NATIONAL INSTITUTE OF PUBLIC HEALTH AND THE ENVIRONMENT
BILTHOVEN, THE NETHERLANDS

Report no. 620110 005

A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework

W. Slob, M.N. Pieters

July 1997

This study was conducted by order of the Directorate-General for Environmental Protection, Directorate for Chemicals, Safety and Radiation Protection, within the framework of project number 620110, Human Effect Assessment I.

National Institute of Public Health and the Environment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands, tel: 030-2749111, fax: 030-2742971

MAILING LIST

1. Directie SVS/S
2. Plv. Directeur-Generaal Milieubeheer, Dr Ir B.C.J. Zoeteman
3. Dr A.G.J. Sedee, DGM
4. Dr D.W.G. Jung, DGM, SVS/S
5. Dr J. van Zorge, DGM, SVS/S
6. Prof. J.J. Sixma, Gezondheidsraad
7. Drs C.A.J. Denneman, DGM/BO/WK
8. Dr Ir P.C. Bragt, VWS, HIGB
9. Dr Ir G. Kleter, VWS, HIGB
10. Dr H. Roelfzema, VWS, Gezondheidsbeleid/Consumentenveiligheid en omgevingsrisico's
11. Drs P.R.M. Kerklaan, SoZaWe, ARBO/AMIL
12. Dr P.C. Noordam, SoZaWe, ARBO/AMIL
13. Ing H.M.A. Géron, SoZaWe, ARBO/AMIL
14. Depot Nederlandse Publikaties en Nederlandse Bibliografie
15. Dr M.L. Dourson, Toxicology Excellence for Risk Assessment, 4303 Hamilton Avenue, Cincinnati, Ohio 45223, USA
16. Dr B. Meek, Room 217, Environmental Health Centre, Tunney's Pasture, Ottawa Ont. K1A 0L2, Canada
17. Dr J. Gift, US-EPA (MD-52), Research Triangle Park, NC 27711, USA
18. Dr R. Hertzberg, Waste Management, US-EPA-Region 4, 100 Alabama Street SW, Atlanta GA 30303-3104, USA
19. Dr B. Tesh, ChemRisk, 1685 Congress Street, Portland Maine 04102, USA
20. Dr P.S. Price, 129 Oakhurst Road, CapeElizabeth ME 07107, USA
21. Dr P.E.T. Douben, Royal Commission on Environmental Pollution UK, Room 648, Church House, Great Smith Street, London SW1P 3BZ
22. Dr R. Kodell, FDA/NCTR, 3900 NCTR Road, Jefferson AR 72079, USA
23. Dr E.M. Faustman, University of Washington,USA, Dep. Environm. Health, 4225 Roosevelt Way NE, #100, Seattle, Washington 98105-6099, USA
24. Dr A.M. Jarabek, US-EPA, Natl Ctr for Env Assessment, Research Triangle Park, NC 27711, USA
25. Dr. S.H. Moolgavkar, Fred Hutchinson Cancer Research Center, Public Health Sciences Division, MP-665, 1124 Columbia Street, USA
26. Prof Dr M. Jakubowski, Nofer Institute of Occupational Medicine, 8 Teresy St., P.O. Box 199, 90-950 Lodz, Poland
27. Dr H.J. Clewell, ICF Kaiser Engineers, Inc., 602 East Georgia Avenue, Ruston, LA 71270, USA
28. Prof. M. Razzaghi, Bloomsburg University, Dep. of Mathematics, Bloomsburg, PA 17815, USA
29. G. Kowalczyk, Dames And Moore, Blackfriars House, St. Mary's Parsonage, Manchester, M3 2JA, UK
30. Dr. D.A. Belluck, Minnesota Pollution Control Agency, 520 Lafayette Road, St. Paul, MN 55155-3898
31. Anne-Marie Lafortune, Ministhre de l'Environnement et de la Faune, 360 Rue Frauet, Bureau 40, Sainte-Foy, Quibec, Canada

32. Dr J. Katoh, Mitsubishi Chemical Safety Institute LTD, 1000 Kamoshida-Cho, Aoba-Ku, Yokohama, 27 Japan
33. Drs J.H.M. Schobben, RWS-RIKZ, PO Box 20907, 2500 EX Den Haag
34. Prof dr S.A.L.M. Kooijman, Theoretische Biologie, VU, Amsterdam
35. Dr F.X.R. van Leeuwen, WHO, Bilthoven
36. Dr W.F. Passchier, Gezondheidsraad
37. Mw Dr ir P.W. van Vliet, Gezondheidsraad
38. Ir C. Hoeksema, Gezondheidsraad
39. Dr C. Lally, N.V. Procter & Gamble ETC, Strombeek-Bever, België
40. Dr J Bedaux, Theoretische Biologie, VU, Amsterdam
41. Dr D.J.J. Heederik, LUW
42. Prof. Dr P. Th. Henderson, Rijksuniversiteit Limburg, Maastricht
43. Dr G. de Jong, Shell International Petroleum Maatschappij, Den Haag
44. Dr W.F. Tordoir, Shell International Petroleum Maatschappij, Den Haag
45. Dr J. Molier-Bloot, Academisch Medisch Centrum (AMC), Amsterdam
46. Dr J.J.A.M. Brokamp, Sociaal Economische Raad, Den Haag
47. Dr T. Smid, KLM Health Safety and Environment, Schiphol
48. Dr H.G. Verschuuren, Dow Europe, Horgen (Zwitserland)
49. Dr G.M.H. Swaen, Rijksuniversiteit Limburg, Maastricht
50. Dr A.A.E. Wibowo, Coronel Laboratorium, Amsterdam
51. Dr F. de Wit, Arbeidsinspectie, Den Haag
52. Dr A.C. Aten (RCO), Willibrorduslaan 70, 5552 HE Valkenswaard
53. Dr V.J. Feron, TNO-Voeding, Zeist
54. Dr Ir J.P. Groten, TNO-Voeding, Zeist
55. Dr Ir B.C. Hakkert, TNO-Voeding, Zeist
56. Ir P.M.J. Bos, TNO-Voeding, Zeist
57. Ir H. Stevenson, TNO-Voeding, Zeist
58. Dr W. Seinen, RITOX, Utrecht
59. Dr J. Hermens, RITOX, Utrecht
60. Ir H.J. Kramer, RITOX, Utrecht
61. Dr G.H.E. Nieuwdorp, Universiteit Utrecht
62. Dr W. ten Berge, DSM
63. Directie RIVM
64. SBD/Voorlichting & Public Relations
65. Prof Dr Ir D. Kromhout, SB2
66. Dr Ir A.H. Havelaar, MGB
67. Dr P.F.M. Teunis, MGB
68. Dr E. Lebret, CCM
69. Drs A.E.M. de Hollander, CCM
70. Dr P.G.N. Kramers, VTV
71. Ir F. Langeweg, SB5
72. Drs A. van der Giessen, LBG
73. Ir R. van den Berg, LBG
74. Dr F. Swartjes, LBG
75. Drs T. Aldenberg, LWD
76. Ir A.H.M. Bresser, LWD
77. Drs A.C.M. de Nijs, LWD
78. Dr W. Sloof, MNV
79. Dr J.A. Hoekstra, MNV

80. Dr Ir G. de Mik, SB4
81. Dr W.H. Könemann, hoofd CSR
82. Prof Dr C.J. van Leeuwen, CSR
83. Dr G.J.A. Speijers, CSR
84. Ing P.J.C.M. Janssen, CSR
85. Drs A.G.A.C. Knaap, CSR
86. Dr W.C. Mennes, CSR
87. Drs T. Vermeire, CSR
88. Dr J. de Bruijn, CSR
89. Drs J.H. Canton, ECO
90. Dr O. Klepper, ECO
91. Drs D.T. Jager, ECO
92. Dr J. Notenboom, ECO
93. Drs. T.P. Traas, ECO
94. Dr. H.P. Leenhouts, LSO
95. Dr H. Slaper, LSO
96. Dr M.J.M. Pruppers, LSO
97. Dr J.W. van der Laan, LGM
98. Dr ir H.J.G.M. Derks, hoofd LBO
99. Dr M.P. van Veen, LBO
100. Dr S.H. Heisterkamp, LBO
101. Prof. Dr J.G. Vos, hoofd LPI
102. Dr R.B. Beems, LPI
103. Dr H. van Loveren, LPI
104. Dr P.W. Wester, LPI
105. Dr W.H. de Jong, LPI
106. Dr J. Garssen, LPI
107. Dr A. Opperhuizen, Hoofd LEO
108. Dr G. Mohn, LEO
109. Dr C. van Kreijl, LEO
110. Dr. E.D. Kroese, LEO
111. Dr L. van Bree, LEO
112. Dr Ir E.H.J.M. Jansen, LEO
113. Dr Ir M.J. Zeilmaker, LEO
114. Dr A. Piersma, LEO
115. M.P. Schaddelee, LEO
116. S. Groebe, LEO
117. Dr. H.F.R. Reijnders, LBG
118-119. Auteurs
120. Bureau Rapportenregistratie
121. SBD/Voorlichting & Public Relations
122. Bibliotheek RIVM
123-173. Bureau Rapportenbeheer

CONTENTS

ABSTRACT

The standard method for deriving a safe or acceptable intake or exposure limits for humans such as the acceptable daily intake (ADI), tolerable daily intake (TDI) or Reference Dose (RfD) ignores the uncertainty in the No-observed-adverse-effect level (NOAEL) as an estimate of the "true" no-adverse-effect level in the animal. The adjustment of this level by multiplication of uncertainty factors, on the other hand, may be viewed as a conservative method of taking all uncertainties into account. In this paper a general framework is discussed that quantifies both the uncertainties in the estimated no-effect level in the animal and the uncertainties in the various extrapolation steps. The resulting distribution reflects the overall uncertainty associated with the no-adverse-effect level in the sensitive human subpopulation. A lower percentile of this distribution may be regarded as an acceptable level (ADI, TDI, RfD) that takes account of the various uncertainties involved in a nonconservative fashion. The same methodology may also be used as a tool to derive a distribution for possible human health effects at a given exposure level.

