Benchmark dose method: Derivation of health-based recommended exposure limits in new perspective

Gezondheidsraad **President**
Health Council of the Netherlands

To
State Secretary of Housing, Spatial Planning
and the Environment

Subject : report on the assessment of substances - benchmark dose method
Your reference : -
Our reference : U 382/WP/MK/442-L3
Enclosure(s) : 1
Date : March 20, 2003

Mr State Secretary,

Please find enclosed an advisory report on the assessment of the risks of exposure to chemical substances – benchmark dose method. It has been prepared on my request by a committee of the Health Council and was reviewed by the Council's Board on Health and Environment. Also the Dutch Expert Committee on Occupational Standards—a committee of the Health Council of the Netherlands that derives health-based recommended occupational exposure limits—reviewed a draft of the report. The report was also presented today to the Minister of Health, Welfare and Sports and the State Secretary of Social Affairs and Employment.

The advisory report is part of a series of publications on the scientific assessment of the risks of exposure to chemical substances. Predecessors of the present report are *Exposure to combinations of substances: a system for assessing health risks* (2002/05) and *Toxicity testing: a more efficient approac*h (2001/24). The benchmark dose method is a tool to reduce the uncertainties in health based recommend exposure limits to be used as a criterion for standard setting. The method was developed in the last 25 years; TNO and RIVM provided important contributions to this development.

The committee recommends implementing the method in national policy; I strongly support this recommendation. Optimal implementation requires adjustment of the guidelines for toxicity testing. Such adjustments have to be the result of international deliberation as an efficient management of the risks of occupational and environmental exposure to chemical substances presupposes international harmonization of national policies. Given the available expertise in our country I suggest that The Netherlands takes some first steps in this respect.

P.O Box 16052
NL-2500 BB The Hague
Telephone +3 (70)340 6262
Telefax +31 (070)340 75 23
email: ja.knottnerus@gr.nl

Visiting Address
Parnassusplein 5
NL-2511 BX The Hague
The Netherlands

Gezondheidsraad
Health Council of the Netherlands

Applying the benchmark dose method in the derivation of health based exposure limits does not circumvent the extrapolation from research and tests data to the situation of the population at risk. At present, this step generally consists of the introduction of safety or uncertainty factors. As indicated by the 2003 Work Programme of the Health Council I will establish in the course of the current year a committee that will study improvements of this extrapolation step. It is my intention to present the results of that study in the Spring of 2004.

Sincerely Yours,

(signed)

Prof. dr JA Knottnerus

P.O Box 16052
NL-2500 BB Den Haag
Telephone +3 (70)340 6262
Telefax +31 (070)340 75 23

Visiting Address
Parnassusplein 5
NL-2511 BX Den Haag
Telefax (070) 340 75 23
email: ja.knottnerus@gr.nl

Benchmark dose method: Derivation of health-based recommended exposure limits in new perspective

to:

the State Secretary of Housing, Spatial Planning and the Environment

the Minister of Health, Welfare and Sport

the State Secretary of Social Affairs and Employment

Nr 2003/06E, The Hague, March 20, 2003

The Health Council of the Netherlands, established in 1902, is an independent scientific advisory body. Its remit is "to advise the government and Parliament on the current level of knowledge wiht respect to public health issues...." (Section 21, Health Act).

The Health Council receives most requests for advice form the Ministers of Health, Welfare & Sport; Housing, Spatial Planning & the Environment; Social Affaris & Employment; and Agriculture, Nature Preservation & Fisheries. The Council can publish advisory reports on its own initiative. It usually does this in order to ask attention for developments or trends that are thought to be relevant to government policy.

Most Health Council reports are prepared by mulitidisciplinary committee of Dutch or, sometimes, foreign experts, appointed in a personal capacity. The reports are available to the public

Preferred citation:
Health Council of the Netherlands. Benchmark dose method: health-based recommended exposure limits in new perspective. The Hague: Health Council of the Netherlands, 2003; publication no. 2003/06E.

ISBN: 90-5549-481-X

Contents

Executive summary

Exposure to a chemical via air, water or food can result in the impairment of health, depending on the degree of exposure and the toxicity of the substance. Health-based recommended exposure limits for a substance correspond with the highest estimated dose or concentration that does not lead to health impairment. To support the implementation of policy on industrial safety, environmental management and food safety, the Health Council derives such limits for specific substances. The Council also studies the methods employed to derive recommended exposure limits, and from time to time it advises modifying these methods to bring them into line with the latest developments in scientific knowledge.

This document, drawn up by the Health Council's Committee on the Derivation of Health-Based Recommended Exposure Limits, forms one such methodological advisory report. It discusses the use of the 'benchmark dose' or BMD method in deriving health-based recommended exposure limits. This method is an alternative to the approach generally employed in the Netherlands and elsewhere, in which either animal experiments, studies with volunteers or epidemiological research are used to determine the highest level of exposure which does not lead to adverse health effects (the No Observed Adverse Effect Level, or NOAEL). An 'uncertainty factor' is then applied to allow for the differences between experimental animals and humans, differences in sensitivity between human individuals, and research data deficiencies. The resulting exposure value becomes the health-based recommended exposure limit.

Background

The reason that an alternative is being sought lies in the shortcomings of the NOAEL method, the most important of which is that an NOAEL value strongly depends on the quality of the available research data. The less precise this data is, the larger the corresponding NOAEL value tends to become. However, for the purposes of public health protection one would prefer such data imprecision to imply a lower NOAEL value rather than a higher one. Evaluation studies have shown that, in practice, exposure to levels equal to (or even below) the NOAEL does not rule out the occurrence of adverse health effects. The fact that health-based recommended exposure limits do indeed provide public health protection is due to the magnitude of the uncertainty factors which are used to derive the recommended exposure limit from the NOAEL. The way in which such factors are determined is rather ad hoc.

This last argument also demonstrates why an apparently simple approach ought to be replaced by a mathematically and statistically complex one. The simplicity of the NOAEL method conceals considerable uncertainties, which are explicitly addressed in the BMD method. The BMD method can therefore yield health-based recommended exposure levels having a lower inherent uncertainty than the figures generated using the NOAEL approach.

The BMD method

The BMD method sets out to analyse the data on the effects of a chemical on animal or human health in order to determine, as accurately as possible, the relationship between a given exposure level and the likelihood of its detrimental effects (the so-called response). The statistical uncertainty to which this data is invariably subject is incorporated into the calculations. The figures are then used to yield a ‘benchmark dose’ or BMD: this is the dose which corresponds with a given statistical likelihood of health impairment in the exposed population—for instance, 1 per cent or 10 per cent. The BMD is then divided by an uncertainty factor to yield a health-based recommended exposure limit.

As has already been mentioned, the BMD method takes better account of research data uncertainties than does the NOAEL method. Moreover, the NOAEL is by definition one of the experimental doses applied, while the BMD is a quantity derived from all experimental values. Finally, the BMD method also holds out the prospect of obtaining information about the risks associated with exposure exceeding the health-based recommended exposure limits.

In 1996 the Committee recommended that it be investigated whether these theoretical advantages also existed in practice. In the Netherlands, the Ministry of Housing, Spatial Planning and the Environment and the Ministry of Social Affairs and Employment commissioned the TNO Nutrition and Food Research and the National Institute of Public Health and the Environment to carry out this research. Outside the Netherlands, too, useful experience has been gathered on the BMD method, stimulated in particular by the US Environmental Protection Agency. The Committee has assessed the findings of this research and now presents the following conclusions and recommendations.

Feasibility of the BMD method

Research in recent years, both in the Netherlands and elsewhere, has shown that the BMD method offers clear and tangible advantages over the existing NOAEL method. The Committee is of the opinion that if these advantages are to be fully exploited, the protocols for toxicological studies should be modified. It also notes that comparatively little attention has so far been given to the uncertainty factor that must be applied to the BMD in order to yield a health based recommended exposure limit. Although this aspect was outside the scope of its remit, the Committee recommends further study of the extrapolation from BMD tot exposure limit.

Another point requiring closer attention is the choice of a model function to describe the form of the dose-effect (or dose-response) relationship. These functions are currently still strongly determined by statistical considerations. The Committee holds that it would be desirable to strengthen the biological basis of this choice. The same applies to the choice of the degree to which an effect is deemed no longer not to impair health. For example, is a 5 per cent weight gain in the liver with respect to the average liver weight in a non-expose population evidence of damage to health, or would 10 per cent still be compatible with good health? This choice should be made on biological and toxicological grounds wherever possible. However, this is easier said than done. The Committee there-fore indicates a route by which, for the time being at least, this choice can be made on the basis of pragmatic considerations.

Notwithstanding the need for its further development, the Committee considers the BMD method to be a useful technique for the derivation of recommended exposure limits. Where toxicological data makes its application possible, the Committee prefers the BMD method above the NOAEL approach. The BMD method yields improved foundations for health-based recommended exposure limits and, when deriving these limits, it curtails dependency on uncertainty factors that have been arrived at on the basis of qualitative considerations. The Committee recommends that the Dutch government accept the derivation of recommended exposure limits using the BMD method as the basis for limit values laid down in law or governmental policy.

Estimating effects

The Committee is less sanguine about the prospects of estimating the effects of exposure exceeding the health-based recommended exposure limit. It considers that considerable uncertainties are attached to such estimates, and that they should therefore be used with extreme caution in concrete cases of exposure to toxic substances. Given the need for such estimates, the Committee recommends that the government support further research into the reduction of these uncertainties.

Additional advice

In order to further develop the BMD method and its application in determining recommended exposure limits, the Committee considers additional advice to be required in three areas in particular:

Uncertainty factors

The Committee recommends that the Health Council assess the customary values of the uncertainty factor and new approaches for describing this factor. It considers that an international panel of experts is called for.

Protocol development

The Committee notes that existing protocols for toxicity studies are not well matched to BMD methodology, although their application does not make it im-possible. It recommends that steps be taken to modify the protocols accordingly. If the Dutch government adopts the Committee's views on the usefulness of the BMD method, it should also stimulate discussion of this matter at OECD and EU level.

Choice framework for BMD parameters

The BMD method demands that three choices be made: the statistical likelihood of an effect underlying the determination of the BMD; the dividing line between an effect size deemed to be benign and one deemed to be non-benign; and the choice of a model function with which to describe the relationship between dose and effect. As things stand, these choices have to be made and justified on a substance-by-substance basis. The Committee considers it desirable that this reasoning be carried out within a framework

of selection criteria. It proposes that the Health Council or other authoritative body initiate the creation of such a framework, preferably at an international level.

Chapter 1

Introduction

1.1 Request and Committee

In 1996 the Health Council *Derivation of Health-Based Recommended Exposure Limits* Committee, the current members of which are listed in Appendix , issued an advisory report about developments relating to the determination of 'toxicology-based recommended exposure limits for substances'.[1] Toxicology-based recommended exposure limits are scientifically-based policy instruments for the protection of individuals against exposure to hazardous substances. Using these values is intended to prevent or limit health effects.

The 1996 advisory report set out a programme for further research and evaluation. The aim was to acquire data in order to answer the questions raised. The programme included the derivation of health-based recommended exposure limits. These are toxicology-based recommended exposure limits at which and under which it is not reasonable to expect adverse effects. In the summary of its advisory report, the Committee had the following to say about this area:

The formulation of health-based recommended exposure limits

In the Netherlands and elsewhere, health-based recommended exposure limits are normally formulated using the NOAEL-uncertainty factor method. […]A health-based recommended exposure limit is calculated by dividing the NOAEL [no observed adverse effect level] for the substance in question by a factor which takes account of uncertainties in the available data and in the extrapolation from that data of findings perti-

nent to the population group to which the limit is to apply. No other methods are in widespread use. The Committee believes that limits should be formulated using a method which makes systematic use of data on the relationship between exposure and response; the so-called 'benchmark dose' (BMD) method is felt by the Committee to be particularly promising. However, before this method can be regarded as a viable alternative to the established method, further practical evaluation is required. The Committee therefore recommends that new health-based recommended exposure limits be formulated using the BMD method for various substances which have already been assessed for occupational health and safety and environmental protection purposes. In this way, it would be possible to identify the data required to apply the method in practice and the issues relevant to the extrapolation of health-based recommended exposure limits from BMDs.

