From THE INSTITUTE OF ENVIRONMENTAL MEDICINE
Karolinska Institutet, Stockholm, Sweden

RISK ASSESSMENT OF ENDOCRINE DISRUPTING COMPOUNDS

Anna Beronius

Karolinska Institutet

Stockholm 2013

Cover illustration by Estella Larsson Beronius

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet. Printed by University Service, US-AB, Nanna Svartz väg 4, 171 77 Solna

ISBN 978-91-7549-144-8

To my children

ABSTRACT

During the past decade a growing number of chemicals have been identified as having endocrine disrupting properties in laboratory studies. Also, associations between exposure to such substances and endocrine-related health effects in the general population, as well as in wildlife, have been increasingly reported. This implies that past chemical regulation has failed to adequately protect human health and the environment. Endocrine disrupting compounds (EDCs) have lately been identified as substances of very high concern that should be phased out in new European (EU) legislations for e.g. industrial chemicals, plant protection products and biocides. There is thus an increased pressure on regulatory agencies to be able to efficiently and reliably identify, characterize and risk assess EDCs.

However, risk assessment of EDCs has proven complicated, in part due to the complex toxicity exhibited by substances that can interact with the endocrine system, and also because there are currently no generally agreed upon criteria within the EU or internationally that direct how to specifically identify compounds with endocrine disrupting properties.

The aim of this thesis project has been to identify how scientific uncertainties concerning the toxicity of EDCs can be reduced or handled to make health risk assessments of EDCs more transparent, systematic, and reliable. To that end literature studies were conducted that investigated the risk assessment process for EDCs within different regulatory frameworks in the EU, as well as the underlying toxicity data available to risk assessors and how the use of all available toxicity data can be improved. The much debated EDC bisphenol A (BPA) was used for a case study in a large part of this work.

A comparison of different regulatory frameworks within the EU showed that the regulatory risk assessment process, including underlying policies, criteria and requirements may differ for EDCs belonging to different regulatory groups, e.g. industrial chemicals, plant protection products or pharmaceuticals. The investigations within this project also showed that non-standard research studies, i.e. studies not conducted according to standardized regulatory test guidelines, fill data gaps and contribute information that could be particularly important for the identification and risk assessment of EDCs. However, non-standard studies were often criticized for having methodological limitations or being insufficiently reported, limiting their use in regulatory risk assessment. Regulatory agencies commonly gave more weight to standard than non-standard studies in risk assessment of BPA, despite the growing amount of research indicating that toxic effects at low doses were being overlooked.

A framework of criteria and guidelines intended to enable transparent and systematic evaluation of non-standard research studies, as well as guidance for how to report *in vivo* research to meet the requirements for regulatory risk assessment, was proposed. These tools are intended to facilitate the use of non-standard research studies in regulatory risk assessment and hopefully improve the reliability of risk assessment conclusions for EDCs.

LIST OF PUBLICATIONS

I. Beronius A, Rudén C, Hanberg A and Håkansson H (2009) Health risk assessment procedures for endocrine disrupting compounds within different regulatory frameworks in the European Union. Regul Toxicol Pharmacol 55:111-22.

II. Beronius A, Rudén C, Håkansson H and Hanberg A (2010) Risk to all or none? A comparative analysis of controversies in the health risk assessment of Bisphenol A. Reprod Toxicol 29:132-46.

III. Beronius A, Johansson N, Rudén C and Hanberg A (2013) The influence of study design and sex-differences on results from developmental neurotoxicity studies of bisphenol A, implications for toxicity testing. Toxicology. Published online ahead of print: http://dx.doi.org/10.1016/j.tox.2013.02.012.

IV. Beronius A, Molander L, Rudén C and Hanberg A. Facilitating the use of non-standard *in vivo* studies in health risk assessment – a proposal to improve evaluation criteria and reporting. *Manuscript*.

CONTENTS

LIST OF ABBREVIATIONS

ACTH	Adrenocorticotropic hormone
ADI	Acceptable daily intake
AF	Assessment factor
BMDL	Lower confidence limit of a benchmark dose
BPA	Bisphenol A
bw	Body weight
CRH	Corticotropin releasing hormone
DES	Diethylstilbestrol
DNEL	Derived no effect level
DNT	Developmental neurotoxicity
EC	European Commission
ECHA	European Chemicals Agency
EDC	Endocrine disrupting compound
ER	Estrogen receptor
EU	European Union
LOAEL	Lowest observed adverse effect level
MOA	Mechanism of action
MOE	Margin of exposure
MOS	Margin of safety
NOAEL	No observed adverse effect level
NTP-CERHR	National Toxicology Program Center for the Evaluation of Risks to Human Reproduction
OECD	Organisation for Economic Co-operation and Development
PC	Principal component
PCA	Principal component analysis
PLS	Partial least squares projection to latent structures
PoD	Point of departure
PPP	Plant protection products
QSAR	Quantitative structure-activity relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RfD	Reference dose
TDI	Tolerable daily intake
TG	Test guideline
UNEP	United Nations Environment Programme
US EPA	United Stated Environmental Protection Agency
WHO	World Health Organization
WoE	Weight of evidence

1 INTRODUCTION

This is a doctoral thesis within the field of regulatory toxicology that aims to identify and discuss weaknesses in the health risk assessment process for endocrine disrupting compounds (EDCs) and how this process can be improved.

Humans and wildlife are continuously exposed to a very large number of chemicals present in our environment. Risk assessment is the process of evaluating whether that exposure constitutes a risk to human health (health risk assessment) or the environment (environmental risk assessment). This thesis focuses on health risk assessment.

Concern regarding potential adverse health effects from EDCs in the human population is increasing (UNEP/WHO 2012). Over the last decade a growing number of chemicals that can be found in the environment have been identified as having endocrine disrupting properties in laboratory studies. There are also reports of increasing trends of endocrine-related disorders, such as some cancers and reproductive, metabolic and neurobehavioral disorders, in the human population, as well as observations of endocrine-related effects in wildlife. In their recent reports, the United Nations and World Health Organization (UNEP/WHO 2012) and European Environment Agency (EEA 2012) state that exposure EDCs may be a contributing factor to these observed effects. This implies that past chemical regulation has failed to adequately protect human health and the environment.

EDCs have been identified as substances of very high concern in the new European Union (EU) legislation for industrial chemicals (REACH) (EC 2006) and are also specifically mentioned as substances that should not be put on the market in the new EU legislations for plant protection products (EC 2009) and biocides (EC 2012). This increased legislative focus has put pressure on regulatory authorities to be able to efficiently and reliably identify, characterize and risk assess EDCs.

However, risk assessment of EDCs has proven complicated, in part because there are currently no generally agreed upon criteria within the EU or internationally that direct how to specifically identify compounds with endocrine disrupting properties or how to distinguish adverse effects of EDCs from normal regulation and function of the endocrine system. Risk assessment of EDCs has also been hampered by large scientific uncertainties because of the complex toxicity exhibited by hormonally active substances, such as varying and multiple mechanisms of action (MoA), different effects at high and at low doses, potential delayed on-set of effects and non-monotonic dose-response relationships (e.g. UNEP/WHO 2012; Kortenkamp et al. 2012). This complex toxicity challenges the methods by which chemicals are traditionally tested and evaluated for adverse health effects.

The primary focus of this thesis has been to identify and address factors that contribute to scientific uncertainties in the health risk assessment process for EDCs, such as assumptions and principles traditionally used in toxicity testing and risk assessment, and ways to reduce this uncertainty.

2 BACKGROUND

2.1 PRINCIPLES OF HEALTH RISK ASSESSMENT

Health risk assessment of chemicals entails identifying and characterizing the risk of adverse health effects occurring at current exposure levels and is commonly conducted by national or international regulatory agencies, such as the Swedish Chemicals Agency (KemI) or the European Food Safety Authority (EFSA), or expert organs, such as the World Health Organization (WHO). Although it is performed with the overall objective to protect human health, the primary purpose of the risk assessment may vary depending on, for example, the type of substance, its intended use and the human exposure scenario. The aim of health risk assessment may be e.g. to establish health based guidance values for contaminants in food and compare these values to the estimated exposure levels. In other cases the purpose may be to evaluate human health risks as basis for authorization or restriction of the use of a chemical.

It is not within the scope of this thesis to describe in detail the process and principles of health risk assessment but some of the most fundamental aspects are summarized below.

2.1.1 Components of health risk assessment

Different organizations describe the structure of risk assessment in slightly different ways but commonly containing the same components (e.g. ECHA 2011a; WHO/IPCS 2010). Simplified, health risk assessment can be said to consist of three main parts: hazard assessment (including hazard identification and hazard characterization), exposure assessment and risk characterization (Figure 1). These parts are briefly described below based on information from guidance documents provided by the European Chemicals Agency (ECHA 2011b; ECHA 2012) and the WHO (WHO/IPCS 2010, 2009). The work done within this PhD-project is primarily relevant for the hazard assessment part of risk assessment.

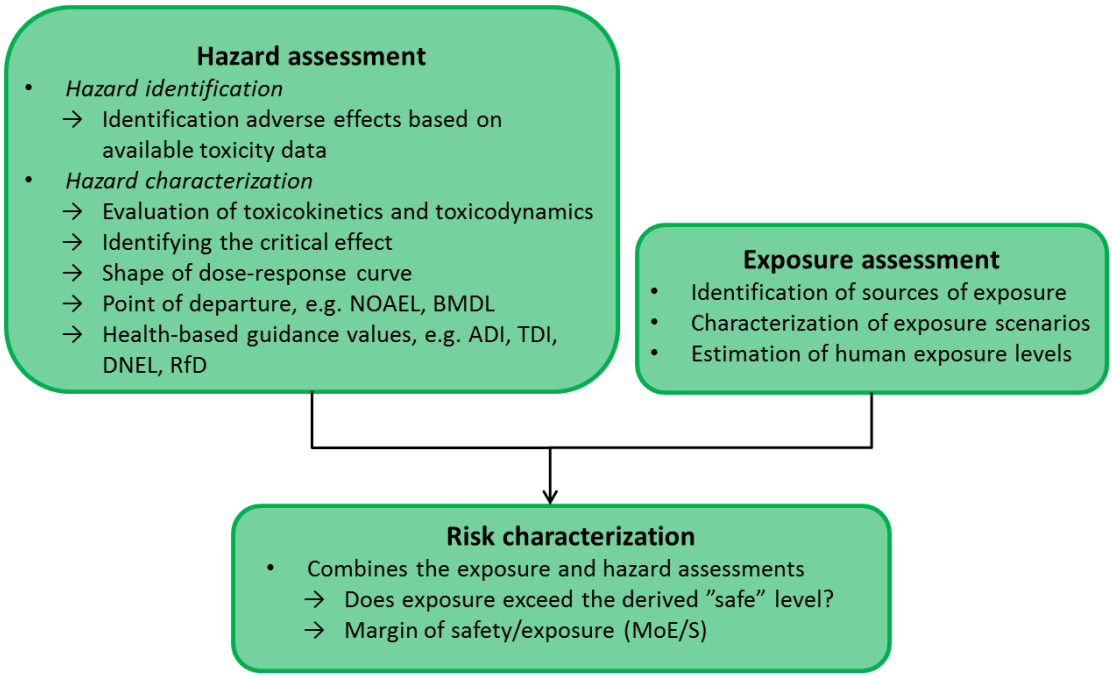

Figure 1. A simplified scheme of the components of health risk assessment.

2.1.1.1 Hazard assessment

The hazard assessment aims to identify and characterize the adverse effects of a compound and involves evaluating existing data, such as quantitative structure–activity relationship (QSAR) data, *in vitro* and *in vivo* toxicity studies, as well as epidemiological data (see section 2.1.2 for further discussion on data evaluation). Commonly, toxicity studies conducted in animals (*in vivo*) provides the primary basis for hazard assessment. Hazard assessment can be divided into two steps, hazard identification and hazard characterization.

In the *hazard identification* step the type and nature and of potential adverse health effects of the compound are identified.

The *hazard characterization* step entails further describing the toxicity of the compound. This includes evaluating the toxicokinetics, i.e. the absorption, distribution, metabolism and excretion of the substance, and its toxicodynamics, i.e. the molecular events at the target tissue, as well as the relevance of these aspects to humans.

In hazard characterization the most sensitive adverse effect that is relevant to human health should be identified. This is often referred to as the critical effect. An important aspect is to characterize the dose-response relationship and identify a point of departure (PoD) for the critical effect, such as a no observed adverse effect level (NOAEL) or a lower confidence limit of a benchmark dose (BMDL), i.e. the highest dose considered not to have resulted in significant adverse effects in test animals. In some risk assessments the PoD is used as basis for deriving health-based guidance values considered “safe” for humans, such as an acceptable or tolerable daily intake (A/TDI),

a derived no effect level (DNEL) or a reference dose (RfD), by applying assessment factors (see section 2.1.4). As far as possible the mechanism of action (MoA) for the critical effect, i.e. the molecular and cellular events that the compound initiates and that subsequently leads to the critical effect, should be identified and described.

2.1.1.2 Exposure assessment

In the exposure assessment possible sources of exposure as well as any exposure scenarios relevant to humans are identified and exposure levels in human populations are estimated.

The human exposure levels resulting from different scenarios can be calculated based on concentrations of the substance e.g. in food, dust and air, and data on the intake of certain food items, respiratory rate, etc. Alternatively, the total exposure to a substance can be estimated based on concentrations of that substance, or its metabolites, in human urine, blood or other tissues, if such data is available.

The exposure assessment also has to take into consideration life stage-specific exposure. Exposure patterns and scenarios differ between, for example, adults and infants due to differences in the main sources of exposure, e.g. toys, food and cosmetics, as well as differences in behavior and physical parameters, such as food intake per kg body weight or respiratory rate.

2.1.1.3 Risk characterization

In the risk characterization step the hazard and exposure assessments are combined to draw conclusions about the risk to human health. One method is to see if, and by how much, estimated human exposure levels exceed the derived "safe" levels, e.g. the ADI, TDI or DNEL. Another approach is to divide the PoD (e.g. the NOAEL or BMDL) by the estimated exposure to calculate margins of safety (MoS), sometimes also referred to as margins of exposure (MoE). The sufficiency of the MoS is determined on a case-by-case basis, but typically a MoS $\geq$ 100 is considered sufficient.

2.1.2 Evaluation of toxicity data for risk assessment

As mentioned above, different types of data may be used in health risk assessment, e.g. QSAR-data or information from studies conducted in cells or tissues (*in vitro* data), animals (*in vivo* data), as well as epidemiological studies investigating associations between chemical exposure and health effects in humans. Understandably, epidemiological studies have the potential to deliver very relevant information for health risk assessment, as they investigate the association between exposure and health effects in humans. However, epidemiological studies have a number of limitations. For example, human populations are exposed to an uncontrolled mixture of chemicals and other environmental factors, so it may be difficult to draw conclusions about the exposure to a single compound. Also, human data is lacking for many compounds.