SAMENVATTING

De standaardmethode voor het afleiden van een 'veilig'of acceptabel inname niveau zoals de ADI (acceptable daily intake), TDI (tolerable daily intake) of RfD (reference dose) houdt geen rekening gehouden met onzekerheden aanwezig in de schatting van het 'no-adverse-effect' niveau. Het gebruik van het produkt van onzekerheidsfactoren bij de afleiding van de ADI, leidt tot een opeenstapeling van 'worst case' aannames. Dit laatste introduceert een conservatief element in de normstelling. Hoewel voor de preventieve risicoschatting (de normstelling) het conservatieve karakter geen probleem hoeft te vormen, staat het ontbreken van inzicht in de onzekerheden rond de gestelde norm een accurate actuele risicoschatting in de weg. In dit rapport wordt een algemeen raamwerk besproken waarin alle onzekerheden zo goed mogelijk worden gekwantificeerd en vervolgens door een waarschijnlijkheidsbenadering (probabilistische benadering) worden gecombineerd. Dit resulteert in een onzekerheidsverdeling voor het "no-adverse-effect" niveau in de gevoelige humane subpopulatie. Deze benadering kan worden gebruikt voor het afleiden van normen door b.v. het eerste percentiel van de verdeling te definiëren als een ADI. Bovendien kan hetzelfde probabilistische raamwerk gebruikt worden om een onzekerheidsverdeling af te leiden voor mogelijke gezondheidseffecten bij de mens voor een gegeven blootstellingsniveau.

Enkele belangrijke voordelen van deze probabilistische benadering zijn de volgende:
1. Er wordt rekening gehouden met de van de kwaliteit van de dierstudie afhangende onzekerheden in de schatting van het "no-adverse-effect" niveau in het dier.
2. Het conservatieve karakter van de standaardbenadering, namelijk het met elkaar vermenigvuldigen van veiligheidsfactoren, vervalt.
3. Er ontstaat de mogelijkheid om de routinematige afleiding van een ADI in de toekomst beter te onderbouwen, door de vooralsnog gepostuleerde onzekerheidsverdelingen voor de diverse extrapolatiestappen te baseren op data, b.v. uit historische gegevensbestanden.
4. Het vergelijken van de verschillende onzekerheden die een rol spelen maakt een beter gefundeerde uitspraak mogelijk over de noodzakelijke omvang van een dierstudie.

SUMMARY

The standard method for deriving an acceptable intake or exposure limit for humans such as the acceptable daily intake (ADI), tolerable daily intake (TDI) or Reference Dose (RfD) ignores uncertainties associated with the estimation of the 'true' no-adverse-effect level. Multiplying uncertainty factors with each other may be viewed as a piling-up of worst case assumptions, and therefore introduces a conservative element in the standard setting procedures. This conservative character may not be unwanted for deriving exposure limits. However, the degree of conservatism in the RfD in any particular assessment is unknown, which hampers risk managers to appraise possible health risks against other (e.g. economic) interests.

This report discusses a general framework in which all uncertainties are quantified in the best possible way, and then combined in a probabilistic approach. This results in an uncertainty distribution for the no-adverse-effect in the sensitive human subpopulation. The probabilistic approach may be used for the derivation of exposure limits, e.g., by defining the distribution's first percentile as an RfD. In addition, the same probabilistic framework can be used to derive a uncertainty distribution for possible human health risks at any given exposure level.

The following are some important advantages of this probabilistic approach.

1. It takes into account the uncertainties in the estimated no-adverse-effect level in the animal, depending on the quality of the particular studied from which the data are used,
2. It has lost the conservative character of the standard approach, i.e., the multiplication of uncertainty factors,
3. It gives rise to the perspective of an improved routine method for deriving RfDs in the future, by basing the as yet postulated uncertainty distributions for the various extrapolation steps on real data, e.g., from historical files,
4. It offers the possibility to compare the various uncertainties involved in a typical risk assessment, so that a better funded decision can be made on the required size of toxicological studies.

1. INTRODUCTION

Basically, the standard method for deriving acceptable human limit values, such as RfD, ADI, or TDI, from animal study data consists of two steps. First, from the available data the (highest) dose level is determined that did not result in statistically significant adverse effects in the animal (in any of the appropriate studies available). This dose level, or No-Observed-Adverse-Effect-Level (NOAEL) is considered as the dose that an animal can tolerate without suffering from adverse effects. The NOAEL must then be extrapolated to the situation of human exposure, in particular considering specific sensitive subpopulations. Therefore, the NOAEL is usually divided by an uncertainty factor of ten to account for possible differences between human and animal, and by another uncertainty factor of ten to account for differences in sensitivity between humans. In addition to these uncertainty factors, other factors may be applied, e.g. for the extrapolation from subchronic to chronic exposure if only data from subchronic studies are available, or from LOAEL to NOAEL if all dose levels applied show significant effects compared to the controls.

Because of the application of various uncertainty factors that are multiplied with each other, the standard method for deriving acceptable human limit values is generally considered to be conservative. Indeed, when each individual uncertainty factor by itself is regarded to reflect a worst case situation, their product, i.e. the overall uncertainty factor, will tend to be overly conservative.

While the use of these factors is meant to cover all extrapolation steps and uncertainties that appear relevant in a particular assessment, the uncertainty in the value of the NOAEL is typically ignored. It should be kept in mind, however, that the NOAEL is only an estimate of the 'true' no-adverse-effect-level (NAEL) in the animal, and that it might be a quite poor estimate of it, depending on study design (dose levels, sample size) and experimental variation. In typical toxicological studies this uncertainty may be substantial (see, e.g., Leisenring and Ryan, 1992), and ignoring it introduces an anti-conservative element in the derivation of acceptable exposure limits.

In the case of standard setting procedures (preventive risk assessment), the conservative character implied by multiplying uncertainty factors may be acceptable. After all, there is always the possibility that, say, the intraspecies variation in fact exceeds a factor of 10 for a particular compound. In such cases the conservatism in the multiplication of safety factors may act as a buffer and still result in protective exposure limits. However, problems arise when exposure limits are actually exceeded or when the costs for realizing the limits are high. In those cases the complete lack of knowledge on the *degree* of conservatism involved impairs a rational weighing of possible health risks against other interests.

In short, the current approach of deriving an acceptable exposure limit (ADI, TDI, RfD) has as its major limitation that it results in a single, uncertain value, while the magnitude of that uncertainty cannot be quantified. In this report we discuss a probabilistic approach that aims to estimate the safe human exposure level in the form of a distribution expressing the uncertainty in that estimate. In this way an exposure limit can be derived from an *a priori* desired level of conservatism. Further, the same probabilistic approach can be used to estimate the possible health effects in the

sensitive human population at a given actual exposure, again in the form of an uncertainty distribution. In this way plausible lower and upper bounds for human health risks can be quantified for situations that exposure limits are exceeded. We present this methodology as a general framework, in which particular choices are still open for discussion. Our main goal is to open that discussion, and to show what kind of data and toxicological insights are needed to further improve the risk assessment process.

2. GENERAL APPROACH

In the standard method acceptable intake or exposure limits are obtained by dividing the NOAEL by a number of uncertainty factors:

$$ADI, TDI, RfD = \frac{NOAEL}{UF_1 \; x \; UF_2 \; x \; UF_3 .....} \qquad (1)$$

In other words, the RfD (ADI,TDI) is defined in a purely operational way, i.e., in terms of how to assess its value. This operational definition implies that it results in a single value, the quality (uncertainty) of which cannot be quantified. We propose an alternative approach for defining the RfD that does not have this limitation. Instead of deriving a single value we aim for a range of values, or a set of lower and upper margins, that reflect the uncertainties in our scientific knowledge and in the available data. In this way an RfD can be based on a desired level of conservatism.