Since that time, in the Netherlands, the National Institute for Public Health and the Environment (RIVM) and TNO Nutrition and Food Research have examined ten substances that had already been assessed for toxic effects. The aim of this assessment was to determine whether the benchmark dose method is, given the available data, a practicable approach for identifying the difficulties associated with the application of the method and for establishing the quality of the results by comparison with the standard approach.[2,3] Furthermore, experience has been acquired outside the Netherlands in recent years with the benchmark dose method.[4] For example, the American Environmental Protection Agency encourages the use of the method by providing the required software free of charge.[5]

This experience justifies a return to the discussion by the Committee of the benchmark dose method for the derivation of health-based recommended exposure limits. This discussion is situated in the context of the series of advisory reports about the toxicology-based appraisal of substances.[1,6,7] The present advisory report contains that discussion and concludes with a recommendation about the use of the new method.

1.2 Concepts: exposure, effect and response

The concepts of exposure, effect and response play an important role in the derivation of recommended exposure limits. Definitions of these concepts follow here. Where possible, they are consistent with the conceptual framework of the Committee's 1996 advisory report.[1]

Exposure

Contact between an organism and a substance, such that the substance may affect functions of the organism.

Organisms are usually exposed as a result of inhalation, ingestion or skin contact. The terms dose and dosage are commonly used to quantify exposure. The Committee makes no distinction between the two and prefers the term dose, meaning the administered or ingested amount of a substance per unit of body mass. The Committee believes that numerical values should be followed by units such as g/kg or mg/kg.* Other exposure measures are often used instead of 'dose' for exposure via the skin or the air. For example, in the case of exposure via inhalation, the concentration of a substance in the air is in widespread use.

The consequences of exposure are designated using the concepts of 'effect' and 'response'.

Effect

The reaction of an organism—over the short or long term—induced by exposure to a substance.

Effects may include changes in morphology, physiology, physical and mental development or life-span.

The reaction of the organism to exposure to the substance may be local or systemic (or both). In the first case, the effect will be confined to the site of the contact, for example the skin, the windpipe or the lungs. In the second case, the substance is absorbed into the body and can induce effects in a range of organs.

Response

The proportion of an exposed population group or group of animals in which a given effect is induced by exposure to a given substance.

1.3 Health-based recommended exposure limits

In its 1996 advisory report, the Committee introduced the concept of 'toxicology-based recommended exposure limit'.[1]

* In toxicology, the amount of a substance is usually expressed in units of mass. Furthermore, 'dose rate' (dose per unit of time) is also often shortened to 'dose'. The context will make clear what is meant.

Toxicology-based recommended exposure limit

The level of exposure to a substance that, taking due account of factual lacunae and uncertainties regarding the interpretation of toxicity data, may reasonably be believed not to have more than a specified chance of affecting the health of a specified population group in a specified way.

This definition implies that a toxicology-based recommended exposure limit must always specify the following:

- the relevant health effect
- the maximum chance of the substance having the specified effect
- the population group to which the recommended limit applies.

It is also necessary to clarify the nature of 'the level of exposure': whether the exposure concerned is short-term or chronic, occurs only once or is intermittent, etc. To prevent the toxicity of a substance being underestimated, the Committee has endeavoured to frame a definition which will ensure that the inherent uncertainties are properly accounted for.

Health-based recommended exposure limit

A 'health-based recommended exposure limit' is a specific toxicology-based recommended exposure limit that specifies the level of exposure at or below which the risk of any adverse health effect may reasonably be expected to be nil.

The government uses toxicology- or health-based recommended exposure limits derived by scientific experts as the basis for setting standards. Social considerations such as economic feasibility may lead to a difference between the standard and the recommended exposure limit, with the standard usually being higher than the limit.

In the Netherlands, many MAC* values—limits for exposure via the air to substances in the workplace—are underpinned by health-based recommended exposure limits set by the Health Council. In the context of the environment policy, the Health Council and RIVM have adopted health-based recommended exposure limits for maximum permissible risk levels. The policy for limiting exposure to carcinogenic substances is based on the risk of contracting cancer later in life. The associated level of exposure can be described as the toxicology-based recommended exposure limit.

* Maximum Accepted Concentration

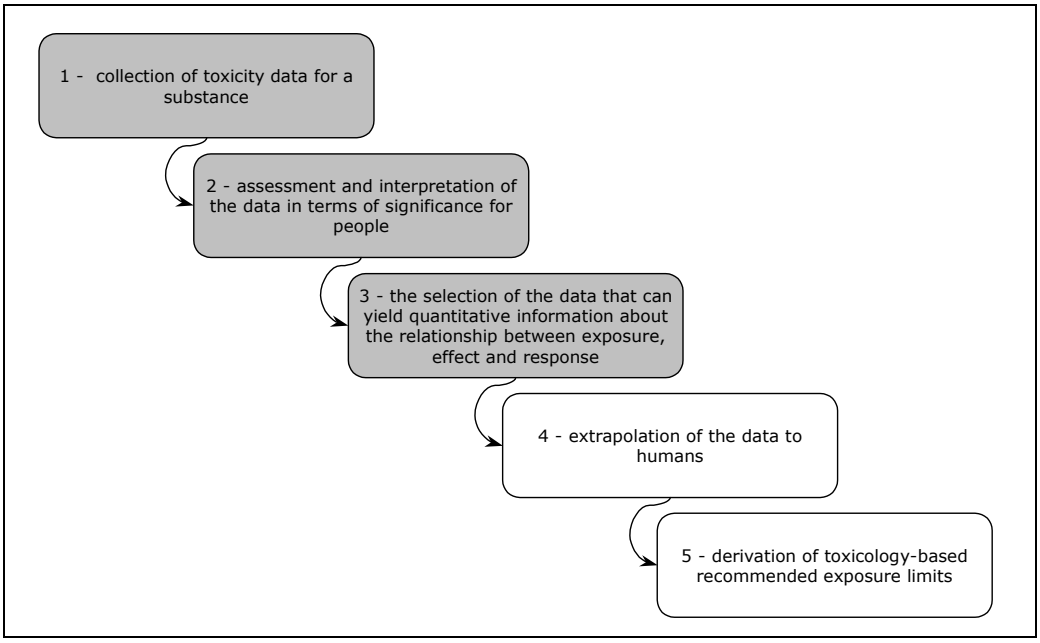

Figure 1 Derivation of toxicology-based recommended exposure limits as a phased process.[1] Steps 4 and 5 are not discussed in this advisory report.

The derivation of toxicology- and health-based recommended exposure limits is a phased process (Figure 1). In many cases, especially when the substances in question are 'new', data relating to toxicity—step 1—is collected from international standardised trials with cells or tissue (*in vitro* studies) or experimental animals (*in vivo* studies). Data about effects in humans (from reports of poisoning, volunteer studies or epidemiological studies) is not available at this stage. Step 2 involves the selection of the data relevant for effects in humans. It covers the quality of the research, as well as similarities and differences between the experimental animals used and humans. Step 3 involves the selection and interpretation of the information about the relationship between exposure, effect and response. The data is then extrapolated to the population group requiring protection (step 4). The final step (5) results in the toxicology- or health-based recommended exposure limit. In this step, margins of uncertainty in the data and variations in people's sensitivity are included in the calculations.Figure 1 Derivation of toxicology-based recommended exposure limits as a phased process.[1] Steps 4 and 5 are not discussed in this advisory report.

This advisory report on the benchmark dose method is limited to the first three steps in Figure 1. The emphasis is on the process of the derivation of *health-based* recommended exposure limits and the discussion focuses on the use of data from *in vitro* and *in vivo* studies. There is reasonable world-wide consensus about the way in which

toxicity data is obtained at present. However, in recent years, it has been suggested that new approaches are required given the current scientific state-of-the-art.[7]

There would appear to be a threshold value for many effects of toxic substances. At and under the threshold exposure value, the substance has no observable adverse effect.[8] Toxicity studies therefore generally concentrate on acquiring information about that threshold (step 2 in Figure 1).

1.4 Structure of the advisory report

Chapter 2 contains a short description of what has, until now, been the usual method for deriving health-based recommended exposure via the determination of a No Observed Adverse Effect Level. In Chapter 3, the Committee describes the benchmark dose method. Subsequent chapters take a closer look at two key elements of the benchmark dose method: the determination of the link between the toxicological effects in a population exposed to a substance and the level of exposure (Chapter 4) and the criteria for deriving a benchmark dose from that link (Chapter 5). Chapter 6 follows with a description of activities undertaken by RIVM and TNO in order to identify problems with the application of the new method. The concluding chapter (7) sets out the conclusions of the Committee.

The scientific basis for the advisory report consists of the literature published since 1995 about the benchmark dose method. The Committee used the on-line literature search facility PubMed* to search for publications with 'benchmark dose' in any 'field'. The articles with an appraisal of the method were then selected from the resulting list. If necessary, the Committee also consulted any literature sources referred to in those publications. This corpus of literature also included the reports from TNO Nutrition and Food Research and RIVM about the benchmark dose method and about the derivation of health-based recommended exposure limits in general.

* Internet: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

Chapter 2

NOAEL method

In this chapter, the Committee looks at the NOAEL method for the derivation of health-based recommended exposure limits. The focus is on substances for which toxic effects only emerge above a threshold exposure value.

2.1 Method

A estimate of the threshold value referred to in 1.3 is arrived at in most cases by comparing the results for groups of experimental animals exposed to a range of doses with those for a reference group that has not been exposed to the substance in question.* That estimate, the No Observed Adverse Effect Level (NOAEL), is defined as:[8]

> The highest level of exposure in toxicological or epidemiological studies at which there is no statistically or biologically significant increase in the frequency or severity of adverse effects between the exposed population and a suitable reference population.

The NOAEL may therefore include observable effects but they will not be considered to be harmful or to be harbingers of adverse effects.

* Alternatively, in the case of substances found normally in, for example, food or air, the comparison will be with groups not subjected to additional exposure.

In some trials, there may still be a significant effect or a significant response in the group exposed to the lowest dose. This dose is known as the Lowest Observed Adverse Effect Level or LOAEL.[8] A trial of this kind will not therefore result in a NOAEL value.

When the effects can be described using a threshold model, the most usual method used at the Health Council and elsewhere for determining health-based recommended exposure limits for substances is to divide the NOAEL for the critical effect by an uncertainty factor (UF*). The 'critical effect' is the adverse effect with relevance from the point of view of human health that is first observed as exposure increases. The UF is a product of constituent factors. It is meant to ensure that, when humans are exposed to doses below the health-based recommended exposure limit (that is equal to NOAEL/ UF), adverse effects will not occur in healthy individuals.

The UF constituent factors focus on:[1,8,9]

- The nature of the effect
  Sometimes, there may be uncertainty about the exact positioning of the NOAEL in the case of an effect considered to be severe, such as a congenital defect. In such cases, an additional uncertainty factor can be used to establish a recommended exposure limit that can be viewed with greater confidence in terms of the absence of an effect.
- Completeness of the set of toxicological data
  Protocols have been developed for toxicity studies that are intended to safeguard the quality and completeness of toxicity data. A complete dataset of this kind is not always available. An additional uncertainty factor can be used to take account of the resulting uncertainty.
- LOAEL instead of NOAEL
  If the toxicity data is adequate for the derivation of a LOAEL but not a NOAEL, an additional uncertainty factor is usually introduced.
- Differences between species
  Many toxicity studies use rodents, in particular rats. The results of such research, as well as those of studies using other experimental animals, has to be extrapolated to humans. An uncertainty factor takes account of the fact that a human adult is not 'a 70 kilogram rat'.
- Variations in sensitivity between people
  Data from experimental animal trials generally relates to animals that are very similar in terms of sensitivity. In a human population, there will be more variation, depending on the nature of the population requiring protection. This is the result of differences in inherited traits, different stages of development (babies, children,

---

* Some authors prefer 'assessment factor' (for example [33]); terms such as 'extrapolation factor' and 'safety factor' are also used.

adults, elderly people), differences caused by illness and the influence of the living environment. An uncertainty factor is intended to ensure that also the more sensitive individuals will be protected by the use of health-based recommended exposure limits.

The last two constituent factors are often referred to as, respectively, the '*interspecies* factor' and the '*intraspecies* factor'.

The determination and scientific justification of the UF and its constituent factors constitute a distinct field of research.