In contrast, experimental toxicity studies are not conducted in humans but have the advantage of allowing the investigator to control the study population (i.e. the cells or animals), the exposure to the chemical under study, as well as any environmental and confounding factors which may influence toxicity. *In vivo* studies conducted in animals

are commonly considered especially useful for health risk assessment since they investigate effects in intact organisms. Health risk assessment is therefore, in most cases, primarily based on data from *in vivo* toxicity studies, i.e. animal data provides the basis for conclusions about the critical effect and dose-response relationships.

2.1.2.1 Reliability and relevance

Evaluation of the reliability and relevance of toxicity studies for hazard identification and characterization is an integral part of the risk assessment process.

Reliability indicates the "quality" of the study, e.g. the reproducibility of results and degree of certainty in these results, while relevance relates to how appropriate the study is in relation to the human health effect and exposure scenario under investigation (ECHA 2011c).

There are different methods available for evaluating the reliability and relevance of (eco)toxicity studies (e.g. Durda and Preziosi 2000; Hobbs et al. 2005; Klimisch et al. 1997; Küster et al. 2009). However, most are primarily for evaluation of reliability with less focus on relevance. Further, application of these different methods have been shown to result in different conclusions concerning reliability of the same study (Ågerstrand et al. 2011).

The use of the Klimisch-method (Klimisch et al. 1997) when evaluating toxicity data for health risk assessment has been commonly promoted by regulatory agencies, such as the European Chemicals Agency (ECHA 2011c) and the US EPA (USEPA 1999). However, this method puts a lot of emphasis on the application of standardized guidelines and Good Laboratory Practices (GLP) and research studies not adhering to these standards can at best be categorized as "reliable with restrictions". In practice it means that if standard studies are available, they will always be given more weight than non-standard studies in risk assessment.

2.1.2.2 Standard and non-standard toxicity studies

Toxicity studies that are conducted for the purpose of regulatory risk assessment, e.g. in connection with authorization processes for use and putting substances on the market, are generally required to comply with standardized, internationally validated and accepted test guidelines, such as the Organisation of Economic Co-operation and Development (OECD) test guidelines (TG). Standardized TGs give detailed directions on how to design, execute and report studies for different types of toxicity, including which animal models and methods to use and what endpoints to measure. Further, these studies should follow GLP, a set of standards for study execution and reporting. Standardized TGs and GLP standards are intended to guarantee high reliability of toxicity studies.

A major disadvantage of standard methods is that they do not always represent the most relevant or sensitive testing approach given the type of compound or endpoint under investigation. This limitation of current standardized testing paradigms has been pointed out for EDCs (e.g. Kortenkamp et al. 2012; Zoeller et al. 2012) and is further discussed in section 2.2.5.2.

Research studies conducted at academic and research institutions often do not comply with standardized TGs. Such studies are hypothesis-driven and their study design and execution are aimed at investigating specific questions related to that hypothesis (Myers et al. 2009). Research studies also commonly utilize novel and sensitive methods measuring endpoints considered specifically sensitive and relevant to the study hypothesis. Another reason is that standardized TGs require the use of a large amount of animals, something that is specifically difficult to attain ethical permits for in academic research (Myers et al. 2009). Nonetheless, non-standard research studies[1] undergo a peer-review process in connection with publication in scientific journals and should fulfill general quality criteria for scientific investigations, e.g. the control of relevant variables, comparison to appropriate control groups, and proper reporting of the results etcetera. However, in the regulatory setting, e.g. for the purpose of chemicals risk assessment, the reliability of non-standard studies is often questioned for reasons such as suffering from methodological limitations and/or being poorly reported (Alcock et al. 2011; Hengstler et al. 2011).

2.1.2.3 Weight of evidence

In health risk assessment it is often stated that a "weight of evidence approach" was used when summarizing available toxicity data for hazard assessment. Exactly what is meant by weight of evidence (WoE) and the methods and criteria used when this approach was applied are, however, seldom defined. In the scientific literature the term WoE has been used to imply a number of different methods or concepts. The most common use of WoE in risk assessment is as a general concept for summarizing, synthesizing and interpreting a body of evidence, but the term has also been used to describe other ideas, such as (more or less clearly defined) quantitative methods for applying different weights to individual toxicity studies (Weed 2005).

2.1.3 The dose-response relationship

A central principle in toxicology is the paradigm of the dose-response relationship. This principle is traditionally based on two main assumptions about the dose-response of a toxic compound: 1) there is (in most cases) a threshold dose for effect, and exposure to doses below this threshold is assumed to not induce any adverse effects, and 2) the effect will increase with increasing dose until a maximum response is reached, generating a sigmoidal dose-response curve (Figure 2A).

Genotoxic (or mutagenic) carcinogens are often exempted from the assumption regarding threshold for effect (ECHA 2011b; US EPA 2005). The primary reason is that the MoA of these compounds, i.e. the induction of genetic damage by, for example, adduct formation or strand breaks, implies that exposure to one single molecule could potentially initiate events that could ultimately result in tumor formation and cancer. In other words, based on knowledge concerning the MoA of these compounds it can be argued that, theoretically, there is no threshold for effect and the dose-response is linear at low doses (Figure 2B).

[1] NOTE: In some parts of this thesis work non-standard studies are referred to as non-guideline studies.

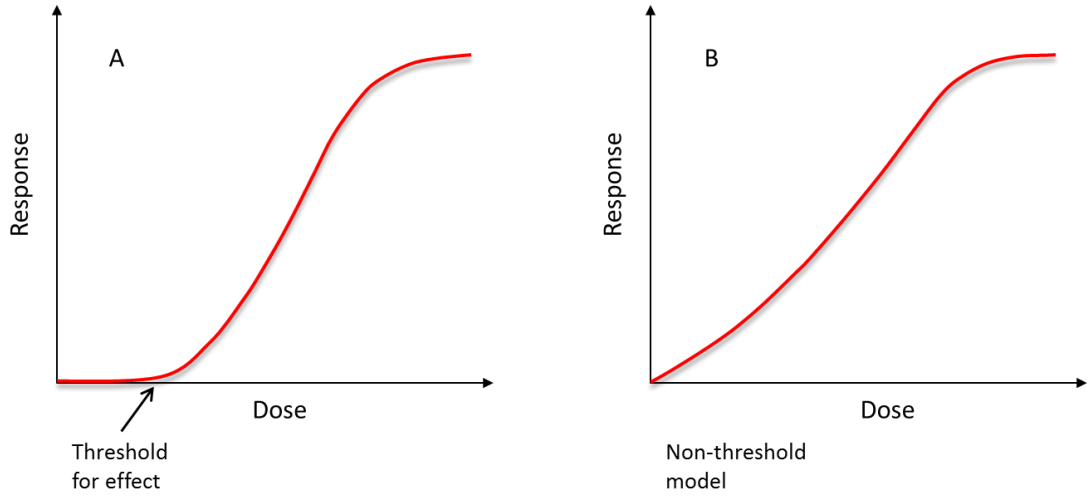

Figure 2. The traditional dose-response relationship (A) has a sigmoidal shape indicating that there is a threshold dose for effect below which no (adverse) effects occur. Above the threshold the effect will increase with increasing dose until reaching some maximum response. For genotoxic carcinogens a non-threshold model (B) is commonly assumed.

Toxicity studies do commonly not include more than three dose groups and a negative control group (with zero exposure to the compound), often less. Conclusions regarding the shape of the dose-response can thus not be made solely on experimental observations but are heavily reliant on the assumptions stated above.

In risk assessment of non-genotoxic substances the assumption of a threshold and monotonic dose-response curve is used to identify e.g. a NOAEL and LOAEL and to determine a PoD for the critical effect, on which conclusions regarding "safe" levels of human exposure can be based.

It is important to note that a NOAEL is no true "no effect level" below which no adverse effects occur. The NOAEL is the highest dose administered in a toxicity study that did not result in observations of *statistically significant* adverse effects compared to the response in unexposed control animals. The presence or absence of a threshold for *any* substance can never be experimentally proven (Slob 1999). All experiments have a limit of detection below which statistically significant effects cannot be shown, i.e. no conclusion regarding the shape of the dose-response curve can be made below this detection limit. Also, to generate an exact dose-response curve would require an infinite number of doses and infinitely precise measures. In other words, the determination of a threshold, e.g. a NOAEL, relies on the statistical power of the study, as well as the choice and spacing of dose levels, and subtle effects or effects in sensitive individuals below the NOAEL are expected (Davis et al. 2011).

2.1.4 Extrapolation from animal data to human health

As mentioned above, experimental toxicity studies in animals where the exposure to the substance under study, as well as the surrounding environment, can be controlled is most often used as basis for health risk assessment. A fundamental default assumption of health risk assessment is that findings in animal studies are relevant for humans, i.e. that the observed effects can be suspected to occur also in humans, unless the opposite is proven to be true (e.g. Boobis et al. 2008; JRC 2013).

Extrapolation from animal data to human health, for example to calculate health-based guidance values, such as an ADI or TDI, or to calculate a MoS, is commonly done by dividing the PoD derived from animal studies, e.g. a NOAEL, by assessment (or uncertainty) factors (AF) (Falk-Filipsson et al. 2007; Kalberlah et al. 2003). AFs are used to account for differences in sensitivity between species and between individuals or, importantly, the lack of such knowledge. Historically, the default assessment factor for health risk assessment has been 100, consisting of a factor 10 for extrapolating from the test animal to humans and a factor 10 to account for differences between individuals. In other words, in the absence of data proving otherwise, humans are assumed to be more sensitive than the test animals. The assessment factor may however be adjusted (up or down) depending on available knowledge about, for example, species differences in toxicokinetics, and confidence in the data material and additional assessment factors may be added to account for other considerations, such as uncertainty in the NOAEL or lack of data, the nature and severity of the effect(s), duration of exposure or route-to-route extrapolation (Falk-Filipsson et al. 2007).

2.1.5 Adversity

Traditionally, risk assessment is applied to protect against *harmful* effects of chemical substances. The underlying principle is that an organism may respond physiologically to exposure to a compound in a way that can be considered adaptive and not lead to detrimental (adverse) health effects (e.g. ECHA 2011b; WHO/IPCS 2009). Hazard assessment thus entails identifying the *adverse* effects of toxic compounds. The International Programme on Chemical Safety (IPCS) of the WHO (WHO/IPCS 2009) has defined an adverse effect as:

"Change in the morphology, physiology, growth, development, reproduction or lifespan of an organism, system or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences."

However, there are no generally accepted criteria for adversity and the distinction between adverse and non-adverse health effects is dependent on expert judgment (see section 2.1.6 below). For example, in the case of EDCs an important issue of discussion is the distinction between endocrine disruption, an implied adverse event, and endocrine modulation, which can be regarded as a compensatory, or adaptive, event (EFSA 2013), but there are no clear criteria for when these modulatory events become adverse.

2.1.6 Expert judgment

Although there are certain rules and criteria set up for the risk assessment process by different authorities and organizations, for example for how to select and evaluate toxicity studies and how to determine the critical effect, (e.g. ECHA 2011b; WHO/IPCS 2009), risk assessment is inevitably reliant on the expert judgment of the risk assessor(s) (Weed 2005). Expert judgment is dependent on the knowledge, views and experiences of the of the risk assessor.

Since the properties, uses and exposure scenarios vary for different chemicals, a risk assessment process that is too rigid and does not allow for the individual expertise of risk assessors to influence conclusions may not be able to account for all the relevant aspects of the substance that is being assessed. The use of expert judgment allows the assessment process to be flexible enough to potentially evaluate the most relevant risks and be as protective as possible. The identification of adverse effects and the evaluation of the reliability and relevance of toxicity studies for risk assessment are examples of aspects of the risk assessment process where expert judgment plays an important role.

On the other hand, the use of expert judgment introduces value-based assumptions to the assessment, and it is thus of key importance that these assumptions are transparently described and justified (Wandall 2004).

2.2 ENDOCRINE DISRUPTING COMPOUNDS

Compounds that are hormonally active, meaning they can interact with the endocrine system in one way or another, can be both naturally occurring, e.g. phytoestrogens, or manmade. Such substances are present for example in food (as natural constituents, pesticide residues or migrating from contact materials), toys, cosmetics, textiles, medical equipment and construction materials, meaning that we are exposed to many of them every day throughout our lifetime (UNEP/WHO 2012).

However, different definitions of what constitutes an EDC have been proposed (Table 1). The main difference between these definitions is the importance attributed to plausible causality between an endocrine MoA and an adverse health effect. The definitions proposed by the European Commission (EC 1996) and the WHO/IPCS (2002) require that an adverse health effect occurs as a result of endocrine action, while the US Environmental protection Agency (EPA) and Endocrine Society definitions imply that interaction with the endocrine system in itself can be considered as endocrine disruption.

Table 1. Definitions of an EDC proposed by different organizations.

Organization (year)	**Definition**	**Reference**
US EPA (1996)	An exogenous agent that interferes with the production, release, transport, metabolism, binding action or elimination of natural hormones in the body responsible for the maintenance of homeostasis and the regulation of developmental processes.	Kavlock et al. 1996
European Commission (1996)	An exogenous substance that causes adverse health effects in an intact organism, or its progeny, secondary to changes in endocrine function.	EC 1996
WHO/IPCS (2002)	An exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub) populations.	WHO/IPCS 2002
The Endocrine Society (2012)	An exogenous chemical, or mixture of chemicals, that interferes with any aspect of hormone action.	Zoeller et al. 2012

The definition proposed by the Endocrine Society is especially open for interpretation. It builds on the definition from the US EPA but was intentionally made even more general in order "to account for current and future information about the range of actions through which chemicals may influence the endocrine system" (Zoeller et al. 2012). Further, this definition implies that observations *in vitro* are enough to identify a substance as an EDC, as no requirements of effects *in vivo* are clearly stated.

2.2.1 Basic features of the endocrine system

The endocrine system regulates the development and function of essentially all cells, tissues and organs in an organism throughout its lifetime (Molina 2010; Tortora and Grabowski 1996). As such it controls various vital processes, such as reproduction, growth and development, metabolism and mood. The endocrine system also helps maintain homeostasis if an organism is subjected to any type of stress, such as infections, trauma, emotional stress, dehydration, starvation, hemorrhage, temperature extremes, etc. Perturbations to normal endocrine function leads to dysregulation of these processes and may result in a wide spectrum of endocrine-related diseases, such as goiter, diabetes, growth inhibition, certain types of cancer and reproductive problems. Some major endocrine glands and tissues, as well as examples of hormones, are illustrated in Figure 3.