To achieve that goal we first need to make an explicit distinction that is quite common in science: that between the definition of a concept or entity on the one hand, and how to assess or estimate its value from available information on the other. Thus, the RfD (ADI, TDI) should be defined on two levels: (i) on the conceptual level, say, as the (maximum) dose that does not lead to adverse health effects in the sensitive human population, and (ii) on the operational level by specifying how to estimate that dose using animal or epidemiological data and using extrapolation methods. The conceptual definition at level (i) should be both based on toxicological considerations (e.g., What physiological changes should be regarded as adverse health effects?, or, What sensitive subpopulations can be distinguished in the human population?), and on risk management choices (e.g., What sensitive subpopulation do we want to protect?, or, What health effects do we regard as unwanted?). The operational level (ii) in the definition is largely a matter of science, as it addresses the issue of what data or extrapolation methods are appropriate to derive quantitative statements. Since chemical risk assessments typically rely on limited data and poorly based extrapolation methods, a point estimate of the dose as defined in (i) does not suffice. To be of full practical use for risk management, it would be very helpful if the uncertainty margins of that estimate could be quantified as well. Therefore, in the definition of the RfD the quantification of the uncertainties involved in step (ii) is indeed essential, and in this report we focus on this aspect.

For the time being, let us assume that all toxicological and risk management issues in the conceptual definition of step (i) has been settled, and has been termed the "no-adverse-effect level in the sensitive human", or $NAEL_{sens.human}$. The specifics of this definition will not have consequences for the general framework that we propose for step (ii); it has only consequences for the quantitative implementation of it. Since we aim to estimate the $NAEL_{sens.human}$ from animal data, we define the interspecies *extrapolation factor* (*EF*)

$$EF_{interspec} \equiv \frac{NAEL_{animal}}{NAEL_{human}} \; . \qquad (2)$$

Similarly, the intraspecies *EF* is defined as

$$EF_{intraspec} \equiv \frac{NAEL_{human}}{NAEL_{sens.human}} , \qquad (3)$$

where the $NAEL_{animal}$ and the $NAEL_{human}$ are again conceptually defined as the *true*, but unknown no-adverse effect levels in the animal and in the average or typical human, respectively. Thus, the $EF_{intraspec}$ refers to the *true* but unknown ratio that holds for a particular compound, expressing the *true* difference (or rather ratio) in sensitivity between animal and man. Clearly, for a particular compound we have

$$NAEL_{sens.human} = \frac{NAEL_{animal}}{EF_{interspec}\, EF_{intraspec}} \qquad (4)$$

which follows from the definitions (2) and (3). Although expression (4) has the same appearance as the standard expression (1), it fundamentally differs in interpretation: all entities in (4) refer to true but unknown values.

We now come to the definition of step (ii): How to estimate the $NAEL_{sens.human}$ and quantify the uncertainty in that estimate? Basically, the answer is simple: estimate the entities in the right-hand side of (4) and quantify the uncertainties in each of them. To achieve that, we propose as a general approach to quantify each separate entity in (4) by way of an uncertainty distribution. By Monte Carlo calculations these distributions may then be used to derive the uncertainty distribution of the entity of interest: the $NAEL_{sens.human}$. In the following sections we further elaborate on the separate distributions and then illustrate the approach as it may be used for deriving an RfD as well as for estimating human health risks at an actual exposure level.

3. ESTIMATION OF THE NO-ADVERSE-EFFECT LEVEL IN THE ANIMAL

The NOAEL as it is derived from animal toxicity studies, is defined as the highest dose level at which no statistically significant effects occur (for all endpoints that are considered toxicologically relevant). It is important to keep in mind that the NOAEL is not the same as the "true" no-adverse-effect level (NAEL). Suppose there is a (true but unknown) threshold dose below which the agent does not evoke any adverse effects. Then, depending on the study design used (dose levels, number of animals), the NOAEL resulting from a statistical analysis of the data can be lower or higher than this true threshold dose. The potential deviation of the NOAEL from the true threshold dose, however, cannot be quantified. Or, stated differently, although the NOAEL could be considered as an estimate of the true threshold dose, the quality (precision) of the estimate cannot be assessed.

There are several other objections against the use of the NOAEL, which have been discussed extensively elsewhere (e.g., Crump, 1984). The obvious solution to all these objections is to omit the idea of statistically testing separate dose groups against a control, and instead consider the dose-response relationship as a whole. In statistical terms this implies an approach of fitting a regression function to the whole set of response data, and use that function to estimate the dose at which adverse effects start to arise. Crump (1984) suggested to appoint a nonzero effect level, small enough to be considered as an acceptable or nonadverse effect and estimate the dose associated with that effect level. In this case the uncertainty (precision) of the estimated dose can be quantified, and Crump introduced the "benchmark dose level", defined as the lower 95%-confidence limit of the dose associated with the postulated effect size.

Making the benchmark approach operational implies that one needs to postulate a *critical effect size (CES)* below which there is no reason for concern. The problem that arises here is that current toxicological and biological knowledge does not provide sufficient basis to unequivocally establish the breaking point between nonadverse and adverse effect size for most endpoints. Lacking clear scientific arguments for the establishments of critical effect sizes, consensus needs to be reached on the magnitude of the critical effect(s) that are considered to be 'acceptable'. To avoid this arbitrary element, one might propose to always use a zero critical effect size. However, the estimation of a threshold-dose from typical dose-response data leads to severe statistical problems (Slob and Pieters, 1997). In addition, considering any nonzero effect as adverse seems unnecessarily conservative for many endpoints.

3.1. Quantification of effect-size

The need of postulating nonzero critical effect-sizes requires an operational definition of effect-size (ES): How should it be expressed? The first thing to be considered is that this definition will depend on the measurement scale, i.e. are the response data quantal, ordinal or continuous? We will discuss quantal and continuous data here. For ordinal data, e.g. histopathological data that are expressed in classes such as slight, moderate, marked and severe, the effect size may be defined in a way similar to that for quantal data.

3.1.1. Quantal data

Quantal data refer to the situation where each animal is recorded to have responded or not. The dose-response function denotes the fraction, $P(d)$, of animals responding at each dose d. The ES may be measured as "extra"risk or as "additional"risk (Crump, 1984):

$$\text{additional risk} \equiv P(d) - P(0),$$

or

$$\text{extra risk} \equiv \frac{P(d) - P(0)}{1 - P(0)},$$

where $P(d)$ denotes the fraction of responding subjects at dose d, and $P(0)$ the fraction at dose zero.

At first sight, it is obvious to use the additional or extra fraction of responding individuals as a measure of effect: this is in fact what we want to know for the human population. However, for quantal data, the steepness of the dose-response function $P(d)$ as observed in data from an animal study reflects experimental error on the one hand, and the variation between animals on the other. More specifically, the more homogeneous the animals, the steeper the dose-response $P(d)$, and the higher the benchmark dose (see dotted lines in lower panels of Fig. 1). Theoretically, if one could reduce both the variation between animals and the experimental error to nil, $P(d)$ would reduce to a step function, changing from 0 to one at the ED50.

In other words, the use of additional or extra risk would only be relevant if the variation between animals were equal to that of the human (sub)population we wish to protect. Quite opposite to that, laboratory animals are usually bred in a way so as to minimize the interindividual variation, so that it is most unlikely to represent the variation in humans. Therefore, it is more appropriate to derive the ED50 as the dose at which the "average" animal studied responds, and then extrapolate the ED50 to the dose at which the "average" human would respond (using EF_{inter}), and finally extrapolate the latter to the dose at which the "sensitive" human would respond (using EF_{intra}). Since the variation in the test animals is not relevant for that in the human population, information on EF_{intra} must be obtained from other sources. Consequently, for quantal data there is no need to postulate a CES: the ED50 can be used as the CED. The critical effect size is in fact hidden in the mind of the experimental observer when deciding if a particular observation should be classified as a response or not. In many cases, e.g. with histopathological observations, the response is classified in more categories, such as slight, mild, moderate, etc. In that case the ED50 for the relevant (highest nonadverse) category may be taken as the CED.

If human data are available, and the observed individuals are thought to be more or less representative for the population to be protected, one might yet consider to use additional risk or extra risk for defining a CES, and derive the CED associated with that level of response. Even then one should realize however that the steepness of the fitted dose-response may be influenced by the "experimental error", e.g. measurement

errors, or inherent within-individual variation (individual responses often fluctuate in time), or varying circumstances at which responses were measured. This makes the dose-response less steep, and, as a consequence, the CED (benchmark dose) related to a certain extra (or additional) risk in a human data set most likely to be a conservative estimate of the true value for that particular subpopulation. Of course, one might argue that this a desirable feature, in view of the possibility that other subpopulations are more sensitive than the population studied. Alternatively, one might estimate the ED50 for the human data set, and try to quantify possible differences in sensitivity with other subpopulations explicitly, based on other information.