No threshold

If the toxicological effects induced by the exposure of an organism to a substance cannot be described in terms of a threshold, the NOAEL method for the determination of health-based recommended exposure limits is inappropriate. That is assumed to be the case, for example, with a certain category of carcinogenic substances, the stochastic genotoxic carcinogens.[10] In these cases, standards for health protection are generally based on toxicology-based recommended exposure limits that correspond to a given response. The standard for the carcinogenic substances referred to is, for the general population in the Netherlands, 1 in one million exposed people given lifelong exposure.[1,10]

2.2 Evaluation

An advantage of the NOAEL method for the derivation of health-based recommended exposure limits is its conceptual simplicity. Laboratory research (or volunteer studies, or epidemiological studies) is used to establish a level of exposure at which and under which no harmful effect of the substance in question is observed. The introduction of an uncertainty factor then results in a level of exposure that, in practice, affords a substantial degree of protection.

However, this conceptual simplicity is offset by a number of drawbacks (c.f. [11-14]), that give rise to the question of whether the derived health-based recommended exposure limit does not mainly depend on the use of a sufficiently large uncertainty factor. In other words: it is unclear in many cases what the relationship is between a NOAEL and the real No Observed Adverse Effect Level in the study in question. For example, Faustman estimated that actual response levels associated with the NOAEL can vary from 5 to 20 per cent (and are therefore larger than the assumed 0 per cent).[13]

The NOAEL corresponds to one of the exposure levels in a toxicological test. The results at other exposure levels are not used to acquire information about variability in

the data or, therefore, to establish a clearer picture of the precision of the NOAEL that has been found. Nor is the shape of the exposure-response curve used for that purpose.

Furthermore, the NOAEL depends on the design of the toxicological study, and in particular on the total number of experimental animals used, the number of exposure levels and the intervals between the levels. Whether a statistically significant difference in response or average effect can be demonstrated between an exposed group of animals and the reference group depends on the number of animals in each group. In general, with a given set of exposure levels, smaller numbers of experimental animals per group may result in a higher value for the NOAEL. In any case, the NOAEL will not be lower. From the protection perspective, one would wish the opposite to be the case: a less reliable trial in statistical terms should result in a lower value.

Chapter 3

The BMD method

The above objections (2.2) to the NOAEL method have led to suggestions for alternative approaches. The Committee drew attention to one of the main ones, the benchmark dose method, in its 1996 advisory report.[1] This chapter describes the method and the next two chapters examine it in greater detail. The main reason for a more detailed review of the BMD method is that it does greater justice to the available research results. As a result, it reduces the uncertainty surrounding recommended exposure limits.

3.1 Method

The BMD method was described for the first time by Crump in 1984.[11] He defined the BMD as the lower 95% confidence interval for the dose that induces a given response for a toxicological effect above the background incidence (in the international literature, this is now usually referred to as the BMD-L; see below). In the determination of health-based recommended exposure limits, the BMD/BMD-L is competing with the NOAEL.

The BMD method has been studied extensively in the last twenty years.[4,5,12,13,15-30] This has not yet resulted in a distinct preference for a particular variant. However, in principle, all the variants are based on the selection of two parameters: the benchmark response (BMR) and the critical effect size (CES).

Benchmark dose (BMD)

The BMD for a toxic substance is the exposure* to that substance at which the estimated response, in other words the chance of a given toxic effect occurring in the exposed population (an effect that is larger than the CES), corresponds to the value selected in advance for the benchmark response (BMR). In some cases, this definition requires some elaboration, for example if the effect also occurs in a population which has not been exposed to the substance. It follows from the definition that, in principle, every toxic effect has a BMD and that the value of the BMD depends on the selection of the CES and of the BMR.

Critical effect size (CES)

The CES is the size of a toxicological effect above which it is considered to be harmful. Smaller effects will therefore be classified as benign and larger effects will, at least in principle, be considered to be harmful. In the case of toxic effects that may or may not be found in an exposed individual (also known as 'quantal effects'), there are usually no differences of opinion about the CES: any effect size is harmful. This is the case, for example, with malignant tumours. Other examples are mortality within a certain time after exposure to a toxic substance, and birth defects. Other effects, by contrast, are more gradual in nature. Examples are growth impairment expressed as a lower rate of increase in body weight, and changes in haemoglobin concentration. Here, a consensus is more problematic when deciding on the CES as the dividing line between 'good' and 'bad', or rather between 'not bad' and 'possibly bad'.[12]

Benchmark response

The BMR is a particular value that is chosen to designate the response (incidence) for an effect considered to be harmful—an effect in excess of the CES—in the exposed population. The dose expected to induce a response equal to the BMR in that population is therefore the benchmark dose or BMD. The required data here generally comes from research with experimental animals. Epidemiological data, data from research with cells and organ systems (*in vitro* studies) or with volunteers can be used.

The selection of the BMR is relatively *ad hoc* in nature. The usual values are 1, 5 or 10 per cent[11], although 50 per cent has also been proposed[23,31]. In the latter case, the average** effect at the dose that corresponds to the BMR is equal to the CES.

---

* For example, in the form of a daily dose or a concentration of the substance in the air.

** Strictly speaking, the median effect. When the distribution is symmetric, the average and median correspond.

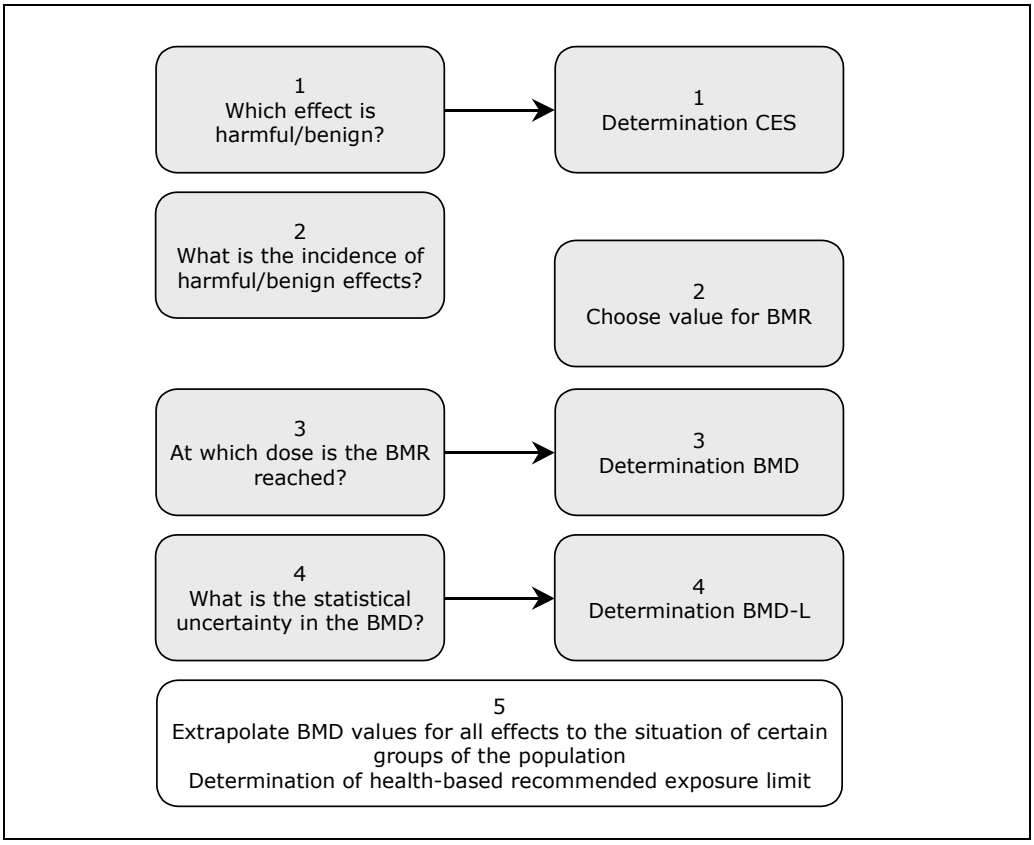

Figure 2 Flow chart for the derivation of health-based recommended exposure limits using the benchmark dose method. Step 5 is not discussed in this advisory report. For abbreviations, see the glossary at the back of this report. In principle, the CES is the result of biological and toxicological considerations. By contrast, the selection of the BMR is much more ad hoc.

From BMD to health-based recommended exposure limit

In order to account for the variation of the research data, the BMD is not taken as the point of departure for the determination of a recommended exposure limit. The preferred point of departure is the value that corresponds to the lower 90% or 95% confidence interval for the BMD: the BMD-L. Figure 2 shows how the BMD method is used to derive health-based recommended exposure limits.

As with the NOAEL method, the health-based recommended exposure limit is obtained by dividing the BMD-L by an uncertainty factor, UF. The value of the UF is not automatically equal to the value in the 'standard' method.[28] The constituent factor that, in the case of data from experimental animal trials, accounts for *interspecies* variation, will generally correspond to the one used for the NOAEL method. It should be

pointed out that there have been alternatives proposed, precisely in conjunction with the application of the BMD method, for division by a UF.[30]

In general, toxicity experiments generate information about several parameters. In so far as the values of those parameters in the dose range studied are actually influenced by exposure to the substance, it is possible to determine a BMD for each of those parameters. In general, the lowest value in the set of BMDs will serve as the point of departure for the derivation of the health-based recommended exposure limit, unless the UF to be applied is not the same for each toxicity parameter. If the latter is the case, a recommended exposure limit has to be derived for each BMD. The lowest is then taken as *the* health-based recommended exposure limit.

3.2 Experience

In the Netherlands in recent years, partly in response to the Committee's 1996 report[1], TNO and RIVM have made a joint contribution to the development of the BMD method.[2,3,12,23,25,30,32-36] Their work will be discussed in more detail later in this advisory report.

The BMD method is the most widely-studied and widely-used method for reproduction toxic effects.[19,37] In 1995, Foster and Fleet concluded that the use of the BMD method in reproduction toxicology has advantages compared to the NOAEL approach.[38] Piersma *et al* also arrived at the conclusion that the BMD method was actually very suitable for demonstrating the specific embryotoxicity of a model compound (butyl benzyl phthalate).[34] BMD values have now also been determined for a range of other substances. Alternatively, health-based recommended exposure limits have been derived using the BMD method.[37] Gephart *et al* studied the pros and cons of the BMD method for subchronic toxicity data (90-day studies).[39] They believe that the method is still too substance-specific to be used generally. An extensive study was recently published that looked at the significance of the BMD method for recommended exposure limits relating to mortality after acute exposure.[24] The authors calculated a benchmark concentration for a large number of substances and concluded that the BMD method was a feasible option in many cases.

The US EPA* has already used the BMD method to determine *a reference dose* (RfD) or *reference concentration* (RfC) for various substances ** (for an overview, see [28]). Sand *et al* recently published an analysis of the models proposed by the US EPA,

* US EPA - Environmental Protection Agency of the US.

** RfD and RfC can be thought of as equivalents of the Dutch health-based recommended exposure limits.

looking at the influence of choices relating to the BMR.[29] The OECD* is discussing the method but has not yet implemented it.

3.3 Handling data

Data about the consequences of exposure to a toxic substance can be found in numerous forms. Here, the Committee will discuss quantal, continuous and ordinal data.

