# To a harmonized system for classification of chemical substances. How to weigh the evidence for reproductive toxicity?

## Report of the CGC/NVAB meeting on June 10th, 2010

*Paul Scheepers*[1]

### Introduction

In the Netherlands, the first attempt to identify reproduction toxicity of industrial chemicals was the "Indicative Reprotox List" by dr. Anne Stijkel published in 1983 (1) and updated in 1999 (2). Such a list is attractive for health professionals, who use it for hazard identification and risk characterization purposes. But what does the word 'indicative' in the title of this list mean? To what extent is inclusion of substances in this list evidence-based? The underlying evidence for classification and the procedures followed to perform a classification was the topic of a seminar organized in 's-Hertogenbosch by the Contact Group Health and Chemistry (CGC) and the Netherlands Society of Occupational Medicine (NVAB). The focus of this seminar was on classification of chemical substances for their reproduction toxic hazard in the USA and The Netherlands. The evaluation processes and criteria used for classification of reproductive hazards were reviewed and the use of the new hazard statements under the globally harmonized system (GHS) was discussed.

### Evaluation process and classification under the National Toxicology Program in the US

The National Toxicology Program's (NTP) Center for the Evaluation of Risks to Human Reproduction (CERHR) is one program in the US that conducts classifications based on reproductive and developmental toxicity. The NTP is an interagency program established in 1978 to coordinate toxicology testing programs within the US Federal government, strengthen the science base in toxicology, develop improved testing methods, and provide information on potentially toxic chemicals to experts and the public (3). CERHR publishes monographs that evaluate the evidence on whether substances cause adverse effects on reproduction and development, and whether there should be any degree of concern for these types of effects given the extent of human exposure. The NTP is not a regulatory agency such as OSHA, FDA, California EPA or US EPA, but regulatory agencies often consider the conclusions presented in CERHR monographs. However, regulatory agencies do not have to adopt CERHR conclusions. For example, the state of California publishes a list of chemicals known to cause reproductive or developmental toxicity as part of Proposition 65 (4). Chemicals can be proposed for listing on Proposition 65 by a variety of ways, including proposals based on an evaluation conducted by an authoritative body such as CERHR.
The Center performs a non-quantitative risk characterization based on human data and animal data. For this meeting, the acting director of CERHR, Dr. Kristina Thayer, was invited to explain about the selection of substances considered for CERHR evaluation, the evaluation process that leads to classifications of chemicals as reproductive or developmental toxicants, and to highlight some of the dilemmas encountered in weighing the evidence.

CERHR nominations can be made by private persons, enterprises, non-governmental organizations, and also by governmental bodies. Nominations can also be developed internally. The scientific part of the evaluation process (Figure 1) starts with preparation of an expert panel report (the draft NTP Monograph), a document usually covering several hundreds of pages. At this stage, this report is not suitable for public review and also contains studies that may be rejected in a later stage of the evaluation process. There are several opportunities to include views from the public and the industry, such as in listening sessions. Once the committee reaches consensus, a so-called draft NTP brief is prepared and published for public review. This is a concise document that contains the conclusions of the committee. This document is then finalized and submitted for interagency review. The final document is published on the website of CERHR (3).

Communication about reproduction toxic risks (dr. Kristina Thayer, CERHR)
In the communication about reproductive toxicity to the public, the NTP uses five standard phrases indicating 'level of concern' (see Figure 2). These phrases are based on the intrinsic hazard, i.e. adverse developmental and reproductive effects in humans and animals, but are also related to the extent of human exposure. This adds a risk characterization component to the level of concern. For each substance, an evaluation can have multiple conclusions in terms of level of concern, each for different endpoints, life stages, or levels of exposure. This makes the framework for communication much more detailed and gives a much broader scope than the EU classification, which is limited to intrinsic reproductive toxicity (hazard) and limited to only three categories (see below).
In addition to the level of concern classification, a grading system for the weight of evidence is also used, involving a

