used as appropriate, in well justified cases (see Chapter R.4).

Human data differ from animal data in that they are mostly derived from observational (non experimental) studies in contrast to controlled experimental animal studies. This has profound consequences for the reviewing and handling of data. In experimental studies data quality is controlled a priori by the experimental study design while the relevance to humans needs to be assessed a posteriori. In observational human studies, data quality control is done in connection with data analysis (a posteriori) with focus on the validity of the data. Furthermore, animal studies are done in inbred strains, whereas epidemiological studies are done on heterogenic populations. This implies that the process to arrive at a dose descriptor from human studies is somewhat different from that of obtaining a dose descriptor from animal studies. Quality considerations in particular differ from those for experimental studies and the accuracy of the exposure information is an important issue. APPENDIX

[1] Historical Human Data is a term used by REACH. It refers to already available human data.

R.8-15 gives guidance for all the phases to arrive at a dose descriptor and to derive DNEL/DMEL from human data (see Figure R.8-x in APPENDIX R.8-15). The process leading to the identification of the leading health effect and the associated levels of exposure is also described in APPENDIX R.8-15. There are specific uncertainties that deserve attention when using human data. These include the influence of bias, confounding from mixed exposures and other risk factors and accuracy of the exposure information. Therefore special expertise is needed when using epidemiological or other human data for obtaining DNELs/DMELs.

The term "dose descriptor" is used to designate the exposure level (dose or concentration) that corresponds to a quantified level of risk of a health effect in a specific study. In animal studies common dose descriptors for threshold effects are NOAEL (No Observed Adverse Effect Level) or LOAEL (Lowest Observed Adverse Effect Level), while examples of dose descriptors of non-threshold effects are T25 and BMD10. For epidemiological or other human data, typical dose descriptors for threshold effects are exposure levels for which health effects are not observed or are observed (NOAEL or LOAEL, or NOAEC etc.) as well. For non-threshold effects the dose descriptors are often expressed as levels of exposure that are associated with a Relative Risk (RR) or comparable relative risk metrix such as Odds Ratio (OR), Standardised Mortality Ratio (SMR) or Standardized Incidence Ratio (SIR). Nevertheless, levels of exposure associated with such relative risk metrics can also be used as dose descriptors for threshold effects.

Human data are in principle the most relevant source of information on human toxicity (see Chapter R.4). Since there may be limitations in reliability of human studies (e.g. problems in study design, analysis and reporting as well as limited coverage of the different target organs), they are normally considered together with animal and other data. For many chemicals, both human data and animal data are available (see Money 2007 for a summary on the use of human data under the Existing Substances Regulation). Therefore an integrated approach is required. This also applies to DNEL/DMEL derivation. APPENDIX R.8-15, Phase 9, provides advice on the integration of animal and human data. In this approach, the criteria for the selection of useful data are the *quality* and *relevance* of the data and the *level of the DNELs/DMELs* obtained from human versus animal data. It is assumed that before the integration phase is started, the relevant animal studies have been examined and the DNELs/DMELs have been derived from them according to the guidance given in Sections R.8.2 to R.8.6.

An important element of the CSR is the justification and documentation of the choices made in the DNEL/DMEL derivation, in particular, when choosing the assessment factors, dose descriptors and the leading health effect.

Appendix R.8-15 Use of human data in the derivation of DNEL and DMEL

Introduction

Human data have been used in the risk assessment of many chemicals; e.g. under the Existing Substance Regulation (EEC 793/93), for setting Occupational Exposure Limits (OELs), and for cancer risk assessment by the International Agency for Research on Cancer (IARC). The merits of the use of human data as a source of hazard information are that the mode of action is usually relevant, i.e. the same mechanism applies as in the larger population that should be protected and no inter-species safety or assessment factors are needed. In addition to that, human data in most cases originate from exposure levels comparable to those in the target population and relate to a pertinent route of exposure.

Under REACH human data, when available and relevant, are used in the human hazard assessment and as part of the Chemical Safety Assessment as described in Annex I of the REACH Regulation. Furthermore, according to the provisions of Annex XI of the REACH Regulation "Historical Human Data"[1] can be used to adapt the standard testing requirements of Annexes VII to X provided that the quality of the data is properly assessed and judge to be appropriate.

The human health hazard assessment as defined in Annex I of REACH comprises four steps, the last one of which is the derivation of DNELs:

Step 1: Evaluation of non-human information.

Step 2: Evaluation of human information.

Step 3: Classification and labelling.

Step 4: Derivation of DNELs.

The above-mentioned steps 1 to 3 shall be undertaken for every effect for which information is available. Step 4, derivation of DNELs, shall be undertaken by integrating the results of steps 1-3. If justified by the exposure scenario(s), a single DNEL may be sufficient. However, taking into account the available information and the exposure scenario(s) of the CSR it may be necessary to identify different DNELs for each relevant human population (e.g. workers, consumers) and possibly for certain vulnerable sub-populations (e.g. children, pregnant women) and for different routes (oral, inhalation, dermal) and durations (acute, long-term) of exposure.

It may not always be possible to derive a DNEL for an end-point. This may be the case when a substance exerts its effect by a non-threshold mode of action (e.g. carcinogenicity through genotoxic mechanism) or the threshold cannot be reliably identified (e.g. sensitisation and irritation). For such carcinogens, and assuming that there are data allowing it, the registrant may develop a DMEL as stated in Section R.8.1.1.

The process of deriving a DNEL/DMEL from human data is divided in nine phases as shown in Figure R.8-x below. Each of these phases is described in detail in the following sections. Phases 1 to 5 are identical for DNEL and DMEL derivation and therefore both derived effect levels are dealt with together. However, Phases 6 to 8 differ significantly in case a DNEL or a DMEL is derived. For these

[1] Historical Human Data is a term used by REACH. It refers to already available human data.

phases a different section has been developed for DNEL and DMEL with the specificities of each case. In Phase 9 all effect levels (derived either from animal data (AD) or from human data (HD)) are brought together for selection of the critical DNELs/DMELs to be taken to the risk characterisation.

Phases 1 to 5 are common for all types of human data, independent of the type of the effect. However, some specific guidance for DNEL derivation from human data on acute toxicity, irritation/corrosion and sensitization are given in Appendices R.8-8, R.8-9, and R.8-10.

For using purely qualitative human data, which is not within the scope of this Appendix, the reader is referred to Sections R.8.6 and Guidance E.

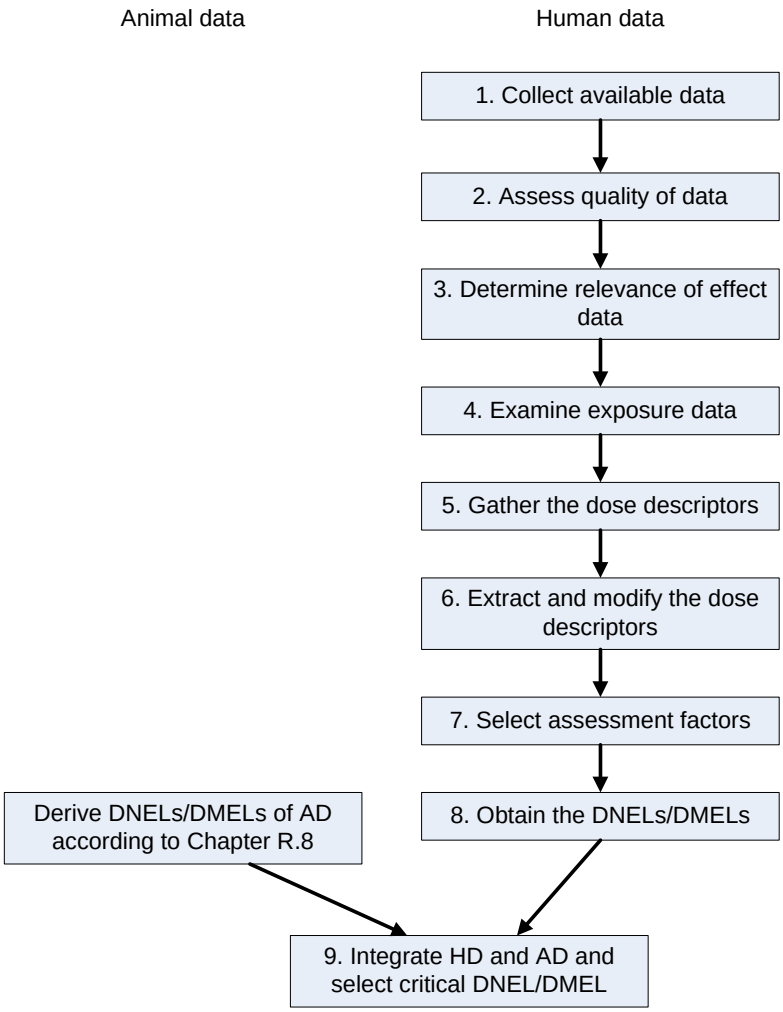

Figure R 8-x. Illustration of the process for DNEL/DMEL derivation from human data.

The use of human data for risk characterisation requires a high level of expertise in relevant scientific fields (e.g. epidemiology, industrial hygiene, risk assessment). It is not possible to provide simple and explicit guidance on every detail; this guideline rather aims to provide the general framework as a basis for expert judgment. This guidance has been made as practical as possible bearing in mind the needs of an "average" user. In order to make the structure and logic of the guidance as clear as possible, each section includes a short reminder of what should be available when the phase is completed. With these reminders the user of the guidance can make sure that the phase is completed and adequate recording has been done.

Interpretation and assessment of epidemiological and other human data necessitate a good understanding of the inherent methodological issues. It is recommended that an epidemiologist by training or another person with relevant expertise on use of human data is involved in the assessment at least in the following situations:

1) the derivation of DNEL/DMEL includes an extensive amount of epidemiological studies or other human data

2) the derivation of DNEL/DMEL is mainly based on use of human data

3) human data not previously published in peer-reviewed scientific journals are being applied

4) the derivation of DNEL/DMEL includes adjustments or conversions of the original exposure information or statistical re-processing of original data in order to derive the necessary quantitative exposure estimates.

(Phase 1) Collection of available data

In this section the types of human data that can be used for derivation of the DNEL/DMEL and their availability are briefly described. Assessment of the quality of these data and evaluation of exposure data are covered in the relevant sections below.

The focus of many parts of this guidance is on the epidemiological data. It may be the most reliable type of human data but it would not be available for a very large number of chemical substances. When there are epidemiological data e.g. on carcinogenicity, reproductive toxicity or organ toxicity, before they are used, they should be carefully evaluated for their relevance in the overall risk assessment, including setting the DNEL/DMEL. The basic approach to the use and evaluation of human data is described in Chapter R.4 for all main types of human data, i.e. analytical epidemiology studies, descriptive or correlation epidemiology studies, medical surveillance data and case reports (in very rare justified cases also existing controlled studies in human volunteers). As underlined in Chapter R.4, a weight of evidence (WoE) approach is essential for risk assessment based on epidemiological and other human data. The availability, sources, use and value of human data vary according to the effect. Some endpoint-specific features of human information and its sources as well as evaluation of the human data and information requirements to be fulfilled are outlined in Chapter R.7, in the respective effect-specific sections R.7.2. to R.7.7.

Published epidemiological data can be identified by searching Medline or other Life Science and Biomedical databases. Human data other than epidemiological studies can come from e.g. case reports, clinical studies, occupational disease registries or other occupational surveillance schemes and from poison centre information. In principle all types of toxic effects can be reported in such studies; however, in many cases they address acute and/or local effects. These data can be obtained e.g. from open literature by searching the relevant publication databases (see above), from occupational health units or occupational medicine clinics. Industry sources of unpublished monitoring and surveillance may be important although further efforts are needed to improve uniformity and harmonisation or these data (Ecetoc 2004). IPCS has made an attempt to demonstrate the feasibility of collecting harmonised human data from poisons information centres on a multi-national basis (IPCS 2005a, Ecetoc 2004), which increases the possibility of having these data available in the same form from different countries and with a sufficiently large number of cases. The use of harmonised reporting formats and terminology, developed originally by the IPCS (INTOX Data Management System) has been suggested for aggregating poison centre information. Finally, data collected under other EU or international regulatory processes, e.g. medicines or biocides, could be useful for some chemical substances.

Major chemical companies often have a routine medical surveillance system in place to monitor and manage employee health. Periodic routine medical examinations are offered to employees and these data are maintained in medical files in order to perform clinical practice and provide good quality clinical consultation. Apart from general medical surveillance programs, targeted programs exist to monitor employee health of those with potential exposure to certain chemicals. For example, employees with potential for exposure to benzene may be invited to participate in frequent medical examinations focused on the potential health outcomes such as changes in haematological parameters. This kind of data can also be useful either for setting the DNEL or as an element of the weight of evidence assessment.

The collection of data should be done in a non-discriminatory way as regards the nature of observations made in the individual study. This means that neither negative nor positive results should be preferred in the data collection phase. It is known that the studies with positive results tend to be more easily published than those without an observed effect. Even though statistical methods like the funnel plot have been developed to identify publication bias in meta-analysis (Egger et al 1997) it is difficult to completely overcome this problem when searching the published literature. On the other hand published studies have undergone a review process that aims to ensure a high quality. When using unpublished data (e.g. company surveillance data) an objective approach to collecting data is equally important.

Result of Phase 1: The existing human data on the substance have been collected from the relevant sources. A list of studies has been prepared. It may be useful to indicate for each item in the list: the reference indicating the type of source (published article, published review or report, unpublished data) and the type of health effect observed.