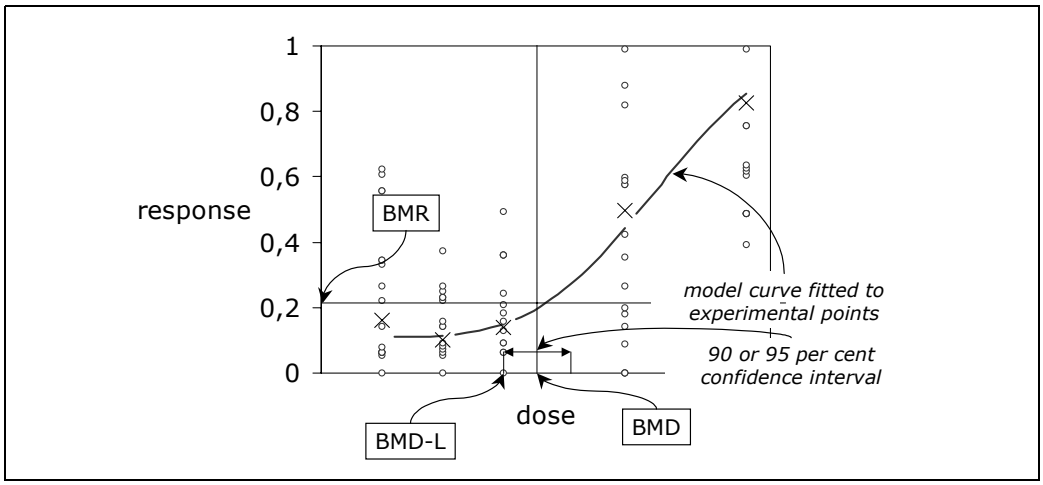

Figure 3 Representation of the derivation of the BMD and the BMD-L. The circles indicate the proportion of prenatal deaths for the different mother animals (response) at different dose values of DEHP (diethylhexaphthalate) administered to the mother animals. The crosses indicate the mean values for each dose group. BMR (benchmark response) refers to the increase in the proportion of still-born animals compared to the proportion of live births in the non-exposed reference group. The value selected here for the BMR is 5%. The BMD-L here is the lower 95% confidence interval for the BMD. Data taken from [37].

Quantal data

All effects like malignant tumours, death or congenital defects will generally be considered harmful or undesirable. Furthermore, with tumours and death, there is generally no dose-dependency: an effect either does or does not occur (tumour incidence is of course dose dependent). In these cases, the CES corresponds to the absence of the effect that is considered to be harmful. The first step in Figure 2 can therefore, in effect, be skipped. The data about substance-induced effects of this kind is shown in the form of values for the responses (incidence) at given exposure values (dose). The effect may also often be found in the absence of any exposure to the substance.

* OECD – Organization for Economic Cooperation and Development.

Fitting a model function to the response data generates a link between response and dose. This can be used for the derivation of the BMD (the dose for the selected BMR value) and of the 90% or 95% confidence interval for the BMD. The lower limit for that interval is the BMD-L. The derivation process is shown schematically in Figure 3. In addition to the determination of a BMR, the procedure therefore requires the selection of a model function that shows the form of the dose-response relationship.

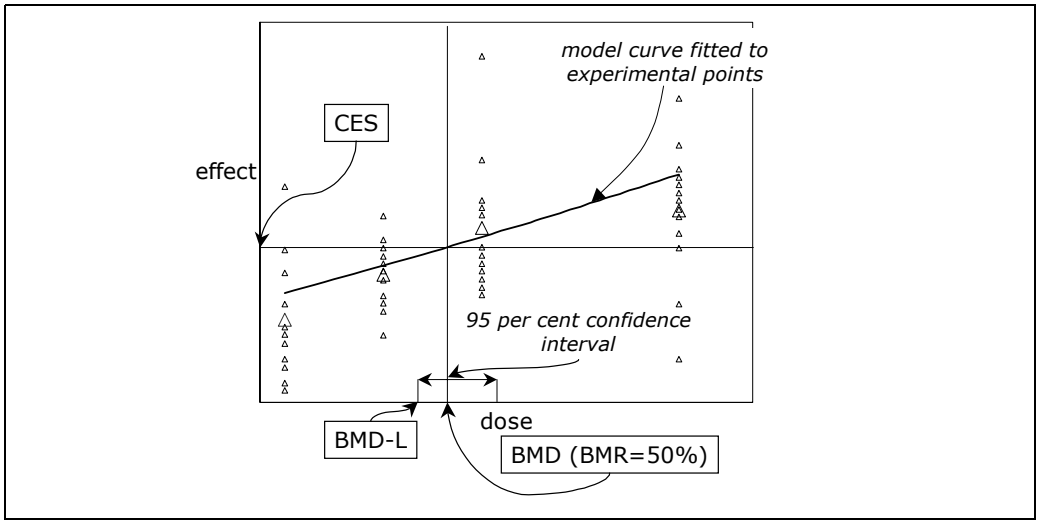

Figure 4 Example of the derivation of the BMD and BMD-L for continuous data. The experimental animals were exposed to three different doses of white spirits. The data on the far left relate to the non-exposed reference group. The small triangles indicate the relative increase in the weight of the kidneys (effect) for an individual animal; the larger triangles indicate the averages for a dose group. The CES is 10% and the BMR 50%. The BMD-L here is the lower 95% confidence interval for the BMD. If a lower BMR was to be chosen, the BMD would also be lower. Data taken from [2].

Continuous data

In the case of continuous data, it is absolutely essential to determine a value for the CES first, in addition to selecting a value for the BMR (see also Chapter 5). A range of arithmetical options are then open. An American 'school'[16,17,28] has tried to establish a link with the derivation of the BMD for quantal data. The researchers from this school draw up a model for the response associated with an effect larger than the CES. They also take into account the uncertainty in the response values that are generated on the basis of the generally wide distribution of effect values as a function of the dose.

In the Netherlands, RIVM and TNO have opted for a different approach.[23,25,30,31] Researchers at these institutions fit model functions to effect data (individual and general) and define the BMD as the dose that, according to the model function that fits the data, results in an effect equal to the CES. In effect, this amounts to selecting a BMR of

50% because the fitted line is the estimate of the median for the effect that occurs at each dose. See Figure 4.

Ordinal data

Data is sometimes available in ordinal form. In other words, it is broken down into categories (classes). An example is the scoring of abnormalities in liver cells in terms of ‘normal’, ‘mild’, ‘moderate’ and ‘severe’. With the benchmark method, extrapolation to quantal data is an option here. The CES will then be set at the dividing line between two categories (between ‘mild’ and ‘moderate’, for example). An alternative is to use the ordinal data as a representation of continuous parameters and to subject them as such to statistical analysis. The calculations from RIVM and TNO in Chapter 6 provide a few examples of this approach.

Chapter 4

Modelling

Calculating the BMD requires, among other things, the determination of the curve that fits the experimental data best: the choice is between dose-response or dose-effect modelling (see 3.3, and, respectively, Figure 3 and Figure 4). In this chapter, the Committee discusses this aspect of the BMD method in greater detail.*

4.1 Quality of the experimental data

Dose-response and dose-effect modelling depend entirely on the quality of the available toxicity data. When using the BMD method, it is desirable to have information for a range of dose values. The result is more detailed information about the relationship between dose, effect and response and therefore a better biological 'fit' between a model function and the experimental data. With a given number of experimental animals in a toxicity study, that implies a preference for a large number of different dose groups and relatively small numbers of animals in each group. Protocols for toxicity experiments are not drawn up with this in mind. Chapter 6 discusses in greater detail the extent to which the figures from studies of this kind can be used in the context of the BMD method.

* The discussion in this chapter is based on laboratory research. Other kinds of data, such as data obtained from epidemiological research, can also be used to determine a BMD.

4.2 Regression model

The determination of the 'best fit' model function is a form of regression analysis, with the dose values being the regressors. Given the fact that a model function is not a simple consequence of biological considerations, various model functions are often tried out before the best fit is chosen.[29] This procedure breaks down into the following stages:

1 Selecting the regression model functions to be included in the analysis
2 Fitting each of the regression models to the experimental points
3 Determining whether, in statistical terms, a fit corresponds adequately to the data and is toxicologically plausible
4 Comparing results for the various regression models.

Model functions (1)

The US EPA has developed software for the application of the BMD method that includes functions for both effect and response modelling.[5] As a part of the work of TNO and RIVM referred to above, Slob has proposed a series of linked functions that are suitable for dose-effect modelling.[30] Depending on the available data, the BMD may be highly dependent on the form of the model function. As a result, the mathematical form may determine the value of the BMD rather than, or to a greater extent than, the toxic properties of the substance. It is important to take this into account already when selecting the model function, and also when assessing the quality of the fit in step 3.

Regression (2)

Over the course of time, various algorithms have been developed that determine, given a specific model function, the values for the model parameters that result in the best fit. The Committee will not discuss them here. The reader is referred to the literature mentioned above and to statistical manuals and software packages.

Realistic fit? (3)

Once software has been used to fit a certain model function to the experimental data—and assuming that this has been successful—checks are required to ensure that the model is a reasonable description of the experimental values. This is particularly important because, as was pointed out above, the calculated BMD can vary considerably, depending on the selected model function.[28] The main concern here is a broader assessment of scientific plausibility. An appraisal of this kind is important because a few

extreme data points ('extreme responses') can have such an effect on the model parameters that the resulting curve is neither biologically nor toxicologically plausible.

Sometimes, none of the model functions are 'realistic'. An example is a set of toxicity data that first indicates an increase in the effect and then, at higher doses, a flattened curve or even a decrease in the effect, while the form of the model functions indicate a steady increase or fall. This kind of levelling off can occur with metabolic saturation. Not including the highest dose group or groups in the modelling could result in an acceptable fit in the dose range that is relevant for the BMD. However, this is generally not an option with data from standard toxicity experiments since so little data is left that it is still impossible to select a suitable model function. The BMD method will not then be practicable (unless a biologically adequate model function can be developed).[21]

Model comparison (4)

It is necessary to formulate criteria for the selection of the final model. When a linked set of functions with increasing complexity is used, a significant improvement in the *goodness of fit* may justify selecting the more complex function as the basis for the BMD.

Goodness of fit tests are not adequate on their own for the selection of the 'best' model function. For example, data containing little 'information' about the form of the dose-response or dose-effect relationship may result in a statistically acceptable fit, but not in discrimination between the various model functions. Nevertheless, the various functions may lead to diverse values for the BMD.[27] The Committee is of the opinion that in such a case the available toxicological data is not suitable for deriving a BMD for the toxicological effect in question.

This shows clearly that the applicability of the BMD method is determined by the availability and the quality of software for fitting model functions to toxicity data and the subsequent calculation of a BMD. The US EPA software is freely available on the Internet.[5] Extensions being made to the scope of the software include the selection of model functions just discussed. An important factor here is the possibility of using a linked set of model functions of increasing complexity.[30]

The analysis of the practicality of the BMD method conducted by TNO and RIVM [2,3] used RIVM's PROAST program, the background of which has not yet been publicized.*

---

* The Committee has been informed that a description of the software is in the preparatory stages. Further information can be obtained from Professor W. Slob, RIVM, Bilthoven, The Netherlands.

4.3 Toxicity parameters for modelling

In a 90-day toxicity study with rodents (usually rats) conducted in accordance with FDA* or OECD guidelines, more than one hundred parameters are measured. They include body weight, consumption of feed and water, symptoms and behavioural abnormalities, neurofunctional tests, haematological and clinical chemical parameters, composition of the urine, organ weights, and abnormalities in body cells and body tissue. The Committee would prefer modelling of the dose-effect or the dose-response relationship for all parameters about which information is obtained in a toxicity study. Exposure to the substance in question in the experimental dose range often has no effect, or hardly any, on a sizeable group of parameters. No BMD can then be established for those parameters.

4.4 Software

The successful use of the BMD method depends entirely on the reliability of the software. Where the method allows for further development, for example in the selection of model functions, in the choice of statistical criteria for fitting the model functions to research data and in the establishment of the confidence interval of the BMD, and in methods for anomalous datasets, the software will have to be kept up to date. The essential requirements relate of course to the statistical quality of the programme modules. Validation is important here. The Committee advocates following the example of the US EPA: the software should be made freely available with a view to encouraging the development of the method.

* FDA - Food and Drug Administration of the US.

Chapter 5

Benchmark response and critical effect size

In Chapter 3, the Committee identified the BMR and the CES as the fundamental parameters for the use of the BMD method. At this point, it will discuss these parameters in slightly greater detail and, where possible, make recommendations for determining them.

5.1 BMR

Values of 1, 5 or 10 per cent above the response in the reference group are generally chosen for the BMR.[20]* Furthermore, the selected value, and therefore the BMD, usually have to be within the range of the available data. It is also important for the result not to be very dependent on the function selected for the modelling of the dose-effect or dose-response relationship (see 4.2).[22]

Slob has advocated the use of a BMR value equal to 50 per cent, at least when using continuous data from animal experiments.[25,30] His argument is that the distribution of the results of a toxicity study at a given dose provides no information about the distribution, in the population requiring protection, of the effects of exposure to the substance in question. The former distribution is, at least in part, the consequence of unintended variations in experimental conditions and genetic differences between the experimental ani-

* The BMR is also defined as the additional response above the number of individuals without adverse effects in the non-exposed population.[28] If there are relatively few, or no, individuals with the harmful effect in the reference group, then both definitions are practically equivalent.

mals. With this BMR, exposure corresponding to the BMD results in an average (properly speaking, median) effect that corresponds to the critical effect size.