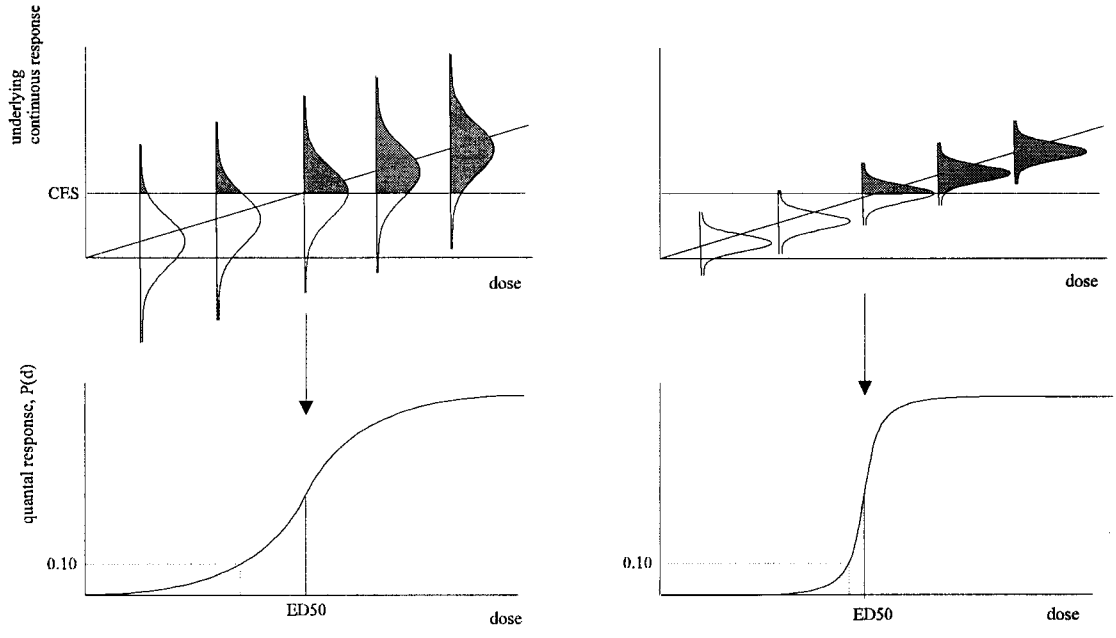

Fig. 1. Relation between quantal response (e.g. fraction of animals with atrophy) and underlying continuous response (degree of atrophy). In the left panels the variation between animals is relatively large, and in the right panels relatively small, while the average (continuous) response is the same. The CES is the level above which the experimental observer classifies an animal as having atrophy. Thus, the shaded areas in the distributions reflect the percentages of animals wih atrophy, which is given in the lower panels as a function of dose.

3.1.2. Continuous data

In the case of continuous response data (e.g. weights, concentrations, counts) the fitted regression function denotes (or rather estimates) the mean response, $f(d)$, of the population at each dose d. One way to appraise the magnitude of the effect is by considering the mean response at dose d and then to see if this value is exceptional in the untreated animals, or not. Thus, the ES may be defined as (Crump, 1984):

$$ES_{rel}\ (\text{relative effect size}) \equiv \frac{f(d) - f(0)}{\sigma(0)},$$

where $\sigma(0)$ denotes the (population) standard deviation of the responses in the control group (which may be replaced by σ in the case of homogeneous residual variance). A basic objection against this definition is that when at dose d the average subject's response falls just within the inter-subject variation in the control group, then non-average subjects will not. Further, one could object against this definition of ES that $\sigma(0)$ does not have a well-defined value to be used as a reference. As already mentioned, one usually aims to minimize the inter-subject variation in animal studies. As a consequence, the degree of success in achieving that goal has a great impact on the value of ES when defined as relative effect size. In addition, since the value of $\sigma(0)$ will be estimated by the sample standard deviation of the observed responses, the value of ES_{rel} will also depend on the accuracy of the measurement techniques used and on the homogeneity of the experimental circumstances of the particular experiment.
The objections against the relative effect size for continuous data are closely related to those mentioned against the use of extra or additional risk for measuring the effect size in quantal data. Not coincidentally, we again come to the conclusion that an animal experiment is only useful for quantifying the response of the average animal, and use the average response of the animal as a model for the average response in the human population. The variation between laboratory animals is not very relevant for the variation in the human population. Besides that, variation between individual measurements not only reflect the variation between individual animals, but also the experimental error in a broad sense.
While in quantal data the response of the average subject can only be quantified by means of a single value, *viz.* the ED50, we can quantify for continuous data the degree of response as a function of dose. And, while in quantal data the CES is already implicitly "defined" in the mind of the experimental observer (see Fig. 1), in continuous data the CES must be quantified explicitly.
Instead of measuring ES relative to the variation in the untreated animals, we advocate an absolute measure for ES, by comparing the difference between mean responses at dose d and dose 0 with the mean response in the controls:

$$ES_{abs}\ (\text{absolute effect size}) \equiv \frac{f(d) - f(0)}{f(0)} = \frac{f(d)}{f(0)} - 1,$$

which, stated differently, measures the proportional change in response of the average animal compared to its response at dose zero.
The problem now is that for each endpoint a critical value for ES_{abs} must be postulated: a single "universal" CES does not seem a realistic option. For example, for enzyme induction one may not be worried about an increase of, say, 25%, while for hemoglobin one might consider a decrease of, say, 5% as unacceptable. Making the CES explicit has the inconvenient drawback that a value must be chosen for each separate endpoint. However, an RfD that is accompanied by an explicitly stated and possibly arbitrarily chosen CES appears more sound than an RfD that was based on a procedure leaving the CES implicit, thereby admitting the possibility that it was in fact unacceptably high without knowing such.

3.2. Uncertainty distributions of no-effect levels in the animal

We now define the Critical-Effect-Size (CES) for a given endpoint as

> CES ≡ value of effect-size below which there is no reason for concern (when occurring in humans)

and the Critical-Effect-Dose (CED) as

> CED ≡ dose at which the average animal shows the (postulated) critical-effect-size defined for a particular endpoint

The No-Adverse-Effect-Level (NAEL) may then be defined as the lowest CED of all endpoints:

> NAEL ≡ minimum of all CEDs.

We re-emphasize that the CED as defined here refers to the true, unkown value, which we can only estimate with a certain degree of precision, if we have data for the endpoint involved. The definition of the NAEL refers not only to an unknown but also to a rather theoretical value, since we do not know to what endpoint it is associated. In practice we can never be sure if we have information on all relevant endpoints for the compound studied. In Chapter 4 we propose an approach how to deal with this problem in practice.

When for a particular endpoint data are available that allow for fitting a regression function, the CED may be estimated. Depending on the quality of the data (experiment) this estimate has a certain degree of imprecision. To take this into account, Crump proposed to calculate the lower 95%-confidence limit of the estimated CED. As indicated in chapter 2, we wish to find the complete uncertainty distribution of this estimate, as part of our probabilistic framework. One way of obtaining this distribution is by calculating the confidence limits for the complete range of confidence levels from zero to unity, instead of only for a single confidence level (e.g., 95%). Thus, the method discussed by Crump for calculating confidence limits for the CED could be extended to obtain the distribution as required. However, this method (likelihood-based confidence intervals) is technically complicated and rather demanding in its implementation. This will only grow worse when extending the method for the whole range of confidence levels as required in our case.

These problems may be avoided by using the bootstrap method. We will use this approach, as long as a systematic comparison between performance and efficiency of the likelihood vs. the bootstrap approach in deriving the uncertainty distribution for the CED, has not indicated to do otherwise. The bootstrap method is conceptually very simple. Once a regression model has been fitted to the data, Monte Carlo sampling is used to generate a large number of new data sets from this fitted regression model, each time with the same number of data points per dose group as observed animals in the real experiment. For each generated data set the CED is re-estimated. Taking all

these CEDs together results in the required distribution. An example of a distribution for the CED resulting from a toxicological data set is given in Fig. 2. The resulting distribution can be seen as an extension of Crump's Benchmark.