(Phase 2) Assessment of the quality of the human data

Historical human data, such as epidemiological studies on exposed populations, accidental or occupational exposure data and clinical studies, shall be considered in this phase. The value of the data for a specific human health effect depends, among other things, on the type of analysis and on the parameters covered and on the magnitude and specificity of the health effect and consequently the predictability of the effect. Annex XI of the REACH Regulation sets out the following general criteria for assessing the adequacy of the data in view of using human data for the adaptation of the standard testing requirements:

1) the proper selection and characterisation of the exposed and control groups;

2) adequate characterisation of exposure;

3) sufficient length of follow-up for disease occurrence;

4) valid method for observing an effect;

5) proper consideration of bias and confounding factors; and

6) a reasonable statistical reliability to justify the conclusion.

It is recommended to use the same assessment criteria for the use of human data in the DNEL/DMEL derivation.

Epidemiology is the study of how often and why diseases occur in different groups (Coggon et al. 1997 bmj-online). Comprehensive guidance on both evaluation and use of epidemiological evidence for risk assessment is provided by Kryzanowski et al. (WHO 2000). Specific guidance on criteria to assess the quality of epidemiological studies can be found in text books of epidemiology (Checkoway et al. 2004, Hernberg 1992, Rothman and Greenland 1998).Chapter R.4 and Chapter R.7 of the Guidance on Information Requirements and Chemical Safety Assessment also describe some general and effect-specific quality aspects. The most important points to be addressed when assessing and documenting

the above-mentioned six adequacy criteria are described below. Please note that the focus in this section is on the assessment of the quality of epidemiological studies. Although many of the quality assessment aspects (e.g. reliability of exposure and outcome information) are similar to other types of human data, there are also important differences, for example when case reports are assessed. This is further discussed later on in this Phase. APPENDICES R.8-8 to R.8-10 also contain useful information on the use of human data concerning acute toxicity, irritation and sensitisation.

1) The proper selection and characterisation of the exposed and control groups

Most commonly epidemiological studies are categorised into cohort (longitudinal), case-control (case-referent) and cross-sectional studies. Due to their non-experimental nature they are almost invariably subject to bias of one sort or another. A crucial aim of the study design is to keep bias to the minimum. The main types of bias to be avoided are selection bias (e.g. the so called healthy worker effect[1]) and information bias (e.g. recall bias). Both the selection and recruitment of the study subjects and the collection of information concerning their exposures and diseases should be done in such a way that they do not introduce bias in the difference in the disease rates between the exposed and unexposed groups (cohort design) or the difference in the occurrence of exposure between the cases and the controls (case-control design). In a cross-sectional design, an important aspect is the temporal relationship between exposure and health effect. Many types of effects are known not to occur immediately after exposure and original members of the study population may leave the study during the latent period of the health effect. Cross-sectional studies are considered to be suited to study acute effects or effects that do not lead to serious overt disease that would result in affected subjects leaving the exposure environment. A good epidemiological study describes and justifies the selection and recruitment procedures of the study subjects (exposed/unexposed cohorts or cases and controls), sampling, number of study subjects, non-response/non-participation, completeness of follow-up and the measures undertaken to ensure a comparable ascertainment of exposure and disease status between the different study groups.

2) Adequate characterisation of exposure

The relevant exposure parameter (mean level, peak level, duration, cumulative dose) depends on the health outcome and exposure setting and should be justified. In the analysis exposure can be considered as a continuous or categorised variable and the choices made should be described and justified. Multiple measurements will increase the accuracy of the exposure information especially in long-term exposures or when the exposure variation is high (see Phase 4). Errors in measuring exposure can be an important source of bias in epidemiology, especially concerning exposures further in the past and mixed exposures. In a good epidemiological study the methods to assess exposure are clearly described and their validity is assessed in case previously non-validated methods are used. Furthermore, exposure data should be quantitative and exposure categories well defined, in order to allow a DNEL/DMEL to be obtained from that study.

[1] The healthy worker effect results from three main phenomena: the healthy hire effect refers to the fact that the healthiest individuals are the most likely to get a job and the healthy worker survivor effect refers to the healthiest individuals being the most likely to continue in the job. The third phenomenon refers to changes in life that are related to the employment.

Comparing results from several studies may necessitate conversion of exposure parameters. Sometimes more qualitative exposure categories need to be converted to quantitative estimates by using exposure data from other sources. Combining results of several studies may involve additional statistical modelling and new analyses. In all the above situations the choices and compromises made should be clearly described and their potential impact preferably assessed by sensitivity analyses using varying parameter values. The uncertainty resulting from such approximations needs to be assessed and may need to be taken into account when applying the assessment factors (see Phase 4 and Phase 7).

3) Sufficient length of follow-up for disease occurrence

The latency time between exposure and the occurrence of the health outcome varies from less than one day (e.g. acute irritation) to decades (e.g. most malignant diseases). A good epidemiological study describes and justifies the length of follow-up in longitudinal studies and the exposure time windows used in case-control studies.

4) Valid method for observing an effect

Measuring disease occurrence in populations requires diagnostic criteria. For practical reasons the criteria used in epidemiological studies usually differ at least somewhat from the criteria used in clinical practice. Many health outcomes represent more of a continuum rather than a dichotomous phenomenon and therefore standard predefined criteria need to be established for classifying the disease status of the study subjects. The types of health effects for which human data exist vary from acute effects to long term effects. The occurrence of health effects can be determined in various ways. Self filled-in questionnaires, clinical examinations or queries from already existing databases (e.g. causes of death databases or cancer registries) can be used. The quality of the health effect data depends on the data collection methods used. Standardised and validated data collection or diagnostic techniques with satisfactory sensitivity and specificity should have been used. It is crucial that the reliability of health effect data collection techniques is the same for the exposed and non-exposed populations. Errors in measuring disease status can be an important source of bias in epidemiology. In a good epidemiological study the diagnostic criteria used are clearly described and their validity is assessed in cases where previously non-validated criteria are used.

For a given chemical substance, several health effects may be relevant. The studies addressing these different health effects may have used quite different methods to ascertain the occurrence of that health effect. Each of these should be assessed for the validity in order to be sure that adequate information is available in the later phases of the DNEL/DMEL derivation process.

5) Proper consideration of bias and confounding factors

Bias should be minimised in the study design and the biases that cannot be avoided should be identified, assessed for their potential impact and taken into account when interpreting the results. Confounding occurs when the exposed and non-exposed populations have different background disease risks. Confounding could occur, for example, if the difference in the occurrence of the health effect results from differences in the age or gender distributions or life style rather than from differences in

the chemical exposure between the exposed and unexposed populations. If taken into account in the study design, the effect of confounding factors can be controlled in the analysis. A good epidemiological study describes the confounders that could be controlled in the analysis, and how they could be controlled , and estimates the potential impact of confounders that could not be controlled.

6) A reasonable statistical reliability to justify the conclusion

Usually the results of epidemiological studies are expressed in terms of risk estimates characterising the difference in risk of disease between the exposed and unexposed populations. Parameters like relative risk (RR), odds ratio (OR), standardised incidence or mortality rate (SIR or SMR) are used. More generally, the statistical analysis aims to describe the study results of individual observations as meaningful numerical values (mean, median etc.). Control of confounding is also an important element of the statistical analysis.

Even after biases have been minimised and confounding controlled, these mean values or risk estimates may be unrepresentative just by chance (random error). In general terms the statistical precision of the risk estimate, i.e. the narrowness of its confidence intervals, is inversely proportional to the number of cases observed, i.e. the larger the study size the more precise the risk estimate in statistical terms. In epidemiology, it is preferable to base the statistical inference on the point estimate of the risk and its confidence intervals. P-values from statistical testing are also used. Nevertheless, p values have inherent problems due to being dependent not only on how much the result deviates from the so called null hypothesis, but also on the sample size in which this deviation is observed. In addition to the point estimate and its statistical confidence intervals, the actual methods of statistical analysis and the statistical models used should be known and justified. In addition to the statistical precision and significance, the internal consistency of the results is an issue (e.g. did all the analyses support an association, was it observed in all subgroups, was there a dose-response relationship?).

Although the above six quality criteria are formulated especially from the point of view of the assessment of the adequacy of a single epidemiological study, they can be applied to assess the adequacy of the relevant aspects of other types of historical human data, e.g. accidental exposure data and occupational exposure data (medical surveillance data). For example major chemical companies have implemented routine health surveillance programs to monitor and manage employee health. Although the information is not specifically collected for hazard assessment purposes, it can provide a useful database provided that the relevant points of the above-mentioned six quality criteria are met, i.e. unpublished surveillance data should meet the same quality criteria as published reports. It is especially important to address and document in a report the attendance rates and other selection mechanisms, quality of the health effect and exposure data, confounding and biases and the proper analysis of data.(see von Elm et al. 2007 for guidance on proper reporting of observational human studies). A drawback often reducing the usefulness of company surveillance data is the lack of an unexposed comparison group that would have been followed with an identical protocol.

Case reports can also provide crucial qualitative information or even quantitative information on the dose descriptor, but they generally pertain to less complex situations and instances where the causality of exposure and effect are immediately obvious. This is the case for example, when the health effect is acute, specific or preferably both. The quality criteria above can still be applied for the relevant parts. The quality criteria for study design do not apply as such but bias and confounding should,

nevertheless, be considered and even very simple but reliable exposure information and health outcome information can suffice.

Bias and confounding are challenges in human data and therefore an important issue when assessing the quality of the data. Their potential influence, nevertheless, varies depending both on the specificity of the exposure-effect relation and on the latency period between exposure and effect. Therefore, in practice the quality requirements for specific effects (e.g. asbestosis) are different from those for multifactorial effects (e.g. lung cancer). The same applies for the difference between acute effects and long-term effects. These differences are also reflected in the requirements concerning the study design. Swaen (2006) and ECETOC (2009) have produced guidance on the quality aspects and use of information according to the type of effect and type of study. APPENDICES R.8-8 to R.8-10 also contain useful information on the use of human data concerning acute toxicity, irritation and sensitisation.

In all cases adequate and comprehensive enough documentation shall be provided to justify the use of human data in the derivation of DNEL/DMEL.

Result of Phase 2: Each human study has been characterised for its quality. Some studies with borderline quality, for example from the point of view of the quantitative nature of the exposure data, may be taken to the weight of evidence analysis, depending on the expert judgment. The studies with inadequate/low quality are unlikely to be of significant value in the DNEL/DMEL derivation.

(Phase 3) Evaluation of the relevance

An assessment of the likelihood of a **causal association** should be made for all endpoints or health effects identified in Phases 1 and 2. The best available guidance on causal inference are the criteria described by Hill (1965). These criteria should not be used in a stringent manner in the sense that they all must be met. A too stringent causal inference approach will lead to false negative conclusions. A too loose application of the Hill criteria will lead to false positives. A practical approach to application of Hill's criteria with an analysis based on actual data has been introduced by Swaen and Amelsvoort (2009. As underlined in Chapter R.4, a weight of evidence approach is essential for risk assessment based on epidemiological data. The specific features of evaluation of human information are outlined in Chapter R.7 in the respective effect-specific Sections R.7.2. to R.7.7.

Due to its non-experimental nature, human data, unlike animal experiments, very seldom relate to exposure to a pure, clearly defined chemical substance. **Confounding from concomitant other exposures** as well as non-specific characterisation of the chemical substances in question are common challenges that need to be assessed when judging the relevance of the human data.

Epidemiological studies may differ in the extent to which they are focussed on testing a specific hypothesis. Studies targeting testing of a specific hypothesis with a specific exposure, a specific effect and an *a priori* specified statistical analysis protocol should be given more weight than studies with a more exploratory character. In general studies designed to test a specific hypothesis tend to have more extensive and reliable quantitative exposure information than studies with a more general hypothesis or studies with an exploratory aim.

The evaluation should include an assessment of whether **the available human information addressing the endpoints of interests is sufficient and consistent** with the tonnage driven data requirements necessary to fulfil the REACH obligations, or whether the knowledge provided by the human information still presents data 'gaps'. For example, in the case where sperm quality has been analysed in a group of male workers, it would be incorrect to regard that study as covering all the reproductive toxicity parameters. In many cases there are gaps in the human studies. Nonetheless, animal data often complement human data. The whole database available (animal and human data) should be sufficient to address the endpoints compliant with the data requirements specific for the tonnage level. Where the human data set is incomplete, in terms of covering all the relevant effects, a DNEL/DMEL should not be established based upon that data alone, when additional relevant animal data is available. In case the study is incomplete but of sufficient quality it should be taken to Phase 9 below.

Result of Phase 3: Each relevant finding in the human studies has been characterised for its relevance, i.e. the degree of certainty concerning the causal relationship between the exposure and effect, completeness in terms of coverage of relevant effects, and association of the effect with a specified chemical in the case of multi-exposure.

(Phase 4) Examination of the exposure data

Exposure data have already been addressed as one element of the quality assessment of Phase 2. In Phase 4, quantitative measures of exposure are identified, which are later used in extraction of the dose descriptors In principle, only human studies with quantitative information on exposure are useful in the process of setting a DNEL/DMEL. If the exposure data are instead of a descriptive and qualitative nature, the study results can be relevant for hazard identification, but usually not for quantitative risk assessment; (however, see Section on “Quantitative exposure data by modeling” below.). The exposure information varies according to the type of the study. In case reports or studies with a limited number of individuals, as for acute effects, the exact doses and other exposure characteristics may be known. In epidemiological studies on long-term effects, data are often less accurate and can contain relatively old measurements or other exposure data, which are difficult to validate. Mixed exposure to several agents is often a problem, which needs to be taken into account in the study design, and in the analysis and interpretation (see Phases 2 and 3).

Exposure conditions can vary substantially. The number of exposure measurements needed depends on the variability of the exposure conditions. If exposure is stable, with no significant variation over the workday, the season or between time periods, a few sample points can be adequate to characterise the exposure situation. However in reality, exposures usually vary from place to place and, from task to task. They may change over time (short-term and long-term) due to differences in production process, exposure reduction measures, and use of personal protective equipment.