The Committee agrees with Slob that opting for 50 per cent for the BMR results in a value for the BMD that is relatively insensitive to unintended variations in experimental conditions. It would wish to point out that opting for 50 per cent rather than, for example, 5 per cent has implications for the choice of the uncertainty factors when determining the health-based recommended exposure limit on the basis of the BMD. A decision about a definitive preference is therefore only possible in conjunction with a discussion of steps 4 and 5 in Figure 1 or step 5 in Figure 2.

Although it is often possible to achieve a response value of 50 per cent in a toxicity study, this is not the case with the results of epidemiological studies. The requirement that the BMR should preferably be located in the range of the available data means that lower values must be used.

5.2 CES

The selection of the CES is a clear issue in the case of continuous and ordinal data (see 3.1 and 3.3). It consists of two components. First of all, there is the *measure* for the effect in question. Then we have the value of the effect thought to be 'critical' that is expressed by that measure. This is the effect considered to be the dividing line between benign and non-benign. In terms of the choice of a measure, for example influence on liver weight, possible options are the absolute weight, the relative weight compared to the average for the reference group, the absolute difference with the average for the reference group and the relative value of that difference compared to the average body weight of the reference group. Research conducted by Dekkers *et al.* shows that there is no consensus about the best way to quantify the CES.[12] Furthermore, biomedical and toxicological know-how is generally inadequate to determine what is, or what is not, harmful in the long and short terms for the functioning and the health of the organism.[33] Dekkers *et al.* therefore conclude that a default CES cannot be determined at present for all toxicological parameters.

The selection of the CES is complicated by the fact that, in a non-exposed population, there can also be effects that can be considered to be harmful, or at least abnormal. A criterion sometimes chosen for the CES during benchmark dose calculations based on epidemiological data is that there should be an abnormal effect in a certain proportion (for example 5 per cent) of the non-exposed population.[40]

Use of the BMD method is not feasible without a clear criterion for the selection of a CES for a given effect. Figure 5 contains the Committee's suggestions about how to determine a CES. However, in anticipation of further developments, its position is that a reasoned approach will be required in each case for the selection of the CES, unless

there is broad consensus about the critical effect size. This is for example the case with the bonding of carbon monoxide to haemoglobin: a blood concentration for carboxyhaemoglobin (relative to the total haemoglobin concentration) of 5% is not considered to be harmful in healthy adults.[41]

In the ideal case, the CES can be determined on the basis of biomedical and toxicological considerations. An assessment by experts will then be inevitable. In addition, some discussion between medics and toxicologists will be required for many effects before a consensus is reached since the effect parameters for which information is available will not always be directly related to a disease or handicap.

If there is little or no biomedical and toxicological information available, the Committee suggests that it would be appropriate to find out, for the toxicity parameter in question, whether there is a consensus about what is harmful or, at least, abnormal for the *non-exposed* population. If there is a consensus, this can be used to select the CES.

If there is not, the effects at a default value (for example, 5 per cent of the population) may be classified as abnormal and used as the basis for the CES. The Committee does believe that it is sensible to determine whether such a value for the CES generates observable variances in average effect in a well-designed toxicity experiment.

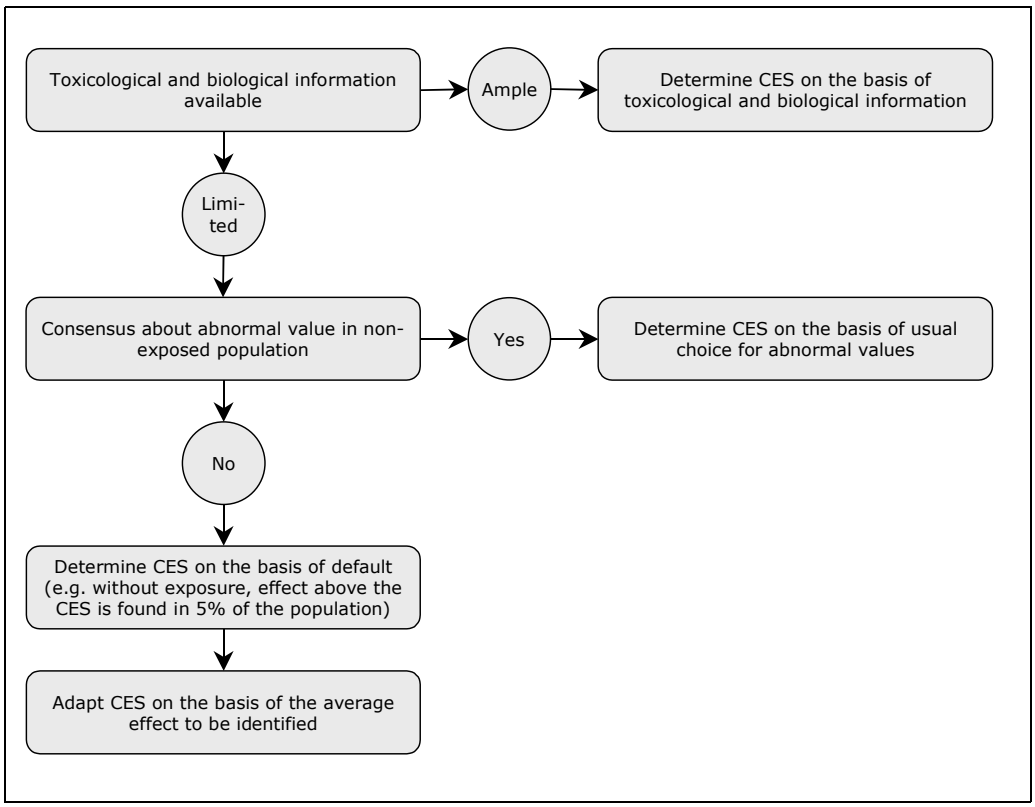

Figure 5 Flow chart for use as a checklist when determining the critical effect size.

The variance that can be observed with a given level of precision between the values for a toxicity parameter in a dose group and a reference group depends on the size of the groups and the statistical variation in the calculated variance. The larger the number of experimental animals in each dose group, the smaller the statistically significant variances that can be determined. Appendix lists the observable variances for a number of parameters in the usual values for the coefficient of variation. The Committee has also rounded off to multiples of 5 per cent the observable variances when groups of 20 experimental animals of the same gender are used. It advocates accounting for these values when determining the CES for the parameters in question (see Table 1). Similar calculations are possible for parameters not listed in the table or for a more precise breakdown according to age or gender.

The figure of 20 animals per group was taken by the Committee from the FDA protocol for a 90-day study. The studies conducted in accordance with FDA guidelines are the most 'conservative' of the usual protocol studies in terms of the size of the detectable effect: they provide the highest statistical possibility of tracing an effect variance between the non-exposed reference group and a dose group. Variances shown in this way are generally considered to be toxicologically relevant.

Table 1 Suggested approach for choosing a value for the critical effect size (CES) for a number of toxicity parameters in animal trials based on statistically observable variances in a toxicity study (see also Figure 5, bottom left, and also Annex B). Variances of this kind are generally considered toxicologically relevant.

Toxicity parameter	Proposed CES (%)
Red blood cells	5
Haemoglobin	5
Thrombocytes	10
Prothrombin time	5
White blood cells	30
Alkalic phosphatase	30
Alanine aminotransferase	25
Aspartate aminotransferase	20
Relative weight of the adrenal glands	20
Relative weight of the kidneys	10
Relative weight of the spleen	10
Relative weight of the liver	5

Chapter 6

Practical test

6.1 TNO and RIVM project

In 1999, RIVM and TNO initiated the project *Evaluation of the benchmark dose approach in practice*. The original plans included—in accordance with the proposals of the Committee submitted prior to the project—the evaluation of five different substance categories: 'new substances', 'existing substances', pesticides, substances used in the workplace and pharmaceutical compounds. In the end, financial resources were only released for the study of substances in the workplace and new substances. In 2001, the two institutions reported on the first group (substances in the workplace).[2] The report on the study of new substances was available in draft form as this advisory report was being written.[3]

The aim of this project was to evaluate the usefulness of the BMD method and to identify its possible limitations when used for:

- results from studies of substances in the workplace on the basis of which health-based recommended exposure limits had already been derived using the NOAEL method
- results from studies of 'new substances' (substances introduced to the market after 1981 in the EU).

In addition, the two institutions looked at the influence of the number of dose groups and group size on BMD quality. To do so, the toxicity study previously conducted for

Rhodorsil Silane in accordance with the usual protocols was repeated with more dose groups.[36]

The following substances were studied in the 'substances in the workplace' category: butyl acetate, captan, acetone cyanohydrin, lyorthol and white spirits. Research data from studies with reasonable numbers of dose groups was available for these sub-

Table 2 Practical applications of the BMD method and comparison with NOAEL (LOAEL) values. These examples use only the 'critical' effect parameters for the derivation of the BMD, in other words parameters in which changes emerge first as doses increase. The values of the BMR and the CES were 50% and 5% respectively (unless stated otherwise). The BMD-L is the lower 90% confidence limit of the BMD.

Substance	NOAEL mg/(kg.dag)*	LOAEL mg/(kg.dag)*	BMD-L mg/(kg.dag)*	Basis for the BMD calculation
Substances in the workplace				
Butyl acetate	2420 mg/m^3	-	3122 mg/m^3	reduced increase in body weight
Captan	-	0.13 mg/m^3	0.16 mg/m^3	extra minimal squamous-cell hyperplasia of the laryngeal epithelium
Acetone cyanohydrin	35 mg/m^3	-	Not determined	Data was inadequate for a reliable BMD
Lyorthol (a mixture of o-benzyl-p-chlorophenol—BCP and o-phenylphenol—OPP)				
BCP	-	30	53	increase in relative kidney weight
OPP	40	-	69	extra hyperplasia of the renal transitional epithelium
White spirits	-	2000 mg/m^3	1357 mg/m^3	increase relative liver weight
New substances				
NC1	50	-	2.3	increase in cholesterol level in blood plasma
NC2	30	-	4.5	increase in relative testis weight
NC3	50	-	20	increase in absolute spleen weight
NC4	100 (♂)	100 (♀)	9	increase APTT†
NC5		15	81	increase in relative liver weight
Rhodorsil Silane 198 (methyl,2-butoxy ethyl silicate).				
	50		79	increase in RBC numbers‡
	50		106	increase of HB§ level
	50		19	increase in relative liver weight
	150		53	increase in relative spleen weight

* - Dose in mg per kg body weight a day. If another unit is stated, exposure is expressed as the concentration of the substance in the air in mg/m^3.

† - APTT: activated partial thromboplastin time

‡ - RBC - red blood cells

§ - HB – haemoglobin

stances and mixtures, including effects considered to toxicologically relevant. In the 'new substances' category, five substances were selected from 501 candidates. They were, for reasons of confidentiality, encoded as 'New Chemical 1' to 'New Chemical 5' (NC 1 – NC 5).

The results of all toxicity studies were analysed again using the BMD method and software developed by RIVM (PROAST; see Regressiemodel). With regard to the continuous variables, a relative change of 5 per cent in the parameter in question, like CES and a BMR of 50 per cent, was the criterion for deriving the BMD. The categories 'mild effect' or 'minimum effect' were used as the CES for ordinal (histopathological) data.

The results of the various studies can be found in Table 2.

6.2 Substances in the workplace[2]

The usual first step in the NOAEL method for the derivation of health-based recommended exposure limits is the selection of the 'critical' effect, the effect with the lowest NOAEL value. The research data for this critical effect is used to derive a BMD. Despite the fact that the toxicity studies for each of the five substances were conducted with a limited number of dose groups, the researchers concluded that the BMD method could be used in all cases without major difficulties.