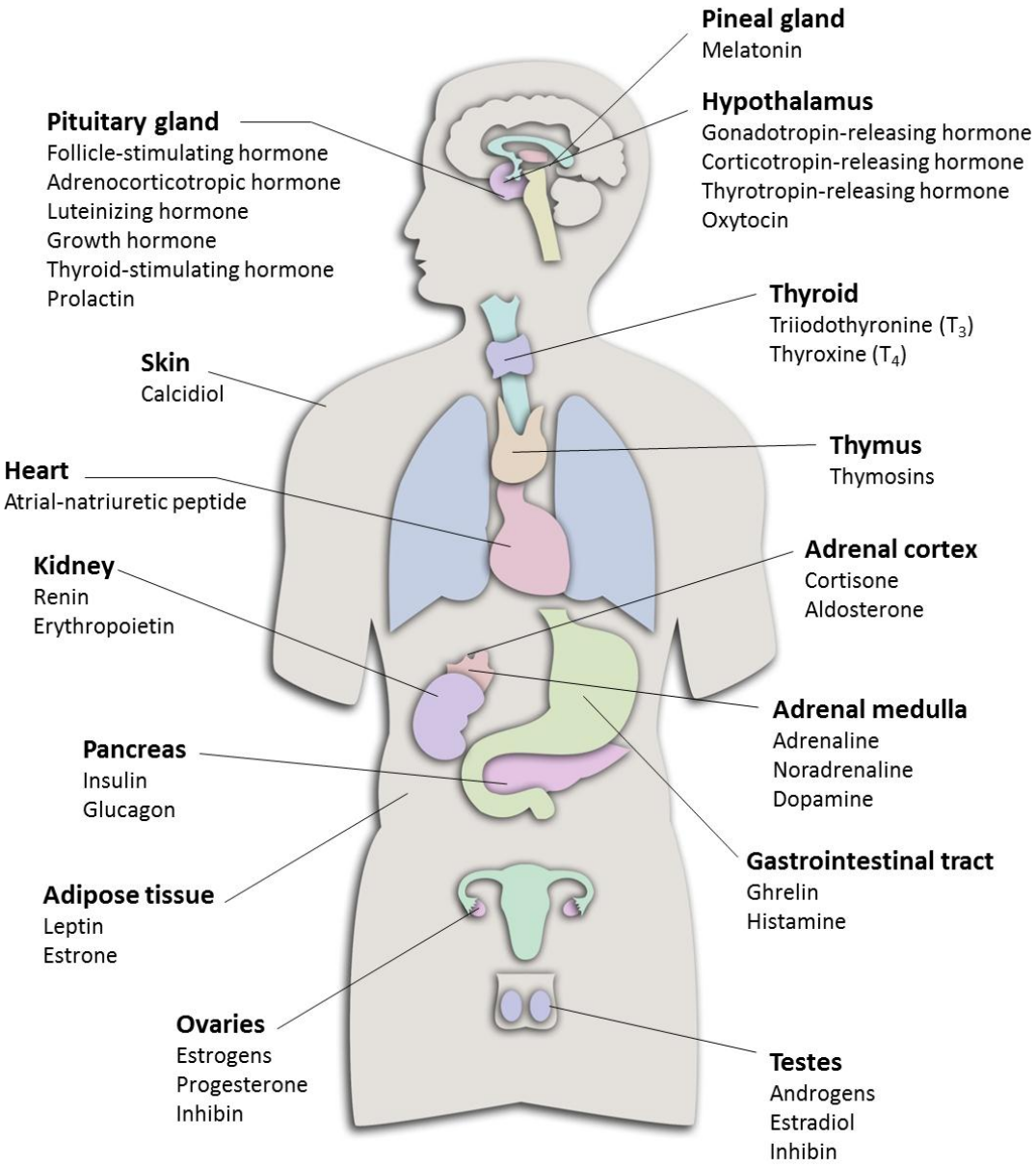

Figure 3. Some of the major endocrine glands and tissues and examples of hormones produced and released at each site. (Illustration by Viktoria Berglund.)

The endocrine system acts via chemical messengers (hormones) that are released from endocrine glands and tissues into the bloodstream to travel to target tissues. There are many different hormones, which are commonly divided into four groups: 1) steroids, such as estrogens and androgens, 2) biogenic amines, such as the thyroid hormones (T_3 and T_4), histamine and serotonin, 3) peptides and proteins, such as insulin and oxytocin, and 4) eicosanoids, such as prostaglandins (Tortora and Grabowski 1996).

Hormones exert their effects by binding to specific receptors on the surface of cell membranes or intracellular receptors in target tissues (Molina 2010) as illustrated in Figure 4. Interactions with membrane-bound receptors start off cascades of signaling events within the cell, e.g. kinase activation and phosphorylation of proteins, which lead to specific biological responses of the cell. This chain of events can also lead to the

activation of transcription factors and transcription of specific target genes. Hydrophobic hormones, e.g. estrogens and androgens, can cross the cell membrane and interact with intracellular receptors. These receptors are ligand-regulated transcription factors and the hormone-receptor complex initiates transcription of target genes.

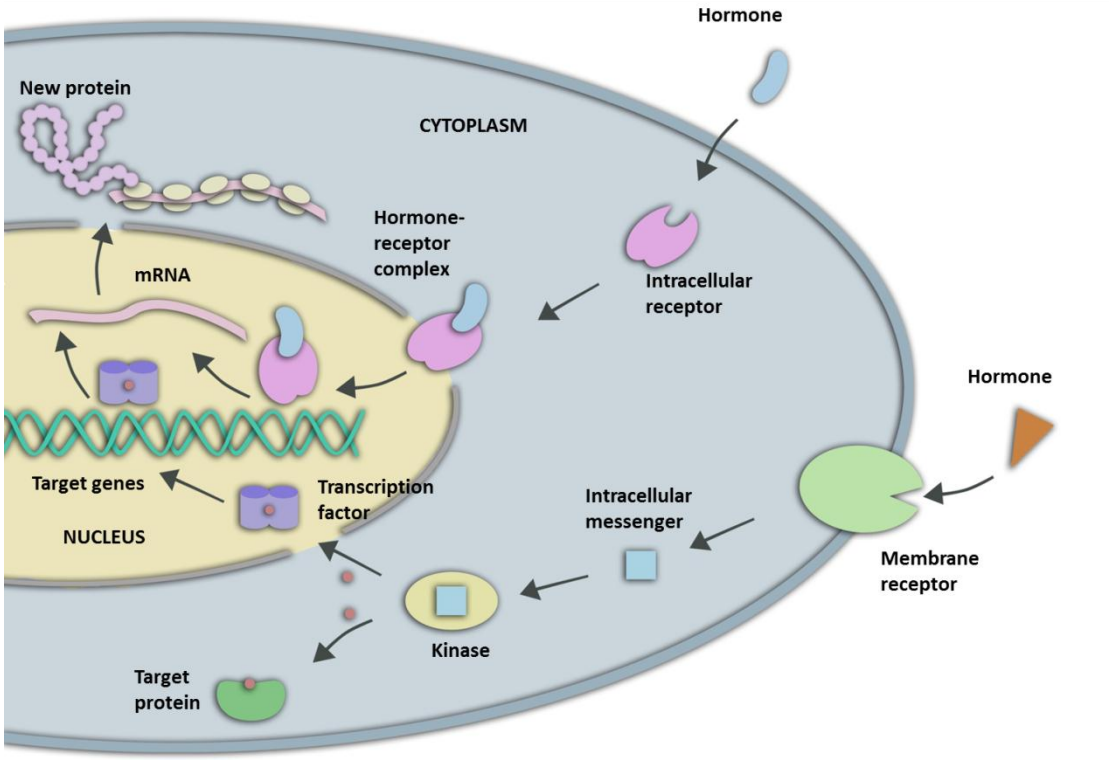

Figure 4. Schematic figure illustrating some general cellular pathways of hormone action resulting from hormones interacting with intracellular receptors or receptors on cell membranes. (Illustration by Viktoria Berglund.)

A single hormone may be able to bind to and activate different types of receptors. For example, estrogens can act via different types of intracellular receptors, e.g. estrogen receptor (ER) α and β, as well as via receptors on the cell membrane (Nadal et al. 2001). The response to a single hormone may differ between tissues, depending on the type of receptors that are expressed. Also, the responsiveness of tissues to hormone signaling may be controlled by up- or down-regulation of hormone receptors.

Endocrine signaling is controlled by positive and negative feedback loops between endocrine glands and tissues (Molina 2010). For example, emotional stress causes the hypothalamus to release corticotropin releasing hormone (CRH), which stimulates the pituitary to release adrenocorticotropic hormone (ACTH). In response to ACTH the adrenal cortex produces and releases cortisol, adrenaline and noradrenaline. These hormones bind to receptors in many different tissues and regulate blood pressure, metabolism, mood, immune system and many other functions in response to stress. Increased levels of cortisol, adrenaline and noradrenaline in the blood stream are also detected by the hypothalamus and pituitary and work by negative feedback to decrease the production and release of CRH and ACTH.

Some important features of the hormone system related to the issue of endocrine disruption are (Molina 2010; Zoeller et al. 2013):

- The binding of hormones to receptors. As a result, hormone action is saturable. It also means that the effect of a hormone is dependent both on its affinity for the receptor and the number and type of receptors present in the target tissue.
- Maximum effects are reached at concentrations well below receptor saturation.
- Endogenous hormones act at very low concentrations.
- Potency of a hormone is not equal to its binding affinity to hormone receptors.
- The same hormone can have different functions during different life-stages.
- Hormones may act as agonists or antagonists of each other, or in a synergistic fashion.

2.2.2 Mechanisms of endocrine disruption

Endocrine disruption is not a MoA in itself; rather it is a collection of different possible mechanisms which can lead to perturbations of the normal function of the endocrine system, i.e. binding to hormone receptors or interfering with the production, transport or metabolism of endogenous hormones (Diamanti-Kandarakis et al. 2009; Kortenkamp et al. 2012; UNEP/WHO 2012). As such, an "endocrine *mode* of action" is often discussed, which is a more general concept than MoA and that does not intend to describe in detail the molecular basis of a toxic effect (e.g. EFSA 2013; JRC 2013).

Some EDCs are hormone receptor agonists, meaning they bind to hormone receptors and activate transcription of endocrine-responsive genes. Such substances will mimic the characteristics of endogenous hormones. However, their actions will not necessarily result in the same patterns of molecular events or effects as endogenous hormones, depending on their potency and the complexity introduced e.g. by multiple MoA (Zoeller et al. 2012). EDC can also act as hormone receptor antagonists, i.e. "blocking" a hormone receptor and subsequently gene transcription.

Importantly, knowledge is lacking regarding the MoA for EDCs as well as the relationship between these molecular events and subsequent adverse health effects.

2.2.3 Complex toxicity of EDCs

EDCs display a complex toxicity which challenge traditional toxicological assumptions and contribute to making toxicity testing and health risk assessment difficult (Kortenkamp et al. 2012; UNEP/WHO 2012; Zoeller et al. 2012), for example:

- Varied and multiple MoA, e.g. a single EDC may have both estrogenic and anti-androgenic properties, or may be metabolized into compounds with different MoA from the original compound. Further, since hormone receptors are differentially expressed in different tissues the MoA of an EDC may differ between tissues.
- Large differences in sensitivity between and within species. For example, some rat strains have been reported to be particularly insensitive to estrogens

(Hossaini et al. 2003; Long et al. 2000) making the choice of animal model used for toxicity testing critical.

- Delayed onset of effects, i.e. effects that appear long after exposure and sometimes only in subsequent generations, as exemplified by the estrogen-replacement drug diethylstilbestrol (DES). DES was administered to pregnant women from the 1940's to the 1970's to alleviate pregnancy-related complications and reduce the risk of miscarriages. No apparent adverse side-effects in mothers or babies were noted. However, many daughters born to women treated with DES were later afflicted with a rare form of vaginal cancer, which did not manifest until after puberty (Newbold 2004).
- Effects that occur at very low doses and non-monotonic dose response relationships. These aspects have been extensively debated internationally during the last couple of years and are discussed in more detail below.

2.2.3.1 Low-dose effects of EDCs and the question of a threshold

As mentioned above, endogenous hormones act at very low concentrations and slight fluctuations may result in biological response (Tortora and Grabowski 1996). EDCs act by the same mechanisms and against a background of endogenous hormones, adding to the actions of these hormones and increasing the response of already ongoing biological processes. It therefore stands to reason that EDCs are capable of inducing effects at very low doses, especially during particularly sensitive and critical windows of development. Indeed, there are many reports of effects occurring after administration of low doses of EDCs both in *in vitro* and *in vivo* toxicity studies (Richter et al. 2007; Vandenberg et al. 2012).

There is, however, no generally accepted definition of "low dose", and the term has been used with slightly different meanings in the literature, implying e.g:

- doses below those commonly used in standardized regulatory toxicity studies, or below a previously established regulatory NOAEL for the compound,
- doses at or below a health-based guidance value, e.g. a TDI, or
- doses resulting in exposure corresponding to human exposure levels measured e.g. in blood or urine.

General agreement on a single definition of low dose may not be necessary but, given the different possibilities for interpretation, it is important to clearly define what is meant whenever low-dose effects are discussed.

Based on arguments concerning the shared MoA between endogenous hormones and EDCs, i.e. receptor-binding, and the concept of "additivity-to-background" many experts claim that a threshold for effect for EDCs cannot be assumed (Kortenkamp et al. 2012; Vandenberg et al. 2012; Zoeller et al. 2012). These arguments are especially strong for early molecular events, such as the first interaction of and EDC with the receptor and gene-transcription activation or suppression (reviewed in Beronius and Hanberg 2013). However, the complexity of a biological system, including interactions between a myriad of different signaling pathways and the presence of compensatory mechanisms intended to maintain homeostasis, could mean that thresholds exist for

"higher" endpoints such as behavior, reproduction, organ weights and growth (Conolly and Lutz 2004).

Importantly, sensitivity to EDC toxicity is expected to vary in the general population, as is the pattern of exposure to mixtures of EDCs and other substances. Consequently, even if a threshold for a certain effect does exist it could be "masked" by individual variation (White et al. 2009).

2.2.3.2 Different effects at low and high doses and the issue of non-monotonicity

Due to the nature of receptor-binding, endogenous hormones, as well as EDCs, can exhibit different effects at high and at low doses and so-called non-monotonic dose-response relationships (Diamanti-Kandarakis et al. 2009; Zoeller et al. 2012). Non-monotonicity means that the dose-response curve changes direction at least once over the dose-range, resulting e.g. in a U- or inverted U-shape or even a bi-phasic curve, as illustrated in Figure 5.