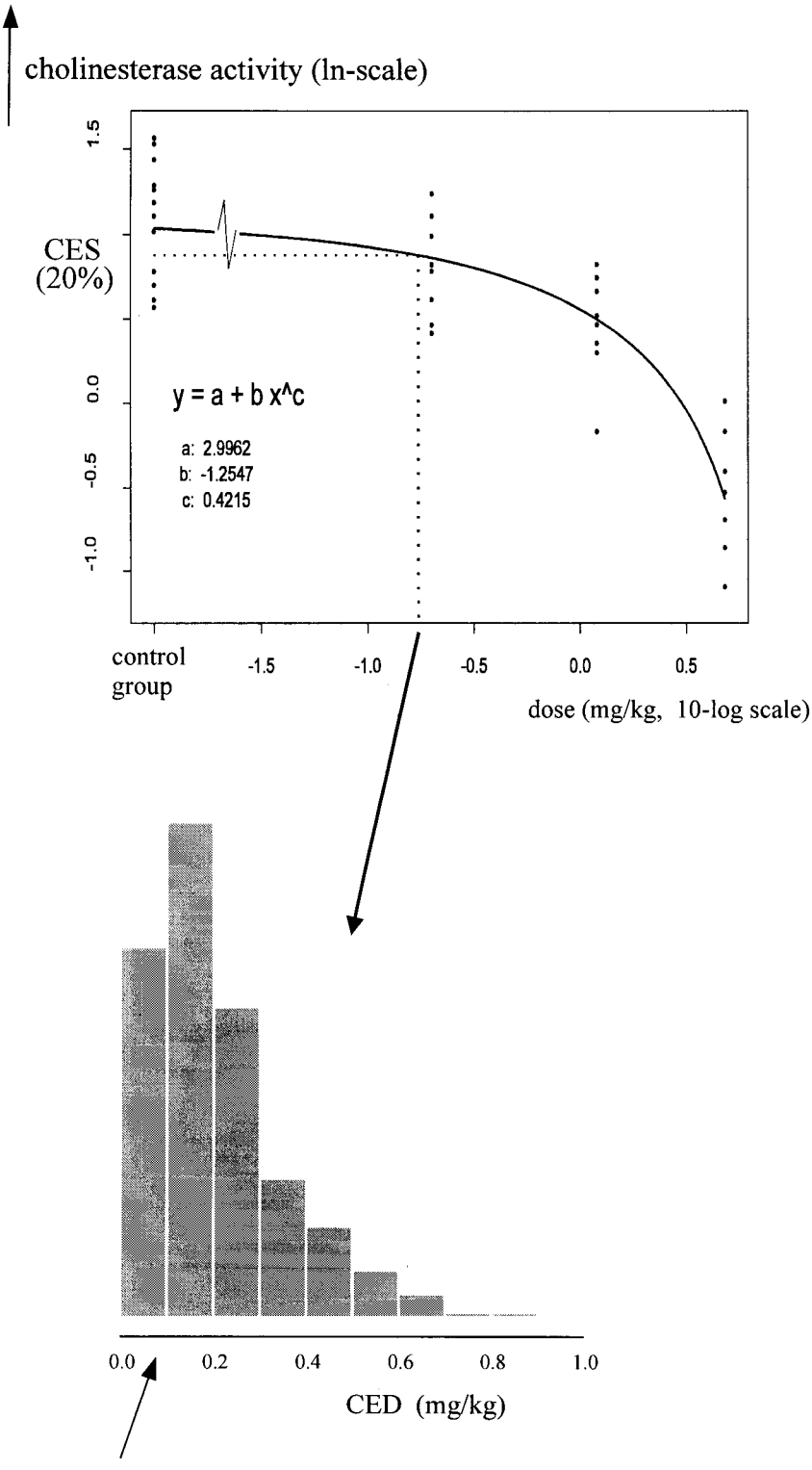

Fig. 2. Upper panel: cholinesterase activity inhibition (µmole/ml, ln-scale) in erythrocytes as a function of ^{10}log-dose (dots refer to individual animals), with fitted regression function, and the estimated CED (value: 0.17 mg/kg) at a CES of 20% cholinesterase inhibition. Lower panel: the associated uncertainty distribution (obtained from 500 Monte Carlo runs from the fitted regression model) for the CED. The lower 5th percentile of this distribution (0.04 mg/kg) is comparable to the benchmark dose.

4. PROBABILISTIC EXTRAPOLATION FACTORS

The use of uncertainty factors in standard setting procedures reflects deficits in our knowledge. To correct for interspecies and intraspecies variation, factors of ten are generally applied. The use of default factors reflects the believe that these factors will be sufficient to cover the necessary extrapolation steps for almost all compounds possible. It is thus considered unlikely that humans are more than ten times as sensitive than the (most sensitive) experimental animal used. Similarly, it is assumed to be unlikely that the most sensitive human differs more than a factor of ten from the average human. As a rule, various uncertainty factors are applied to obtain an overall uncertainty factor ranging from 100 to 10,000. However, multiplication of uncertainty factors implies a piling up of worst case assumptions: the probability of simultaneous occurrence of worst case situations for the same chemical will be smaller than that of a single worst case situation to occur. Therefore, the more extrapolation steps are taken into account, the higher the level of conservatism.

The probabilistic approach as introduced in chapter 2 intends to avoid the piling-up of worst-case assumptions resulting from the multiplication of various uncertainty factors. In this approach we defined the extrapolation factor as an alternative for the uncertainty factor. For example, the interspecies extrapolation factor is defined as the "true" ratio of the no-adverse-effect level in the animal to that in man, *for a particular compound.* Clearly, the value of this factor is unknown for any specific compound that needs to be assessed on the basis of animal data. However, the interspecies extrapolation factors for the universe of all compounds must have a specific distribution, and we might be able to estimate that distribution from historical data (e.g. from drugs). Similarly, we might attempt to find the distributions of the other extrapolation factors. If we succeed in doing that, we can derive the uncertainty distribution of the $NAEL_{sens.human}$ from animal data in a particular assessment, in the way discussed in chapter 2.

To make this approach workable in practice, we first need to further specify the definition of an EF. In chapter 2 we defined an EF as the ratio of the NAEL in one situation (e.g., subchronic exposure) to the NAEL in another (e.g., chronic exposure). In chapter 3 the NAEL was defined as the lowest of all CEDs (for a particular situation), which is in fact the entity of interest when it refers to the (sensitive) human population. However, such a theoretical definition is hard to work with in practice. For example, one can never be sure to have information on all endpoints that may be affected by a particular compound. Further, the lowest CED in the two situations (e.g., animal vs. human) may not refer to the same endpoints. For example, rats may be most sensitive to endpoint A, but humans to endpoint B. Therefore we suggest the following approach to be used in practice. Instead of defining the EF as the ratio of NAELs we define it as the ratio of CEDs for a particular endpoint (at the same CES). Then we assume a certain distribution for this EF, expressing the variation of EFs between all endpoints and all substances. When we know this distribution, we can extrapolate any CED (for any endpoint and substance) from one situation to the other. Thus, in assessing a particular compound, we may derive, for all observed endpoints showing a dose-response relationship, the distribution of the associated CED in the sensitive human. Then this complete set of $CED_{sens.hum}$ distributions can be considered and used as a basis for deriving an RfD, for example by choosing the lowest of each

distribution's first percentile. This procedure takes account of the fact that the data for the various endpoints do not have the same quality, e.g., due to differences in study designs (e.g., sample size, duration of exposure), or to differences in data types (e.g., quantal vs. continuous data).

In this chapter we address the quantification of the EF distributions, by using historical data, if available. Unfortunately, quantitative knowledge on the distributions of the various extrapolation factors is scarce. We will postulate a preliminary distribution for each EF, based upon data in as far as available, and such that the postulated distributions are consistent with the current default uncertainty factors. Therefore, these distributions should be regarded as a first educated guess only. Our primary goal here is to illustrate their use in the general framework presented, and to stimulate efforts to find data that may be used as a basis for updating the EF distributions.

For each EF it will assumed that it is lognormally distributed. This is based on the observation that ratios of NOAELs appeared to be well described by the lognormal distribution (e.g. Kramer et al. 1996). Therefore, the assumption that ratios of CEDs and thus EFs will be lognormally distributed is plausible. Furthermore, the current uncertainty factors of 10 are regarded as worst case assumptions, and therefore will be taken to correspond with the 99th percentile of the EF distributions.

To make extrapolation factors useful in risk assessment, it is necessary to assume that the value of any extrapolation factor for a particular compound and a given endpoint does not depend on the critical effect size that is postulated. As a consequence, the two relevant dose-response relationships (e.g. animal vs. average human, average human vs. sensitive human, subchronic vs. chronic exposure) are assumed parallel on a log-dose scale, at least in the lower dose range. For quantal data this assumption refers to the underlying continuous response, to ensure that the ratio $ED50_{human}$ to $ED50_{animal}$ is independent of the value of CES (as inherently defined in the mind of the experimental observer).

4.1. Interspecies EF

The interspecies EF is defined as

$$EF_{interspec} \equiv \frac{CED_{animal}}{CED_{average\ human}}$$

for any particular endpoint. Interspecies differences may result from toxicokinetic differences on the one hand, and toxicodynamic differences on the other. The rat is the most common experimental animal used in toxicological studies. As an average value for toxicokinetic differences (based on differences in basal metabolism) between rat and man a factor of around five has been suggested, when dose is expressed as amount per kg body weight (e.g., Peters-Volleberg 1994). For toxicodynamic differences there seems to be no *a priori* reason to assume that humans will, on the average, be more sensitive than the animal tested, nor the other way around. To achieve that the

distribution has the property of exceeding a factor of ten with low probability (1%), we need to assume that toxicodynamic differences will usually be less then a factor of two. In this way the resulting distribution (Fig. 3) is consistent with the current UF of ten. However, one might wonder if this distribution is not too narrow (see discussion), or, equivalently, if the default UF of ten is not too small.

Similar to the procedure described above it is possible to postulate distributions for $EF_{interspec}$ when experimental animals other than rats are involved (Baird et al, 1996). One might even consider separate distributions for strains. However, in practice it will not be easy to quantify these distributions, since relevant information on the strain level will be hard to find.

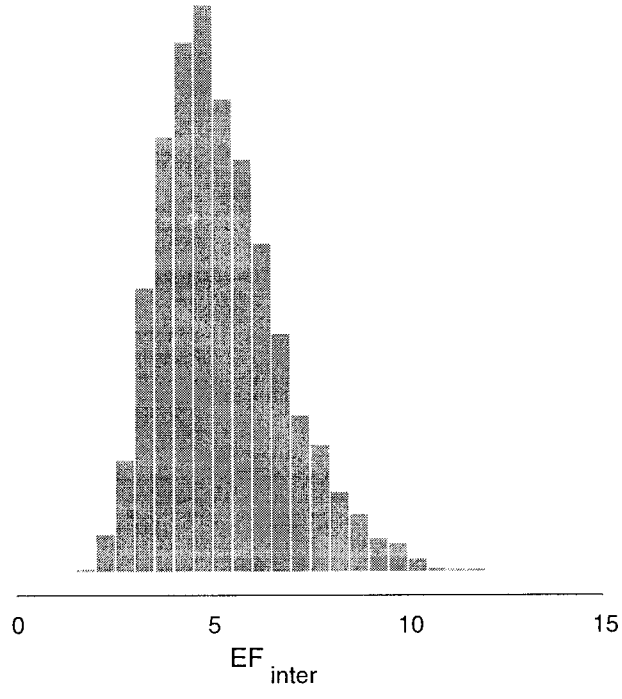

Fig. 3. Proposed (lognormal) distribution of $EF_{interspec}$, the extrapolation factor between CED_{rat} and CED_{human}, with median = 5, first percentile = 2.5, and 99th percentile = 10.