The type of exposure information also varies. Sometimes the only known exposure parameter is that a person has been employed in a particular industry. More specific information would be the type of job the person has been doing in that industry and over which time period. Quantitative exposure characterisation can be made if industrial hygiene measurements are available. In general, industrial hygiene measurements can be done for various purposes. They can be done e.g. to identify sources of

release or tasks with high exposure. In the latter case the results constitute an overestimate of general exposure at the workplace. Industrial hygiene measurements can also be conducted to provide a reliable picture of the exposure conditions at a specific work place. If the exposure measurements are collected by means of a systematic approach they are more valuable. It should be clear under which circumstances samples have been taken.

The precision of exposure measurements in estimating true exposure is not only determined by the number of measurements but also by the variability of exposure. Two aspects of exposure data are important for final interpretation of the findings. First, the internal validity should be satisfactory, meaning that the exposure data adequately describe the actual exposure situation. Internal validity depends on the sampling strategy and sampling frequency. Second, external validity should also be satisfactory. It relates to the comparability between the exposure conditions under investigation and the exposure conditions in other situations.

Money and Margary (2002) described a number of core principles to derive reliable and robust exposure assessments. They essentially describe three types of exposure data: actual data, analogous data and predicted exposure data derived from suitable validated models collected in a systematic manner. All three types of data can vary in quality and reliability.

Exposure data in a human study can be e.g.

- Measured data, which refers to ranges/categories of exposure (e.g. 0-10 ppm; 11-50 ppm, above 50 ppm). If these ranges are very wide, they may not be adequate for obtaining a DNEL/DMEL. When the sampling strategy and validity of exposure data is documented, it can be used for obtaining the DNEL/DMEL. How dose descriptors are derived from exposure categories is explained in Phase 5 below.

- Qualitative exposure categories such as “no exposure”, “low exposure”, “medium and high exposure”. As such this kind of data is not useful for setting a DNEL/DMEL, because no quantitative measure can be extracted from it.

- In case where biomonitoring values are available, where specific biomarkers can be clearly associated with the effects observed, they can be taken as dose descriptor. More guidance on the use of biomonitoring data in DNEL/DMEL derivation is given in Section R.8.1.2.7 and APPENDIX R.8-5.

- Measured analytical values associated both with effective dose/concentration and non-effective dose/concentration; e.g. for irritation, corrosion or, in some rare cases, for sensitisation. When representative and valid, this data can be used for obtaining a DNEL/DMEL.

Quantitative exposure data by modelling

In case qualitative exposure categories have been used in the original study, it may be possible on a case-by-case basis to obtain a quantitative estimate of exposure. More notably, there may be sufficient information in the human study on those exposure parameters that are needed for modelling the exposure. For example, the modelling tools referred to in Chapters R.14 and R.15 can sometimes be used. More sophisticated exposure modelling tools can also be used (IPCS 2005b). Exposure modelling in epidemiological or other human data requires specific expertise.

The information on the exposure parameters for modelling may be available from the same human study, or from a different study describing the operational conditions on that particular sector of industry or in those work tasks. Expertise on occupational hygiene is necessary for evaluation of relevance and reliability of this type of secondary data source.

It is emphasised that compensation of missing quantitative data by modelling results should only be done when that human study (with qualitative exposure data) is of good/sufficient quality. In case there are also other concerns (in addition to the exposure data) in relation to the quality, the study should not be used for setting the DNEL/DMEL. The aim of the modelling exercise is to "upgrade" or complement those human studies, which provide crucial evidence on a health effect that cannot be identified or adequately assessed based on other available animal or human studies. Thus, only studies with high relevance should be subject to this type of *ad hoc* exercise in the risk assessment.

Use of a job-exposure matrix (JEM) can in certain cases provide a reasonably robust tool for assessing the quantitative exposure levels linked to job titles or complete job title histories (Hoar 1983). In longitudinal epidemiological studies the construction of a job-exposure matrix has been shown to be a valuable means of using exposure information. It is only useful to construct a job-exposure matrix if (semi-)quantitative exposure information is available. The job-exposure matrix is based on homogeneous exposure groups, consisting of those jobs that are thought to be characterised by comparable exposure conditions. For each homogeneous exposure group the exposure intensity is estimated. Historical changes in the production process or work practices, resulting in changes in exposure, are taken into account and form a dimension of the matrix. The job-exposure matrix allows calculating cumulative exposure, but can also serve to stratify the study groups into subgroups with certain exposure characteristics, such as those exposed at least once over a certain concentration.

It is strongly recommended that in case modelling or application of a JEM is used *a posteriori* in order to generate the missing quantitative exposure data, that is done by an occupational hygienist or a similar expert, who has comprehensive knowledge of the relevance and use of various exposure parameters and is familiar with the modelling tool or JEM, which is used. The parameter choices made should be justified and documented in a clear and transparent manner. Sensitivity analyses should be performed in order to assess the effect of the parameter choices made and to justify the validity of the choices.

In human studies other than analytical epidemiological studies (case reports, medical surveillance data, etc) valuable quantitative exposure information can also be generated from more qualitative data (e.g. job titles, occupational histories). In such study designs a JEM is not needed, while access to the company information or more general industrial hygiene information concerning exposure levels is required (if not already used in the report available).

Result of Phase 4: The respective exposure levels (concentration or dose) have been assigned to each relevant health effect observed in the human studies. Additional information e.g. on the physicochemical properties of the substance, the pattern of use and the work tasks performed which affect exposure should be recorded here, as appropriate. It is recommended that the type of each exposure data is indicated: e.g., "measured worst case scenarios/high exposures", "representative measure data" or "modelling data". The studies which only have qualitative exposure data, and for which modelling cannot be applied are identified as such. They are put

aside for the next phases and when relevant and crucial, considered again in Phase 9 in the weight of evidence analysis.

(Phase 5) Gathering the dose descriptors

Provided that according to Phases 1 to 4 there are sufficient human data on a health effect associated with exposure to a certain chemical, the most reliable dose descriptors for each health effect will have to be identified.

If for a given health effect, only one human study with quantitative dose descriptor information has been identified in Phases 1 to 4, the selection of the dose descriptor is straightforward, i.e. the dose descriptor is the exposure concentration or dose that has been assigned to that health effect observed in the human study. It is to be underlined that, if overall only one human study is available, the quality of the data needs to be of high standard in order to justify its use, unless animal or other test data give supporting evidence (see Phase 9). In case several studies are available for a single health effect the most reliable should be selected using a weight of evidence approach. In case comparable and good quality data are available from more than one study, a summary measure could be created by a meta-analysis or pooled analysis. A weight of evidence approach is essential for risk assessment based on epidemiological data to (a) assess (sources of) heterogeneity across the studies and (b) increase statistical stability of the risk estimates. A meta-analysis of published studies or a pooled analysis of original raw data provides the best basis for deriving an overall dose-descriptor. Meta- and pooled analyses can also take into account small studies, which - on their own - are not suitable for deriving dose-descriptors due to statistical instability. If a good summary of all evidence is not available, using an individual relatively large study may be an acceptable, but statistically less accurate alternative (in comparison to a meta-analysis or pooled analysis using all the studies). For some substances, a dose descriptor on the dose-response curve may be derived from a single good quality epidemiology study, if this is the only adequate study. Once the dose descriptors have been gathered, they should be collected in a table (see Table R.8-14 of APPENDIX R.8-1). Please note that table R.8-14 will often contain dose descriptors expressed in ranges/categories of exposure. This is because epidemiological studies often relate the health effects to exposure categories. These categories are further processed in Phase 6.

Before actually deriving DNELs/DMELs on the basis of the derived dose descriptors, it is important to determine whether the substance exerts its effect by a threshold or by a non-threshold mode of action.

If the substance exerts its effect by a threshold mode of action, one or more DNELs will have to be derived for the different threshold endpoints, based on the most relevant dose descriptors for these endpoints. For non-threshold effects, for which in principle any level of exposure carries a risk, one or more DMELs could be derived instead (if data allow) on the basis of their most relevant dose descriptors.

The mechanism to derive DNELs differs substantially from the DMEL approach. For this reason, Phases 6, 7 and 8 of the DNEL/DMEL derivation process are described separately for each effect type (.i.e. threshold or non-threshold).

Result of Phase 5. The dose descriptors derived from relevant human data have been gathered (see Table R.8-14 in APPENDIX R.8-1). A decision has been made on the substance mode of action (threshold or non-threshold).

A.- DNEL DERIVATION FOR THRESHOLD EFFECTS

(Phase 6-A) Selection and modification of the relevant dose descriptors

6.A.1 Selection of the relevant dose descriptors

For threshold effects, i.e. health effects induced only above a certain exposure level, the aim is to find a NOAEL or LOAEL, more or less analogous to the procedure using animal data.. Some differences exist in the nature of such data from experimental vs. observational studies. In experimental data (e.g. animal tests) the doses are predefined and concern more or less exact values apart from each other (e.g. 100 mg/kg, 300 mg/kg, 1000 mg/kg) and the experiment usually does not provide observations on the occurrence of effects between these values. The highest dose without observed effect is identified as the NOAEL and the lowest dose with an observed effect as the LOAEL. The true threshold apparently lies somewhere between the two values. In the case of observational (human) data with exposure categories forming a continuum, i.e. the next category normally starting from where the previous category ended (e.g. 0- 5 ppm, 5 – 25 ppm, >25 ppm), the situation is a bit different from animal data. As NOAEL and LOAEL are based on absence or occurrence (e.g. a statistically significant increase in RR) of an "observed effect", the NOAEL category would be the highest category where an effect is not yet observed and LOAEL category as the lowest category where an effect is observed. Nevertheless, the true threshold value does not lie between the categories (as they form a continuum) but in one of the categories.

As a consequence, many study reports with quantitative exposure and effect data do not directly allow to establish the exact NOAEL, but only to approximate the exposure range the NOAEL lies in. If the exposure categories form a continuum, the upper exposure limit of the range of exposures in the no-effect category is the same as the lower limit of the range of exposure in the lowest category showing an effect. In the absence of more details on the distribution of exposures this value (i.e. the boundary between the two categories) should be used as a point estimate of NOAEL. In cases where the number of individuals in the NOAEL category is small or if there is indication that the exposure distribution is skewed towards the lower end of the category, a more conservative NOAEL may be justified. In that case, the average of the lower and the upper limit value of the NOAEL category could be used, based on expert judgement that should be explained in the dossier. If sufficient data are available, the average exposure of the individuals or the median exposure value of the NOAEL group may be a better choice for describing the exposure of the group.

This procedure applies to acute as well as long-term health effects (i.e. effects with a longer latency period). In case only a LOAEL can be identified, that value should be carried to the next Phases with an indication of the fact that it is a LOAEL and an adequate assessment factor should be applied in Phase 7. Using LOAEL with an assessment factor should also be considered when the identification of the "NOAEL category" is uncertain, for example if the number of observations in the "NOAEL category" is low.

The above considerations are written from the point of view of epidemiological studies. In acute specific effects (e.g. irritation), the data are sometimes reported also as simple frequency data over

exposure ranges. The identification of NOAEL and LOAEL could then follow a similar reasoning. Case reports may also contribute to defining LOAEL values.

In case there was no effect at any of the exposure ranges, the study should not be used for derivation of the NOAEL, because there is no need to set a DNEL. However, see section on negative studies under Phase 9.

6.A.2 Modification , when necessary, of the dose descriptors to the correct starting point

In a few situations, the exposure situation from which the dose descriptor is obtained is not directly comparable to the exposure situation for which the DNEL is being derived in terms of exposure route, units and/or dimensions. In these situations, it is necessary to convert the dose descriptor into a correct starting point. This applies to the following situations:

1. If epidemiological data derive from another exposure route than the route to which the risk assessment has to be applied, a route-to-route extrapolation is necessary.
2. Differences in exposure conditions between the source population and the target population, e.g. differences in respiratory volumes, or intermittent versus continuous exposures etc.

In principle situation 1 above is rare in the case of human data as these data more or less by definition deal with a route of exposure relevant to humans. Nevertheless exposure routes of consumers and those exposed in the occupational setting (often the origin of studies available) may differ. If needed, the principles described in Section R.8.4.2 apply to both situations 1 and 2 above. It should be noted that modification is usually not needed in cases where human exposure is evaluated based on biological monitoring data (internal dose metric). If valid human data that relate the effect directly or indirectly to a biomonitoring metric are available, the calculation of $DNEL_{biomarker}$ values can be straightforward. See APPENDIX R.8-5 for more guidance on the derivation of DNELs using biomonitoring data.

After modification where necessary of the relevant dose descriptors, the corrected starting points should be collected in a table (see Table R.8-15 of APPENDIX R.8-1)

Result of Phase 6–A: For each threshold effect, one or more dose descriptors have been selected. The dose descriptors, after modification (if required), are collected in a table (see Table R.8-15 APPENDIX R.8-1)

(Phase 7-A) Selection and justification of the Assessment Factors

In the use of human data for DNEL derivation, assessment factors (AFs) associated with e.g. intraspecies variation, dose-response relationship and differences in exposure conditions are considered. Contrary to the case with experimental animal findings, there is no need to consider interspecies variation when using human data for DNEL development. Where human data are considered a suitable starting point for the derivation for a DNEL, then a partly similar set of considerations can be identified as those applied to experimental data (Section R 8.4.3.1 and Table 8-6). These aspects will be discussed under the following headings:

1. intraspecies differences;
2. differences in duration of exposure;
3. issues related to dose-response;
4. quality of human data

7.A.1 Intraspecies Differences

When human studies are used for derivation of the DNEL, intraspecies assessment factors are needed to account for the variability in the target population, which can be anticipated to be usually larger than that in the study sample. For example, there can be differences in toxicokinetics due to slower excretion of the substance or due to a higher rate of transformation of the parent substance to a more toxic substance in some sub-populations, but also due to differing absorption rates or due to differences in toxicodynamics. The source population may comprise only/mainly healthy workers and the target population may include also e.g. sensitive target populations such as very young children, elderly people and persons having diseases (e.g. diabetics, people with kidney diseases). This would mean in practice that usually in the target population there are more sensitive people than in the source population and therefore the effect level in the target population could be significantly lower. Obviously, workers may develop diabetes, cardiovascular and other diseases just like the general population, but it is generally acknowledged that selection of workers will lead to a worker population, which is either healthier and/or more resistant to the physical and chemical stress factors of the work. Sensitivity of the human sub-population should be taken into account when establishing the DNEL. For example the possibility of higher sensitivity of children and pregnant women should be considered.