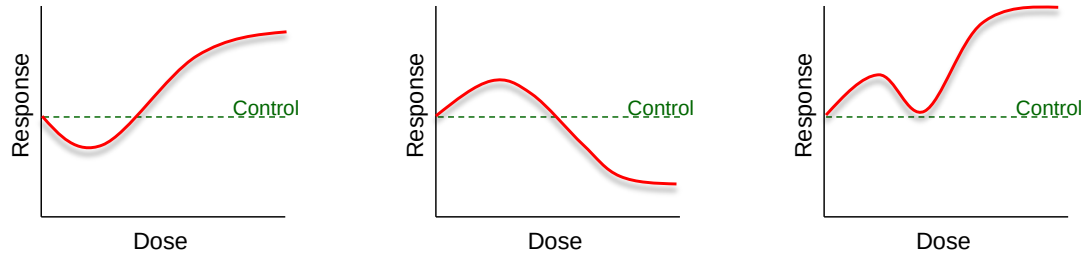

Figure 5. Illustrations of U-shaped, inverted U-shaped and bi-phasic non-monotonic dose-response relationships.

Non-monotonic dose-response curves have been reported both *in vitro* and *in vivo* for several EDCs (reviewed in Vandenberg et al., 2012), such as phthalates (Ge et al. 2007), pesticides (Brodeur et al. 2009; Palanza et al. 2001), PCBs (Love et al. 2003) and dioxins (Fan et al. 1996). A common and well characterized example of non-monotonicity is the "tamoxifen flare" phenomenon. Tamoxifen is an anti-estrogenic drug used to treat certain breast cancers by inhibiting estrogen-dependent proliferation of cancer cells. However, at low doses, i.e. below the therapeutic dose, tamoxifen actually *induces* cell proliferation in estrogen-dependent cells (reviewed in Howell 2001; Vandenberg et al. 2012). This is manifested by growth of breast tumors, i.e. a "flare", during the first two weeks of administration before therapeutic circulating concentrations are reached, which was described already in the 1970's (Plotkin et al. 1978).

There are several mechanisms, especially involving receptor-binding, that can explain non-monotonicity (Diamanti-Kandarakis et al. 2009; Vandenberg et al. 2012; Zoeller et al. 2012), for example:

- Competing mechanisms, e.g. at low doses estradiol initiates a proliferative response in target cells but at high doses estradiol becomes cytotoxic resulting in a subsequent decrease in viable cell number.

- Over-stimulation of hormone receptors that leads to down-regulation of receptors resulting in a decrease in response at high doses.
- Different genes being activated in response to high vs. low concentrations of a receptor ligand.
- Receptor-selectivity, e.g. at low doses a compound may bind primarily to membrane receptors but at higher doses intracellular receptors are activated.
- Competition for receptors. An EDC with low potency compared to the endogenous hormone may at low doses add to the effect of the hormone. But at higher concentrations the EDC out-competes the hormone for receptor-binding sites but leads to a decrease in response, due to the lower potency of the EDC-receptor complex and thus reduced gene-transcription.

2.2.4 Sensitive windows of exposure

Endogenous hormones have different functions, i.e. produce different effects, during different periods of an organisms' life cycle (Tortora and Grabowski 1996; Zoeller et al. 2012). During fetal and early life stages the endocrine system is responsible for the development of tissues, organs and their functions, while in adulthood it generally regulates functions of tissues and organs. Exposure to an EDC during early development can thus be specifically detrimental and lead to permanent effects in an individual that can be manifested at much later life stages (Diamanti-Kandarakis et al. 2009; Zoeller et al. 2012). The developing organism is also more sensitive to EDC toxicity than the adult, i.e. developmental effects occur in offspring at doses much lower than those causing effects in the mother (e.g. UNEP/WHO 2012). Thus, the timing of exposure is an important determining factor of EDC toxicity.

The roles of estrogens, androgens and thyroid hormones are specifically important during fetal development.

2.2.4.1 Estrogens

Estrogen receptors are present in a large number of tissues in both males and females but the functions of estrogen in all these structures are not known (Molina 2010). Estrogens play an important role in the development as well as maintenance of female reproductive structures, and are also responsible for the development of secondary female characteristics (Tortora and Grabowski 1996).

The actions of estrogens, particularly estradiol, are also crucial in early development of the brain, both promoting and preventing synaptogenesis, i.e. the formation of synapses between neurons (reviewed in McCarthy 2009). Estrogens (and androgens) are also involved in sexual differentiation of certain areas of the brain, as well as the development of sexually dimorphic behaviors in animals (primarily rodents) (reviewed in McCarthy 2010; Simerly 2002). However, the development of sexually dimorphic behaviors in humans seems to be less well studied and established.

Developmental exposure to estrogenic compounds, such as ethinyl estradiol, DES and BPA, have been reported to result in a wide array of effects in animal studies, e.g. effects on male reproductive organs (vom Saal et al. 1997) and behaviors (Jones et al.

2011), increased anxiety (Ryan and Vandenbergh 2006) and cancer (reviewed in Newbold 2004).

2.2.4.2 Androgens

Testosterone and dihydrotestosterone (DHT), drive and mediate the development of the male reproductive system, including the development of external genitalia and descent of the testes, and regulate the development of secondary male characteristics (Tortora and Grabowski 1996). As mentioned above, androgens are also important in the development of sexually dimorphic areas of the brain and behavior (reviewed in McCarthy 2010; Simerly 2002).

Developmental exposure to anti-androgenic substances, such as PCBs or the fungicide vinclozolin, have, for example, been shown to adversely affect spermatogenesis (Anway et al. 2006) and the development of male reproductive organs (reviewed in Kavlock and Cummings 2005) , and also affect sexually dimorphic behaviors (Colbert et al. 2005; Dickerson et al. 2011).

2.2.4.3 Thyroid hormones

Thyroid hormones, especially thyroxine (T_4), are critical for the regulation of normal brain development in the fetus (Molina 2010; Patel et al. 2011). Thyroid hormones bind to receptors in the fetal brain and activate transcription of target genes that are involved in different aspects of brain maturation, such as myelination and cell differentiation. Perturbations to thyroid hormone signaling during pregnancy, such as insufficient supply of thyroid hormones, iodine, transporters or enzymes involved in the production and metabolism of thyroid hormones, may result in various neurological disorders e.g. mental retardation or deafness (reviewed in Patel et al. 2011).

2.2.5 Regulatory aspects

During the past few years new chemical regulations have been implemented in the EU, e.g. the REACH legislation and the new directive for regulation of plant protection products, which identify EDCs as specifically problematic compounds that should be phased out or subjected to strict authorization processes. However, there are currently no generally agreed procedures that specify how substances with EDC characteristics are to be identified or risk assessed, and there is a lack of appropriate standardized toxicity tests with regulatory acceptance.

Within the EU and internationally there are current on-going activities to improve the knowledge on EDCs as well as identification and characterization of these compounds, test methods and risk assessment and management strategies. Several reports have been published as a result (EEA 2012; EFSA 2013; JRC 2013; Kortenkamp et al. 2012; UNEP/WHO 2012). Recently, the EU Parliament approved a resolution saying that current legislation and practices for regulating EDCs should be closely examined with the goal to update or propose new legislation by June 2015 (EU Parliament 2013).

2.2.5.1 Criteria for identifying EDCs

One critical issue in terms of regulatory measures for EDCs is the lack of generally established criteria for identifying compounds as EDCs. The new EU legislations for

plant protection products and biocides state that a definition of EDCs should be agreed upon in 2013 (EC 2009, 2012). In 2012 the European Commission requested EFSA to advise on such criteria.

The development of criteria for EDCs is inherently dependent on which definition of EDC is used. EFSA (2013) adopted the WHO/IPCS definition (see Table 1) and concluded that in order for a substance to be identified as an EDC it has to 1) have an endocrine MoA, 2) cause adverse effects in an intact organism or (sub)populations, and 3) that there is a causal link between the endocrine MoA and the adverse effect.

Thus, the issue of distinguishing between adverse effects of endocrine disruption and compensatory endocrine effects becomes critical. However, as discussed above, there are no clearly defined criteria for adversity and, at the time being, we lack the knowledge to draw up general criteria for what constitutes and adverse endocrine effect and distinguishes it from normal modulations of the endocrine system (EFSA 2013; JRC 2013; Kortenkamp et al. 2012;). Adversity, and consequently the identification of EDCs, thus has to be determined on a case-by-case basis based on expert judgment.

In contrast, others, e.g. the Endocrine Society (2012) advocate that endocrine activity in itself could be regarded as endocrine disruption. Reasoning along those lines would not require a causal link between the endocrine MoA and downstream adverse effects. However, it would probably mean that a very large number of substances would be identified as EDCs and would lead to a regulatory situation that would require other types of criteria to guide decisions on which compounds that constitute a health risk and when to apply risk management strategies.

2.2.5.2 Toxicity testing

Toxicity testing according to standardized and internationally validated test guidelines has traditionally been an important basis for regulatory risk assessment of chemicals (see section 2.1.2.2). However, current standardized test methods and batteries, e.g. the OECD test guidelines, have been criticized for being insufficient to identify and test the complex toxicity of EDCs (e.g. EFSA 2013; JRC 2013; Kortenkamp et al. 2012; Zoeller et al. 2012).

The issues of concern are, e.g:

- Current standardized tests do not include the most sensitive endpoints relevant to endocrine disruption
- There is no test in mammals available intended for investigating *in utero* or early developmental exposure and effects at later life stages.
- Important sensitive windows of exposure, e.g. before and during mating are not adequately covered in many tests.
- The inclusions of few dose groups, and requiring that relatively high doses are administered to observe statistically significant effects, means that effects at low doses are not adequately evaluated.
- Endocrine disruption entails a wide spectrum of different hormone actions that could possibly be affected; it is therefore not likely that a single test is sufficient to identify all EDCs.

Work is on-going e.g. at the OECD (OECD 2012) to enhance guidelines for the purpose of testing and assessment of EDCs. But these efforts has mainly focused on compounds interacting with estrogen, androgen and thyroid signaling or steroidgenesis and development of tests for identifying and evaluating other types of endocrine activity has received less attention.

2.2.5.3 Risk assessment methodology

It is apparent that the application of traditional risk assessment principles and assumptions becomes problematic in the evaluation of EDCs. For example, non-monotonic dose-relationships does not allow for extrapolation from high to low doses, i.e. drawing conclusions about the nature and risk for toxic effects in humans at environmentally relevant (low) exposure based on observations in animals exposed to relatively high doses, the way that it is traditionally done. Further, if a threshold for effect for EDCs cannot be assumed no "safe" dose for humans can be derived.

There is always uncertainty in risk assessment (Kalberlah et al. 2003), e.g. uncertainties concerning species-differences and the relevance of a certain effect to humans, uncertainties in measurements and in default values used. Some of these may be handled by applying default AFs, as discussed above. However, risk assessment of EDCs is especially riddled by uncertainties due to their complex toxicity and our incomplete understanding of it. In addition, for many EDCs (as for other substances) there may be a lack of data further hampering the understanding of their toxicity. There is thus a need to handle these large uncertainties in order to ensure reliable risk assessment and sufficient regulation of EDCs that is protective for the entire population.

Currently, standard toxicity studies are not expected to provide all relevant toxicity data for EDCs (e.g. EFSA 2013; JRC 2013; Kortenkamp et al. 2012; Zoeller et al. 2012). Thus, reliable and relevant non-standard exploratory research has an important role to fill information gaps and potentially reduce uncertainty in regulatory risk assessment of these compounds. However, methods for evaluating the reliability and relevance for health risk assessment of toxicity studies, which allow for potentially attributing the same weight to non-standard as to standard studies, are lacking.

2.3 BISPHENOL A

Bisphenol A (BPA) is and EDC for which risk assessment has proven especially complicated and it has been used for a case study in parts of this thesis work. BPA was first synthesized at the end of the 19th century. Reports of its estrogenic potential were published in the 1930's (Dodds and Lawson 1936) and it was for a while considered as a candidate substance in the development of estrogen replacement therapies (Vogel 2013). However, other more potent estrogenic substances, such as DES, were discovered during this period and since the 1950's BPA has primarily been used in the production of epoxy resins and polycarbonate plastics.

The production volume of BPA has increased steadily and today it is one of the most highly produced industrial chemicals globally. Polycarbonate and epoxy are very

versatile materials and are used in a wide variety of consumer products, such as plastic bottles and containers, sports equipment, building materials and electronics (Beronius and Hanberg 2011). Measured concentrations of BPA in human blood, urine and other tissues confirm that exposure is widespread in the general human population (Calafat et al. 2008; Vandenberg et al. 2007). It is generally believed that consumer exposure to BPA occurs primarily via food in contact with BPA-containing materials, such as polycarbonate baby bottles, tableware and food containers as well as food and beverage cans lined with epoxy resins. It has also been shown that BPA can be transferred to the skin from certain types of thermal printing paper, such as some types of cashier's receipts, in significant amounts (Biedermann et al. 2010).

The estimated BPA-exposure in the human population is 0.01 – 4.5 µg/kg body weight (bw)/day (FAO/WHO 2011). Infants and small children are estimated to have the highest exposure, which can be explained by a large intake of food per kg bw, as well as high intake of foods polycarbonate feeding bottles and canned foods. Studies measuring urinary BPA-concentrations in the general population seem to confirm that children have a higher BPA-exposure than adults (Calafat et al. 2008; Vandenberg et al. 2007).

It is well known that BPA can interact with nuclear and membrane-bound estrogen receptors and the toxicity of BPA is very well studied compared to many other chemicals. Still, there is disagreement among scientists as well as regulators as to the nature and size of the health risks posed by this compound. The NOAEL for BPA established for regulatory purposes in Europe and the US is currently 5 mg/kg body weight (bw) and day (EFSA 2006; US FDA 2008). This NOAEL was identified from two multigeneration standard studies in rats and mice investigating reproductive and developmental toxicity (Tyl et al. 2002 and 2008). The current TDI for BPA has been based on this NOAEL and an AF of 100 and has consequently been calculated to 50 µg/kg bw/day. However, there are a large number of non-standard research studies available reporting effects of BPA exposure at doses well below 5 mg/kg bw/day, sometimes around only a few µg/kg bw/day (Richter et al. 2007). This so called "low dose controversy" has made the risk assessment of BPA particularly difficult and has led scientists and others to question the regulatory NOAEL and the sufficiency of the current TDI for BPA.

Exposure to low doses of BPA, i.e. below the NOAEL, especially during early development, has also been reported to result in non-monotonic dose-response relationships for several endpoints, e.g. male reproductive behavior (Jones et al. 2011), female fertility (Cabaton et al. 2011) and learning and memory (Xu et al. 2011).