[1]*Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre*

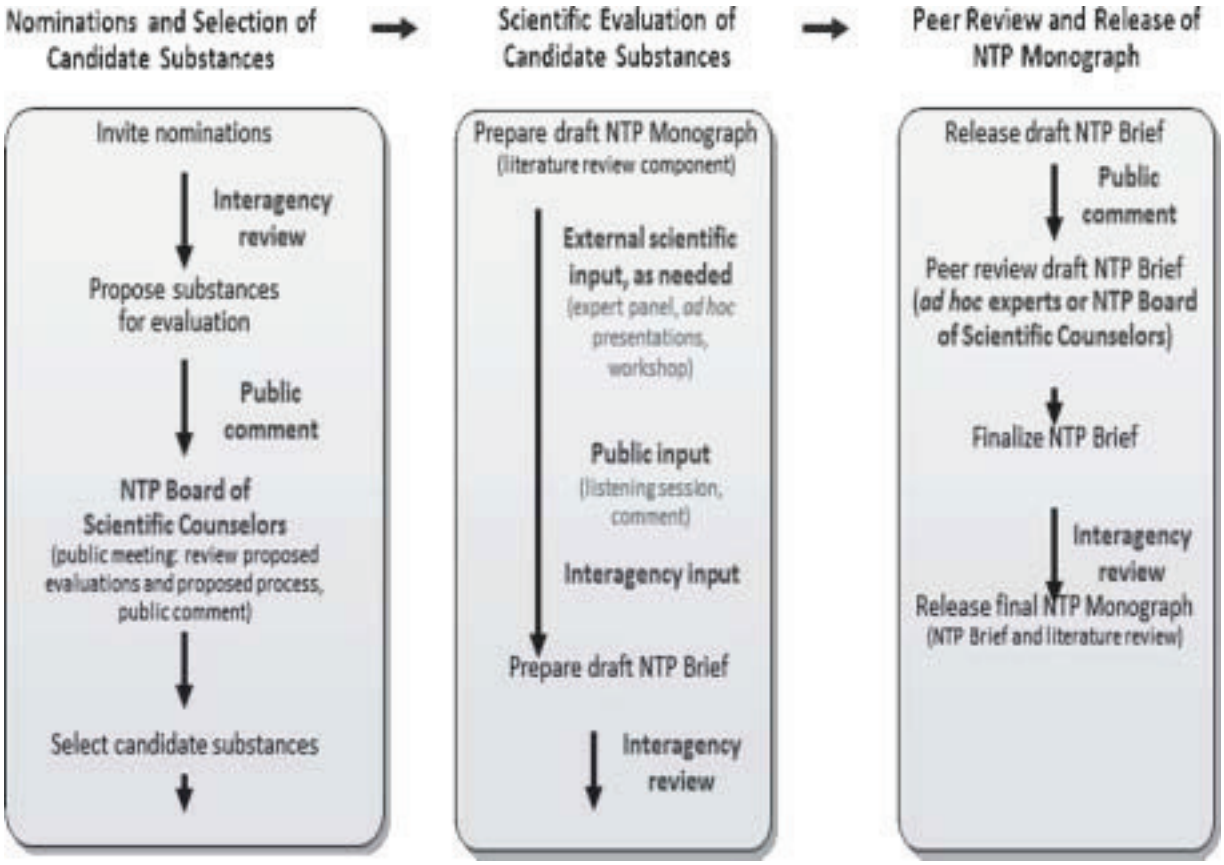


*Figure 1: CERHR evaluation process. Source: CERHR.*

seven point hazard identification scale. Weights of evidence are designated separately to human data and animal data (Figure 2).

A limitation of the current levels of concern system is that there are no explicit scientific criteria for each level of concern. Also, there can be confusion in the scientific community and by the public on what a specific level of concern conclusion means. At the moment, the CERHR is in the process of evaluation the weight of evidence and level of concern descriptors currently used.