The substances studied were selected on the basis of the availability of toxicity research with changes in the most sensitive effect parameter* in as many dose groups as possible. For some parameters, effects were only observed in a single dose group. The derived BMD is, in that case, not very reliable (other dose-response models that also generate a good 'fit' with the experimental data can produce widely varying results) because the data does not allow for an accurate determination of the dose-effect relationship. In any case, under these circumstances, the BMD method is not a genuine alternative for the NOAEL method. During the analysis, any doubts arising about the accuracy of the selected model can therefore be attributed to the shortage of dose groups in which a change was observed in the toxicity parameter in question. This problem can be solved by increasing the number of dose groups. Given the wish to use as few experimental animals as possible, this can only be done with fewer experimental animals per group. However, these smaller groups do not constitute a problem in theory for the application of the BMD method (see 6.5).

It follows from Table 2 that, in so far as application of the BMD method was feasible, the value of the BMD-L did not substantially diverge from the NOAEL (or from the no adverse effect level estimated on the basis of the LOAEL). A definitive assessment of

* Parameter for the 'critical effect', the harmful effect that becomes apparent first as exposure increases. Not to be confused with critical effect size (CES).

what this means for the health-based recommended exposure limits can only be made after the selection of the uncertainty factor. The discussion of this issue is outside the scope of this advisory report. If one selects the same value for the UF in the NOAEL method and the BMD method in the present cases, the recommended exposure limits derived with the BMD method are of the same order of magnitude, and generally slightly higher than the 'usual' values.

In the case of one substance (acetone cyanohydrin), it was debatable after the model curve had been determined whether there actually was a dose-dependent effect. In this example, the fact that the BMD method accounts for all the data means that it results in a different conclusion from the NOAEL method: the second method results, on the basis of the effect in question, in a health-based recommended exposure limit but the BMD method does not.

An interesting point with white spirits is that it was possible here to determine a BMD but not an NOAEL. This fact led the DECOS Committee of the Health Council to use the LOAEL as the basis for the derivation of health-based recommended exposure limits and (at least for the time being) to apply an additional uncertainty factor of 6.[42]

6.3 New substances[3]

The BMD method also proved suitable for the 'new substances'. Here, it should be pointed out that the selection criterion meant the effects of each substance could be observed in at least two dose groups.

In the case of NC 1, the BMD was based on an increase of 5 per cent in the level of blood cholesterol. If the approach discussed in CES (see 5.2) had been followed, a higher CES could have been chosen and that would have resulted in a higher BMD and BMD-L.*

The basic assumption for the BMD of NC 2 was an increase of 5 per cent in relative testis weight. The 28-day study—that is the basis for the calculations for NC 2—did prove sensitive enough to demonstrate the presence of a difference of this kind between a dose group and the non-exposed reference group. The NOAEL determined previously is therefore, by definition, too high.

The basic assumption for the BMD of substance NC 3 was an increase of 5 per cent in absolute spleen weight. It is however questionable whether a change in the absolute spleen weight is the correct measure for the toxicity parameter in question because spleen weight is linked to the growth of the animal. In the present case, no dose effect was observed on relative spleen weight (as compared to body weight). The Committee is

* For this purpose, the calculation in Annex B for the toxicity parameter in question should be carried out with variation data from a 28-day toxicity study.

of the opinion that the researchers in this case did not fully consider the toxicological relevance of the result (see 4.2).

In the case of NC 4, the BMD is based on an increase of 5 per cent in the activated partial thromboplastin time (APTT). As with NC 2, the toxicity experiment in this case is not sensitive enough to detect a difference of this magnitude between the dose and reference groups and this results in a value for the NOAEL that is too high.

The BMD for NC 5 is based on an increase of 5 per cent in relative liver weight. However, the LOAEL (see Table 2) was based on a histopathological finding (minimal hepatocellular hypertrophy). If it is decided to take the transition from ‘minimal’ to ‘minor’ as the dividing line between benign and harmful, the BMD-L will be higher than for liver weight. However, if the transition from ‘no abnormality’ to ‘minimal’ is adopted, the BMD-will be lower than the value in Table 2 (and will be about the same as the stated LOAEL). This example demonstrates the importance of the choice of the critical effect size, especially with ordinal data where correct classification requires a lot of experience.

Table 2 shows that, given the choices made in four of the five cases (NC 1 – NC 4), the BMD-L is more than an order of magnitude lower than the NOAEL. For NC 5, the BMD-L is five times higher than the NOAEL. The Committee pointed out above that the value of the BMD-L depends on the choice of the CES and the BMR. The consequences for the health-based recommended exposure limits are awkward to quantify without making a reasoned selection of a UF.

6.4 Substances in the workplace compared to new substances

It is striking that the BMD-L is, in most cases, equal to or higher than the NOAEL in the category ‘chemical substances in the workplace’, while the BMD-L in the category 'new substances' is usually lower than the NOAEL. On the BMD side, the choice of the criti cal effect size may play a role here, as discussed by the Committee in 6.3. It happened, partly as a result of the chosen CES, that the lowest BMD was based on a different effect than the NOAEL. In addition, the number of experimental animals in the toxicity experiments will have played a role in the apparently systematic difference. The number of experimental animals in the studies of substances from the working environment was, in all cases, 20 or more per dose group, whereas the studies with new substances were all conducted with 10 experimental animals per dose group. Smaller groups may result, all else being equal, in a higher NOAEL and a lower BMD-L.

6.5 Rhodorsil-Silane[36]

A 28-day toxicity study conducted in compliance with OECD 407 guidelines with the aim of determining an NOAEL will in many cases not be adequate for determining a BMD. A study design with more dose groups is more suitable for the BMD method, because more precise dose-response and dose-effect curves can be derived. Woutersen *et al* conducted a sub-acute 28-day study with 7 groups of 10 female rats that were exposed, using a stomach tube, on a daily basis (7 days a week) for 4 weeks to Rhodorsil-Silane 198 (methyl,2-butoxy ethyl silicate) dissolved in corn oil. The doses administered were 50, 150, 300, 450, 600 and 750 mg per kg body weight a day. The reference group received corn oil only. The researchers used the resulting data to determine a BMD-L in three ways:

1 From 7 dose groups of 10 animals per group (all data: '7x10')
2 From 7 dose groups of 5 animals per group ('7x5')
3 From 4 groups of 10 animals per group, in accordance with the research design of the OECD protocol: ('4x10'). The dose groups were 0, 50, 150 and 450 mg per kg body weight a day.

The '4x10' analysis also served to derive an NOAEL; this was 50 mg per kg body weight a day for changes in red blood cells, haemoglobin level and the relative liver weight as critical effects. The BMR was 50 per cent.* A few of the results can be found in Substances in the workplace22.

As expected, the '7x10' analysis resulted in dose-effect relationships with the smallest statistical distribution for the toxicological parameters studied: red blood cells, haemoglobin level, ALAT**, ASAT***, spleen and liver weights, and spleen and liver pathology. The '4x10' analysis resulted in 'reasonable' dose-effect relationships, something that the researchers attribute to the linear form of the relationship in the dose area studied.

The criteria for the CES were: 5 per cent for red blood cells, haemoglobin and the spleen and liver weights, and 30 per cent for ALAT and ASAT. For the spleen and liver abnormalities, the transitions from 'minimal' to 'minor' (in the case of RES**** cells in the liver) or from 'minor' to 'moderate' (brown pigment and extramedullary haematopoiesis in the spleen) were considered to be the dividing line between benign and harmful. The BMD-L value for the relative liver weight is lowest in each of the

* The researchers used the PROAST software from RIVM referred to above.
** ALAT – alanine aminotransferase
*** ASAT – aspartate aminotransferase
**** RES - reticulo-endothelial system

three analyses (see Substances in the workplace22) at 27, 19 and 4 mg per kg body weight a day for, respectively, the ‘7x10’, ‘7x5’ and ‘4x10’ analyses. These values are all below the NOAEL of 50 mg/kg a day.

The researchers note that, with the NOAEL, the best estimate of the change in the relative spleen weight is more than 10 per cent, a level that can be considered harmful.

The main conclusion of this study is that the BMD method generates sound results with data from a 28-day study matching the OECD protocol if the number of dose groups is increased to 7 (instead of 4). The result remains ‘reasonable’ if the number of animals is reduced to 5 per dose group (instead of 10). It is necessary to select a dose range such that at least a few parameters show an effect of exposure in various dose groups.

Chapter 7

Conclusions

7.1 Comparison of NOAEL and BMD

A comparison of the NOAEL method and the BMD method is only meaningful to a limited extent. First of all because there is no ‘gold standard’ against which the results of the two approaches can be measured. Furthermore, even in the BMD method, the extrapolation of the BMD-L for a toxicological parameter measured in a particular experimental animal (or an item of data from research of a different kind) requires extrapolation for the human population in question (see Figure 1). The considerations that play a role here are the same for both methods but the values to be used for the uncertainty factor may be different.

First of all, the Committee wishes to make an observation about the transparency of the two methods. The strength of the NOAEL method would appear to be its conceptual simplicity. On the basis of research, one determines when no more damage to health can be observed. One then divides that level of exposure by factors in order to account for differences between the research conditions and the situation of the group of people requiring protection. The result is a ‘safe’ level of exposure: the health-based recommended exposure limit. The BMD method, on the other hand, requires the selection of apparently complex mathematical functions, the use of advanced statistics and unusual choices of response levels and critical effect sizes. Why should we replace an apparently simple approach with complex calculations?

The reason is that the results of the NOAEL method are less clear than they appear. In many cases, it is uncertain whether, in the case of the NOAEL, adverse effects are actually absent; evaluation research shows that this is generally not the case. The protection afforded by observing the health-based recommended exposure limits therefore depends mostly on the uncertainty factors being large enough. Many believe that they are but the evidence is necessarily sparse. The complexity of the BMD method reduces, at least to some extent, the uncertainty associated with the derivation of health-based recommended exposure limits. However, that complexity does require techniques that are not easy to understand. Arbitrariness can then be prevented by providing each BMD-L and recommended exposure limit with a reasoned description of the choices made when implementing the method.

The main substantial differences between the two methods are, in effect, the reason for this advisory report. First of all, this has to do with the dependence on the design of toxicity experiments. When determining the NOAEL, a smaller number of animals per group and the same series of test doses result in greater statistical uncertainty and possibly, therefore, in a higher NOAEL. This is undesirable from the point of view of health protection. With the BMD method, on the other hand, greater statistical uncertainty actually results in a lower value. In addition, the software uses all available information to determine the BMD-L as precisely as possible. The NOAEL is, strictly speaking, determined only by the group with the highest dose in which there is no statistically significant, harmful, effect compared to the reference group. The other information does play a qualitative role in the selection of uncertainty factors for the derivation of a health-based recommended exposure limit.

Some authors point to the need, with the BMD method, to determine what is harmful or benign. However, even with the NOAEL method, it is necessary to decide what effect should, and what effect should not, be considered harmful. A difference is that, with the BMD method, the dividing line between non-benign and benign has to be quantified explicitly. With the NOAEL method, on the other hand, the designation of the groups in which the observed effect is considered to be harmful can suffice (something that is generally only done on statistical grounds: is there a significant difference?).

Conclusion

The considerations in this advisory report do not detract from the benefits that the Committee saw in the BMD method.[1] On the contrary, all studies conducted in recent years, both in the Netherlands and elsewhere, show that those benefits are genuine. The Com-

mittee does recommend a further discussion and appraisal of the extrapolation of the BMD-L to a health-based recommended exposure limit (see 7.6).

7.2 Feasibility of the BMD method

The main aim of this advisory report is to arrive at a conclusion about the feasibility of the BMD method, and in particular the derivation of a benchmark dose. An important step is the adequate modelling of the dose-effect or dose-response curve.

High-quality software is indispensable for the BMD method, especially when routine application is involved. The computer programs available at present remain quite cumbersome, especially if the Committee's recommendation to subject all the parameters in a toxicity experiment to a BMD analysis is followed. The heart of the software is the modelling of the toxicity data. The Committee considers it desirable for an approach to be further developed for the systematic testing of the applicability of a linked set of models, as in the work of Slob *et al*.[23,25,30] It will be important here to avoid a situation in which the software becomes a *black box* for average users. It was pointed out in 4.2 and it has also been argued by others[37] that it is absolutely essential to assess the results of the analyses on their biological and toxicological merits. The steps sketched out in 4.2 should therefore be elaborated as guidelines (see 7.6).