[The following paragraphs, which are marked, will be kept for the consultation period; it is suggested that they will be deleted thereafter]

Regulatory and scientific background

There has been a general agreement that assessment factors established in the last update of the Technical Guidance Document applied in the implementation of the Existing Substances Regulation could not be reconsidered during the preparation of the REACH CSA guidance. Therefore the part on use of human data does not reconsider the concepts agreed under development of Chapter R.8.

Traditionally a default assessment factor of 10 has been used to account for human variability in the risk assessment of chemicals (re e.g. Lehman and Fitzhugh 1954). The same intraspecies assessment factor of 10 has been also used by some national agencies, such as US-EPA, Danish EPA and TNO/Netherlands (see Table R.8-19) as well as by SANCO (2006). Furthermore, under the directive of Plant Protection Products (91/414/EEC) a guidance has been given (point 3.2) that recommends the use

of factor 10 as a default to cover inter-individual variability. ” Some recent literature concerning the human variability and safety/assessment factors is summarized below.

Differences between neonates/infants and adults relevant to assessing chemical risk include: immaturity (lower synthesis or absence) of metabolizing enzymes, immaturity of renal function, higher ventilation rate, immature blood-brain barrier, and immature and developing nervous system (Renwick et al. 2000). Renwick also examined whether a higher factor than 10 would be necessary as intraspecies factor to cover higher sensitivity of neonates/infants, but no consistent evidence was found to support generic deviation from AF 10 for infants.

There have been some further attempts to examine whether the conventional intraspecies assessment factors cover the human variability. Hattis et al. (1987) collected individual measurements of key pharmacokinetic parameters for specific substances for groups of at least five healthy adults. Inter-individual variability in elimination half-lives, maximal blood concentration and AUC (area under the curve) were measured for about 100 substances. The median values of these parameters given as log10 geometric standard deviations in the range of 0.11-0.145. This implies that ten-fold difference in pharmacokinetic parameters would correspond 7-9 standard deviations, which clearly is unnecessarily conservative. For a lipophilic chemical variability was 0.4 (as log10 GSD) and for that substance ten-fold difference would correspond to only about 2.5 standard deviations. The authors point out that the study excluded most human variability due to age and illness. Thus, it can be concluded that in case where AFs higher than 5 and 10 would be proposed – it would lead to unnecessarily conservative risk assessment for several chemicals, whereas clearly lower AFs would not take into account the substance-specific issues in toxicokinetics, neither would a lower AFs account for factors like age and illness.

Scientific basis of intraspecies uncertainty factor was also examined by Burin et al. (1999). Based on clinical trials of pharmaceuticals they conclude that most of the evidence shows that AF of 1-10 is protective for 99% of human populations including subgroups such as children.

For substances to which, due to their wide use (e.g. in paints, glues, textile colouring, adhesives), a large section of the population is extensively exposed, high level of protection should be applied. This is necessary as even a low fraction of the general population exposed that may be affected would constitute a very significant number of people.

Use of an assessment factor of 10 to cover human variability has recently been considered and confirmed in the IPCS document on chemical-specific adjustment factors for interspecies differences and human variability (IPCS, 2005c). In the IPCS document, subdivision of factor 10 to 3.16 for toxicokinetic and also 3.16 for toxicodynamic parameter is suggested. It is furthermore proposed that “The default sub-factors for toxicokinetics or toxicodynamics for human variability could be replaced by data that has defined the variability in the relevant parameter estimates in healthy human adults, including the influence of any functional **genetic polymorphism**, as well as the variations between different potentially **susceptible subgroups as appropriate.**” This modification of above mentioned

factors in case there is **substance specific data** on human variation, is in accordance with the approach suggested below in chapter "Modification of standard assessment factors**",** paragraph (ii).

Concerning the effect of genetic polymorphism, a recent analysis (Renwick and Lazarus, 1998) demonstrated that the default AF 10 for human variability is adequate providing that the compound is not metabolized by a single pathway, which shows genetic polymorphism.

7.A.1.1. Selection of assessment factors

Human studies normally cover at least some of the human inter-individual variability. Use of AFs strongly depends on the human data that is available for obtaining the DNEL. Therefore, a pragmatic approach is described below for using appropriate intraspecies assessment factors, which are based on the specific human studies available. Adequate justification of selection of any AF should always be given.

In case specific intraspecies assessment factors cannot be justified with the human data available, the values of Table R.8-6 in Section R.8.4.3.3. should be used.

7.A.1.2.Use of standard assessment factors

According to Table R.8-6, the standard intraspecies assessment factors would be 5 for workers and 10 for the general population. These AFs are the same as those applied when using animal data as a starting point.

A standard assessment factor would be appropriate when the human study is small and the sample in the study is homogenous and therefore no significant part of human variability could be regarded as covered.

Examples of cases, where the use of standard AF is necessary are when:

(i) there are **one or two case studies/reports** with low number of individuals observed, or

(ii) there is **a small occupational surveillance study** with a sample of 10-20 workers who might have been selected so that healthy worker effect applies.

7.A.1.3 Deviation from standard assessment factors

It should be always examined whether there are substance specific data to justify deviation from the standard assessment factors. Some cases, where this deviation could be justified are specified below.

(i) In some cases, e.g. when a high inter-individual variation of susceptibility has been identified, assessment factors higher than the standard assessment factor may be needed. This could be the case for example when genetic polymorphism leads to a high variation in the level of the metabolizing

enzymes. Those rare cases where an unusually large assessment factor is necessary for protection of children are described in Section R.8.4.3.1.

In case the human study is small and/or the sample is only representative of a particular sub-population, it should be considered whether that study should be used for the derivation of the DNEL. Obviously, a well conducted and relevant human study should not be rejected only, because the sample size is small Use of expert judgment by the registrant is necessary when that type of study is used for setting the DNEL. See a relevant example on hydrogen peroxide in APPENDIX R.8-16.

(ii) Use of AFs lower than the standard assessment factors is appropriate when it can be shown that some of the **factors that cause the intraspecies variation** in the target population, such as gender, age, nutritional status, health, susceptibility and genetic polymorphism have been covered in the study population. When this is the case, a value lower than the standard assessment factor should be selected and justified based on expert judgment.

(iii) In some cases, **substance specific information** might be available that can be used to justify special assessment factors. This information could be from **toxicokinetic and/or toxicodynamic studies** where variation in the human population has been measured. For example, when measurements in sufficient number of humans have shown that toxicokinetic and toxicodynamic factors, taken together, can be accounted by an AF between 2 and 5/10, that value can be used instead of "standard" or "lower" AFs. It should be acknowledged that the number of substances for which this information is currently available seems limited. It is also noteworthy that when substance specific information is obtained from studies where the sample size (number of people) is small (10-30), it is not justified to set a low AF, since the effects of human variability cannot be fully observed in a study with a relatively small sample size. In principle, the intraspecies variability for workers can be addressed in a smaller study sample, in comparison with a study that aims to cover the human variability in the general population. Guidance for the use of substance specific data and some examples are provided in the IPCS document 'Chemical-specific adjustment factors for interspecies differences and human variability' (IPCS, 2005c). Furthermore, as specified in APPENDIX R.8-4, **PBPK modeling data** can aid in the quantification of intraspecies variability, which may be caused by variation in anatomical, physiological and biochemical parameters with age, gender, genetic predisposition and health status. PBPK models can be used to quantify these, which would result in possible modification of AFs. IPCS is about to finalize guidance on these issues (for the progress of the project see http://www.who.int/ipcs/methods/harmonization/areas/pbpk/en/index.html).

(iv) There can be cases where the sample (i.e. the population) in a good quality human study is so heterogeneous and well characterised for different "aspects" of intraspecies variation that the use of a lower than the standard AF, i.e. 1-2, is justified. In the current experience, the number of substances with that kind of human data is not high. When e.g. the **sample size adequately takes into account the frequency of the effect and the study group is heterogeneous and the surveillance/study has an adequate duration**, it may be concluded that most of the intraspecies variation has been covered in that study. In an optimal case, justification for a low assessment factor could be based for example on a description of the demographic data of the study sample, such as age distribution, gender and diseases.

In addition , if the NOAEL is obtained from a study where a susceptible group of people has been specifically addressed (e.g. the registration of respiratory effects in a group of persons with asthma or hyper-reactive airways) a reduced intraspecies AF may be more appropriate. In cases where e.g. children, elderly or sick people or people having a special diet were not represented or were excluded from the study sample, the use of a low AF would not be justified.

Use of low AFs should be considered on a case-by-case basis and it is acceptable only when this is supported by appropriate data, which is given in a transparent way. In APPENDIX R.8-16 an example is given of a substance, for which the human database enables the use of low intraspecies assessment factors.

(v) When the effect seen in humans is associated with **biomonitoring** data such as urinary or blood level of a compound or its metabolite, again the toxicokinetic factor of intraspecies variation is accounted for and the AF could be the remaining 3.16 for the toxicodynamic variability. Biomonitoring data reflect the internal exposure and thus, toxicokinetic parameters influencing the internal/systemic bioavailability do not play a role. The factor of 3.16 would then be applied to the dose descriptor obtained from the biomonitoring study. (See also APPENDIX R.8-5)

See APPENDIX R.8-16 for examples of modification or deviation from the default intraspecies assessment factors.

7.A.1.4 Study Size

It is not possible to define minimum and/or recommended size of the study population, since (i) it will depend on the study type/aim and (ii) because study size as such does not provide assurance that sensitive subpulations and factors causing variation in the human population have been covered in the source study. Thus it would not be scientifically justified to give accurate sample sizes (to cover variability in the target population). It is notable that for e.g. substances covered under Existing Substance Regulation (ESR) (see APPENDIX R.8-16), for which the hazard evaluation primarily relied on human data, the epidemiological studies (for carcinogenicity or organ toxicity) generally had large sample sizes, i.e. thousands or tens of thousands of individuals. Smaller studies have been used in some of the ESR cases, for the purposes of hazard identification and classification and labeling.

Therefore, sample size of the source study should be considered together with an evaluation of whether the different “factors” that cause intraspecies variation have been addressed in the study. This means in practice that e.g. even large worker surveillance, when done in a homogenous group of workers, does not cover intra-species variation among general population. The homogeneity of the worker population could, for example, result from the healthy worker effect which can play a role both at hire and during the career (see Phase 3). Similarly a sample of general population that is limited to a region next to a source of the release of a substance, might be rather homogenous e.g. in terms of dietary habits, and ethnic background.

The issue of study size can also be addressed within the overall study quality (see Phase 2). In case the sample size is so small that it compromises the statistical power of the study or the reliability of the

study, it would usually be more appropriate to use other data (for setting the DNEL) instead of that study, rather than try to correct the weakness of the study by additional AFs.

7.A.2 Duration of exposure

Provided that the information in a human study covers a sufficient time span, there is generally no need to introduce an assessment factor to account for differences in the duration of exposure for the study population and scenario under consideration (target population). Instead, the duration of the human study should be compared with the exposure situation in the target population. When considering the time span aspects of a study, one should check that from the point of view of the end-point studied, it covers both a sufficient duration exposure time and a sufficient follow-up time to observe the effect. The time of follow-up should be sufficient to cover the latency period between the exposure and the manifest disease or organ damage. This is especially important in case of cancer epidemiology and chronic organ toxicity studies

Whether the duration of a human study can be considered sufficient will also depend on the type of effect under consideration. Acute effects, e.g. effects on the central nervous system caused by solvents or skin and eye irritation/corrosion can usually be observed within a few days. The reversibility of the effects can also be observed during some more days. Development of most signs and symptoms of target organ toxicity can usually be observed if the exposure duration has been several years. A longer follow-up and/or a longer duration of exposure is necessary when carcinogenicity is studied. This is due to the latency period, which depending on the type of cancer, can even be of many decades.

If a reliable NOAEL for a chronic endpoint is available, this is the preferred starting point for a $DNEL_{long\text{-}term}$ and no assessment factor for duration extrapolation is needed regardless of whether the information is applied to workers or consumers. In some cases, the duration of exposure falls between traditional acute and chronic studies (such as depression of blood counts following days/weeks of exposure, i.e. they are observable effects of possible pre-clinical significance and serve as a surrogate measure for serious effects) and where a $DNEL_{long\text{-}term}$ must be derived. In such cases, an AF of 2 is suggested to be applied to the NOAEL (and which is consistent with past practice in this area[1]). A NOAEL for an acute endpoint (NOAEL following short term exposure only) should not be used as the basis for the derivation of a $DNEL_{long\text{-}term}$. If the study design does not allow to adequately address any latency of the observed effect, then these data should not be used for deriving a $DNEL_{long\text{-}term}$.

[1] Steven Fairhurst (1995) "The Uncertainty Factor In The Setting Of Occupational Exposure Standards" Ann. Occupational Hyg., 39: 375 - 385.

7.A.3 Dose-response relationship

The dose-response relationship and the shape of the dose-response curve for the endpoint of interest are important elements to be considered in the derivation of the DNEL.