However, many of the studies reporting low-dose effects of BPA have so far not been considered adequate to serve as the basis for the derivation of health-based guidance values, or the evaluation of MoS, in regulatory risk assessments of BPA. The reasons given are often that they suffer from methodological flaws, such as only using one or two dose groups and inappropriate statistical methods, and/or are poorly reported, which limit their reliability for risk assessment purposes.

2.4 AIM

The overall purpose of this project was to contribute to making health risk assessments of EDCs more transparent, systematic, and reliable. As discussed above, risk assessment of EDCs has proven particularly complicated and is often hampered by large scientific uncertainties. The aim of the studies presented in this thesis was to identify how these uncertainties can be reduced or handled in order to improve risk assessment of EDCs. To that end this work has endeavored to make detailed scrutiny of the strengths and weaknesses of the risk assessment process and to identify important scientific and policy-related aspects that influence this process for EDCs.

The specific aims of the different studies have been:

I. To investigate the lack of regulatory coordination for EDCs by comparing the risk assessment processes within legislative frameworks for different regulatory groups of chemicals.

II. To investigate to what extent the conclusions vary between the available risk assessments for BPA and what might be the scientific and/or policy-related reasons for these differences, with the overall aim to contribute to the understanding of the risk assessment process for EDCs and the factors that influence this process.

III. To systematically investigate how the results in behavioral and functional parameters differ between DNT studies of BPA and if any factors of study design, such as choice of test species or test method, could explain the differences in results from these studies and what the implications are for DNT testing of BPA and other EDCs.

IV. To propose criteria and guidance for the evaluation of reliability and relevance of non-standard in vivo studies, which could be used to facilitate systematic and transparent evaluation of such studies for health risk assessment. And to propose user friendly guidance for reporting of non-standard studies intended to promote an improvement in reporting of studies that could be of use in risk assessment.

3 METHODS

Identification and investigation of scientific and policy-related aspects, as well as interactions between these two, that influence the risk assessment process for EDCs require careful scrutiny of relevant regulatory and scientific documents. This PhD-project was thus carried out as a series of literature studies, using the methodology developed by Rudén (2001a, b). Legislative and guidance documents, risk assessments, as well as toxicological studies and other published investigations have constituted the working materials, which have been systematically compared and analyzed in the different studies.

A database approach was used for **Papers I, II and III,** which entailed collecting detailed information from the materials in databases constructed in Microsoft Word or Access, depending on the amount and complexity of the data and analyses to be made. Key questions to be investigated were then formulated for each study. This approach enabled systematic and detailed comparisons within as well as between documents.

Legislative, guidance and risk assessment documents were identified via internet searches or contacts with authorities. Toxicity studies and other relevant scientific literature were generally identified via searches in PubMed or from the reference lists of risk assessment documents or other key literature.

Model compounds were used in three of the studies in order to investigate how risk assessment of EDCs has been conducted in practice, in addition to studying how this process is described and regulated in legislation. In **Paper I** three EDCs were used as model compounds to represent three different regulatory groups within EU legislation; BPA was used to represent existing industrial chemicals, dioxins as environmental pollutants in food, and vinclozolin as an existing active substance in plant protection products. The different regulatory frameworks were compared in terms of the scope of respective EU Regulations or Directives relevant for the risk assessment process and requirements for toxicity testing and risk assessment stated therein, as well as the availability and scope of guidance documents for risk assessment.

One recent EU risk assessment report for each model compound, produced in accordance with each of the regulatory frameworks investigated, was identified from the websites of responsible EU authorities (BPA and dioxins) or provided by the Swedish Chemicals Agency (vinclozolin). Key questions were formulated to allow comparison between these reports in terms of e.g. the toxicological data on which they were based, conclusions regarding the critical effect, as well as the toxicological principles used to determine human relevance of the identified endocrine MoA.

In **Papers II and III** the case of BPA was further investigated. BPA was considered relevant and suitable as a model substance of a focused case study mainly because 1) its toxicity has been well studied and thus there are a large amount of toxicity data available, 2) there is wide-spread and continuous exposure in the general population making its safety a relevant area of research, 3) there are several recent risk assessment documents available from different national and international authorities and expert

groups, and 4) risk assessments come to different conclusions regarding the risk to human health.

In **Paper II** risk assessment documents for BPA were scrutinized. A database collecting key information from the different risk assessment documents as well as the critical toxicity data on which they were based was constructed. Available risk assessments were identified from the websites of relevant authorities as well as via internet searches. The aim was to include all available and recent risk assessments of BPA addressing human health risks to the general population.

Key questions were formulated to specifically investigate differences between risk assessment documents concerning conclusions regarding risk to the human population at current exposure levels, identification of critical study and critical effect, how scientific uncertainty was handled and assumptions and arguments used in determining the significance for health risk assessment of non-standard studies reporting effects at low doses of BPA, i.e. below the previously established regulatory NOAEL.

In **Paper III** a data base was constructed based on the requirements and recommendations for DNT-testing according to OECD TG 426. Information from available DNT-studies of BPA was collected in this data base enabling systematic comparisons between different DNT-studies, as well as to the requirements and recommendations in TG 426. In total, 47 studies investigating behavioral endpoints in offspring of rats or mice exposed to BPA prenatally and/or during lactation were identified from risk assessment documents as well as via searches in PubMed. Three of these studies were excluded due to being judged too insufficiently reported based on predefined criteria. Consequently, 44 DNT-studies were included in analyses.

In addition to the database approach for systematic comparisons, Principal component analysis (PCA) and Partial least squares projection to latent structures (PLS) modeling were conducted to identify any systematic information in the data collected from DNT-studies and explore how different factors of study design (independent variables) may have contributed to differences in results between studies (dependent variables). PCA is a technique that is used to extract the most relevant information from a multivariate data set, i.e. a data set where several observations are described by several variables, with the aim to make it easier to interpret (Abdi and Williams 2010). The goal is to identify the variables that contribute to the most variability in the data set, i.e. that carry the most information. To that end, PCA transforms the large number of variables (which are possibly correlated) into fewer, un-correlated, principal components (PC). Imagining the multivariate data set as a "cloud" of data points, each PC represents an axis passing through the center of the data in one dimension of the data set. The first PC is the axis with the largest variance i.e. that accounts for as much of the variability in the data as possible. Each subsequent PC is orthogonal to the one before and accounts for as much of the remaining variability as possible. The observations and variables can then be plotted in regard to the PCs and their patterns can be analyzed (Abdi and Williams 2010). The aim is to explain a data set that includes a large number of variables by using a much smaller number of PCs.

PLS is a multivariate regression method that can be used to model the relationship between two matrices, e.g. X and Y. Specifically, the advantage of using PLS is that it simultaneously models the variation among numerous and strongly correlated X-variables and the variation among several Y-variables, as well as the relationship between them (Eriksson et al. 2008; Wold et al. 2001). Another strength of PLS is that it can handle missing data. It is thus a very useful method for investigating the relationship between numerous independent and dependent variables in complex data sets.

In **Paper IV** five OECD TGs for different types of *in vivo* toxicity testing were scrutinized to identify requirements and recommendations for *in vivo* toxicity testing that have been internationally accepted. Previously proposed methods for study evaluation were also identified from the open literature and reviewed. Based on this information a framework containing criteria and guidelines for evaluating the reliability and relevance for health risk assessment of non-standard *in vivo* studies was developed. In addition, a checklist for reporting *in vivo* research in order to meet the requirements of regulatory risk assessment was proposed. In order to ensure their scientific soundness, relevance and user-friendliness, feedback on the criteria and guidelines for study evaluation as well as the reporting checklist was requested from experts within the field of toxicity testing and risk assessment from research institutions in Europe and the US, the Swedish Chemicals Agency and the US FDA.

4 RESULTS AND DISCUSSION

This section is intended to summarize and discuss the main results of this thesis work in a wider perspective.

4.1 HEALTH RISK ASSESSMENT PROCEDURES FOR EDCs WITHIN THE EU

Four different legislative frameworks[2] regulating 1) existing industrial chemicals, 2) environmental pollutants in food, 3) existing active substances in plant protection products and 4) pharmaceuticals within the EU were investigated in **Paper I**.

At the time, there were no requirements within any of the investigated frameworks to specifically investigate endocrine disrupting potential or, consequently, for the identification of EDCs.

The requirements for toxicity testing within these four frameworks, as well as the availability and scope of guidance documents intended to guide the risk assessor, varied substantially. Considerably more toxicological information was required for pharmaceuticals and plant protection products than for the other two groups. In particular, for pollutants in food, such as dioxins, for which there is no intended use and no manufacturer, there are currently no data requirements for risk assessment stated by any EU legislation. Risk assessment of such compounds is conducted only if risks to human health are known or suspected and must then be based on data that is available in the literature. For compounds such as dioxins where there have been incidents of human exposures such as the Seveso accident (Pesatori et al. 2003), and for which there is a strong research interest, the existing data material may be vast. But this cannot be expected to be the case for other pollutants that may be an issue for concern.

This lack of a generally agreed procedure under any of the investigated regulatory frameworks that directly specifies how substances with EDC characteristics should be identified or risk assessed, what end-points are crucial to investigate, or how the results of such investigations are to be interpreted, means that the regulatory risk assessment process, as well as underlying policies, criteria and requirements may differ for different EDCs. Indeed, in **Paper I** it was concluded that if the only data available was the data required by legislation, only the plant protection substance vinclozolin was likely to be identified as an EDC. The endocrine disrupting properties of the other two model compounds, dioxins and BPA, would have gone undetected had they not been previously known or suspected due to a large research interest and the academic research literature available.

The increased focus on EDCs in REACH (EC 2006) and the new EU regulation for plant protection products (EC 2009), as well as the work within e.g. the OECD to develop tests and construct test batteries aimed at identifying compounds with endocrine disrupting properties and evaluating their toxicity, is a step forward to

[2] The regulatory frameworks for industrial chemicals and plant protections products have since been replaced by the REACH Regulation (EC) No 1907/2006 and the new EU Regulation (EC) No 1107/2009 for plant protection products, respectively.

closing the net on EDCs. However, there are several aspects remaining which hamper risk assessment of these compounds, e.g. disagreements among researchers, authorities and other stake holders concerning the criteria for identifying EDCs and what effects to consider adverse as opposed to adaptive. The development and validation of appropriate standardized test methods also progresses slowly due to extensive validation and harmonization procedures.

4.2 THE BPA CASE

The purpose of the investigations conducted for **Papers II and III**, using BPA as a case study, was to investigate and systematically compare practices, principles and assumptions in the risk assessment of BPA and in DNT-testing, respectively. It was not to discuss the WoE in this case or to draw conclusions about whether or not BPA poses a risk to human health.

4.2.1 Factors influencing the lack of consensus regarding the health risks of BPA

At the time of the investigation in **Paper II** ten risk assessments evaluating BPA were available from different national and international authorities and expert groups (AIST 2005; ECB 2003, 2008; EFSA, 2006, 2008; Health Canada 2008; NTP-CERHR 2008; SCF 2002; US FDA 2008; vom Saal et al. 2007). These had all been conducted within a six-year period between 2002 and 2008. Conclusions regarding health risks of BPA vary between these assessments from "there is no risk to any part of the population" to "there is risk to the entire population", with a couple of them (Health Canada 2008; NTP-CERHR 2008) stating that there is too much scientific uncertainty in the case to make any strong and definite conclusions concerning health risks.

In most of the BPA risk assessments compliance with internationally standardized and validated test guidelines was considered a mark of quality or even a quality criterion. The majority of the assessments agreed that the two standard reproductive toxicity studies conducted by Tyl et al. (2008; 2002) provided the most reliable data that could serve as key evidence in health risk assessment. Commonly, a NOAEL of 5 mg/kg bw/day and a LOAEL of 50 mg/kg bw/day for were identified from these studies.

The main reason for differences in risk assessment conclusions seemed to have been differences in the evaluation of the reliability and relevance of non-standard research reporting effects at low doses of BPA, below the LOAEL established in the studies by Tyl et al. While the most of the assessments concluded that this low-dose literature was of questionable reliability and/or relevance to humans one, the expert group convening in Chapel Hill in 2006, stated that "There is extensive evidence…that low doses of BPA have persistent effects on brain structure, function and behavior in rats and mice" and that "The wide range of adverse effects of low doses of BPA in laboratory animals…is a great cause for concern with regard to the potential for similar adverse effects in humans" (vom Saal et al. 2007). The Chapel Hill assessment was conducted by researchers with extensive prior expertise in the field of BPA which may explain

why their evaluation of the available toxicity data differed from that of the other risk assessments. However, some of the other assessments stated that, although not sufficiently reliable or relevant, the low dose studies could not be entirely dismissed as insignificant for human health risk assessment. This reasoning seems to have led to the expression of uncertainty in the risk assessments of Health Canada and the National Toxicology Program Center for the Evaluation of Risks to Human Reproduction (NTP-CERHR).

It can thus be argued that, in the case of BPA, the fact that the amount of toxicity data available has increased significantly during the last decade has contributed to scientific uncertainty in risk assessment conclusions rather than to leading to more certain assessments. One reason may be that a lot of the research published for BPA has suffered from insufficient reporting and that there is a lack of agreed upon methods on how to evaluate the reliability and relevance of non-standard research for risk assessment purposes.

4.2.2 Developmental neurotoxicity of BPA – contributions of non-standard studies

One issue of disagreement between risk assessors has been the potential of BPA to cause DNT at low doses. The studies by Tyl et al., which were used as key evidence in most risk assessment of BPA, did not investigate DNT. However, several of the BPA risk assessments have evaluated other studies investigating DNT concluding, in many cases, that available DNT-studies were not sufficiently reliable or relevant to set a NOAEL below that established based on the studies by Tyl and co-workers. However, the Chapel Hill experts judged these studies as relevant for evaluating the risks to human health. Also, in the assessments by NTP-CERHR and Health Canada it was specifically stated that effects on neurobehavioral development may prove important for the assessment of BPA and that further research in this area is needed (Health Canada 2008; NTP-CERHR 2008). The DNT issue also became a point of disagreement between member states in the 2008 update of the assessment from European Chemicals Bureau (ECB 2008). Denmark, Sweden and Norway argued along the lines of Health Canada and the NTP-CERHR, that the data on DNT did indeed raise concerns about the health risks of BPA. However, the official conclusion from the ECB assessment was that there was no concern regarding any health risks form BPA and the Nordic countries' opinion was included as a footnote in the report. Since then, a large DNT-study adhering to OECD TG 426 has been conducted in an attempt to settle this dispute (Stump et al. 2010). The conclusion from this study was that BPA does not cause DNT, either at high or at low doses.