The state of California adopted the CERHR evaluation and published a list of chemicals known to cause developmental toxicity as part of Proposition 65 (4). The CERHR evaluation procedure was illustrated by three examples: propylene glycol, bisphenol A and di(2-ethylhexyl)phthalate.

*Propylene glycol*

Propylene glycol is used as coolant and antifreeze. It is also used in personal care products like deodorant and is approved as food additive (E1520). The weight of evidence conclusions regarding reproductive or developmental toxicity for propylene glycol were 'insufficient to reach a conclusion' for humans based on too few studies. Animal studies performed in four different species did not detect any reproductive or developmental toxicity. Thus, the weight of evidence in laboratory animals was considered to be 'clear evidence of no adverse effects'. For humans, the exposure situation was not considered excessive and propylene glycol has a short half life, indicating rapid detoxification and excretion. There was also some indication that humans may be less sensitive than laboratory animals, because saturation of the propylene glycol metabolism occurs at a lower dose in humans compared to animals. This is considered protective because propylene glycol exhibits less indications of general toxicity than its metabolites. Taken together, the CERHR committee concluded that there is 'negligible concern' for an adverse reproduction toxic effect of propylene glycol.

*Bisphenol A*

Bisphenol A (BPA) is used as an additive in plastics such as polycarbonate, e.g. in feeding-bottles for infants. Since only few human studies were available, the human data were rated as 'insufficient evidence'. The weight of evidence for developmental toxicity based on animal studies was less clear at 'low' doses (≤ 5mg/kg), compared with high doses (> 5 mg/kg). The effects at low doses were evaluated by the NTP as 'limited evidence' to support developmental toxicity in animals. The 'positive' low dose studies were sometimes difficult to fully interpret, but could not be discounted either, because a number of these studies were considered of high technical merit. By way of example, two types of brain and behavior studies were discussed: an open field study and a light/dark chamber study. The conclusion from both studies was that BPA exposure may lead to increased anxiety response and less exploration behavior in mice. In addition, there appeared to be a loss of sexual dimorphisms for non-reproductive behaviors. These types of findings did not support a stron-


*Figure 2: Phrases used by the NTP in communication about adverse effects: levels of concern (left panel) and weight of evidence (right panel). Source: CERHR*

ger weight of evidence conclusion because the extent to which these findings in mice can be extrapolated to humans is difficult to judge.
In contrast, there was 'clear evidence' for developmental toxicity at higher doses. Decreased fetal survival was observed in rats at doses exceeding 500 mg/kg, but at these doses maternal body weight was reduced as well. At doses > 50 mg/kg, puberty was delayed and growth and survival were reduced. Taken together, the CERHR evaluated bisphenol A to raise 'some concern for adverse effects', specifically effects on brain and prostate in fetuses and infants.

*Di(2-ethylhexyl)phthalate*
Phthalates form another large group of chemicals, which are among others used as plasticizers. Di(2-ethylhexyl) phthalate (DEHP) is an important and much used representative of this group. In a multigeneration study in Sprague-Dawley rats, 14-23 mg/kg was established as the lowest adverse effect level (LOAEL) for reproductive organ malformations in male offspring. A dose of 3-5 mg/kg was established as the no adverse effect level (NOAEL). This study implied that for DEHP the weight of evidence from animal studies was rated as 'clear evidence of adverse effects', whereas the human data set was considered 'insufficient' to contribute to the weight of evidence for adverse effects. Interestingly, the CERHR concluded that the level of concern for male infants undergoing extensive medical procedures should read 'serious concern'. This classification is based on the risk characterization for neonates and infants undergoing extensive medical procedures, for which a dose above the NOAEL reported for animal studies is predicted. This illustrates how CERHR involves risk in its final evaluation, a method that is not customary in the evaluations carried out by the Health Council of the Netherlands, which are limited to evaluation of hazards.