As with animal data, consideration should be given to the uncertainties in the dose descriptor (NOAEL, benchmark dose…) as the surrogate for the true effect threshold, as well as to the extrapolation of the LOAEL to the effect threshold (in cases where only a LOAEL is available) and the extent and severity of the effects at the LOAEL.

Unlike in experimental animal data, in many human studies the response/effect will not be displayed at discrete exposure concentrations but within exposure categories/ranges. The cut-off points of these categories should be set in a transparent and scientifically sound manner instead of trying to create exposure categories so that a favourable result (i.e. a high NOAEL/LOAEL) can be obtained. Exposure categories and ranges of the original study should usually be kept.

In case the exposure data comes from e.g. worker surveillance studies or from case studies, the data would normally be more accurate than just wide exposure categories. It is recommended that the reliable exposure measures are compiled and that those, which best represent the conditions where a health effect was identified are used when setting the dose descriptor.

It is proposed that in the absence of more detailed information the lower boundary/limit of the lowest exposure category in which the most sensitive effect is still observed should be considered the LOAEL and the upper boundary/limit of the exposure range in which no statistically or biologically significant effect is observed should be considered the NOAEL. For further consideration of NOAEL and LOAEL values see also Phase 5 and Phase 6.A.1.

Some of the uncertainties associated with the reliability/accuracy of the dose-response relationship of a substance, such as dose/exposure spacing, group sizes and statistical methods, cannot be dealt with using formalised assessment factors. These uncertainties have to be addressed qualitatively. In cases where the uncertainties are major, the study should not be used for derivation of the DNEL. See more guidance on these issues e.g. in Phase 2 (data quality). The only major uncertainty in the dose-response relationship that is traditionally addressed with the application of assessment factors is the extrapolation of the LOAEL to the NOAEL when only a LOAEL is available.

It is proposed that when the starting point for the DNEL calculation is a LOAEL, an assessment factor ranging from 3 (as minimum/majority of cases) to 10 (as maximum/exceptional cases) is applied. An AF of 3 may be more appropriate for instance in situations, where the effects at the LOAEL are mild, or the LOAEL represents the lower boundary of the exposure range in which the effect is observed. Higher numerical values should be considered in situations where the effects at the LOAEL are severe and irreversible, or the shape of the dose-response curve is shallow or the quality of the study (e.g. group sizes, statistical methods, study design, exposure data) gives rise to uncertainties about the reliability of the identified LOAEL. It is especially important to apply a high assessment factor to a shallow dose-response curve, when dealing with incidence data[1]. This is because a large decrease in dose is needed to make sure that no individual is affected by a serious effect , e.g. cancer.

[1] In some studies you have yes/no data and a dose-response characterised by incidence (e.g. for carcinogenicity), whereas in other studies there is continuous data and a dose-effect relationship (e.g., RDT-data such as the effect of cadmium on urinary protein excretion).

Case reports e.g. on acute effects must be evaluated within their context, i.e. against the background information concerning the exposure levels in this setting at the time when no cases were reported. A report of an unusual case of a disease or health effect would indicate that earlier exposure was below the NOAEL and that the exposure of the reported case is above the NOAEL. From such a report it can be concluded that the NOAEL must lie between the normal exposure condition and the unusual exposure condition leading to the induction of the effect. Appropriate assessment factors need to be applied in these instances and the $DNEL_{acute}$ will generally lie between the normal exposure condition and the exposure level responsible for the effect. If a serious health effect is reported (other than irritation or rash for example) the application of an assessment factor could lead to a DNEL below the normal exposure condition.

7.A.4 Quality of human data (including exposure data)

In principle, significant flaws concerning the quality criteria set in Phase 2 will lead to rejection of that individual study in the process of setting a DNEL Application of an additional AF may be necessary when a relevant and valuable set of human data with limited quality is used. For example data from a human study with qualitative (e.g. job title and factory information) or semi-quantitative exposure parameters was converted into quantitative exposure data with the help of some additional external exposure information. This conversion of exposures was, nevertheless considered to contain so much uncertainty that the study would not have been used as "a stand alone piece of information". Nevertheless further qualitative data from other sources gave additional support for this study and it was concluded being the best available basis for derivation of the DNEL. In such a case it should be considered to add an additional assessment factor for quality of human data when using the dose descriptor from this study in the derivation of DNEL.

As described earlier, also combining results of several studies may involve additional statistical modelling and new analyses. In such situations the choices and compromises made should be clearly described and their potential impact preferably assessed by sensitivity analyses using varying parameter values. Typically the uncertainty resulting from such approximations needs to be taken into account, case-by-case, when setting assessment factors.

7.A.5 Overall assessment factor and its application to the correct starting point

The overall AF is obtained by simple multiplication of individual assessment factors discussed in the previous paragraphs. Care should be taken to avoid double counting several aspects when multiplying the individual factors.

Result of Phase 7-A: Assessment factors for intraspecies variation, duration of exposure, dose-response and quality of the human data are assigned to each dose descriptor. The justification of the assessment factors is documented. The overall assessment factor for each dose descriptor is calculated by simple multiplication of the individual assessment factors of that dose descriptor.

(Phase 8-A) Obtaining the DNEL

This phase describes how the DNELs are obtained from dose descriptors and assessment factors.

Once the relevant dose descriptors have been selected for each endpoint and modified to the correct starting point (see Phase 6-A) and the overall assessment factor calculated for each of them (see Phase 7-A) an endpoint-specific DNEL will be derived by dividing each dose descriptor by its overall assessment factor.

All derived DNELs are collected in a table (see Table R.8-16 in APPENDIX R.8-1). In case there are more than one DNEL per endpoint, all of them are taken to Phase 9.

Result of Phase 8-A: For each relevant dose descriptor selected in Phase 6, a DNEL is calculated. This is done by dividing the dose descriptor by its overall assessment factor. All DNELs are summarised in a table (see Table R.8-16 in APPENDIX R.8-1) and taken to Phase 9.

B- DMEL DERIVATION FOR A NON THRESHOLD CARCINOGEN

Two quantitative risk assessment formats can be followed to derive a DMEL for a non-threshold carcinogen: the 'Linearised' approach, or the 'Large Assessment Factor' approach. Both formats are based on the same principal elements of risk extrapolation or risk evaluation using a dose-descriptor related to a risk estimate (an RR or a comparable measure such as an OR or an SMR). Because of different perceptions of the uncertainties involved in quantitative risk assessment and risk evaluation and of different approaches to risk communication, there may be a preference for one of these formats .

In the following sections of this document, Phases 6, 7 and 8 of the DMEL derivation process are explained in detail for the 'Linearised' approach. Additionally the general principles of the 'Large Assessment Factor' approach will be outlined. However, due to the lack of experience in the use of this method for the derivation of DMELs from human data, no further explanation will be provided on how to proceed on each phase.

The 'Linearised' approach

Some regulatory agencies including the US EPA, the Danish EPA and the Dutch Health Council basically follow this approach (US EPA, Danish EPA 2004, Dutch Health Council, 1989; *see also* Goldbohm et al., 2006). The aim, when using this approach, is to identify an exposure level that gives rise to a risk which is considered to be of very low concern. A review of carcinogenicity risk levels used or discussed by different organisations, countries and committees is given in APPENDIX R.8-14.

(Phase 6 B) Extraction and modification of the relevant dose descriptors

6.B.1 Selection of the relevant dose descriptors

For non-threshold effects, i.e. notably carcinogenicity through a genotoxic mechanism, the dose descriptor is usually derived from cohort or case-control studies reporting Relative Risks (RR) or comparable measures to describe a dose-response association. The RR is the ratio between the risk of the health effect in the exposed population divided by the risk in the unexposed population. Comparable measures are the standardised ratio, such as standardised mortality ratio (SMR) or standardised incidence ratio (SIR), which are conventionally used in cohort studies if the unexposed reference group is the general population. The odds ratio (OR), which is derived from case-control studies, is also a measure of relative risk. The dose descriptor of interest for derivation of a DMEL is the exposure level related to a RR (or comparable measure). In its most simple form, the dose descriptor represents the exposure level related to a relative risk observed in an exposed compared to an unexposed population. Ideally, it is based on the slope of the exposure-response function derived for the whole range of exposure levels observed in the study or based on pooled data from all available adequate studies by modelling. As default a linear relative risk model should be used. In this way, only a single RR per unit of exposure (i.e. slope factor) is obtained for a substance. Occasionally, a non-linear exposure-response model may be fitted to the data and used to derive the dose descriptor. When this is the case, the selection of the dose-response model should be clearly justified. A background and further explanation can be found in Goldbohm et al (2006).

6.B.2 *Modification , when necessary, of the dose descriptors to the correct starting point*

In a few situations, the exposure situation from which the dose descriptor is obtained is not directly comparable to the exposure situation for which the DMEL is being derived in terms of exposure route, units and/or dimensions. In these situations, it is necessary to convert the dose descriptor into a correct starting point. This applies to the following situations:

1. If epidemiological data derive from another exposure route than the route to which the risk assessment has to be applied, a route-to-route extrapolation is necessary.
2. Differences in exposure conditions between the source population and the target population, e.g. differences in respiratory volumes, or intermittent versus continuous exposures etc.

In principle situation 1 above is rare in the case of human data as these data more or less by definition deal with a route of exposure relevant to humans or a combination of routes. Nevertheless exposure routes of consumers and those exposed in the occupational setting (often the origin of studies available) may differ. If needed, the principles described in Section R.8.4.2 apply to both situations 1 and 2 above.

The exposure metric most often used in the analysis of the epidemiologic data is a cumulative exposure value including years of exposure, e.g. 'ppm-years'. For genotoxic carcinogens cumulative dose is thought to be the more relevant exposure metric than exposure concentration. Hence, a correction for *duration* of exposure is not needed if an adequate cumulative dose exposure metric is used.

It must be noted that in many instances epidemiological data on long term cancer risks from chemicals are derived from epidemiological studies on occupationally exposed cohorts. These risks need to be converted to continuous (24 hours per day, 365 days per year and 75 years long) exposure for the general population.

After modification, where necessary of the relevant dose descriptors, the corrected starting points should be collected in a table (see Table R.8-15 of APPENDIX R.8-1)

Result of Phase 6-B (Linearised approach): For each non threshold effect, one or more dose descriptors have been selected. The dose descriptors, after modification (if required), are collected in a table (see Table R.8-15 APPENDIX R.8-1)

(Phase 7-B) Selection and justification of the Assessment Factors

The next step in the derivation of a DMEL is to address uncertainties in the extrapolation of the study data to the real human exposure situation, taking into account variability and uncertainty. Clearly, the use of epidemiological data has advantages over the use of animal data since there is no need for interspecies extrapolation. Furthermore the extrapolation from high exposure (study data) to low exposure (level of exposure/risk of low concern) is usually done over a narrower range of exposure levels. Nevertheless, some assessment factors still need to be considered.

For DMEL derivation based on epidemiological studies, the following assessment factors will still need to be considered:

1. Intraspecies differences
2. Quality of the database (amount and quality of available information)

7.B.1. Intraspecies differences

Part of the population is suspected to be more susceptible to cancer due to differences in toxicokinetics (ADME) and to genetic properties (such as having specific polymorphisms). If it can be documented that these properties are equally distributed across the relevant population subgroups (e.g. workers vs. general population or age groups, men/women, healthy/ill), there is no need to use AFs for extrapolation of a DMEL derived from one subgroup (e.g. a worker population) to a DMEL for the general population. For non-threshold effects such as carcinogenicity through genotoxic mechanism, it is often assumed that different large population groups have similar susceptibility. Usually the human dose-response data is based on reasonably large epidemiological studies. However, exceptions may arise, e.g. if the human data are derived from populations with a different genetic background. The principles described in Phase 7 of DNEL Derivation (Section 7.1) should be applied to justify the selection of the intraspecies assessment factor. If there are data on some risk-related parameters that allow comparison of dose-response (relative risk estimates) between the general population and susceptible individuals, the additional analyses may be performed to adjust the general population estimate for susceptible individuals.

The ‘Linearised’ approach intrinsically takes into account that individuals may be exposed during different time periods and at different exposure levels during life. In many epidemiological studies, in particular occupational studies, cumulative exposure (cumulative exposure = exposure level * exposure duration, e.g. ppm years) is used as exposure metric. As the dose-response is, in general, considered to be constant over all age groups, extrapolation does not need an AF. However, exceptions to this general rule may be encountered. For example, it is known that breast tissue is more susceptible to unrepaired genotoxic damage in the period between menarche and first pregnancy, as during pregnancy breast cells differentiate. In this case, RRs may be higher depending on age. This should be solved by applying the life table method (see Phase 8), incorporating the higher RRs during adolescence and young adulthood and the lower RRs during the remaining life periods.

Duration of follow-up is addressed in Phase 2. It is assumed that only studies with a sufficient follow-up time will be used for DMEL derivation. Therefore an assessment factor for duration of follow-up is not needed.

7.B.2 Quality of the database

The quality of the individual studies available is assessed according to the criteria of Phase 2. When assessing the quality of the overall human data consisting of these individual studies, the same issues should be summarised. Especially the following should be carefully considered:

- The *amount* of available data, i.e. the size of the study (or studies) that are used when extracting the dose descriptor, determines the amount of random error in the risk estimates related to the dose descriptor. This uncertainty is usually represented by the confidence intervals that are routinely derived for such estimates. A pooled analysis or meta-analysis, when based on a substantially large database, has relatively small confidence intervals. An assessment factor larger than 1 may be applied if the selected risk estimate has wide confidence intervals (i.e. the uncertainty concerning the risk estimate related to the dose descriptor is large).