One reason that the reliability of the available research on DNT of BPA has been questioned is the varying and sometimes contradictory results reported from studies in this area. The purpose of **Paper III** was thus to investigate to what extent results in behavioral and functional parameters differ between available studies investigating DNT of BPA in rodent studies, and what could be the reasons for these differences. The

studies were also compared to the requirements for designing and conducting a DNT-study according to the standardized guideline OECD TG 426 (OECD 2007).

OECD TG 426 states that pregnant dams (preferably of a common rat strain) should be administered the test substance orally from the first day of gestation until weaning of the pups. Offspring are evaluated for effects in different functional and behavioral endpoints, as well as other physical and developmental landmarks, such as body weight and sexual maturation, before weaning, at adolescence and young adulthood. Brain weight and neuropathology data are collected at weaning and at termination. The required behavioral endpoints include evaluations for behavioral ontogeny, motor and sensory function, motor activity and learning and memory.

However, many non-standard DNT-studies investigating BPA have evaluated other types of behavioral effects, such as anxiety, exploration and social and sexual behaviors. Forty-four DNT-studies were identified from the open literature and deemed sufficiently well reported to be included in analyses. Only one, the study by Stump et al., had been carried out in accordance with OECD TG 426. Studies in both mice and rats were included. Evaluations of behavioral effects conducted in the DNT-studies were categorized into either: 1) motor activity, 2) learning and memory, 3) anxiety-related or exploratory behaviors, or 4) other behaviors, including e.g. social, sexual and maternal behaviors. Systematic comparisons showed that, indeed, very varying and sometimes contradictory results were reported, especially for the required endpoints motor activity and learning and memory. Also, effects were more often observed in endpoints that are *not* required according to OECD TG 426 while relatively few studies reported effects on e.g. motor activity (**Paper III, Fig. 2**). This is not very surprising since behaviors are linked to hormonal state as well as hormonal mechanisms (Cory-Slechta et al. 2001; Zoeller et al. 2012). These behaviors may thus be particularly relevant for the evaluation of the neurotoxic actions of EDCs in general.

Another observation in **Paper III** was that non-standard research studies often lacked information about the research aim, design, performance or results which hampered the interpretation and evaluation of study results.

4.2.3 Sex-differences

Given the estrogenic potential of BPA it is reasonable to assume that exposure, especially during early development, may give rise to different effects in males and females. Indeed, this has often been observed in toxicity studies of BPA, e.g. in regard to sexually dimorphic behaviors (Carr et al. 2003; Gioiosa et al. 2007; Rubin et al. 2006). Sex-differences in sensitivity, i.e. whether one sex is more sensitive overall to the toxicity of BPA, have however not been evident, nor has this issue been discussed in BPA risk assessments (**Paper II**).

In **Paper III** it was observed that behavioral effects after developmental exposure to BPA have more often been investigated in male than in female offspring. This is problematic since sex-differences in effects are to be expected and extrapolations between the sexes, i.e. drawing conclusions about risk to females based on toxicity data

conducted in males, may be difficult. It was also observed in **Paper III** that the non-required behavioral endpoints, especially social and sexual behaviors, seemed to have been particularly important in identifying DNT-related effects in female offspring. These types of behavioral effects were observed in females in about 70% of the studies where they were investigated. In contrast, effects on motor activity or learning and memory in females were only observed in about 30% of the studies where these parameters were investigated (**Paper III, Fig. 2**).

4.2.4 Implications for toxicity testing and risk assessment

The investigations of the BPA-case in **Papers II and III** raise issues that could have implications for toxicity testing and risk assessment of EDCs.

Primarily, conclusions from these studies concur with on-going discussions that standardized test guidelines may not contain the most sensitive and relevant endpoints and up-to-date methods needed to evaluate EDCs (Kortenkamp et al. 2012; Zoeller et al. 2012). Since standard studies are traditionally given more weight than non-standard research studies in regulatory risk assessment there is thus a chance that sensitive effects of BPA and other EDCs are not being adequately considered to ensure a risk assessment that is protective of even the most sensitive individuals in the human population. As the case of BPA shows, even when sensitive effects at very low doses are strongly implied from a large amount of non-standard research studies their relevance for health risk assessment is questioned in the presence of data from standard studies that contradict these findings.

Work is being carried out e.g. at the OECD to develop new standardized test strategies suitable for identifying and testing EDCs (OECD 2012). However, this process has proved challenging, in part due to the complex toxicity of EDCs previously described. As discussed in **Paper III**, effects in social and sexual behaviors were often observed for BPA and are likely to be sensitive effects of many EDCs in general since such behaviors are linked to hormonal state as well as hormonal mechanisms (Cory-Slechta et al. 2001). However, the standardization of tests to evaluate social and sexual behaviors is hampered by the complexity of these behaviors. For example, they must often be interpreted before they can be quantified, which means that they are difficult to automate and require a high level of expertise and trained investigators (Cory-Slechta et al. 2001). It is therefore challenging to incorporate tests for these endpoints in standardized test batteries for neurotoxicity.

The development of new standardized methods is also inherently slow due to the extensive validation and harmonization procedures for standardized test methods. In the meantime, a lot of new research is published in the area of EDC toxicity. It therefore seems important to be able to use non-standard studies in a reliable and transparent manner in risk assessment in parallel to the work of developing new sensitive and relevant standards for toxicity testing.

4.3 FACILITATING THE USE OF NON-STANDARD STUDIES IN HEALTH RISK ASSESSMENT OF EDCs

Conclusions from **Papers I, II and III** indicate that non-standard research studies contribute data that may be important to fill knowledge gaps and improve risk assessment conclusions for EDCs. Also, it seems counterproductive that a lot of academic research is conducted on the toxicity of chemicals, which can ultimately not be used to inform risk assessment. However, as the case of BPA illustrates, 1) the use of non-standard studies in risk assessment is often hampered by perceived methodological limitations and insufficient reporting, and 2) the reliability and relevance of the same studies may be judged differently by different evaluators.

In **Paper IV** the aim was thus to propose a framework of criteria and guidelines intended to facilitate systematic and transparent evaluation of the reliability and relevance of non-standard studies for health risk assessment. One additional but important purpose was to suggest a checklist of information that should be reported, which could be used as guidance for authors when preparing manuscripts for publication.

The desired outcome of these efforts was to contribute to reducing scientific uncertainty in health risk assessment conclusions and, in extension, to better targeted policy decisions for health risk reduction. However, more research and discussions in this area is needed.

4.3.1 A framework for evaluating reliability and relevance of non-standard studies

The framework for data evaluation presented in **Paper IV** allows for potentially attributing equal weight to non-standard and standard studies in health risk assessment. Several of the previously published methods for study evaluation, especially the preferred method of many regulatory agencies proposed by Klimisch et al. (1997), attributes higher reliability to studies conducted according to standardized test guidelines by default. The importance of the relevance of the study is emphasized in the framework proposed in **Paper IV**. Few of the previously published methods for study evaluation provide guidance for the evaluation of relevance and focus mainly on reliability. In addition, it provides clearly defined and detailed criteria and guidelines for evaluating reliability and relevance.

These criteria and guidelines were primarily based on requirements and recommendations identified from OECD TGs for relevant *in vivo* studies, as well as previously reported methods. Since the OECD TGs have been internationally validated and accepted they were considered an adequate and reliable basis for suggesting criteria for study evaluation.

A two-tiered approach was proposed (Figure 6), where the purpose of the first Tier is to identify studies that are sufficiently well reported and reliable enough to be carried forward to a more thorough evaluation of reliability in Tier II in parallel with evaluation of relevance.

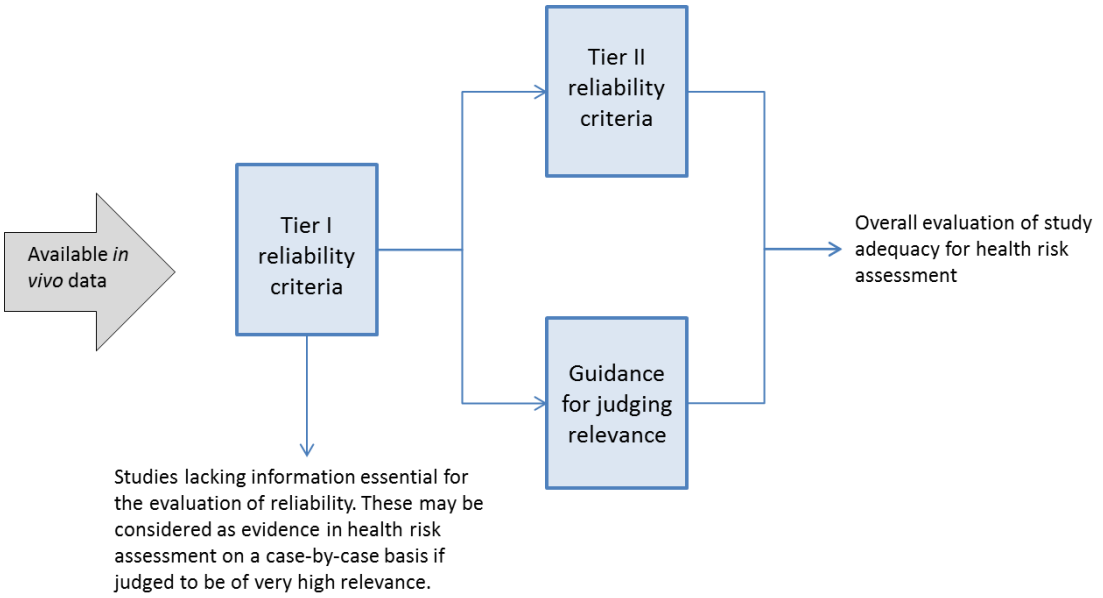

Figure 6. The structure of the proposed framework for study evaluation.

A web-based color-coding tool was developed for the purpose of applying the criteria in Tier II. Using this tool the evaluator can mark each criterion as green for "fulfilled", orange for "partially fulfilled", red for "not fulfilled" or white for "not applicable". The tool generates a color chart for the study in an exportable excel-sheet, which can be used as a basis for determining whether the study is of high, sufficient or low reliability. I.e. if the chart is dominated by red the conclusion may be that the study is not reliable enough to be used for risk assessment, while if it is mainly green and/or orange the study may be considered of high or sufficient reliability and adequate to serve as key or supporting evidence in risk assessment. Depending on the substance being investigated and the type of study, e.g. chronic or reproductive toxicity, some criteria may be considered to be more critical for the reliability of the study than others. Visualizing if and to what extent each criterion has been fulfilled using a qualitative method of color-coding, rather than attributing a numerical value as proposed by Schneider et al. (2009), opens up for a more flexible evaluation of the overall reliability of the study and is a transparent method for applying expert judgment.

The goal is that the proposed criteria and guidelines, as well as the color-coding tool, should be publically available and free to use on-line in the near future.

4.3.2 Guidelines for reporting animal research

Based on the proposed criteria and guidelines for study evaluation a checklist for reporting *in vivo* studies was constructed. The list contains items considered important to that should be reported from *in vivo* studies to ensure that the study can be evaluated and considered as evidence in regulatory risk assessment.

However, the amount of information presented in published research articles is usually restricted by space limitations. Therefore, the checklist can alternatively be used as a template for providing supplementary information in cases where the information is too

extensive to include in the manuscript text or inclusion of all details is considered to prevent a clear description of main results and conclusions related to the study hypothesis.

5 CONCLUSIONS

The main conclusions from this thesis work can be summarized as follows.

- In the absence of generally agreed procedures for how substances with EDC characteristics are to be identified or risk assessed, what end-points are crucial to investigate, or how the results of such investigations are to be interpreted, the regulatory risk assessment process, as well as underlying policies, criteria and requirements may differ for different EDCs.
- Because of the complex nature of endocrine disruption, test methods, principles and criteria for data interpretation traditionally used might not be directly applicable to EDCs and further research within this field is needed.
- Standardized test guidelines, such as the OECD test guidelines, do not always include the most sensitive endpoints relevant for the evaluation of EDCs.
- Non-standard studies, i.e. research studies not conducted according to any standardized test guidelines, could fill information gaps and contribute information that could be particularly important for the risk assessment of EDCs.
- The work of developing new standardized tests for EDCs is progressing slowly, in part due to the complex toxicity of EDCs but also due to the extensive validation and harmonization procedures for standardized test methods. It is therefore important to be able to use non-standard studies in a reliable and transparent manner in risk assessment in parallel to the work of developing new sensitive and relevant standards for toxicity testing.
- Tools are needed that facilitate systematic and transparent evaluation of non-standard studies for the purpose of risk assessment. These tools should allow for potentially giving equal weight to non-standard and standard studies in risk assessment.
- Information that is crucial for reproducibility and the evaluation of study reliability is often missing from non-standard research studies published in scientific journals. Reporting of non-standard studies needs to be improved in order to meet the requirements of regulatory risk assessment.

6 POPULÄRVETENSKAPLIG SAMMANFATTNING

Vi omges dagligen av en blandning av olika kemikalier från till exempel tillsattser och föroreningar i vår mat, textilier, leksaker, kosmetika och byggnadsmaterial. Därför är det viktigt att ha en effektiv kemikaliekontroll, som säkerställer att de ämnen vi exponeras för inte leder till oönskade hälsoeffekter. Hälsoriskbedömning av kemikalier görs som en del i kemikalieregleringen för att bedöma om deras användning innebär någon risk för människors hälsa. Hälsoriskbedömning innebär att man utvärderar vilka toxiska effekter ett ämne kan ge upphov till, ofta baserat på information från toxicitetsstudier i djur, och vid vilka halter det kan tänkas ge upphov till skadliga hälsoeffekter hos människor. Man kan sedan bedöma huruvida människors exponering överstiger de halter som kan anses säkra.

Oron för hormonstörande ämnen och de effekter de kan ha på människors hälsa och i miljön har ökat under de senaste årtiondena. Hormonstörande ämnen påverkar hormonsystemets normala funktioner, till exempel genom att härma kroppsegna hormoner eller genom att störa hur dessa produceras, bryts ner eller transporteras i kroppen. Samband mellan sådana ämnen och hormonrelaterade sjukdomar, som vissa typer av cancer, försämrad fertilitet och hjärt-kärlsjukdomar i den allmänna befolkningen, liksom effekter i miljön och djurliv, har rapporterats i ökande grad under de senaste decennierna. Detta tyder på att tidigare kemikaliereglering inte har lyckats skydda människors hälsa och miljön tillräckligt.