## Evaluation process and classification in the Netherlands and the EU

The classification in the Netherlands is carried out by the Subcommittee on the Classification of Reproduction Toxic Substances (Committee 543) of the Dutch Expert Committee on Occupational Safety (DECOS) of the Health Council of the Netherlands (5). This committee currently uses the Directive 93/21/EEC of the European Union to classify chemicals at the request of the Minister of Social Affairs and Employment. Dr. Nel Roeleveld, associate professor of reproductive epidemiology at the Radboud University Nijmegen Medical Centre and a member of this subcommittee, presented the evaluation process and the methodology of classification, and explained the dilemmas encountered when considering evidence from observational studies in humans to support classifications of reproductive hazards.

Evaluation process (dr. Nel Roeleveld, RUNMC)
The subcommittee consists of toxicologists, a clinical geneticist and pediatrician, and a reproductive epidemiologist (NR), is chaired by a reproductive toxicologist, and is supported by two staff members from the Health Council secretariat. The committee drafts a report for public review, based

on an initial collection of published data from peer reviewed human and animal literature, that is prepared by an expert who is not a member of the committee. The scope of the literature search is on male and female fertility endpoints (libido, sexual behavior, spermatogenesis/oogenesis, hormonal activity, capacity to fertilize, fertilization, and development of the conceptus through implantation) and on developmental toxicity endpoints (embryotoxic/fetotoxic effects, such as reduced body weight, developmental retardation, organ toxicity, death, abortion, structural and functional defects, peri- and postnatal defects, and impaired postnatal development through puberty). In addition to the literature study provided, members of the committee consult original literature and have the possibility to add publications to increase confidence in the evaluation process. Extensive discussion and weighing of the available evidence eventually leads to a consensus classification, which is described in a draft report. This report is submitted for internal review within the Health Council and for public review thereafter. Once the report in finalized, a substance is, dependent on the classification given, or is not added to the list of reproduction toxic substances. This list is legally enforced by the Minister of Social Affairs and Employment. It should be noted that in addition to the work of committee 543, substances that have been classified as a reproduction toxic hazard by the EU, are already part on this list.

Human data (dr. Nel Roeleveld, RUNMC)
Evidence for a human reproduction toxic hazard is usually based on data from observational epidemiologic studies. This implies that there is no controlled exposure situation, as humans are usually exposed to multiple substances simultaneously or in mixtures, while exposure may stem from a multitude of external sources, including the work environment but also the general environment, food, and lifestyle factors. Exposure characterization is often poor and not based on adequate methods of exposure assessment, leading to questions such as: Are study participants truly exposed? And what about exposure level and duration? In addition, for reproductive toxicity endpoints, it is important to know details about the timing of exposure related to the critical time windows of increased susceptibility during development. Therefore, the condition for classification in the highest hazard category ('sufficient evidence to establish a causal relationship in humans') is rarely met, unless several high quality epidemiologic studies on a specific substance all lead to similar results.

Animal data (dr. Aldert Piersma, RIVM)
Dr. Aldert Piersma, reproductive toxicologist at the National Institute of Public Health and the Environment (RIVM) and professor of reproductive toxicology at the University of Utrecht focused on classification based on animal studies. Animal toxicity is evaluated by scrutinizing published developmental toxicity studies (OECD414) and generation studies (OECD416) that are performed according to OECD guidelines. Important for developmental toxicity in rats is administration of the test substance between the 6th and15th day of gestation (GD6-GD15), since implantation of the rat embryo is finalized by GD6 and differentiation of all organs is completed on GD15. For fertility assessment, coverage of the entire reproductive cycle in generation studies is essential. It is the task of the committee to weigh the biological significance of the reproductive toxicity findings against any form of systemic toxicity. This involves expert judgment from the members of the committee and reaching consensus may take a considerable amount of time and effort.