4.2. Intraspecies EF

The intraspecies EF is defined as

$$EF_{intraspec} \equiv \frac{CED_{average\ human}}{CED_{sensitive\ human}}$$

The use of an intraspecies uncertainty factor of ten is meant to protect the most sensitive human subpopulation, in other words it is considered unlikely that these sensitive individuals are more then ten times as sensitive as the average human being.

Though in some cases the intraspecies difference (ratio of sensitive to average individual) is known to exceed a factor of ten, we choose to remain consistent with the factor of ten and take this value as the 99th percentile of the distribution for $EF_{intraspec}$. Since the intraspecies EF cannot be smaller than one, we propose for the time being a lognormal distribution with lower bound one, having median of four, and lower and upper 1%-quantiles of two and ten, respectively (Fig. 4). This distribution arises from a unit shift of a lognormal distribution with median 3 and dispersion factor 3 (the dispersion factor is the ratio between median and lower percentile, which is, in a lognormal distribution, equal to the ratio between higher percentile and median).

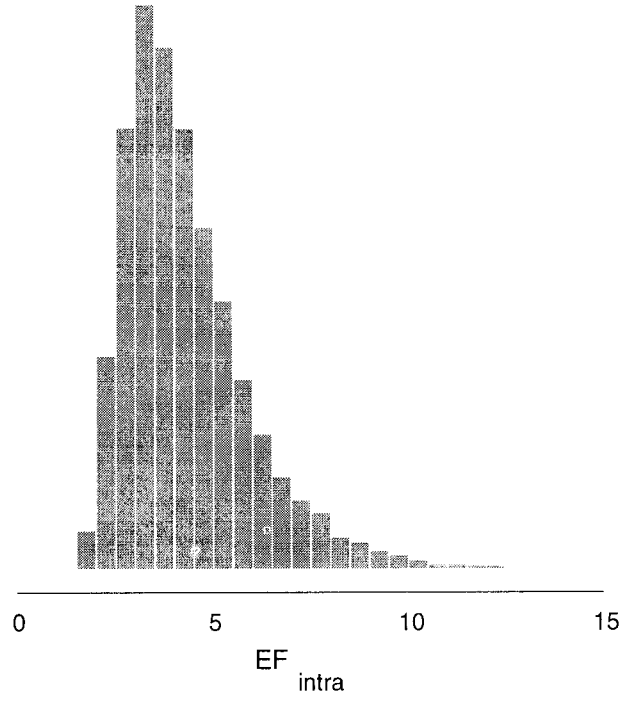

Fig. 4 Proposed distribution of $EF_{intraspec}$, the extrapolation factor between CED_{human} and $CED_{sens.\ human}$. This distribution is obtained by shifting a lognormal distribution (with median = 3, 1st percentile = 1, and 99th percentile = 9) one unit to the right.

4.3. Subchronic EF

When only subchronic data are available an extra uncertainty factor (usually ten) is currently used to extrapolate to chronic exposure. For the distribution of the extrapolation factor

$$EF_{subchronic} \equiv \frac{CED_{animal,\ subchronic\ exposure}}{CED_{animal,\ chronic\ exposure}}$$

several studies comparing NOAELs from chronic and subchronic studies appear relevant (Weil et al. 1963, McNamara 1976, Rulis and Hattan 1985, Kramer et al.

1995). These studies assessed the ratios of observed NOAELs from chronic vs. subchronic studies using historical data for a sample of various compounds.

It should be noted here that subchronic toxicological studies usually have smaller sample sizes compared to chronic studies (typically twice as small). Therefore it may be expected beforehand that NOAELs from subchronic studies will tend to be larger than NOAELs from chronic studies, even if the true dose-response relationships in both studies were identical. Thus, the geometric mean ratios for the NOAELs assessed in the mentioned studies most likely overestimates the median of the distribution of the $EF_{subchronic}$

A second point to be made is that the ratios of the subchronic to chronic NOAELs will have a larger variation than the ratios of the CEDs: the NOAELs are only rough estimates of the true CEDs. Therefore the geometric standard deviation (GSD) of the NOAEL-ratios assessed in the studies mentioned overestimates the variation among the ratios of the true CEDs. Unfortunately, as already stressed, it is impossible to quantify the measurement error of a NOAEL, and therefore it is not possible to correct for this.

This *a priori* notion of overestimation of the variation in CED ratios by the observed variation in the NOAEL-ratios, is confirmed by the data. Consider the histogram of the NOAEL ratios from the Kramer et al. study (see Fig. 5).

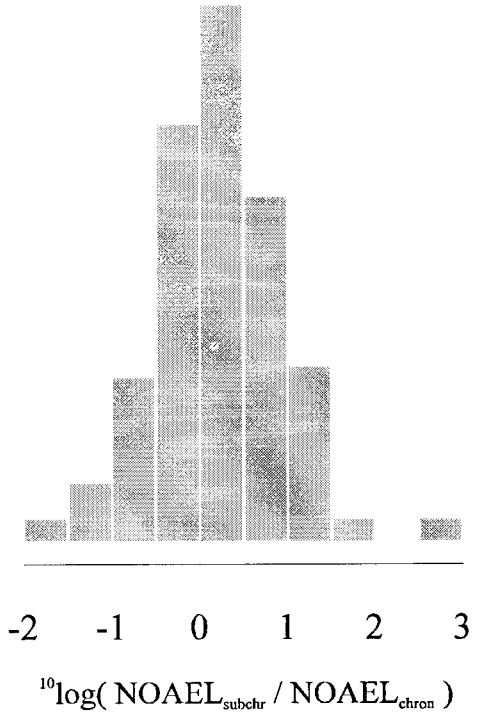

Fig. 5. Histogram of observed ratios $NOAEL_{subchronic}$ to $NOAEL_{chronic}$ (Kramer et al. 1995) on 10-log scale. Of all log-ratios 30% is lower than 0, implying that 30% of the compounds had a lower NOAEL in the subchronic than in the chronic study.

A very large proportion of the ratios (30%) is in fact smaller than unity, i.e. for 30% of the compounds studied the NOAEL in the subchronic study was lower than in the chronic study. Although in some cases long-term exposure might cause some sort of tolerance to the compound, it does not seem likely that so many compounds evoke effects in subchronic studies at lower doses than in chronic studies. Thus, this large percentage may attest of large measurement errors in the NOAELs.

On the other hand, it should be noted that the NOAEL ratios in Fig. 5 were assessed without taking notice of the toxicological endpoints to which the NOAELs referred. Therefore, part of the ratios being lower than unity might be explained by the fact that in subchronic studies usually more endpoints are measured than in chronic studies. For the purpose of getting information on the distribution of the EF_{subchr}, NOAEL ratios (or better: CED ratios) referring to the same endpoints should be considered. This is implied by our definition of an EF distribution, in which each single EF refers to a particular endpoint (and compound).

Only when sufficient toxicological data from both chronic and subchronic studies are analysed by estimating the CED for a given endpoint as well as the uncertainty in that estimate, can the distribution of the $EF_{subchronic}$ be adequately quantified (see Appendix how this may be done statistically). As yet, we can only conclude that the distribution of observed NOAEL-ratios overestimates the variation of the distribution of the $EF_{subchronic}$. Therefore, we propose for the time being a lognormal distribution with median 1.5, and a GSD of 2.3, a value that was chosen to achieve a 1% probability that the EF is larger than ten (see Fig. 6). In this way it is achieved (just as for the EF_{inter} and EF_{intra}) that the distribution for the EF is consistent with the default UF of ten being a worst case value.

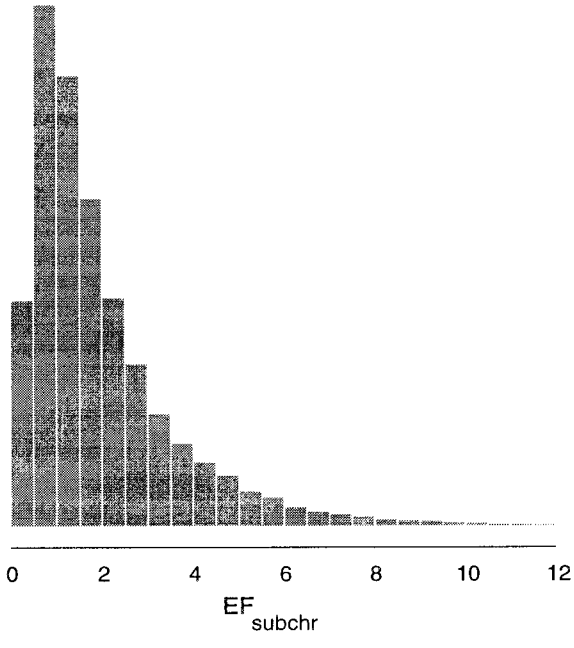

Fig. 6. Proposed distribution of the extrapolation factor between CED_{subchr} and $CED_{chronic}$, with median = 1.5, first percentile = 0.225, and 99th percentile = 10.