- Another source of uncertainty is derived from *uncontrolled biases* (e.g. confounding bias or healthy worker effect) in the data (see Phase 2). Evidently, data likely to be subject to serious bias should not be used for quantitative risk assessment at all. For example if selection bias or information bias (see Phase 2 point 1 and 5) could plausibly explain the main findings of the study. However, in less serious cases, the impact of a possible bias on the dose descriptor may be estimated[1] and compensated by an assessment factor.

- If there is reason to assume that the quantitative exposure-response relationship based on the epidemiological data is probably an underestimate or overestimate of the true association an appropriate assessment factor should be applied. An example of such a situation is when quantitative exposure estimates are lacking from a study and exposure level(s) were estimated from other sources to obtain a dose descriptor.

7.B.3 Overall assessment factor and its application to the correct starting point

The overall AF is obtained by simple multiplication of individual assessment factors discussed in the previous paragraphs. Care should be taken to avoid double counting several aspects when multiplying the individual factors.

Result of Phase 7-B (Linearised approach): Assessment factors for intraspecies variation and quality of the database are assigned to each dose descriptor. The justification of the assessment factors is documented The overall assessment factor for the dose descriptor is calculated by simple multiplication of the individual assessment factors of that dose descriptor.

[1] A practical approach to assess the effect of possible uncontrolled biases on the risk estimate can be to apply sensitivity analyses postulating different levels of bias. A more sophisticated and reliable approach is to use probabilistic simulations to estimate bias, e.g. [Steenland and Greenland, 2004].

(Phase 8-B) Obtaining the DMEL

8.B.1 High to low dose extrapolation

The RR (whether or not corrected in Phase 6-B) must be projected onto the target population (workers or general population) to derive an Excess Lifetime Risk (ELR) at a given level of exposure. I.e. how many excess life time cases in absolute terms will result from a given relative estimate of risk (RR, OR, SMR or SIR) This necessitates the application of the relative risk estimates on actual population data (with person-year data and case occurrence data). There are two options to do this:

i) a simple direct method as described by van Wijngaarden and Hertz-Picciotto (2004) or the Dutch Health Council (1989), and

ii) a more sophisticated method including the use of a life table approach as described by e.g. Steenland *et al.,* 1998.

The direct method calculates the ELR as: ELR = Lifetime Risk * (RR-1). Lifetime Risk is the (background) risk of the relevant health effect in the target population to which the DMEL applies. The simple direct method results in some overestimation of the lifetime risk, in particular if the background risk in the target population is high. This is mostly because the direct method is less accurate in taking into account the mortality from other causes of death. The life-table method calculates and accumulates the ELR for each life year during the lifetime of a virtual cohort (see Goldbohm et al 2006 for an example). It gives a more accurate estimate and can incorporate specific requirements, such as changing exposure patterns over a lifetime, competing risks due to effects of exposure on other endpoints, etc. The life-table method may be used if there is a need to calculate the risk more accurately. A life table should also be used if age-dependent RRs are indicated (see example on breast cancer above). Although the life-table method is the preferred option, the direct method can be used if the background rate of the disease and the potency of the substance are low or if the age for which the risk is considered relevant is relatively young (< 70 years) (Goldbohm et al 2006).

At the end, the ELR estimate linked to the known level of exposure is used to extrapolate the exposure level that corresponds to a given risk level considered from a societal point of view to be of very low concern. i.e. the DMEL. A review of carcinogenicity risk levels used or discussed by different organisations, countries and committees is given in APPENDIX R.8-14. If the RR was calculated from a linear relative risk model, the derived ELR for a given exposure can directly be converted to a DMEL with a linear extrapolation.

If the RR value was based on models other than the linear model, low dose extrapolation should be performed according to one of the following two options. If there is additional evidence (e.g. based upon available experimental data of good quality) that the dose response outside the observable range is non-linear, a non-linear model may be used to assess the risks associated with these lower exposure levels. Otherwise, if there is no information on the shape of the dose-response in the low dose range, as a default, linear extrapolation should be applied. The application of a non-linear model to low dose extrapolation should be performed on a case-by-case basis, and should be extensively documented and justified.

8.B.2 Application of the assessment factors

The exposure level that corresponds to the chosen level of low concern (obtained as described above) is divided by the overall assessment factor (obtained in Phase 7-B) in order to calculate the DMEL.

The derived DMELs are collected in a table (see Table R.8-16 in APPENDIX R.8-1). In case there are more than one DMEL per endpoint, all of them are taken to Phase 9.

Result of Phase 8-B (Linearised approach): DMELs are calculated using human studies of sufficient quality and including adequate exposure data. This is done by extrapolating the study data to a risk level of low concern and dividing this by the overall assessment factor. DMELs are collected in a table (see Table R.8-16 in APPENDIX R.8-1) and taken to Phase 9.

The 'Large Assessment Factor' Approach

The 'Large Assessment Factor' approach was recently presented by the Scientific Committee of the European Food Safety Authority (EFSA SC) when providing guidance for managing risks posed by contaminants in food (EFSA, 2005). When applied to animal data, the approach uses a large assessment factor (10 000 or higher) in order to derive an exposure level of low public health concern from a BMDL10 from an animal study. EFSA SC notes that the benchmark dose approach can also be applied to human data when available. It is considered that presently there is not enough experience on the use of human data and on this approach in the hazard assessment of industrial chemicals and further guidance on such an application is difficult to give.

The method relies on a large assessment factor and the lower confidence limit of a BMDL10, the critical points in applying this approach to human data concerning a non-threshold carcinogen would be the selection of the dose descriptor equivalent to BMDL10 (the lower limit of the confidence interval of the dose related to absolute effect frequency of 10%), and the adjustments of the large assessment factor. Human data are usually described in terms of doses related to relative effect estimates (RR or similar). Apart from very rare situations (e.g. heavily asbestos-exposed worker groups) absolute effect frequencies of 10% are not observed. Therefore the approach would first necessitate an upward extrapolation to the 10% effect level and then a second downward extrapolation to the level of low concern. It is also not clear whether in the absence of the interspecies AF, the "large assessment factor" would anymore ensure a high level of protection. Therefore the use of this approach in the derivation of DMELs from human data needs to be well justified and special attention should be paid to ensure that a sufficient level of protection is reached.

(Phase 9) Integration of human and animal data and selection of the critical DNEL/DMEL to be taken to the risk characterisation

At the start of this phase, the DNELs and DMELs derived from human and animal data have been collected in Table R.8-16 (see Chapter R.8, APPENDIX R.8-1). Data in this table are then integrated in order to arrive at the specific entries of Table R.8-9. Please note that when human data are used for obtaining DNEL(s)/DMEL(s), the guidance given in this phase should be followed instead of that of Section R.8.7.1. Thus, this phase addresses the selection of the leading health effects and the critical DNELs/DMELs, which are subsequently used in the risk characterisation. Integration is based on the quality, relevance, completeness and level of the DNELs derived from different studies as explained below. The same principles and criteria are used, in case there are more than one human study at this phase (for certain endpoint), but no animal data. The selection of the critical DNELs/DMELs to be taken to the risk characterisation should be justified/documented..

The decision on which dose descriptor to use to derive a DNEL based on human or animal data is not straightforward and should be seen in the context of Mode of Action Framework (IPCS 2007a and b). Even when human data are not of adequate quality to derive the DNEL, consideration should be given on their potential use together with animal data for refinement of the risk characterisation. This can be the case for example in the development of PBPK modeling (see Section R.8.4.3.2 and APPENDIX R.8-4). IPCS is about to finalise guidance on these issues (see http://www.who.int/ipcs/methods/harmonization/areas/pbpk/en/index.html for the status of the project).

Within the REACH Regulation the so-called weight of evidence (WoE) approach is a component of the decision-making process on substance properties and thus an important part of the chemical safety assessment (see Chapter R.4). The term WoE neither constitutes a scientifically well-defined term nor an agreed formalised concept characterised by defined tools and procedures. It is based on assigning weights to each available piece of information either in an objective way by using a formalised procedure or by using expert judgment (see Chapter R.4.for more details). Although the use of structured frameworks can be invaluable in promoting harmonization in the assessment of chemical risks (IPCS 2007a), a commonly agreed formalised procedure is not yet available for integrating animal and human data for the various purposes of human health hazard assessment. The weight given to the available evidence will be influenced by factors such as the quality of the data, consistency of the results/data, nature and severity of the effects, relevance of the information for the given toxic endpoint. Tools like the Klimisch scores for experimental toxicity studies or Hills criteria for evaluation of epidemiological data are available for specific factors to be assessed, but the way in which the WoE is implemented to integrate all data remains case-dependent. Some guidance has been developed by IPCS, IARC and ECETOC and these are briefly described below.

The IPCS has developed a procedure, termed the IPCS Human Relevance Framework to make judgments about the relevance (to man) of findings in animal studies for both cancer endpoints and non-cancer endpoints (IPCS 2007a, IPCS 2007b). This procedure involves describing key events leading to the toxicity observed, and establishing the mode of action (MoA) in animals. Each key event in animals is then evaluated for its plausibility in man including both fundamental qualitative aspects and quantitative aspects. This procedure includes several elements that are useful when integrating human and animal data.

When evaluating the human carcinogenicity of agents, the IARC classifies the evidence into five groups: definitive carcinogen, probable carcinogen, possible carcinogen, not classifiable and probably

not carcinogenic. The categorization of an agent is a matter of scientific judgment that reflects the strength of the evidence derived from studies in humans and in experimental animals and from mechanistic and other relevant data (IARC 2006). The human and animal evidence are each first classified into four categories: sufficient, limited, inadequate and evidence suggesting lack of carcinogenicity. The information concerning a given mechanism of carcinogenicity is classified as weak, moderate or strong. A further assessment is then done on whether this particular mechanism is likely to operate in humans. When integrating the above-described categorized human, animal and mechanism information, specific rules are applied to arrive into one of the five groups of overall carcinogenicity.

A more formalised process of integrating the human and animal data would improve the utility and robustness of the risk assessment process. Such an approach has recently been proposed (ECETOC 2009). After having formulated the problem, the proposed method uses five separate category scores to characterise the available human data (for **quality**) and animal data (for **quality and relevance**). The scores are based on a collective weight of evidence assessment of the human data on the one hand and the animal data on the other hand. In case the human data are of equal or better score than the animal data, then human data takes precedence. Otherwise animal data takes precedence. Special care is taken when considering the concordance of animal and human data. In general positive data take precedence over null data. Nevertheless, negative data are also scored, with special emphasis on the confidence intervals of the negative studies. While the principles behind this method (e.g. using quality categories of human data) are appreciated, it is too early to recommend this approach as a systematic tool for the overall integration phase of human and animal data and for the specific purpose of deriving DNELs/DMELs.

9.1. Principles of integrating human and animal data

This phase of derivation of DNEL/DMEL should be transparent and rigorous and the decisions should be made based on the best data available.

The human data carried over to the integration phase include studies in which bias, confounding and chance are ruled out with reasonable confidence and in which the causal relationship between exposure to a specific chemical agent and a certain health outcome has been evaluated. These studies include both studies providing a quantitative estimate for the DNEL/DMEL and possibly studies which do not contribute directly to the quantitative assessment of the DNEL/DMEL, but are still judged to be of qualitative value in the WoE analysis. The available human studies have been assessed for their completeness as regards covering the relevant health endpoints.

The available data need to be assessed for their reliability and consistency across different studies (including available animal data) and endpoints taking into account the quality of the study protocol/methodology, size and power of the study design, biological plausibility, dose-response relationships and statistical association (adequacy of the database). When the human data are robust and of good quality, it should always be considered in this integrative step.

Where human data are inconsistent,they should not generally be used for DNEL derivation, although even in such cases the severity of the observed effect may indicate that the human findings should be considered within the WoE.

A particular point is the use of **negative human** studies. The term "negative" refers to studies where no effect was seen. A single human study can rarely, if ever, be regarded as negative in the sense that it proves an absence of relationship. If a human study is of compromised sensitivity or of compromised completeness,s it could rather be called “inconclusive” than “negative”. For example a study with a relative risk lower than unity, but with an upper limit of the confidence interval being above one, is rather inconclusive as a single study if no other human data are available.

Good quality human data that illustrate the lack of a health effect in a specific exposure range (negative overall data) should be considered in this phase of the process. Although the same quality criteria apply to negative studies as to positive studies, special care must be taken to ensure that the negative outcome is not the result of inadequate sample size (statistical power),or design or measurement error or uncertainty of exposure or effect. In addition the completeness and adequacy of the negative human data for the purpose it is intended to be used should be ensured. In conclusion, a high level of quality is required of negative human data, especially when it is used in this phase to overrule positive evidence from one or more animal studies. The difficulty with negative human data lies in the fact that the only conclusion to be drawn is that the exposure range under investigation is below the effect threshold. Thus, the evidence does not allow estimation of the true effect threshold, but it can allow the inference that the effect threshold is higher than the exposure range investigated. This complicates the derivation of a health based DNEL in that the application of conventional AFs may result in unnecessarily strict exposure limits.

Positive, but non-statistically significant findings should not be regarded as null findings by default. In these instances biological significance is a more appropriate criterion than statistical significance.

9.2 Pragmatic approaches to integrating human and animal data

While the formalised methods to integrate human and animal data are not fully developed and “tested”, more pragmatic approaches which rely on the current experience of the use of Weigh of Evidence analysis, can be useful. Some typical and/or challenging cases of integration are therefore described below. The main challenges of integrating animal and human data concern cases where the available data are inconsistent. This is particularly the case when there are both negative and positive data on the same endpoint/health effect and an unclear mode of action.

9.2.1 Inconsistent data

In cases where **different types of effect** are seen in human and animal studies the possible reasons for inconsistency should be assessed. The WoE analysis should start by characterising the available studies for their quality and relevance. This step should already have been done in the earlier phases of analyzing the human data. As regards animal studies, Klimish score can be taken as a measure of the quality and the relevance. The animal study/studies should be assessed according to the criteria given in Chapter R.4.