In order to determine a BMD with any degree of reliability, data is generally needed from at least three, and preferably more, dose groups with parameter values that are different from the value without exposure (effects must be observable). If this condition has not been met for any parameter at all in a toxicity study, it will not be possible to determine a BMD (the selected NOAEL will, in that case, be equal to the highest test dose). If the BMD method is used anyway, this will have implications for the design of the toxicity experiments: more doses will be required than is usual at present (5 - 7), as well as a more meticulous selection of the dose distribution. The Committee is of the opinion that an adjustment of this kind need not necessarily result in an increase in the number of experimental animals. It may even be possible to reduce the use of experimental animals.* This is a separate development from the substitution of animal trials by, for example, *in vitro* experiments.[7] It should be pointed out that this type of experiment is also, in principle, suitable for determining the BMD.

Conclusion

The experience described in the literature indicates that the BMD method is, in general, practicable. The same applies to substances for which health-based recommended expo-

* According to information provided to the Committee, RIVM is currently studying this issue using computer models.

sure limits have been derived in the past with the NOAEL method and for which, therefore, the specific aim of the toxicity study was not to generate an optimal dataset for the BMD method. There are more routine uses, in particular the assessment of dossiers relating to, for example, 'new substances' and pesticides. Here, it will be desirable to optimise the structure of standardised toxicity experiments for the BMD method (for example by increasing the number of dose groups to between five and seven). On the basis of the data presented in 6.5, the Committee draws the conclusion that this can be done without any real increase in the required experiment workload. Conclusion

The formulae for the dose-effect curves are based mainly on empirical considerations, and only to a limited extent on toxicological factors. The Committee does not expect any changes in this respect in the near future. It believes that efforts should be made to bring such changes about (see also 7.6). A toxicological basis for the dose-effect curves will increase confidence in the BMD method.

Given its positive assessment of the BMD method, the Committee believes that governments, including the Dutch government, should maintain or initiate resourcing for the continued development of the method. Progress in the fields of toxicology and statistics should be reflected in more user-friendly software (see also 4.4). Development will benefit most from the free availability of the computer programs and the underlying source codes. The Dutch government could—together with its EU partners—follow the example of the US EPA.

In cases in which no reliable BMD can be derived, the NOAEL (LOAEL) method remains the best option for the derivation of health-based recommended exposure limits.

7.3 Benchmark response and critical effect size

In Chapter 5, the Committee looked in greater detail at the selection of the BMR and of the CES, crucial aspects of the BMD method. The conclusion of that chapter was that a lot of development work is still required but also that enough experience has now been acquired to make the application of the method possible.

Conclusion

Ideally, the CES should be based on toxicological considerations. Where this is not possible at present, the Committee has formulated an approach for making a practical choice (see Figure 5, Table 1 and Annex B).

A second area requiring attention is how to quantify a given effect (absolute, relative, as compared to another parameter etc.). A standard approach is not possible here. A reasonable choice will have to be made in each case, mainly on the basis of biological and toxicological considerations. Given this fact, the Committee is of the opinion that

anybody using the BMD should state explicitly why a certain CES and a certain definition of a toxicity parameter have been chosen.

The most widely-used values for the BMR are between 1 and 10 per cent. The Committee is of the opinion that those values are practicable, but that it is advisable for the BMR to be located in the range of the experimental values. This will, in animal experiments, generally be closer to 10 per cent than 1 per cent.

Depending on the available data, a BMR of 50 per cent could also be used. This value results in a BMD (and BMD-L) that depends less on experimental variation. The Committee makes no definitive proposals here for the values to be selected, but it does wish to point out that the choice of the BMR affects, in principle, the choice of the uncertainty factor used.

If the Dutch government adopts the positive assessment of the Committee about the practicability of the BMD method, it will have to agree to certain choices for the CES and the BMR (that will initially have to be made on a case-by-case basis by toxicologists and toxicological pathologists). The values of those two quantities cannot be determined entirely on scientific grounds, even though scientific considerations do play a significant role. A reasoned choice and a case-by-case assessment of whether those choices are acceptable in policy terms is an approach that is in accordance with the present state of scientific knowledge.

7.4 Estimating effects

An incidental advantage claimed for the BMD method is that the modelled dose-response and dose-effect relationships could be used to establish a quantitative understanding of the risks associated with exposure levels above the health-based recommended exposure limits.[1] In principle, this appears to be possible for exposure values within the modelled range since the data allow for an estimate of the response in the exposed population. However, the great unknown is the extrapolation of that response into a risk for the group of people requiring protection. There are difficulties associated with the use of the same uncertainty factor as for the derivation of the health-based recommended exposure limit from the BMD. First of all, this factor does not need to be independent of the dose. Furthermore, the aim of a health-based recommended exposure limit is to determine with a reasonable level of certainty that there are no health effects. At high levels of exposure, the aim will generally be more to make the best possible estimate of the nature and extent of the health effects in the population. In order to make this possible, a probabilistic approach has been proposed.[23]

Conclusion

The dose-response or dose-effect curves determined in the BMD analysis may be suitable for use when estimating health risks associated with exposure above the health-based recommended exposure limit. However, estimates of this kind are associated with so many uncertainties that they should be used with considerable caution in concrete situations in which there is exposure to hazardous substances. Given the necessity for such estimates, the Committee advises the government to support further research into the reduction of these uncertainties.

7.5 Selection of the relevant BMD

Toxicity experiments conducted in accordance with standard protocols are a particular source of information about numerous parameters. Which BMD-L value should therefore be used for the derivation of the health-based recommended exposure limits? If there are arguments in favour of the assumption that the uncertainty factor to be used is not dependent on the toxicity parameter in question, the lowest BMD-L can be used as the basis for the subsequent steps (Figure 1). If that is not the case, a recommended exposure limit will have to be derived for each BMD-L before choosing the lowest.*

Conclusion

Effective application of the BMD method requires the inclusion of all toxicity parameters about which information has been acquired in a BMD analysis. It is only after the modelling stage that it is possible to decide whether the information about a given parameter is relevant or not.

The Committee is of the opinion that it is not possible, at present, to adopt without discussion the lowest BMD-L of a series of studied toxicity parameters as the point of departure for the derivation of a health-based recommended exposure limit. Every derivation procedure will have to indicate which uncertainty factor is applied to the BMD value for a certain parameter and what the thinking is behind differences or similarities.

7.6 Further advice

In the last twenty years, considerable expertise and experience have been acquired relating to the BMD method. This is an adequate basis for a policy requiring limits in the

* In a certain sense, there will then, once again, be a loss of information. However, this is inherent to the definition of a health-based recommended exposure limit, which is based on the absence of a harmful effect.

Netherlands to be established by deriving health-based recommended exposure limits with the BMD method. It should be pointed out that government bodies outside the Netherlands, in particular the US EPA, have already taken this road. The Committee is of the opinion that it is desirable to examine, in each individual case, the options for the application of this new method. When both the NOAEL and the BMD method are practicable, it advises using both for the derivation of health-based recommended exposure limits. It will then be possible to make a reasoned choice of the recommended exposure limit to be used as the basis for limit values.

In order to eliminate the experimental nature of the BMD method, additional advice is required in three areas:

1 The choice of uncertainty factors
2 Protocol development for the BMD method
3 A framework for the selection of the BMR, CES and model function.

Uncertainty factors (1)

There is little discussion in the literature about the uncertainty factor to be used with the BMD method. Usually, it is assumed that it corresponds to the factor used with the NOAEL method. This is not necessarily the case. Furthermore, there are authors who believe that statistical modelling of the extrapolation of animal trials to human exposure in practice is preferable to a single figure as the value for an uncertainty factor. The Committee recommends an assessment by the Health Council of the value of the uncertainty factor and of the possibilities for statistical modelling. It believes that international consultations would be appropriate for this purpose.

Protocol development (2)

Existing protocols for toxicity studies are not well matched to BMD methodology, although their application does not preclude this. The Committee considers amendments to the OECD guidelines and EU regulations to be necessary.

Framework for the choice of BMD parameters (3)

Given the fact that the BMD method is still highly experimental, a reasoned choice of benchmark response, critical effect size and model function is required on a case-by-case basis. The Committee considers it desirable for this reasoning to be situated in a framework that includes the considerations determining a particular choice. It proposes the establishment of a framework of this kind, preferably in an international context, by the Health Council or another scientific body.

Literature

(1) Gezondheidsraad: Commissie Afleiding gezondheidskundige advieswaarden. Toxicologische advieswaarden voor blootstelling aan stoffen [Toxicology-based recommended exposure limits. Den Haag: Gezondheidsraad; 1996; Publicatie nr 1996/12 (in Dutch and in English).

(2) Appel MJ, Bouman HGM, Pieters MN, Slob W. Evaluation of the applicability of the Benchmark approach to existing toxicological data. Framework: Chemical compounds in the working place. Bilthoven, Zeist: Rijksinstituut voor Volksgezondheid en Milieu, TNO Voeding; 2001; RIVM Report no. 601930 001, TNO Report no. V99.1098.

(3) Janssen GB, Pieters MN, Slob W. Evaluation of the applicability of the Benchmark approach to new chemicals. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu; 2002; RIVM Report no. 601930 002 in preparation.

(4) Barnes DG, Daston GP, Evans JS, Jarabek AM, Kavlock RJ, Kimmel CA. Benchmark dose workshop: criteria for use of a benchmark dose to estimate a reference dose. Regulat Toxicol Pharmacol 1995; 21(2): 296-306.

(5) Benchmark Dose Software (Help Manual for Benchmark Dose Software 1.3). Washington, DC: US Environmental Protection Agency, National Center for Environmental Assessment; 2002; Report EPA 600/ R-00/014F. Internet: http://cfpub.epa.gov/ncea/cfm/bmds.cfm, consulted 21-8-2002.

(6) Gezondheidsraad. Blootstelling aan combinaties van stoffen: een systematiek voor het beoordelen van gezondheidsrisico's [Exposure to combinations of substances: a system for assessing health risks]. Den Haag: Gezondheidsraad; 2002; Publicatie nr 2002/05 (in Dutch and in English).. Internet: www.gr.nl.

(7) Gezondheidsraad. Onderzoek gezondheidsrisico's stoffen: een gerichtere benadering. Den Haag: Gezondheidsraad; 2001; Publicatie nr 2001/24. Internet: www.gr.nl .

(8) Principles for the assessment of risks to human health from exposure to chemicals. Geneva: World Health Organization, International Programme on Chemical Safety; 1999; Environmental health criteria 210. Internet: http://www.who.int/pcs/risk-assessment-ehc/docs/ehc210_exposure.htm, consulted 22-8-2002.

(9) Gezondheidsraad: Commissie Uitgangspunten voor normstelling. Uitgangspunten voor normstelling. De inzichtelijke opbouw van advieswaarden voor niet-mutagene, niet-carcinogene en niet- immunotoxische stoffen. Den Haag: Gezondheidsraad; 1985; Publicatie nr 1985/31.

(10) Gezondheidsraad: Commissie Beoordeling carcinogeniteit van stoffen. De beoordeling van carcinogeniteit van chemische stoffen. Den Haag: Gezondheidsraad; 1996; Publication nr 1996/26.

(11) Crump KS. A new method for determining allowable daily intakes. Fundam Appl Toxicol 1984; 4(5): 854-871.

(12) Dekkers S, de Heer C, Rennen MAJ. Critical effect sizes in toxicological risk assessment: a comprehensive and critical evaluation. Environ Toxicol Pharmacol 2001; 10(1-2): 33-52.

(13) Faustman EM. Review of noncancer risk assessment: Applications of Benchmark Dose methods. Document prepared for the Commission on Risk Assessment and Risk Management. Seattle, WA: University of Washington; 1996. Internet: http://www.riskworld.com/nreports/1996/risk_rpt/html/nr6aa051.htm, consulted 23-7-2002.

(14) Leisenring W, Ryan L. Statistical properties of the NOAEL. Regulat Toxicol Pharmacol 1992; 15(2 Pt 1): 161-171.

(15) Crump KS, Allen B, Faustman EM. The use of the benchmark dose approach in health risk assessment. Washington DC: Risk Assessment Forum, US Environmental Protection Agency; 1995; Report EPA/630/R-94/007. Internet: http://www.epa.gov/clariton/clhtml/pubtitle.html, consulted 26-5-2002.