Using of the 1993 EU directive in the Netherlands (dr. Aldert Piersma, RIVM)
In the EU system, the weights of evidence from human and animal data are combined (Table 1). In category 1, the evidence leading to the highest ranking of a substance 'known' to be a reproductive toxicant is largely based on human data. In category 2, the weight of evidence is translated into 'presumed' reproductive toxicity and is largely based on experimental animal studies. For those substances for which the evidence based on human and/or animal is not entirely convincing, a category 3 classification is most appropriate. In this currently used system, category 1 and 2 substances were labeled with the hazard statements R60 (May impair fertility) and/or R61 (May cause harm to the unborn child). A category 3 substance was labeled R62 (Possible risk of impaired fertility) and/or R63 (Possible risk of harm to the unborn child). A separate category 'Lactation' was used for substances that 'may cause harm to breastfed babies' (R64). The EU directive does not classify substances for which the data are insufficient (last category 'insufficient data on hazard and/or exposure' in the NTP levels of concern), contrary to the IARC classification of carcinogens ('Group 3 - The agent is not classifiable as to its carcinogenicity to humans', 6). In addition, there is no explicit identification or labeling of substances without a reproduction toxic effect, in line with the IARC group 4 classification ('The agent is probably not carcinogenic to humans') and the NTP system weight of evidence statements ('limited', 'some' or 'clear evidence of no effects', Figure 2).

Conversion to the new GHS compliant classification (dr. Aldert Piersma, RIVM)
The United Nations supported world-wide globally harmonized system (UN-GHS) for classification and labeling of hazards to human health uses a classification to indicate the intrinsic hazardous properties of substances or mixtures (7). The GHS system is also supported by the European Chemicals Agency (ECHA, see guidance document at their website, 8). For harmonization with GHS, the three EU categories are renamed from 1, 2 and 3 to 1a, 1b and 2, respectively (Table 1). Under GHS, new hazard statements are introduced. Category 1a and 1b substances are labeled "may damage fertility or the unborn child" (GHS hazard statement H360). Category 2 substances are labeled "suspected of damaging fertility or the unborn child" (H361). A third hazard statement identifies substances that 'may cause harm to breastfed children' (H362), as the result of impaired lactation, a problem with the quality of the mother's milk, or possible exposure of the child to toxic levels of a chemical through lactation.

*Table 1: Classification and labeling according to the 1993 EU directive (93/21/EEC) and globally harmonized system (GHS).*

| Scientific label | Supportive evidence from | Present (EU, 1993) | | | New (UN-GHS) | | |
|---|---|---|---|---|---|---|---|
| | | Cat | Label | Hazard Statement | Cat | Label | Hazard Statement |
| 'Known'[a] | Largely human | 1 | R60 | May impair fertility | 1a | H360 | May damage fertility or the unborn child |
| | | | R61 | May cause harm to the unborn child | | | |
| 'Presumed' | Largely non-human | 2 | R60 | May impair fertility | 1b | H360 | May damage fertility or the unborn child |
| | | | R61 | May cause harm to the unborn child | | | |
| 'Suspected' | Human or non-human | 3 | R62 | Possible risk of impaired fertility | 2 | H361 | Suspected of damaging fertility or the unborn child |
| | | | R63 | Possible risk of harm to the unborn child | | | |
| n/a | n/a | Lactation | R64 | May cause harm to breastfed babies | Lactation | H362 | May cause harm to breast-fed children |

[a] Sufficent evidence to support a causal relationship between exposure and effect in human populations.

The GHS system covers nearly all hazards that are covered in the present EU system and uses similar or equal criteria for classification. For hazard communication, the GHS system is equivalent to the EU system of hazard statements for single substances that can be used on labels and in material safety datasheets. For mixtures, however, GHS follows a different approach for substances with a human carcinogenic, mutagenic, and/or reproduction toxic hazard (CMR-substances) and for substances with acute toxicity, skin corrosives/irritants, and substances that cause serious eye damage/irritation. For hazards of a mixture, GHS uses lower values for concentrations of ingredients with respect to reproductive toxicity hazards and hazards for irritation/corrosion. The threshold for labeling in mixtures is set at ≥ 0.3 %, meaning that a mixture should be labeled when the concentration of a hazardous component exceeds this value. However, this value is not established scientifically but is the outcome of an administrative proposal.
The GHS system assumes 'reasonably expected use' in processes, technical operations, and household consumer contact, and reasonably foreseeable misuse, but not abuse. A limitation of the hazard characterization and classification of GSH is that it directly feeds into legislation, without considering the type of use (9). This marks the differences with the European policy on use of hazardous substances in REACH.
During the seminar the evaluation process was illustrated by some examples (chloroform and lithium chloride).