In case the human data is of sufficient quality and relevant and the animal data has a Klimish score of 1-2 and is also considered relevant, both human and animal data should be included in the WoE. In that case, DNELs should be obtained from the critical studies and the lower DNEL should be used for the risk characterisation. However, in case the effect/endpoint was not adequately addressed in the human study, use of the animal data is preferable.

It is important to consider why data is inconsistent. It may be due to a different mode of action in humans versus in animals, in which case human data should usually be preferred in the assessment. The human data can be negative simply because the effect observed in animals does not have any relevance for humans as the underlying mechanism is specific to the animal species used in experiment/test. If this can be justified the negative human data is the basis of further evaluation. This has been addressed below in a specific section on "*Negative human or animal data"*. Also a low exposure level and/or compromised study power in the human study may lead to an apparent inconsistency between human and animal data.

Inconsistency between human and animal data may also be due to significantly **lower sensitivity of humans** to the toxic effect (e.g. due to interspecies difference in toxicokinetics). Deriving DNELs/DMELs from both human and animal data and selecting the lower of these values is the recommended approach. However, the human study has to be of sufficient size and there needs to be an understanding of the relevant mode of action, before lower sensitivity in humans can be established and used in the risk assessment.

There may also be inconsistencies within the human data (i.e. between two human studies) that have to be assessed. In such cases it must be determined whether an explanation can be given for the diverging results (were they caused by different kinds of study designs with different sensitivity; different kinds of effect examinations or measuring techniques; not quite comparable groups; different methods for evaluating exposure etc.?). Inconsistencies may not necessarily weaken the evidence if there are good explanations for the diverging results.

9.2.2 Incomplete human data

As a starting point it should be realized that in most instances it is not possible to obtain comparable data sets in humans and experimental animals, e.g. histopathological data are usually not available in human studies. Also, in many cases human studies have not covered as high dose levels as the animal studies and therefore it will not have been possible to observe some relevant effects in the human study. Nonetheless, even incomplete human data can be relevant and should be used at this phase when it is of sufficient quality and gives quantitative information on the exposure.

In cases where the human study did not cover some specific endpoints and animal studies did, the animal study should primarily be used for setting the DNEL/DMEL. An example of this might be an epidemiological study where only a certain type of malformations in human were examined/studied, while both the developmental toxicity study and two generation reproduction toxicity study were carried out animals. In case the DNEL (for malformations) derived from the human study is lower than the respective DNEL from animal studies, DNEL from the human study should be selected for developmental effects. In addition it would be necessary to obtain and report the DNEL concerning the other reproductive toxicity effects from the animal study i.e. from the two generation reproduction toxicity study. In practice, the lowest of these DNELs would normally be taken to the risk characterisation.

Another example could be an occupational surveillance study of limited size and only addressing some of the relevant effects. In that case the animal study might be a more appropriate and reliable starting point in obtaining the DNEL/DMEL. However, also limited occupational surveillance data should be used in the WoE and integration as supporting evidence and/or source of qualitative data.

In order to use occupational survey data instead of good quality animal studies, one should exclude the possibility of chance (i.e. low statistical power), bias (especially selection bias and healthy worker effects) confounders or measurement errors in the study. Assessment of such quality factors is crucial, since unpublished occupational surveillance data have not undergone independent scientific review. Depending on the outcome of the assessment of quality and relevance, occupational surveillance data will either be robust enough to be used instead of animal data or be taken only as supporting study.

9.2.3 Negative human or animal data

Negative and inconclusive human studies are not evaluated in isolation but are taken to WoE analysis together with other relevant human and animal studies, when these are available. Thereby, the whole database and not only individual studies are evaluated.

In case a positive human study is of sufficient quality and relevant, while the animal study is negative, obviously the human study is taken into the risk characterization. It would be useful to explain this inconsistency, when possible, since that would increase the reliability and confidence in the risk assessment outcome. It may be, for example, that the animal study is negative because some observations that were made in the human epidemiological studies are not routinely made in animal studies (e.g. decrease of bone density and increase of fractures caused by cadmium in humans) It is also possible that there is a mechanism of toxicity in humans that is not relevant in animals.

If the animal data are positive and the human data are negative (see also Section 9.1) and both are of good quality the relevance of the animal data becomes crucial. There are two cases:

1. When the **human relevance of the animal data cannot be excluded**, the animal data will be the basis of the dose descriptor, provided the human data do not reasonably exclude the effect shown in the animal data. For example, human study design can be such that all relevant endpoints/health effects have been covered. In a case where, for example, only effects on haematology parameters were covered in the human study, but effects were seen in clinical chemistry or histopathology parameters in the animal study at a lower dose level, the negative human data is not conclusive, and therefore, the animal data has to be used when DNEL is obtained. Also the exposure levels in the human study and study power should be considered. If exposure levels in the human study were low, the effects may not have been observed in the human study and the animal data remains valid.

2. There are a number of **mode of action considerations** that are crucial when integrating negative human and positive animal data. These may lead to a conclusion **that the animal data is not relevant for humans**. The basic concept is that in case the mechanism of toxicity in the animal study is characterised and has been shown not to be relevant in humans, the negative human data have a stronger weight in the analysis made in the integration. An example are the renal neoplasms in male rats, developed coincidentally with α_{2u}-globulin nephropathy, due to accumulation of a specific protein, which are not considered predictive of risk to humans (Doi

et al 2007). In cases similar to those described above, the negative human data are regarded as conclusive and no DNEL/DMEL needs to be obtained for this endpoint. It is to be underlined that it needs to be justified with data that the mode of action of the positive animal data does not apply to humans. Finally one needs to verify that the data concerning other non-threshold effects are negative.

9.2.4 Consistent data

In case data of good or sufficient quality both from human and from animal studies are consistent, i.e. essentially the same effects are seen, DNELs/DMELs from both sources should be considered in the WoE and integration. The lowest of the DNELs /DMELs should be used in the risk characterisation. However, in case the human data are adequate and complete they would take precedence over animal There may be cases where e.g. a biomarker information of an early effect (or e.g. hyperplasia) is obtained from a human study but only gross histo-pathological changes, which take place at higher level of exposure, are seen in the animal study. In that case the most sensitive study could be used for setting the DNEL/DMEL for that effect category based on expert judgment.

For ”data rich” substances there are often human data from several studies. When consistent qualitative and quantitative data come from several independent studies, low intraspecies assessment factors may be applied as described in Phase 7. Furthermore, in risk characterisation an effect seen in several studies should be preferably used instead of another effect, which is poorly characterised or only anecdotally described in literature.

9.2.5 Qualitative data in the integration

As instructed above, qualitative data, i.e. a study that has no dose descriptor (specified levels of exposure) but has valuable and reliable data on the relevant health effects should be considered in the phase of integration, while it cannot serve as a basis of setting the DNEL/DMEL.

Qualitative data should be regarded as a potentially supporting element in the WoE analysis/integration that aims at obtaining DNELs/DMELs. This could in practice imply that when there are more than one DNEL/DMEL or dose descriptor either from human or animal studies, the supporting information concerning the modes of action, should be considered. This may be relevant e.g. in assessment of negative data as described above.

It is important that even after having derived DNELs either on the basis of an animal or human datasets, an assessment be made to verify if the proposed DNELs would be protective for all other endpoints for which a health effect was identified. Similarly in case of DMEL derivation, the proposed DMEL should be assessed to make sure that no other health effect presents a higher risk level.

Result of Phase 9: Having Table R.8-16 as a starting point the available human and animal data are integrated in order to select the critical DNEL (or DMEL) value for the exposure patterns of Table R.8-9. The decisions taken in the selection of the critical DNEL/DMEL have been documented in a

transparent way. It has also been verified that the critical DNEL/DMEL values protect from all the other identified health effects of the substance. Table R.8-9 will be taken to the risk characterisation.

Appendix R.8-16 Examples of modification or deviation from the default intraspecies assessment factors

Below, three examples are given of the use of human data in chemical risk assessment. These cases are based on Risk Assessment Reports prepared by EU Member States within the implementation of the Existing Substances Regulation (ESR). In the guidance, which was applied in the ESR assessments, the instruction to use minimal Margins of Safety (MOSmin) was given. The basic elements of this guidance were the same as these described in Chapter R.8., i.e. the purpose of using MOSmin was the same and also their numerical values were the same as the standard Assessment Factors given in APPENDIX R.8-15.

The following examples are given to illustrate cases where deviation from MOSmins has been adequate and justified. In some of these cases the human studies have been large, heterogeneous and well characterised for human variability. Furthermore, there are examples of both sufficient and insufficient data base of human data for derivation of the DNEL. It is noteworthy that even if human data are insufficient for quantitative derivation of the DNEL, it may be relevant within the Weight of Evidence analysis where all available animal and human data are considered.

Thus, the examples show how in the past in certain cases the human data have been weighed against the animal data and how the default intraspecies factor was adjusted. Examples on how DNEL/DMEL would be derived in concrete cases will be developed and provided in the next phases of this guidance development process.

Cadmium

A relevant example of a situation where deviation from the default AFs is justified, is given in the Risk Assessment Report on cadmium. It is acknowledged that cadmium is a carcinogen; most of the evidence derived from studies where the exposure took place via inhalation and from occupational epidemiology. While deriving a DMEL for carcinogenicity of cadmium is a relevant topic, it is not the item of the example below. Instead the example given below addressed the effect of cadmium on kidney and on bone density (threshold effects) due to long-term exposures mainly via oral route, but also due combined exposure. This example is limited to those target organ effects which have a threshold in order to illustrate how intraspecies variation can be covered.

Among the industrial chemicals, cadmium has one of the largest toxicological data bases, of which human data is a significant part. The two most relevant human studies (Buchet et al. 1990 and Järup et al 2000) had a total sample size of 2720 individuals (workers and general population) in Belgium and in Sweden. Age of the subjects of the studies in the samples was 16-80. In addition, the following independent variables were considered in the analysis: sex, renal disease, diabetes, use of medication, body mass index and renal diseases. There are also several other relevant studies, which represent other European populations, e.g. from the Netherlands and from UK. In conclusion, deviation from the standard AFs in deriving a DNEL for cadmium risk assessment would be justified, because most, if not all, of the factors causing human variability are covered. In the risk characterization, the critical level of 2 µg urinary Cd/g creatinine is used, and this value is taken directly from the most recent, representative and good quality epidemiological studies, in which the renal effects and effects on bone density were seen approximately at this level of excreted cadmium. Depending on the type of effect

(kidney or bone) the mode of calculation and relevant study, several critical urinary levels are given in the risk assessment of cadmium ranging from 0.5 to 5.0 µg urinary Cd/g creatinine.

The text for the "compromise" LOAEL is illustrative, as it says: "Trying to aggregate all these data, a LOAEL of 2 µg urinary Cd/g creatinine is proposed. This figure should be understood as a composite level, based on the association between Cd and not only low molecular weight proteins in urine but also calcium excretion in urine and its possible relationship with bone effects" Margin of Safety of 3 is used in the RC to account the conversion from LOAEL to NOAEL, **but no intraspecies assessment factor** (or MOS accounting for human variability) was used for cadmium.

This example furthermore illustrates that since biomonitoring values are used instead of exposure data, the toxicokinetic factor of intraspecies variation has largely been covered and does not need to be additionally accounted for. The example also suggests that with large heterogeneous populations and the known relevant parameters covered, which cause the intraspecies variation, there appeared to be no reason to use an intraspecies assessment factor.

Hydrogen peroxide

Another (opposite) example is provided in the risk assessment report of hydrogen peroxide (H_2O_2), a strong oxidising agent, which acts as an irritant or corrosive agent, depending on its concentration. A health monitoring study (occupational surveillance) of six aseptic packaging workers was conducted. It involved a 10-month period of high exposure (2-3 mg/m^3 8-hour (time weighted average), peaks up to 11 mg/m^3) due to machine malfunction and, after repairs, a one-year follow up at a reasonably low and stable exposure (0.5-0.7 mg/m^3 8-hour TWA). The results indicated that three of the workers experienced eye and airway irritation, headache, and a uniform course of recurring bronchitis-sinusitis which coincided with the high exposure (Riihimäki et al., 2002). The study did not include specific examinations of the lungs. It was concluded that further data, including human observations, are helpful to characterise and confirm the repeated dose toxicity of hydrogen peroxide by inhalation.

Furthermore, industrial experience from health surveillance of H_2O_2 production workers suggested no exposure-related effects on simple respiratory functions at airborne levels of up to 0.8 mg/m^3 (CEFIC, 1996b) or less than 1.4 mg/m^3 with short-term peaks of up to about 5 mg/m^3 (Degussa-Hüls, 1999). Since these observations were not derived from properly conducted studies, the health data cannot be used as solid evidence for the absence of adverse pulmonary effects.

Because of the uncertainties and/or preliminary nature of the human data, they were not taken into account in the risk characterisation in that risk assessment report. Instead, more robust animal data were used to characterise the repeated dose inhalation toxicity of hydrogen peroxide. Interestingly, the effect concentrations in animal compared to human studies are rather consistent. Whether human data was dealt with in the Weight of Evidence analysis is not explained in the risk assessment report. This example illustrates that a study with small sample size, where all relevant parameters/observations are not covered is not a valid basis for obtaining a NOAEL or dose descriptor.

Toluene

The third example is about human data on reproductive toxicity of toluene. Two studies suggest an increased risk of spontaneous abortions associated with exposure to toluene in the workplace.