(16) Crump KS. Calculation of benchmark doses from continuous data. Risk Anal 1995; 15(1): 79-89.

(17) Gaylor DW. Quantalization of continuous data for benchmark dose estimation. Regul Toxicol Pharmacol 1996; 24(3): 246-250.

(18) Clewell HJ, Gentry PR, Gearhart JM. Investigation of the potential impact of benchmark dose and pharmacokinetic modeling in noncancer risk assessment. J Toxicol Environ Health 1997; 52: 475-515.

(19) Allen B, Gentry R, Shipp A, Van Landingham C. Calculation of benchmark doses for reproductive and developmental toxicity observed after exposure to isopropanol. Regul Toxicol Pharmacol 1998; 28(1): 38-44.

(20) Gaylor D, Ryan L, Krewski D, Zhu Y. Procedures for calculating benchmark doses for health risk assessment . Regul Toxicol Pharmacol 1998; 28(2): 150-164.

(21) Kalliomaa K, Haag-Gronlund M, Victorin K. A new model function for continuous data sets in health risk assessment of chemicals using the benchmark dose concept. Regul Toxicol Pharmacol 1998; 27(2): 98-107.

(22) Murrell JA, Portier CJ, Morris RW. Characterizing dose-response I: Critical assessment of the benchmark dose concept. Risk Anal 1998; 18(1): 13-26.

(23) Slob W, Pieters MN. A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. Risk Anal 1998; 18(6): 787-798.

(24) Fowles JR, Alexeeff GV, Dodge D. The use of benchmark dose methodology with acute inhalation lethality data. Regul Toxicol Pharmacol 1999; 29(3): 262-278.

(25) Slob W. Thresholds in toxicology and risk assessment. Int J Toxicol 1999; 18(4): 259-268.
(26) West RW, Kodell RL. A Comparison of Methods of Benchmark-Dose Estimation for Continuous Response Data. Risk Anal 1999; 19(3): 453-459. privé
(27) Budtz-Jørgensen E, Keiding N, Grandjean P. Benchmark dose calculation from epidemiological data. Biometrics 2001; 57(3): 698-706.
(28) Crump K. Critical issues in benchmark calculations from continuous data. Crit Rev Toxicol 2002; 32(3): 133-153.
(29) Sand S, Filipsson AF, Victorin K. Evaluation of the benchmark dose method for dichotomous data: model dependence and model selection. Regul Toxicol Pharmacol 2002; 36(2): 184-197.
(30) Slob W. Dose-response modeling of continuous endpoints. Toxicol Sci 2002; 66(2): 298-312.
(31) Slob W, Pieters MN. A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. Bilthoven: National Institute for Public Health and the Environment; 1997; RIVM Rapport 620110005.
(32) Vermeire TG, Stevenson H, Pieters MN, Rennen M, Slob W, Hakkert BC. Assessment factors for human health risk assessment: a discussion paper. Zeist, Bilthoven: TNO Nutrition and Food Research, National Institute for Public Health and the Environment; 1998; TNO report no.V97.880; RIVM rapport 620110007.
(33) Vermeire T, Stevenson H, Pieters MN, Rennen M, Slob W, Hakkert BC. Assessment factors for human health risk assessment: a discussion paper. Crit Rev Toxicol 1999; 29(5): 439-490.
(34) Piersma AH, Verhoef A, te Biesebeek JD, Pieters MN, Slob W. Developmental toxicity of butyl benzyl phthalate in the rat using a multiple dose study design. Reprod Toxicol 2000; 14(5): 417-425.
(35) Vermeire T, Pieters M, Rennen M, Bos P. Probabilistic assessment factors for human health risk assessment - A practical guide. Bilthoven/Zeist: National Institue for Public Health and the Environment/TNO Nutrition and Food Research; 2001; RIVM Rapport 601516005, TNO-report V3489. Internet: http://www.rivm.nl/bibliotheek/rapporten/601516005.html, consulted 26-8-2002.
(36) Woutersen RA, Jonker D, Stevenson H, te Biesebeek JD, Slob W. The benchmark approach applied to a 28-day toxicity study with Rhodorsil Silane in rats. the impact of increasing the number of dose groups. Food Chem Toxicol 2001; 39(7): 697-707.
(37) Krewski D, Zhu Y, Fung KY. Benchmark doses for developmental toxicants. Inhal Toxicol 1999; 11(6-7): 579-591.
(38) Foster PM, Auton TR. Application of benchmark dose risk assessment methodology to developmental toxicity: an industrial view. Toxicol Lett 1995; 82-83: 555-559.
(39) Gephart LA, Salminen WF, Nicolich MJ, Pelekis M. Evaluation of subchronic toxicity data using the benchmark dose approach. Regul Toxicol Pharmacol 2001; 33(1): 37-59.
(40) Budtz-Jørgensen E, Keiding N, Grandjean P, Weihe P. Statistical Methods for the Evaluation of Health Effects of Prenatal Mercury Exposure. Copenhagen, Denmark: Department of Biostatistics, University of Copenhagen; 2002. Internet: http://www.pubhealth.ku.dk/bsa/publ-e.htm, consulted 9-2-2003.
(41) Dutch expert committee on occupational standards. Health-based recommended occupational exposure limit for Carbon Monoxide. Den Haag: SDU Uitgeverij; 1993; Report RA 7/92.

(42) Dutch expert committee on occupational standards. Health-based recommended exposure limit for White spirits. The Hague: Health Council of the Netherlands; 1995; Draft report for public comment nom 95-302-4 (final report postponed).

(43) Beynen AC, van Montfoort MAJ. Ontwerp van dierproeven. In: van Zutphen LFM, Baumans V, Beynen AC, editors. Proefdieren en dierproeven. Utrecht: Wetenschappelijke Uitgeverij Bunge; 1991. P. 199-221.

Abbreviations

ALAT	alanine aminotransferase
APTT	activated partial thromboplastin time
ASAT	aspartate aminotransferase
BMD	benchmark dose
BMD-L	lower 90 or 95 per cent confidence interval for the benchmark dose
BMR	benchmark response
CES	critical effect size
FDA	Food and Drug Administration of the United States
HB	haemoglobin
LOAEL	Lowest Observed Adverse Effect Level
MAC	maximum accepted concentration for occupational exposure via the air to a substance
NOAEL	no Obserd Adverse Effect Level
OESO	Organisation for Economic Cooperation and Development
RBC	red blood cells
RES	reticulo-endothelial system
RfC	reference Concentration
RfD	reference Dose
RIVM	National Institute for Public Health and the Environment
TNO	Netherlands Organization for Applied Scientific research

UF uncertainty factor

US-EPA Environmental Protection Agency of the United States

Annexes

Annnex **A**

The Committee

The membership of the Committee on the Derivation of Health Based Recommend Exposure Limits, that prepared the present report, was:

- Dr WRF Notten, *chairman*
  toxicologist, TNO Prevention and Health, Leiden
- Dr WFJPM ten Berge
  toxicologist, DSM, Heerlen
- Dr BJ Blaauboer
  toxicologist, Institute for Risk Assessment Sciences, Utrecht University
- Dr VJ Feron
  professor emeritus of biological toxicology, Utrecht University
- Dr Ir PHM Lohman
  professor emeritus of Radiation Genetics and Chemical Mutagenesis, University of Leiden
- Dr ir G de Mik
  toxicologist, National Institute of Public Health and the Environment, Bilthoven
- Ms Dr ir MN Pieters
  toxicologist, National Institute of Public Health and the Environment, Bilthoven
- Dr GMH Swaen
  epidemiologist, Universiteit Maastricht
- Dr RA Woutersen
  toxicologist and toxicological pathologist, TNO Food and Nutrition Research Institute, Zeist

- Dr JA van Zorge, *advisor*
  environmental health expert, Ministry of Housing, Spatial Planning, and the Environment, The Hague
- Ms Dr Ir PW van Vliet, *scientific secretary*
  toxicologist, Health Council of the Netherlands, The Hague
- Dr WF Passchier, *scientific secretary for the present report*
  physical-chemist, Health Council of the Netherlands, The Hague and special professor of risk analysis, Universiteit Maastricht.

Draft versions of the report were prepared by Dr MJ Appel, toxicologist at TNO Food and Nutrition Research Institute, Zeist. Dr W Slob, statistician (National Institute of Public Health and the Environment, Bilthoven and professor of quantitative risk assessment, Utrecht University) and Dr PJ Boogaard, toxicologist (Shell International BV, Health Services, The Hague) provided the committee with valuable advice.

Ms Way Yee Lee and Ms MFC van Kan provided administrative support.
Lay-out Ms M Javanmardi and Ms AMC van Kan

Annnex **B**

Statistical calculation of observable effect sizes

In 5.2, the Committee proposed establishing, during the determination of critical effect sizes (CES), which differences in the results of toxicity studies are statistically observable (see also 5 and 1). Here, it takes a closer look at the calculation of the values in Table 1.

The statistically observable relative variance between the reference group and a dose group depends on statistical power, the size of the group and the relative standard deviation from the mean for the toxicological parameter (RSD), also known as the coefficient of variance. The RSD is different for each toxicological parameter and can be determined using values from experiments conducted previously. A formula that approximates the link between these three quantities is:[43]

$$n = 2 \times \left[(z_{\alpha/2} - z_{\pi}) \times \frac{RSD}{100} \times \frac{\mu_1}{\mu_1 - 100} \right]^2$$

where:

- n - group size
- z - z score (variance with the mean divided by the standard deviation)
- α - level of significance (here: 5%)
- π - statistical power (here: 90%)
- RSD - relative standard deviation
- μ_1 - the observable relative effect + 100 (%)

This formula allows us to estimate the relative effect μ_1 that can be shown to be statistically significant (α=5%) at a certain statistical power (90%) and a certain group size *n*. In 3, the resulting values are shown for a number of toxicity parameters and different group sizes. The group sizes selected are values for normal study protocols (*n* equal to 5 per gender is usual in a 28-day study, *n* equal to 10 per gender is usual in a 90-day study conducted in accordance with OECD protocols and *n* equal to 20 per gender is usual in a 90-day study conducted in accordance with FDA protocols). The RSD for each parameter is derived from previous toxicity experiments with 20 rats per gender per group.

Table 3 illustrates the influence of the group size on the size of the effect that is statistically detectable. Ninety-day studies conducted in accordance with the FDA guidelines use 20 rats per gender per group. The studies conducted in accordance with FDA guidelines are the most 'conservative' of the usual protocol studies in terms of the size of the detectable effect: they are most precise in terms of the variance between the non-exposed reference group and a dose group. Variances shown in this way are generally considered to be toxicologically relevant. The Committee is therefore of the opinion that they can be included in the grounds used to determine a CES.

The data in Table 3 relate to one type of animal (rat), both sexes and one type of study (90-day study). The RSD for the various parameters will be different for each species and sex, and possibly also for each age group. That can result in a more detailed differentiation for the CES for each species, gender and age group.

Table 3 Relative effect (%) that can be determined for a number of effect parameters at a statistical power π = 90%, a level of significance α = 5% and various group sizes (*n* = 5, 10 or 20 per gender per group). The final column contains a proposal from the Committee for a value that can be used for the determination of the CES.

Effectparameter	RSD (%)	*n* = 5	*n* = 10	*n* = 20	Value to be used in selection of CES (%)
		Relative effect (%) that can be determined			
RBC	4,2	9,4	6,5	4,5	5
HB	5,0	11,4	7,8	5,4	5
Trombocytes	9,5	24,2	16,0	10,8	10
PTT	5,9	13,8	9,4	6,4	5
WBC	21,5	79,5	45,6	28,4	30
ALP	21,0	75,6	43,8	27,4	30
ALAT	18,6	61,6	36,9	23,6	25
ASAT	16,6	51,6	31,7	20,5	20
Relatieve gewicht van de bijnieren	14,3	41,5	26,2	17,2	20
Relatieve gewicht van de nieren	7,3	17,6	11,8	8,1	10
Relatieve gewicht van de milt	10,2	26,4	18,2	11,7	10
Relatieve gewicht van de lever	6,1	14,3	9,7	6,7	5

RBC - red blood cells; HB - hemoglobin; PTT - prothrombin time; WBC - white blood cells; ALP - alkalic phosphatase; ALAT - alanine aminotransferase; ASAT - aspartate aminotransferase.