Hormonsystemet reglerar i stort sett alla kroppens organ, vävnader och celler. Genom specifika signalsubstanser, hormoner, styr hormonsystemet en rad livsviktiga funktioner, så som fortplantning, tillväxt och utveckling, metabolism och humör. Särskilt kritisk är fostertiden då hormonsystemet har en viktig roll i utvecklingen av olika organ och vävnader. Om hormonsystemets normala funktion störs under denna känsliga period kan det i värsta fall leda till allvarliga och permanenta effekter, som till exempel hämmad mental utveckling, missbildningar och ökad risk för vissa typer av cancer. I nya EU-lagstiftningar för t.ex. industrikemikalier, växtskyddsmedel och biocider har hormonstörande ämnen uppmärksammats som särskilt oroväckande ämnen som bör fasas ut eller strikt regleras. Det finns således ett ökat tryck på regulatoriska myndigheter att effektivt kunna bedöma eventuella hälsorisker från dessa ämnen.

Dock råder stor vetenskaplig osäkerhet kring hormonstörande ämnen och de har visat sig särskilt svåra att riskbedöma, bland annat på grund av deras komplexa toxicitet. Till exempel kan ämnen som härmar kroppsegna hormoner ge upphov till effekter vid mycket låga doser. De har också visat sig kunna orsaka olika, och även motsatta, effekter vid höga och vid låga doser i djurstudier. Ofta är effekterna mycket subtila och blir ibland inte tydliga förrän långt efter exponeringen upphört.

Dessa egenskaper strider mot flera av de antaganden och principer som toxikologin och riskbedömning traditionellt bygger på, som att man kan dra slutsatser om hälsorisker vid låga halter av ett ämne baserat på toxicitetsstudier i djur där relativt höga halter har testats.

Syftet med avhandlingsarbetet som presenteras här har varit att undersöka hur den vetenskapliga osäkerheten beträffande hormonstörande ämnens toxicitet kan minskas eller hanteras för att göra hälsoriskbedömningen av dessa ämnen bättre och mer tillförlitlig. Arbetet har byggt på litteraturstudier som undersökt riskbedömningsprocessen för hormonstörande ämnen inom EU, liksom de toxicitetsdata som finns tillgängliga för riskbedömare och hur användningen av all tillgänglig toxicitetsdata kan förbättras. Den omdebatterade substansen bisfenol A (BPA) har använts som en fallstudie i en stor del av detta arbete.

Myndigheter lägger ofta störst vikt vid toxicitetsstudier som genomförts enligt internationellt överenskomna och standardiserade testriktlinjer när de utför riskbedömningar. Standardiserade tester anses vara mycket tillförlitliga men har dock kritiserats av forskare och andra experter vad gäller att kunna fånga upp känsliga effekter av hormonstörande ämnen, bland annat för att de inte tar tillräcklig hänsyn till hormonstörande ämnens specifika egenskaper, som till exempel effekter vid mycket låga doser och fördröjda effekter. Resultaten i avhandlingen visar bland annat att icke-standardiserade forskningsstudier, alltså studier som genererats inom akademisk forskning, kan bidra med information som kan vara särskilt viktig för att få en säkrare riskbedömning för hormonstörande ämnen. Men forskningsstudier kritiseras ofta för att ha svagheter och brister som negativt påverkar deras tillförlitlighet och därför begränsar deras användning i regulatorisk riskbedömning.

Inom detta avhandlingsarbete har också metoder utvecklats som syftar till att kunna öka användbarheten av forskningsstudier i hälsoriskbedömning av kemikalier. Målet är att överbrygga klyftan mellan akademisk forskning och kemikaliereglering och förhoppningsvis bidra till att göra hälsoriskbedömningen för hormonstörande ämnen säkrare.

7 ACKNOWLEDGEMENTS

This work was carried out at the Unit of Environmental Health Risk Assessment at the Institute of Environmental Medicine, Karolinska Institutet. Doing a literature-based project, that aims to bridge the gap between research and chemicals regulation and policy, has been challenging at times and there are many people who have made it possible and, most importantly, enjoyable!

First of all, I have had the best supervisors! Thank you for introducing me to this important area of research and believing in me. I could not ask for better role models. You have managed to support me from every direction while at the same time giving me space to grow and develop independently. I am really proud of how I have developed as a researcher and as a person during the past five years, and I am aware that this is very much thanks to you. I am eternally grateful!

I would especially like to express my gratitude to my main supervisor Annika Hanberg. You have always had time for me, and I have been able to count on you for guiding me along and having my back when I needed it. Thank you for showing me how to stay calm and collected under pressure. Under your tutelage I have always felt respected, secure and confident!

A huge thank you also to my co-supervisor Christina Rudén for providing such a good example of the importance and benefits of doing a project like this, for ironing out all the kinks along the way and for pushing me to be a bit tougher and sharper.

Thank you also to my co-supervisor Gunnar Johanson for your support and for always taking the time to answers my questions whenever I came to you.

I am also grateful to my co-supervisor, and the head of the Unit of Environmental Health Risk Assessment, Helen Håkansson. Thank you for giving me the opportunity to do this work in your unit.

I would also like to extend my sincerest appreciation to the following people:

My external mentor Agneta Ohlsson, for showing an interest in my research and career and sharing your experiences and wisdom.

My co-authors. Linda Molander at ITM, Stockholm University, for contributing sharp wits and creative ideas and being an excellent writer. I look forward to many more breakfast meetings at Vetekatten! Niklas Johansson, for taking the helm on tricky statistics and bringing it to port, and for your patience in explaining it all to me.

Stefan Spulber at the Department of Neuroscience, Karolinska Insitutet, for answering my endless questions about neurotoxicity and behavioral testing, and always with a positive attitude.

Marlene Ågerstrand at ITM, Stockhom University, for sharing your work and experiences and for getting things done. I look forward to more co-operations in the future!

Past and present members of the Environmental Health Risk Assessment Unit, especially the other two members of Team Jönssonligan without whom this mission

would have been much less fun; Daniel, for immaculate planning, great laughs, music and, most importantly, awesome coffee and for never giving up on the idea of the perfectly foamed milk. And Emma, for being a source of fun and creative ideas at the office, for countless wonderful conversations, trips to the book shop, origami and colored pens. Also, thank you Fereshteh for your warm friendship and discussions about everything between heaven and earth! Robert, for all the nice chats and your wonderful sense of humor. Johanna, for organizing really fun and important PhD-courses. And Kina, Maria, Sabina, Ali, Lubna, Krister, Lina, Rachel, Jinyao, Per, Sylwia and Lotta for always being so supportive and providing a really nice working atmosphere!

All my co-workers and friends at Floor 3, thank you for filling the corridors with friendly faces and making IMM a really nice place to go to every day. An especially big hug goes to my buddy Astrid, dear A1, for being a great friend all around, for sending me pretty pictures on Mondays and for having a personality that can give any bad day a silver lining. Thank you Rebecca for always being so helpful and providing scientific expertise, as well pharmaceuticals, when needed. To Christian, Ilona, Ian, Aram, Sandeep, Linda, Åse, Neus, Consol and Hanna and all the rest of the units of Biochemical and Molecular Toxicology for great lunches and "fikas", laughs and conversation.

All my friends who have supported (and distracted) me during this time. Especially Viktoria, thank you for your friendship, for dragging me to England and for all your wonderful artwork. You have made both my life and my thesis more colorful! And to your Lille Billy-Bob Långkalsong, who has definitely made these past weeks very exciting. Carolina, for thousands of lovely get-togethers, trips to garden stores and understanding the importance of a really well decorated Christmas tree!

Matte, the world's best neighbor, for providing our household with a much needed third adult, and for sending computer cables across the country so that I could work during my vacations.

Åsa, for being a kindred researcher-spirit and protecting me against the mean kids when we were children. And for spending valuable time with me and teaching me about life during these past couple of years.

The perpetual social events committee: Lars, Tina, Mats and Astrid (again, because this line-up is not complete without you), thank you for all the great and fun meetings and lovely dinners! I hope to plan many more events with you in the future.

An especially heartfelt thanks goes to Dr. Kilburn in LA, who knew better and, initially against my will, made me realize I wanted to become a toxicologist.

My teachers and professors at HPU for laying a really excellent groundwork for my continued education and research, and to the entire Hawaiian crew, Ida & Dan, Andreas, Jo, Shaggy, Timo & Zenita, Simen, Lisa, Maria, Daniel, Maly and many more who made my time there unforgettable! To Hawaii nei for sorting me out and being my second home, mahalo nui loa!

My large and wonderful family! To my dad and Inger, my mom and Leffe for your endless support, for sending me and Freddie on relaxing weekends and taking care of our children when we had to work, or rest. And to my sisters Anneli, Eva, Marie, Cecilia and Emma, my brother Olle, and your families; I know I can always count on you guys and having you around is one of life's biggest luxuries!

Farmor Lisa, for dragging my sled on the bare ground along Karlavägen to Humlegården when I refused to accept winter was over (I think I was 3), for memories of chestnuts and Shalimar, for endless conversations about everything and anything when I got older, for teaching me good manners at the dinner table, but most of all for loving me and providing me with a sanctuary. I wish you were still here. I miss you.

My grandmother Ulla-Brita for always imagining great things for me.

The entire fabulous Beronius-Sjöman-Espersson clan, and the Ullbergs and Adolfssons, thank you for always caring! You are the foundation from which I grow.

To Lize-Lotte, Thomas, Tobias & Anna, Matilda, Amanda, Ulla-Britt & Roger, Inger & Calle, Rickard & Christel, Jonas och Helena, thank you for being my extended family and for being always supportive.

The people who deserve the most gratitude are Fredrik, Estella and Jack. Thank you for providing me with endless and unconditional love, joy and support, for showing me what really matters in life and giving me perspective. Jag älskar er.

This work was supported by grants from the Swedish research council Formas and the EU Network of Excellence CASCADE.

8 REFERENCES

Abdi H, Williams LJ. 2010. Principal component analysis. WIREs Comp Stat 2(4): 433-459.

Ågerstrand M, Breitholtz M, Rudén C. 2011. Comparison of four different methods for reliability evaluation of ecotoxicity data: a case study of non-standard test data used in environmental risk assessment of pharmaceutical substances. Environmental Sciences Europe 23: 17.

AIST. 2005. Japanese National Institute of Advanced Industrial Science and Technology Risk Assessment Document Series No 4: Bisphenol A. Available on-line at: www.aist.go.jp

Alcock RE, Macgillivray BH, Busby JS. 2011. Understanding the mismatch between the demands of risk assessment and practice of scientists--the case of Deca-BDE. Environ Int 37(1): 216-225.

Anway MD, Memon MA, Uzumcu M, Skinner MK. 2006. Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. J Androl 27: 868-879.

Beronius A, Hanberg A. 2011. Sources of exposure to bisphenol A. IMM report nr 2/2011. Available on-line at: http://ki.se/content/1/c4/91/50/IMMrapport2-2011.pdf.

Beronius A, Hanberg A. 2013. Is it possible to determine thresholds for the effects of endocrine disruptors? – A summary of scientific argumentation from 15 relevant publications on endocrine disruption. Swedish Chemicals Agency report (PM) 2/13. Available on-line at www.kemi.se.

Biedermann S, Tschudin P, Grob K. 2010. Transfer of bisphenol A from thermal printer paper to the skin. Anal Bioanal Chem 398(1): 571-576.

Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J, Vickers C. 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Crit Rev Toxicol 38(2): 87-96.

Brodeur JC, Svartz G, Perez-Coll CS, Marino DJ, Herkovits J. 2009. Comparative susceptibility to atrazine of three developmental stages of Rhinella arenarum and influence on metamorphosis: non-monotonous acceleration of the time to climax and delayed tail resorption. Aquat Toxicol 91(2): 161-170.

Cabaton NJ, Wadia PR, Rubin BS, Zalko D, Schaeberle CM, Askenase MH, Gadbois JL, Tharp AP, Whitt GS, Sonnenschein C, Soto AM. 2011. Perinatal exposure to environmentally relevant levels of bisphenol A decreases fertility and fecundity in CD-1 mice. Environ Health Perspect 119(4): 547-552.

Calafat AM, Ye X, Wong L-Y, Reidy JA, Needham LL. 2008. Exposure of the U.S. population to bisphenol A and 4-tertiary-octylphenol: 2003-2004. Environ Health Perspect 116: 39-44.

Carr RL, Bertasi FR, Betancourt AM, Bowers SD, Gandy BS, Ryan PL, Willard ST. 2003. Effect of neonatal rat bisphenol A exposure on performance in the Morris water maze. J Tox Environ Health Part A 66: 2077-2088.

Colbert NKW, Pelletier NC, Cote JM, Concannon JB, Jurdak NA, Minott SB, Markowski VP. 2005. Perinatal exposure to low levels of the environmental antiandrogen vinclozolin alters sex-differentiated social play and sexual behaviors in the rat. Environ Health Perspect 113: 700-707.

Conolly RB, Lutz WK. 2004. Nonmonotonic dose-response relationships: mechanistic basis, kinetic modeling, and implications for risk assessment. Toxicol Sci 77(1): 151-157.

Cory-Slechta DA, Crofton KM, Foran JA, Ross JF, Sheets LP, Weiss B, Mileson B. 2001. Methods to identify and characterize developmental neurotoxicity for human health risk assessment. I: behavioral effects. Environ Health Perspect 109 Suppl 1: 79-91.

Davis JA, Gift JS, Zhao QJ. 2011. Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1. Toxicol Appl Pharmacol 254(2): 181-191.

Diamanti-Kandarakis E, Bourguignon J-P, Giudice LC, Hauser R, Prins GS, Soto AM, Zoeller RT, Gore AC. 2009. Endocrine-disrupting chemicals. An Endocrine Society scientific statement. Endocrine Reviews 30(4): 293-342.

Dickerson SM, Cunningham SL, Patisaul HB, Woller MJ, Gore AC. 2011. Endocrine disruption of brain sexual differentiation by developmental PCB exposure. Endocrinology 152(2): 581-594.

Dodds EC, Lawson W. 1936. Synthetic oestrogenic agents without the phenanthrene nucleus. Nature 137: 996.

Durda JL, Preziosi DV. 2000. Data quality evaluation of toxicological studies used to derive ecotoxicological benchmarks. Human and Ecological Risk Assessment 6: 747-765.

EC. 1996. European workshop on the impact of endocrine disruptors on human health and wildlife. Available on-line at: http://ec.europa.eu/environment/endocrine/documents/reports_en.htm.

EC. 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union L 396, 30-12-2006.

EC. 2009. Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the

market and repealing Council Directives 79/117/EEC and 91/414/EEC. Official Journal of the European Union L 309, 24-11-2009.

EC. 2012. Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products. Official Journal of the European Union L 167, 27-06-2012.

ECB. 2003. European Union risk assessment report: 4,4'-isopropylidenediphenol (Bisphenol-A). 2003. Available on-line at: http://ecb.jrc.ec.europa.eu/

ECB. 2008. Updated risk assessment of 4,4'-isopropylidenediphenol (Bisphenol-A). Available on-line at: publications.jrc.ec.europa.eu.