Spontaneous abortions among women working in laboratories, (together with congenital malformations and low birth weights of the children) were examined in a retrospective case-control study (Taskinen et al., 1994). The exposure to toluene was assessed on the basis of the reported frequency of the use of the chemical and classified as frequent if the chemical was handled at least 3 days a week and rare if the toluene was handled 1 or 2 days a week. Significant associations with spontaneous abortions were found for frequent exposure to toluene (odds ratio 4.7, confidence interval 1.4 to 15.9) after adjustment for various covariates (206 cases and 329 referents). This study suggests an association between exposure to toluene during early pregnancy and increased risk of spontaneous abortion. The result should be interpreted cautiously because the women were often exposed to several solvents and other chemicals simultaneously. Furthermore, no information on exposure levels is presented. In conclusion the rapporteur (under ESR) considered that the results are of limited use for the risk assessment of toluene.

In another study, rates of late spontaneous abortions were determined using a questionnaire addressing reproductive effects in 55 women with 105 pregnancies exposed to toluene (mean 88 ppm, range 50-150 ppm), 31 women (68 pregnancies) working in the same factory in departments where little or no exposure to toluene occurred (0-25 ppm), and an external community control group of 190 working class women with 444 pregnancies (Ng et al., 1992b). Significantly higher rates for late spontaneous abortions defined as ‘pregnancy loss’ between weeks 12 to 28 were noted in the toluene-exposed women compared with those in the internal and external control groups (12.9% vs. 2.9-4.5%). The differences in the rates of late spontaneous abortions between groups were not likely to be confounded by classical risk factors such as maternal age, gravidity, smoking, or alcohol, which were taken into account both in the study design and the analysis. Information on pregnancy outcomes might be biased by questionnaire interview. The pregnancies and abortions in the factory were not validated by access to medical records or with biological methods. However, relatively unequivocal endpoints were used in the questionnaire, thus excluding doubtful pregnancies and abortions.

In the course of this risk assessment the authors of the study clarified that, since the study was a cross-sectional observational study that relied on a questionnaire, with information obtained by the subject's recall of her recent pregnancy(ies), it was difficult to determine with absolute certainty whether a spontaneous abortion had indeed occurred, especially in the first two months after conception. It is known that foetal loss is a lot more common than is generally supposed especially in the first month immediately after conception. When it occurs, it is often disregarded as a 'missed period', when menstruation resumes a month or two later.

It was concluded in the risk assessment report that the second study cannot be used to establish definitively a causal relationship between late spontaneous abortions and toluene exposure. (To establish a definite relationship, a prospective study of pregnant women exposed to toluene at similar exposure levels (mean 88 ppm, range 50-150 ppm) with individually monitored data on toluene exposure and foetal loss would be needed.) However, based on the current evidence suggesting an increased risk for late spontaneous abortions, exposing pregnant women to such exposure levels would raise serious ethical concerns. Consequently, the results of the second study are used as a basis for the risk characterisation of developmental toxicity in humans.

It is noteworthy that animal inhalation studies provided strong evidence of developmental toxicity (lower birth weight and long-lasting developmental neurotoxicity) in the absence of maternal toxicity. Furthermore, there were also in total about 45 human cases reported in the literature of so-called toluene embryopathy as result of sniffing toluene. (These cases resembled foetal alcohol syndrome, and there might be a common mechanism.)

The human LOAEC of 88 ppm (330 mg/m^3) and the rat NOAEC of 600 ppm (2,250 mg/m^3) were taken forward to the risk characterisation. Different MOSs were applied to these data. Risk characterisation for workers was made for "toxicity to reproduction (fertility and development)" where the animal data was used with a MOSmin of **30**. Another risk characterisation was on "toxicity to reproduction (spontaneous abortions)", where a lower MOSmin of **5** was applied because "the NOAEC for this endpoint is derived from human data". (To correct an error in the report, in fact LOAEC and not the NOAEC was used when the MOSs were calculated.)

The human data was not used for risk characterisation of consumer exposure.

The case of toluene shows that a human study made with only a few hundred individuals may be very important element in the Weight of Evidence evaluation. In this case, the better of the two human studies with relevant exposure data led to use of a specific "dose-descriptor" in the risk characterisation. In fact, the animal NOAEC and human LOAEC are used in parallel, with the respective MOSs and they have led to exactly the same risk assessment result in terms of formal conclusions under ESR, thereby showing the relevance of the human data and increasing the overall robustness of the assessment.

Furthermore, the example shows how important it is to consider the confounding factors. Also the weaknesses of the human study were adequately reported, i.e. the lack of a definite causal relationship and the potential reporting bias.

Assuming that the MOSmin of 5 (for workers, derived from human data) actually covers (i) some of the intraspecies variation and (ii) the step from LOAEC to NOAEL, the case of toluene also demonstrates how human data can be used to modify the intraspecies assessment factor.

References

(Please note that the references in the Appendix R.8-16 have not been listed here; they can be found in the respective Risk Assessment Reports included in the ECB site ***http://ecb.jrc.ec.europa.eu/esis/index.php?PGM=ora****)*

Burin G J, and Saunders D R (1999) Addressing Human Variability in Risk Assessment – The Robustness of the Intraspecies Uncertainty Factor. *Regul Toxicol Pharmacol* **30**: 209-216

Checkoway H, Pearce N, Kriebel D. Research Methods in Occupational Epidemiology. Oxford: Oxford University Press, 2004. ISBN: 0-19-509242-2.

Coggon D, Rose G, Barker DJP. Epidemiology for the uninitiated. Brit Med Journal online. http://www.bmj.com/epidem/epid.html

Danish EPA (2004). Principper for sundhedsmæssig vurdering af kemiske stoffer med henblik på fastsættelse af kvalitetskriter for luft, jord og drikkevand. [Principles for health based evaluations of chemicals for the setting of quality criteria for ambient air, soil and drinking water] in Danish with English summary. Miljøprojekt nr. 974, pp 167.

Doi A, Hill G, Seely J, Hailey J R, Kissling G, Buchler J R (2007) α2u-Globulin Nephropathy and Renal Tumors in National Toxicology Program Studies. Toxicol Pathol **35**, 533-540.

Dourson N L, Andersen M E, Erdreich L S, MacGregor J A (20019 Using Human Data to Protect the Public's Health *Regul Toxicol Pharmacol* **33**: 234-256.

Dutch Expert Committee on Occupational Standards (2003), Formaldehyde, helath based recommended occupational exposure limits. No. 2003/02OSH, The Hague, 27 January 2003.

Dutch Health Council, developed jointly by the members of the committee on the evaluation of carcinogenic substances (1989) Carcinogenic risk assessment of benzene in outdoor air. Regul. Toxicol Pharmacol 9: 175-185

Ecetoc (2009). Framework for the Integration of Human and Animal Data in Chemical Risk Assessment. Technical Report No. 104. Ecetoc, Brussels, January 2009.

Ecetoc (2006). Toxicological Modes of Action: Relevance for Human Risk Assessment. Technical Report No. 99. Brussels, July 2006.

Ecetoc (2004) Workshop on the Use of Human Data in risk Assessment, 23-24 February 2004, Gardiff, Workshop Report No. 3. Ecetoc, Brussels, November 2004.

Ecetoc (2003) Derivation of Assessment Factors for Human Health Risk Assessment, Technical report No. 86. Ecetoc, Brussels, February 2003.

EFSA SC (2005). Opinion of the Scientific Committee on a requat from EFSA related to A Harmonised Approach for Risk Assessment of Substances Which are both Genotoxic and Carcinogenic. The EFSA Journal (2005) 282 :1-31.

Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 ;315(7109):629–624.

Von Elm E, Altman DG, Egger M et al (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370:1453-57.

European Commission, Joint Research Centre, European Chemicals Bureau (2007) European Union Risk Assessment Report, Cadmium oxide and cadmium metal, 3rd Priority List, Volume 72.

European Commission, Joint Research Centre, European Chemicals Bureau (2003) European Union Risk Assessment Report, Toluene, 2rd Priority List, Volume 30.

European Commission, Joint Research Centre, European Chemicals Bureau (2003) European Union Risk Assessment Report, Hydrogen Peroxide, 2rd Priority List, Volume 38.

Goldbohm RA, Tielemans EL, Heederik D, Rubingh CM, Dekkers S, Willems MI, Kroese ED (2006) Risk estimation for carcinogens based on epidemiological data: A structured approach, illustrated by an example on chromium. *Regul Toxicol Pharmacol* **44**: 294-310

Hattis D, Erdreich L, Ballew M (1987) Human variability in Susceptibility to Toxic Chemicals A Preliminary Analysis of Pharmacokinetic Data from Normal Volunteers. Risk Analysis **7**, 415-426

Hernberg S (1992). Introduction to Occupational Epidemiology. Lewis Publishers, Inc. Chelsea, MICH 1992.

Hill AB (1965). The environment and disease: Association or causation? Proc of the Royal Soc Med 58:295-300.

IARC (2006) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Preamble. Word Health Organisation, International Agency for Research on Cancer, Lyon, January 2006.

IARC (2009). A review of human carcinogens—Part F: Chemical agents and related occupations. The Lancet Oncology 10,1143-44.

IPCS. 2005a. Workshop on poison centres and the use of human data in consumer product risk assessment. Report. WHO in collaboration with EC JRC, EAPCCT, ECETOC and the Federal Institute of Risk Assessment, Germany held at the Federal Institute for Risk Assessment, Berlin, Germany, 9 May 2005. International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland.

IPCS. 2005b. Harmonization Project Document No. 3. Principles of characterizing and applying human exposure models. International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland.

IPCS. 2005c. Harmonization Project Document No. 2. Chemical-specific adjustment factors for interspecies differences and human variability: Guidance document for use of data in dose/concentration–response assessment. International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland.

IPCS 2007a. IPCS framework for analyzing the relevance of a cancer mode of action for humans. In Harmonization Project Document No. 4 -Part 1: IPCS framework for analysing the Relevance of a cancer mode of action for Humans and Case-studies. IPCS, WHO, Geneva, Switzerland 2007a. (authors Boobis AR, Cohen SM, Dellarco V, Mcgregor D, Meek ME, Vickers C, Wilcocks D, Farland W, also published in Crit. Rev. Toxicol. 2006;36:781-92).

IPCS 2007b. IPCS framework for analysing the relevance of a non-cancer mode of action for humans. In Harmonization Project Document No. 4 -Part 1: IPCS framework for analysing the Relevance of a cancer mode of action for Humans and Case-studies. IPCS, WHO, Geneva, Switzerland 2007b. (Authors: Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J, Vickers C, also published in Crit. Rev. Toxicol. 2008;38:87-96).

Kilimisch H, Andreae M, Tillman U (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological information. Regul Toxicl Pharm 1997;25:1-5.

Lehman A.J and O.G. FITZHUGH. 1954. 100-fold margin of safety. Assoc. Food Drug Off. U.S. Q. Bull. 18: 33–35.

Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD, Longfellow DG, Pastoor T, Seed J, Patton DE. A framework for human relevance analysis of information on carcinogenic modes of action. Crit Rev Toxicol. 2003;33:591-653.

Money CD, Margary SA (2002) Improved Use of Workplace Exposure Data in the regulatory Risk Assessment of Chemicals within Europe. Ann Occup Hyg 46: 279-285

Money CD (2007). The use of Human Experience Data in the EU Risk Assessment Process. Risk Analysis 27, 387-396.

OECD (2007). Manual for investigation of HPV chemicals. Organisation for Economic Co-operation and Development, Paris, France [www.oecd.org/document/7/0,3343,en_2649_34379_1947463_1_1_1_1,00.html].

Renwick A G, Dorne J L, Walton K (2000) An Analysis of the Need for an Additional Uncertainly Factor for Infants and Children. *Regul Toxicol Pharmacol* **31**: 286-296.

Rothman KJ, Greenland S. (1998). Modern Epidemiology. Lipincott-Raven Publishers Philadelphia USA. ISBN:0-316-75780-2

Steenland K, Greenland S (2004) Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* **160**: 384-392

Sanco (2006) European Commission, health & consumer protection directorate-general. (Working document) Draft guidance for the setting and application of acceptable operator exposure levels (AOELs). SANCO 7531 - rev.10, 7 July 2006

Steenland K, Spaeth S, Cassinelli R, Laber P, Chang L, Koch K (1998) NIOSH life table program for personal computers. Am J Ind Med 34: 517-518

Swaen GMH. A framework for using epidemiological data for risk assessment. Hum Exp Toxicol 2006;25:147-155.

Swaen G, van Amelsvoort L (2009). A weight of evidence approach to causal inference. J of Clin Epidemiol 2009;62:270-7.

US EPA (2005). Guidelines for Carcinogen Risk Assessment. Risk Assessment Forum U.S. Environmental Protection Agency Washington, DC. EPA/630/P-03/001F, March 2005

Vlaanderen J, Vermeulen R, Heedick D, Kromhout H (2008) Guidelines to Evaluate Human Observational Studies for Quantitative Risk Assessment. Environm Health Perspect.,**116**: 1700-1705.

van Wijngaarden E, Hertz-Picciotto I (2004) A simple approach to performing quantitative cancer risk assessment using published results from occupational epidemiology studies. *Sci Total Environ* **332**: 81-87

WHO Working Group report (2000) Evaluation and use of epidemiological evidence for environmental health risk assessment: WHO Guideline Document. Environm Health Perspect.,108: 997-1002.

World Medical Association (2000). Declaration of Helsinki.: Ethical Principles for Medical Research Involving Human Subjects. Adopted by the 18th World Medical Association (WMA) General Assembly Helsinki, Finland, June 1964 and amended by the 29th WMA General Assembly, Tokyo, Japan October 1975, 35th WMA General Assembly, Venice, Italy, October 1983, 41st WMA General Assembly, Hong Kong, September 1989, 48th WMA General Assembly, Somerset West, Republic of South Africa, October 1996 and 52nd WMA General Assembly, Edinburgh, Scotland, October 2000.