4.4. Short-term EF

Although it is not common to extrapolate a short-term NOAEL to a chronic NOAEL in setting toxicological standards, one might consider to do so in the method presented here, using a distribution of an extrapolation factor

$$EF_{short\text{-}term} \equiv \frac{CED_{animal,\ short\text{-}term\ exposure}}{CED_{animal,\ chronic\ exposure}}$$

Kramer et al (1996) found a lognormal distribution for the ratio of $NOAEL_{short\text{-}term}$ to the $NOAEL_{chronic}$, having a median of 4.1, and a geometric standard deviation of 4.4. Of course, just as in the case of the subchronic EF, these data can only indicate a worst case distribution for the distribution of the $EF_{short\text{-}term}$. It can be expected that both the median and the GSD of the $EF_{short\text{-}term}$ will be substantially smaller.

4.5. LOAEL - NOAEL extrapolation

In the standard method an uncertainty factor (default value of 10) is applied in the situation that only a LOAEL is available. In the method we present here, this situation has become non-existent, since an estimate is made of the CED from regression analysis, instead of deriving a NOAEL. Thus the associated uncertainty factor has become irrelevant.

4.6. Incomplete data-base

In situations that not all required studies are available, yet another uncertainty factor is often applied. Distributions for this factor could be assessed by applying multiple regression analysis on historical data. Suppose, for example, that we wish to have a distribution to be used for situations in which a developmental study is lacking. To obtain this distribution, one could compose a database consisting of compounds for which all required studies are available, and calculate for each compound the factor between the observed no-effect-level in the developmental study and the level that is predicted from the other study types by the multiple regression equation. The set of resulting factors may be used to estimate the relevant distribution.

5. ILLUSTRATIVE EXAMPLES

The probabilistic approach discussed here may be used for both deriving acceptable intake or exposure limits (RfD, ADI, TDI) and for estimation of human health risks for a given actual exposure level. We will illustrate both applications for an anonymous compound that acts as a cholinesterase inhibitor.

5.1. Deriving an RfD (TDI, ADI)

Suppose that, in order to derive an RfD for the cholinesterase inhibitor, we have chosen the data given in Fig. 2 as the most appropriate for that purpose. The lowest dose group (0.2 mg/kg) in this study was significantly different from the control group, and the default uncertainty factor would be 1000 (three times a factor of ten, for inter- and intraspecies, and LOAEL - NOAEL extrapolation, respectively), resulting in a classically derived RfD of 0.2 μg/kg.

As an alternative, the probabilistic method may be used. Firstly, the CED_{animal} referring to inhibition of erythrocyte cholesterinesterase activity is estimated from the data by taking 20% reduction as the critical effect size. As shown in Fig. 2, the critical effect dose (CED_{animal}) is estimated at 0.17 mg/kg bw. By using the bootstrap method the distribution of this CED_{animal} is obtained. This data-based distribution is then combined with the postulated distributions of the $EF_{interspec}$ and the $EF_{intraspec}$, as shown in Figs. 3 and 4. This results in the distribution of the $CED_{sens.\ human}$ (Fig. 7), representing the uncertainty associated with the estimate of the true dose resulting in a 20% reduction of cholinesterase activity in the sensitive human. Determining the 1st percentile of this distribution results in an RfD of around 0.8 μg/kg, which is a factor of four higher than the classically derived RfD of 0.2 μg/kg. Note that with the probabilistic approach the LOAEL-NOAEL extrapolation has become obsolete.

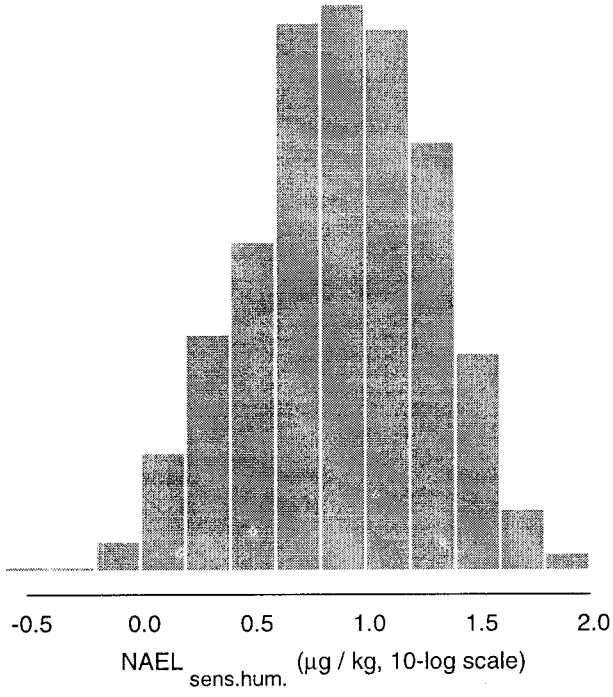

Fig. 7. Uncertainty distribution of the dose level that would evoke a 20% reduction in erythrocyte cholinesterase activity in the sensitive human subpopulation, based on a study in animals (see Fig. 2). The first percentile of this distribution (0.8 µg / kg body weight) might be used as an RfD.

5.2. Estimating actual risk

An additional advantage of the probabilistic method is that it can also be used to estimate the expected effect in sensitive humans for any actual exposure level, including a distribution representing the uncertainties associated with that estimate. In this case the size of the effect in the animal is estimated from the dose-response data obtained from a relevant toxicological study. The uncertainty associated with that estimate due to experimental error is quantified using the bootstrap method, resulting in a distribution for the expected effect in the animal. Then this data-based distribution is combined with the postulated distributions for the relevant extrapolation factors. The resulting distribution reflects the possible effects that might occur in the sensitive human being. A higher percentile of this distribution may be chosen to assess whether or not adverse effects in the (sensitive) human population can be excluded.

As an illustration, suppose that the actual exposure in the human population for the same cholinesterase inhibitor is estimated to be 1 µg/kg, i.e., a factor of 5 higher than the classically derived ADI of 0.2 µg/kg. We can now estimate the possible effect at this exposure level for the sensitive human being, using the probabilistic method. To that end the dose-response data (Fig. 2) and the fitted regression line are used to

estimate the effect-size in the animal. Firstly, the effect (in the animal) associated with an exposure of 1 µg/kg is estimated. Then the distribution that represents the uncertainty in that effect estimate caused by experimental limitations is calculated. Subsequently, this distribution is combined with the distributions for inter- and intraspecies extrapolation, resulting in a distribution for the possible inhibition of cholinesterase in the sensitive human subpopulation (see Fig. 8). From this distribution it can be concluded that it is unlikely that the cholinesterase inhibition in the sensitive human due this particular exposure will be higher than 17% (95th percentile).

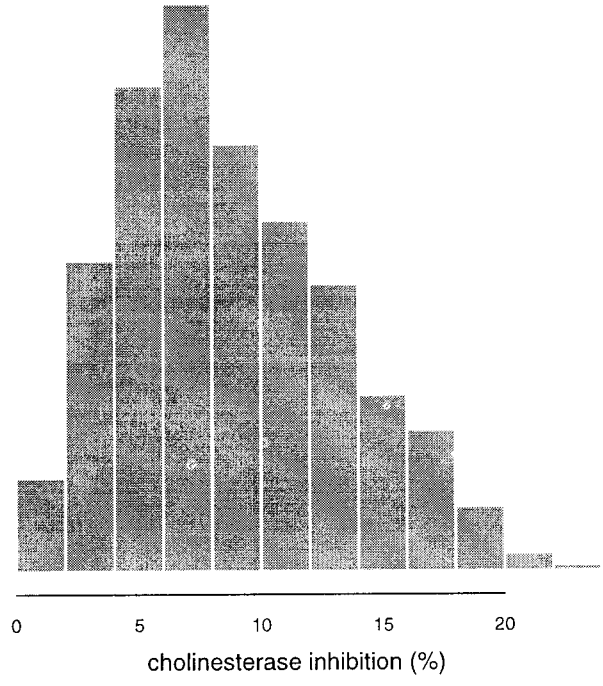

Fig. 8. Uncertainty distribution for the expected cholinesterase inhibition based in the sensitive human subpopulation, for an exposure level of 1 µg/kg/day.

6. DISCUSSION

The standard procedure for deriving human acceptable exposure limits (ADI, TDI, RfD) is usually considered to be conservative. One of the arguments mentioned concerns the procedure of multiplying various uncertainty factors with each other. Indeed, as a procedure, this results in an overly conservative combined uncertainty factor, on the premise that the individual factors are already conservative themselves. We discussed how this *procedural* conservatism can be avoided by taking a probabilistic approach, and use distributions for extrapolation factors instead of single uncertainty factors. This approach forces one to carefully consider how the various extrapolation factors might be distributed in reality. For example, for the interspecies EF we used the pharmacological experience that a rat, on the average, needs a five times higher external dose (mg/bw) than a human being in order to achieve the same internal dose. Since we aimed to remain consistent with the current default UF of ten, only a factor of two was left to cover both toxicokinetic variation (among compounds) around this factor of five, and toxicodynamic differences between rat and human. This simple argument shows that the default uncertainty factor of ten may not be that conservative. For the mouse the situation is even worse: the average toxicokinetic factor between mouse and man as used by pharmacologists is already by itself larger than ten. Similarly, one might question the conservatism of a default factor of ten for intraspecies variation, especially if one aims to protect specific sensitive subpopulations, having for instance congenital (e.g. enzyme) deficiencies, or a particular illness. It may be questioned whether these sensitive subgroups fit in the lognormal distribution with a 99th percentile of ten. Furthermore, the complete disregard of the uncertainty in the NOAEL may also be considered as a nonconservative element in the current procedure. It is hard to see what may result from these conservative and nonconservative elements taken together. Therefore, it appears doubtful to make any firm statement on the conservative or nonconservative character of the current approach in general, let alone in a particular assessment.