ECHA 2011a. European Chemicals Agency. Guidance on information requirements and chemical safety assessment. Part A: Introduction to the Guidance document. Available on-line at: echa.europa.eu.

ECHA. 2011b. European Chemicals Agency. Guidance on information requirements and chemical safety assessment. Chapter R.4: Evaluation of available information. Available on-line at: echa.europa.eu.

ECHA. 2011c. European Chemicals Agency. Guidance on information requirements and chemical safety assessment. Part B: Hazard assessment. Available on-line at: echa.europa.eu.

ECHA. 2012. Guidance on information reguirements and chemical safety assessment. Part E: Risk characterisation. Available on-line at: echa.europa.eu

EEA. 2012. European Environment Agency. The impacts of endocrine disrupters on wildlife, people and their environments. The Weybridge +15 (1996-2011) report. Available on-line at: www.eea.europa.eu.

EFSA. 2006. Opinion of the Scientific Panel on Food Additives, Flavourings, Processing Aids and Materials in Contact with Food on a request from the Commission related to 2,2-bis(4-hydroxyphenyl)propane (bisphenol A). Available on-line at: www.efsa.europa.eu.

EFSA. 2008. Toxicokinetics of Bisphenol A. Scientific Opinion of the Panel on Food additives, Flavourings, Processing aids and Materials in Contact with Food (AFC). Available on-line at: www.efsa.europa.eu

EFSA. 2013. European Food Safety Authority. Scientific opinion on the hazard assessment of endocrine disruptors: Scientific criteria for identification of endocrine disruptors and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment. Available on-line at: www.efsa.europa.eu.

Eriksson L, Johansson N, Kettaneh-Wold N, Wikström C, Wold S. 2008. PLS. In: Design of experiments - Principles and Applications, 3rd Ed. Umeå: Umetrics Academy, pp 323-334.

EU Parliament. 2013. Press release 14 March 2013. MEPs call for clampdown on chemicals affecting hormones. Available on-line at www.europarl.europa.eu.

Falk-Filipsson A, Hanberg A, Victorin K, Warholm M, Wallen M. 2007. Assessment factors--applications in health risk assessment of chemicals. Environ Res 104(1): 108-127.

Fan F, Wierda D, Rozman KK. 1996. Effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on humoral and cell-mediated immunity in Sprague-Dawley rats. Toxicology 106(1-3): 221-228.

FAO/WHO. 2011. Report of Joint FAO/WHO Expert Meeting. Toxicological and health aspects of bisphenol A. Available on-line at: www.who.int.

Ge RS, Chen GR, Dong Q, Akingbemi B, Sottas CM, Santos M, Bernard DJ, Hardy MP. 2007. Biphasic effects of postnatal exposure to diethylhexylphthalate on the timing of puberty in male rats. J Androl 28(4): 513-520.

Gioiosa L, Fissore E, Ghirardelli G, Parmigiani S, Palanza P. 2007. Developmental exposure to low-dose estrogenic endocrine disruptors alters sex differences in exploration and emotional responses in mice. Horm Behav 52(3): 307-316.

Health Canada. 2008. Screening assessment for the Challenge Phenol, 4,4'-(1-methylethylidene)bis- (Bisphenol A). Available on-line at: www.hc-sc.gc.ca.

Hengstler JG, Foth H, Gebel T, Kramer PJ, Lilienblum W, Schweinfurth H, Volkel W, Wollin KM. 2011. Critical evaluation of key evidence on the human health hazards of exposure to bisphenol A. Crit Rev Toxicol 41(4): 263-291.

Hobbs DA, Warne MS, Markich SJ. 2005. Evaluation of criteria used to assess the quality of aquatic toxicity data. Integr Environ Assess Manag 1(3): 174-180.

Hossaini A, Dalgaard M, Vingaard AM, Pakarinen P, Larsen J-J. 2003. Male reproductive effects of octylphenol and estradiol in Fischer and Wistar rats. Reprod Toxicol 17: 607-615.

Howell A. 2001. Preliminary experience with pure antiestrogens. Clin Cancer Res 7(12 Suppl): 4369s-4375s; discussion 4411s-4412s.

Jones BA, Shimell JJ, Watson NV. 2011. Pre- and postnatal Bisphenol A treatment results in persistent deficits in the sexual behavior of male rats, but not female rats, in adulthood. Horm Behav 59: 246-251.

JRC. 2013. Key scientific issues relevant to the identification and characterisation of endocrine disrupting substances. Report of the Endocrine Disrupters Expert Advisory Group. Available on-line at: ihcp.jrc.ec.europa.eu.

Kalberlah F, Schneider K, Schuhmacher-Wolz U. 2003. Uncertainty in toxicological risk assessment for non-carcinogenic health effects. Regul Toxicol Pharmacol 37(1): 92-104.

Kavlock R, Cummings A. 2005. Mode of action: inhibition of androgen receptor function - vinclozolin-induced malformations in reproductive development. Critical Reviews in Toxicology 35: 721-726.

Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, Lucier G, Luster M, Mac MJ, Maczka C, Miller R, Moore J, Rolland R, Scott G, Sheehan DM,

Sinks T, Tilson HA. 1996. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop. Environ Health Perspect 104 Suppl 4: 715-740.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol 25(1): 1-5.

Kortenkamp A, Martin O, Faust M, Evans R, McKinlay R, Orton F, Rosivatz E. 2012. State of the Art Assessment of Endocrine Disrupters - Final Report. Available on-line at: ec.europa.eu/environment.

Küster A, Bachmann J, Brandt U, Ebert I, Hickmann S, Klein-Goedicke J, Maack G, Schmitz S, Thumm E, Rechenberg B. 2009. Regulatory demands on data quality for the environmental risk assessment of pharmaceuticals. Regul Toxicol Pharmacol 55(3): 276-280.

Long X, Steinmetz R, Ben-Jonathan N, Caperell-Grant A, Young PC, Nephew KP, Bigsby RM. 2000. Strain differences in vaginal responses to the xenoestrogen bisphenol A. Environ Health Perspect 108(3): 243-247.

Love OP, Shutt LJ, Silfies JS, Bortolotti GR, Smits JE, Bird DM. 2003. Effects of dietary PCB exposure on adrenocortical function in captive American kestrels (Falco sparverius). Ecotoxicology 12(1-4): 199-208.

McCarthy MM. 2009. The two faces of estradiol: effects on the developing brain. Neuroscientist 15(6): 599-610.

McCarthy MM. 2010. How it's made: organisational effects of hormones on the developing brain. J Neuroendocrinol 22(7): 736-742.

Molina PE. 2010. Endocrine Physiology. 3rd ed. New York: McGraw-Hill Education.

Myers JP, vom Saal FS, Akingbemi BT, Arizono K, Belcher S, Colborn T, Chahoud I, Crain DA, Farabollini F, Guillette LJ Jr., Hassold T, Ho SM, Hunt PA, Iguchi T, Jobling S. Kanno J, Laufer H, Marcus M, McLachlan JA, Nadal A, Oehlmann J, Olea N, Palanza P, Parmigiani S, Rubin BS, Schoenfelder G, Sonnenschein C, Soto AM, Talsness CE, Taylor JA, Vandenberg LN, Vandenbergh JG, Vogel S, Watson CS, Welshons WV, Zoeller RT. 2009. Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A. Environ Health Perspect 117(3): 309-315.

Nadal A, Diaz M, Valverde MA. 2001. The estrogen trinity: membrane, cytosolic, and nuclear effects. News Physiol Sci 16: 251-255.

Newbold RR. 2004. Lessons learned from perinatal exposure to diethylstilbestrol. Toxicol Appl Pharmacol 199(2): 142-150.

NTP-CERHR. 2008. Monograph on the potential human reproductive and developmental effects of bisphenol A. Available on-line at: cerhr.niehs.nih.gov.

OECD. 2012. Information on OECD Work Related to Endocrine Disrupters. Available on-line at: www.OECD.org.

Palanza P, Parmigiani S, vom Saal FS. 2001. Effects of prenatal exposure to low doses of diethylstilbestrol, o,p'DDT, and methoxychlor on postnatal growth and neurobehavioral development in male and female mice. Horm Behav 40(2): 252-265.

Patel J, Landers K, Li H, Mortimer RH, Richard K. 2011. Thyroid hormones and fetal neurological development. J Endocrinol 209(1): 1-8.

Pesatori AC, Consonni D, Bachetti S, Zocchetti C, Bonzini M, Baccarelli A, Bertazzi PA. 2003. Short- and long-term morbidity and mortality in the population exposed to dioxin after the "Seveso accident". Ind Health 41(3): 127-138.

Plotkin D, Lechner JJ, Jung WE, Rosen PJ. 1978. Tamoxifen flare in advanced breast cancer. JAMA 240(24): 2644-2646.

Richter CA, Birnbaum LS, Farabollini F, Newbold RR, Rubin BS, Talsness CE, Vandenbergh JG, Walser-Kuntz DR, vom Saal FS. 2007. In vivo effects of bisphenol A in laboratory rodent studies. Reprod Toxicol 24: 199-224.

Rubin BS, Lenkowski JR, Schaeberle CM, Vandenberg LN, Ronsheim PM, Soto AM. 2006. Evidence of altered brain sexual differentiation in mice exposed perinatally to low, environmentally relevant levels of bisphenol A. Endocrinology 147(8): 3681-3691.

Rudén C. 2001a. Interpretations of primary carcinogenicity data in 29 trichloroethylene risk assessments. Toxicology 169(3): 209-225.

Rudén C. 2001b. The use and evaluation of primary data in 29 trichloroethylene carcinogen risk assessments. Regul Toxicol Pharmacol 34(1): 3-16.

Ryan BC, Vandenbergh JG. 2006. Developmental exposure to environmental estrogens alters anxiety and spatial memory in female mice. Horm Behav 50(1): 85-93.

SCF. 2002. European Commission - Health & Consumer Protection Directorate-General. Opinion of the Scientific Committee on Food on Bisphenol A. 2002. Available on-line at: http://ec.europa.eu

Simerly RB. 2002. Wired for reproduction: organization and development of sexually dimorphic circuits in the mammalian forebrain. Annu Rev Neurosci 25: 507-536.

Slob W. 1999. Thresholds in toxicology and risk assessment. International Journal of Toxicology 18: 259-268.

Stump DG, Beck MJ, Radovsky A, Garman RH, Freshwater LL, Sheets LP, Marty MS, Waechter JM Jr., Dimond SS, Van Miller JP, Shiotsuka RN, Beyer D, Chappelle AH, Hentges SG. 2010. Developmental neurotoxicity study of dietary bisphenol A in Sprague-Dawley rats. Toxicol Sci 115(1): 167-182.

Tortora GJ, Grabowski SR. 1996. Principles of Anatomy and Physiology. 8th ed. New York: HarperCollins Publishers Inc.

Tyl RW, Myers CB, Marr MC, Thomas BF, Keimowitz AR, Brine DR, Veselica MM, Fail PA, Chang TY, Seely JC, Joiner RL, Butala JH, Dimond SS, Cagen SZ, Shiotsuka RN, Stropp GD, Waechter JM. 2002. Three-generation reproductive toxicity study of dietary bisphenol A in CD sprague-Dawley rats. Toxicol Sci 68: 121-146.

Tyl, RW, Myers, CB, Marr, MC, Sloan, CS, Castillo, NP, Veselica, MM, Seely JC, Dimond SS, Van Miller JP, Shiotsuka RN, Beyer D, Hentges SG, Waechter JM. 2008. Two-generation reproductive toxicity study of dietary bisphenol A in CD-1 (Swiss) mice. Toxicol Sci 104, 362-384.

UNEP/WHO. 2012. State of the science of endocrine disrupting chemicals - 2012. Available on-line at: www.who.int.

US EPA. 1999. Determining the adequacy of existing data. Available on-line at www.epa.gov.

US EPA. 2005. Guidelines for carcinogen risk assessment. Available on-line at www.epa.gov.

US FDA. 2008. Draft assessment of bisphenol A for use in food contact applications. Available on-line at: www.fda.gov.

Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jacobs DR, Jr., Lee DH, Shioda T, Soto AM, Vom Saal FS, Welshons WV, Zoeller RT, Myers JP. 2012. Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses. Endocr Rev 33(3): 378-455.

Vandenberg LN, Hauser R, Marcus M, Olea N, Welshons WV. 2007. Human exposure to bisphenol A (BPA). Reprod Toxicol 24: 139-177.

Vogel SA. 2013. Is it safe? BPA and the struggle to define the safety of chemicals. Berkley, California: University of California Press, Ltd.

vom Saal FS, Belcher SM, Guillette LJ, Hauser R, Myers JP. 2007. Chapel Hill bisphenol A expert panel consensus statement: integration of mechanisms, effects in animals and potential to impact human health at current levels of exposure. Reprod Toxicol 24: 131-138.

vom Saal FS, Timms BG, Montano MM, Palanza P, Thayer KA, Nagel SC, Dhar MD, Ganjam VK, Parmigiani S, Welshons WV. 1997. Prostate enlargement in mice due to fetal exposure to low doses of estradiol or diethylstilbestrol and opposite effects at high doses. Proc Natl Acad Sci USA 94: 2056-2061.

Wandall B. 2004. Values in science and risk assessment. Toxicol Lett 152(3): 265-272.

Weed DL. 2005. Weight of evidence: a review of concept and methods. Risk Anal 25(6): 1545-1557.

White RH, Cote I, Zeise L, Fox M, Dominici F, Burke TA, White PD, Hattis DB, Samet JM. 2009. State-of-the-science workshop report: issues and approaches in low-dose-response extrapolation for environmental health risk assessment. Environ Health Perspect 117(2): 283-287.

WHO/IPCS. 2002. World Health Organization/International Programme on Chemical Safety. Global Assessment of the state-of-the-scinence of endocrine disruptors. Available on-line at www.who.int.

WHO/IPCS. 2009. World Health Organization/International Programme on Chemical Safety. Principles and methods for the risk assessment of chemicals in food. Environmental Health Criteria 240. Available on-line at: www.who.int.

WHO/IPCS. 2010. WHO Human Health Risk Assessment Toolkit: Chemical Hazards. Available online at: www.who.int.

Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58: 109-130.

Xu X, Tian D, Hong X, Chen L, Xie L. 2011. Sex-specific influence of exposure to bisphenol-A between adolescence and young adulthood on mouse behaviors. Neuropharmacology 61: 565-573.

Zoeller RT, Brown TR, Doan LL, Gore AC, Skakkebaek NE, Soto AM, Woodruff TJ, Vom Saal FS. 2012. Endocrine-disrupting chemicals and public health protection: a statement of principles from The Endocrine Society. Endocrinology 153(9): 4097-4110.