*Chloroform*
For chloroform, animal studies indicated developmental effects such as missing tail, imperforate anus, and cleft palate that were considered not secondary to maternal toxicity. Resorptions were also observed, but partly at higher doses with maternal toxicity. Human data indicated several adverse effects in offspring (e.g. cardiac defects, neural tube defects, oral clefts, low birth weight, and fetal death) but chloroform was never involved as a single substance. This is often the problem when evaluating organic solvents which are usually part of mixed exposures. Also, the epidemiologic studies involved questionnaire data, which could have introduced recall bias. Therefore, the human data set was considered to be of insufficient quality to assess causality and the subcommittee decided to classify chloroform as a 'Category 1b' substance, presumed to cause developmental toxicity in humans (see Table 1).

*Lithium chloride*
For lithium chloride, human data were available from prescriptions in patients suffering from manic depression. Some case reports described effects on libido and erection but most studies did not report any adverse effects. Overall the committee concluded that the human data were of insufficient quality and quantity to classify lithium chloride as a risk to fertility. For developmental effects, several older studies in patients showed malformations in offspring, but more recent studies did not reproduce these findings. The committee observed that acute therapy of 1,800 mg per day was related to adverse effects, whereas adverse effects were not observed in maintenance therapy at doses of 900 - 1,200 mg per day, suggesting that the therapeutic range of this drug is narrow.
Animal studies were performed to evaluate clinical use at human therapeutic exposure levels, which resulted in general toxicity in most animals. At lower exposure levels, only two studies reported adverse effects on fertility. However, in one study the number of animals tested per dose group was not reported. In the second study, rats received the substance of interest by subcutaneous administration, which is a non-relevant route of exposure in humans. Overall the subcommittee concluded that multiple human studies showed teratogenicity, although more recent studies with slightly lower doses were negative. Therefore, this substance was classified as a 'known to cause developmental toxicity in humans' (Category 1a).

## Discussion

*Predictions of reproductive toxicity using structure information*
The use of structure activity relationships (SARs) in the evaluation of toxicity is an interesting new development, which is often used in the toxicity characterization of mutagenic and carcinogenic substances. In these computer models, chemical structure information is used to predict the toxicity, using, so-called 'in silico' approaches. Such predictions can be used to support decisions about the need for animal testing. According to prof. Piersma such computer models are not considered useful for reproductive toxicity endpoints, because of the complexity of the reproductive system and the development of the fetus, involving multiple toxicity mechanisms. There have been some attempts to prepare SARs for predictions about estrogenic activity, but the use in classification is not feasible at this moment. A good example is the large diversity in reproductive and developmental adverse effects of phthalates. This

large group consists of many different substances with a huge variety in intrinsic reproductive health risks, which are not only dependent on the chain length of the molecule but also on many other, in part unknown, factors.

*Endocrine disruptors*
Concerning the evaluation of endocrine disruptors, prof. Piersma explained that evaluations are carried out considering the relation between dose and relevant adverse endpoints. Knowledge about suggested or confirmed toxicity mechanisms, such as endocrine disruption, is useful but these mechanisms in themselves do not suffice for classification decisions. Therefore, the scientific community does not support a separate procedure for classification of these substances, as was suggested by members of the European Parliament.

*Selection of studies*
Most animal studies are performed according to protocols established by the OECD. However, studies are often published that have one or more weaknesses. These studies may be excluded from the final evaluation. Since both the DECOS and CERHR outsource the preparatory work of collection of published studies to external parties, it is important to define criteria for inclusion or exclusion of studies. Dr. Thayer noted that the CERHR recently stopped using contractors for this work which is now done by NTP staff. In The Netherlands, it is customary to collect and describe all studies and leave it up to the experts of the committee to select studies that are too poor for use in the evaluation.