The important advantage of the probabilistic approach is that it aims at taking account of all the uncertainties involved in a systematic and nonconservative procedure, after quantifying each of the separate uncertainties as accurately as possible. The results are in the form of a uncertainty distribution, so that the degree of conservatism is quantifiable in any particular assessment. As a matter of fact, this approach allows for deriving an RfD as a function of an *a priori* chosen degree of conservatism. In addition, the approach allows for estimating the lower and upper bounds for possible health effects in the sensitive population at a given exposure level.

An obvious objection against the probabilistic approach is that some of the EF distributions that we have proposed do not have a solid scientific basis, and may substantially deviate from real life. Indeed, this could be the case, and the resulting distribution for $CED_{sens.hum.}$ might therefore not be accurate. However, it should be kept in mind that the default uncertainty factors do not have much more ground, other than general acceptance. Further, any estimate of the uncertainty in the $CED_{sens.hum.}$, even if poor, is better than no estimate at all. Therefore, from a theoretical point of view, the probabilistic approach appears an improvement of the current approach. But there are some important practical matters that need to be considered before the probabilistic

approach can be implemented as a standard method for risk assessment. First of all, consensus needs to be reached on default distributions for the EFs, just as at present consensus exists on default UFs. In addition, concensus will be needed on the critical effects sizes (CESs) for all the toxicological endpoints that may be relevant. And finally, the probabilistic approach has implications for the design of toxicological studies. We will briefly elaborate on these three issues.

6.1. Assessing the distributions of EFs from data

In this report we proposed EF distributions based upon common sense, and with the restriction of being consistent with the current default UFs. However, not every expert in the field might agree with our proposals, and it will be hard to find strong arguments for one or the other preferred distribution. Clearly, much would be gained if the distributions could be based on data. A number of studies have gathered historical data to "validate" the default value of ten for several UFs. In these studies the distributions of ratios of NOAELs are examined, e.g. $NOAEL_{subchronic}$ / $NOAEL_{chronic}$, for a sample of compounds. What is needed in the probabilistic method however is the distribution of the ratio of CEDs (related to the same endpoints). As already discussed, it is important to realize that the variation in the ratio of NOAELs overestimates the variation in the relevant CED-ratio, because the NOAELs themselves contain substantial errors (noise).

To more accurately estimate the variation in a particular EF, the relevant CEDs need to be estimated by fitting dose-response functions to the data. For example, for the $EF_{subchronic}$ one should estimate the CED in both a chronic and a subchronic study, for the critical effect size of the associated endpoint. Furthermore, the variance (standard error) of each estimated CED should be assessed, as a measure for the estimation error involved. Data of the estimated CEDs and their estimation errors can subsequently be used to estimate the "true" variation in the $EF_{subchronic}$ (see Appendix 1).

Unfortunately, estimated CEDs, or benchmark doses, are as yet rare. This leaves two options. For a sample of compounds the original data need to be re-analysed by regression analysis. Or, alternatively, we must wait until enough compounds have been assessed by the regression method (or, equivalently, the benchmark approach).

6.2. Choice of critical effect size

The requirement of postulating a critical effect size for each particular endpoint considered, seems to introduce an arbitrary element in the benchmark approach. Again, we stress here, that no less arbitrariness is involved in the use of the NOAEL in the current approach. Although the NOAEL approach appears a formal, objective procedure, the arbitrariness is in fact bigger: the size of the effect that occurs in reality at the NOAEL is usually not considered and left to the chance processes associated

with the particular study design (dose levels applied, number of animals used, experimental noise). We emphasize that the value of a critical effect size should be entirely determined by toxicological and biological understanding and insights. Statistics should only be used to determine how the critical effect size can be estimated and what study design is needed to achieve a desired level of precision.

6.3. Consequences for study design

The approach of estimating the critical effect dose by regression analysis, instead of assessing the NOAEL by significance testing, has implications for the design of experimental studies. The requirement of a minimal number of subjects per dose group to warrant sufficient statistical power can be dropped. As a matter of fact, a study design with more dose groups, but with fewer animals per dose group is recommendable for several reasons. This issue is further discussed in Slob and Pieters (1997).

REFERENCES

Baird, J.S., Cohen J.T., Graham J.D, Shlyakter A.I., Evans J.S. (1996).
Noncancer risk assessment: a probabilistic alternative to current practice.
Human and Ecological Risk Assessment 2:79-102.

Crump, K.S. (1984).
A new method for determining allowable daily intakes.
Fund. Appl. Toxicol. 4:854-871.

Kimmel C.A., Gaylor, D.W. (1988)
Issues in qualitative and quantitative risk analysis for developmental toxicology.
Risk Analysis 8:15-20.

Kramer H.J., Van den Ham W.A., Slob W., Pieters M.N. (1995)
Conversion factors estimating indicative chronic NOAELs from short-term toxicity data.
RIVM, Report No 620110 001

Kramer H.J., Van den Ham W.A., Slob W., Pieters M.N. (1996)
Conversion factors estimating indicative chronic NOAELs from short-term toxicity data.
Regul. Toxicol. Pharmacol. 23: 249-255.

Leisenring W., Ryan L (1992).
Statistical Properties of the NOAEL.
Regul. Toxicol. Pharmacol. 15: 161-171

McNamara, B.P. (1976).
Concepts in health evaluation of commercial and industrial chemicals.
In: Advances in Modern technology, 1, part 1, New concepts in safety evaluation (M.A. Mehlman, R.E. Shapiro and H. Blumenthal, Eds.) pp. 61-140, Hemisphere, Washington D.C.

Peters-Volleberg GWM, De Waal EJ, Van der Laan JW (1994).
Interspecies extrapolation in safety evaluation of medicines in the Netherlands (1990-1992): practical considerations.
Regul. Toxicol. Pharmacol. 20: 248-258.

Rulis, A.M. and D.G. Hattan. (1985).
FDA's priority-based assessment of food additives. II. General toxicity parameters.
Regulatory Toxicology and Pharmacology, **vol. 5**, p. 152-174

Slob W. (1997).
Uncertainty analysis in multiplicative models
Risk Analysis 14:571-576.

Slob W. and Pieters M.N. (1997).
A probabilistic approach for deriving human RfDs and human health risks from toxicological studies: consequences for study design.
RIVM, Report No. 620110 006.

Weil, C.S. and D.D. McCollister. (1963).
Relationship between short- and long-term feeding studies in designing an effective toxicity test.
Journal of Agriculture and Food Chemistry, **vol. 11**, p. 486-491

APPENDIX Estimation of the $EF_{subchronic}$ from estimated $CEDs_{subchronic}$ and $CEDs_{chronic}$ and their estimation error.

By dose-response (regression) analysis of toxicity data the CED can be estimated, given a defined critical effect size,. When for a sample of compounds, subchronic and chronic toxicity studies are analysed, the median ratio of the estimated CEDs may serve as an estimate of the median of the $EF_{subchronic}$. The variances of the individual CED estimates may be pooled (for both the subchronic and chronic studies separately), so that the variation in the observed ratios can be corrected for these estimation errors assuming that

$$\text{var}[\log(\frac{estimated\ CED_{sub}}{estimated\ CED_{chron}})] = \text{var}[\log(\frac{CED_{sub}}{CED_{chron}})] + \text{var}[\log(\varepsilon_{sub})] + \text{var}[\log(\varepsilon_{chron})] \quad (7)$$

where

$$estimated\ CED_{sub} = CED_{sub} \cdot \varepsilon_{sub}$$

$$estimated\ CED_{chron} = CED_{chron} \cdot \varepsilon_{chron}$$

In the report we have proposed for the time being a distribution having median 1.5 and GSD 2.3 (or, equivalently, a variance for the log-EF of 0.69), to obtain a 1% probability that the EF is larger than ten. When we regard the observed variance of the log-ratios of NOAELs (Kramer et al. 1995) of 2.97 (or, equivalently, GSD = 5.6) as an estimate of the variance given by (7), and the NOAELs as estimates of the CEDs, it follows that the variances of both NOAELs must sum to 2.97 - 0.69 = 2.28. In other words, when the postulated GSD of 2.3 for the $EF_{subchron}$ is accurate, the variances of the log-NOAELs must be in the order of unity, or equivalently (Slob 1994), a dispersion factor of around seven. This means that the NOAEL may, with small probability, on the average over- or underestimate the true CED by a factor of seven. Given the use of dose levels differing by a factor of four up to ten, this does not seem an implausible situation.