*Communication about reproduction toxic hazard and risk*
Standard statements are used to express the outcome of the evaluation for communication to the lay public. The audience felt that the CERHR uses too many levels of concern, which might be hard to handle for the public. It seemed difficult to perceive the differences between so many categories, e.g. between 'some' and 'limited' concern. The suggestion was made to use no more than three categories, e.g. 'positive', 'negative' and a third category representing the 'grey zone' in between.
In addition, this would be a good opportunity to harmonize expressions of 'levels of concern' and 'weight of evidence' among different agencies and for different fields in toxicology within the NTP. Such a harmonization will take much more effort to establish, but would help to simplify the communication about health risk to the public. CERHR includes information about the population at risk and also about the magnitude of exposure, which makes the level of concern a reflection of reproductive risk rather than reproductive hazard. In the Health Council subcommittee as well as in the GHS system, the toxicity hazard information is directly transferred into a hazard classification. This approach may also lead to problems in risk communication, such as concerning the reproductive hazard and risk of ethanol exposure in work situations versus alcohol drinking.

*Dealing with uncertainty and gaps in knowledge*
From the presentations and discussions during the seminar, it became clear that there is considerable uncertainty in the classification of reproduction toxic substances. This uncertainty is in part caused by limitations in the available data, e.g. an animal study involving a non-relevant route of administration or poor exposure characterization in epidemiologic studies. Also, relevant studies are often scarce, e.g. second generation studies or good epidemiologic studies. Dr. Roeleveld explained that the quantity of eligible epidemiologic studies adds to the evidence to support a causal relationship. A single epidemiologic study cannot be the basis for a judgment of causality because of the substantial influence of uncontrolled and most often unknown external risk factors. Multiple studies showing the same results will increase the weight of evidence to support a category 1 classification. This is different from animal data that use well-controlled experimental study designs.
Since CERHR is part in the NTP, research needs are fed back to other parts of the organization that have the ability to take such gaps in knowledge into account, when planning and funding toxicology studies. Expert committees have to work with what is available and sometimes have to combine the weight of different studies which each have limitations. At this point, a classification is always the result of consensus within a committee, involving expert judgment from the committee members.

*Further reading*
In the knowledge file CMR-substances of the knowledge base Arbokennisnet.nl, the classifications of CERHR and the Health Council are presented for CMR-substances (updated until the end of 2008) with reference to some of the underlying evidence (10).

## References


Stijkel A. (1992) Een reprotox instrument. Hulpmiddel voor bedrijfsgezondheidsdiensten ter advisering van werkgevers en werknemers inzake het beheersen en voorkomen van reprotoxrisico's in arbeidssituaties. Arbeidsinspectie, Ministerie van Sociale Zaken en Werkgelegenheid, Den Haag, S138-2

Netherlands Centre for Occupational Diseases (1999-2001) Guidance for registration of occupational diseases. Reproductive disorders resulting from occupational settings: www.beroepsziekten.nl/datafiles/K001.pdf

CERHR website: http://cerhr.niehs.nih.gov/

List of CMR classified substances in California: http://www.oehha.org/prop65/prop65_list/ Newlist.html

Health Council of The Netherlands: http://www.gezondheidsraad.nl/en

Preamble to the IARC Monographs: http://monographs.iarc.fr/ENG/Preamble/index.php

GHS system: http://www.unece.org/trans/danger/publi/ghs/ghs_rev02/02files_e.html)

ECHA guidance document: http://guidance.echa.europa.eu/docs/guidance_document/clp_en.pdf?vers=20_08_09

Potential effects on EU Downstream legislation: http://ec.europa.eu/enterprise/sectors/chemicals/files/ghs/ghs_sc_study_final_and_addendum_101207_en.pdf
Knowledge database (in Dutch): http://www.arbokennisnet.nl