From the INSTITUTE OF ENVIRONMENTAL MEDICINE (IMM)
Karolinska Institutet, Stockholm, Sweden

APPLICATION OF THE BENCHMARK DOSE-RESPONSE MODELLING APPROACH FOR RISK CHARACTERIZATION OF CHEMICALS

Antero Vieira da Silva

Karolinska Institutet

Stockholm, 2022

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2022

ISBN 978-91-8016-561-7

BENCHMARK DOSE-RESPONSE MODELLING FOR HAZARD AND RISK CHARACTERIZATION
THESIS FOR DOCTORAL DEGREE (Ph.D.)

by

Antero Vieira da Silva

This thesis is publicly defended in the Andreas Vesalius room at the Karolinska Institutet, Stockholm, Sweden, on the 20th of May, 2022, at 09:00 – and virtually via the Zoom meeting platform.

Principal Supervisor:
Associate Professor Mattias Öberg, ERT
Karolinska Institutet
Institute of Environmental Medicine
Unit of Integrative Toxicology

Co-supervisors:
Elin Törnqvist, PhD
Karolinska Institutet
Institute of Environmental Medicine
Unit of Integrative Toxicology

Professor Annika Hanberg, ERT
Karolinska Institutet
Institute of Environmental Medicine
Unit of Biochemical Toxicology

Opponent:
Professor Elaine Faustman, DABT
University of Washington
Department of Environmental and Occupational Health Sciences

Examination Board:
Professor Anders Glynn
Swedish University of Agricultural Sciences
Department of Biomedical Sciences and Veterinary Public Health

Associate Professor Maria Kippler
Karolinska Institutet
Institute of Environmental Medicine
Unit of Metals and Health

Professor Margareta Törnqvist
Stockholm University
Department of Environmental Science (ACES)

To the world citizens, in general

And to my family, in particular

All things are poison and nothing is without poison.

Solely the dose determines that a thing is not a poison.

Philippus Aureolus Theophrastus Bombastus von Hohenheim (Paracelsus), 1538

Não sou nada.

Nunca serei nada.

Não posso querer ser nada.

À parte isso, tenho em mim todos os sonhos do mundo.

Trecho do poema "Tabacaria" de Álvaro de Campos,

heterónimo de Fernando Pessoa, Portugal, 1933.

I'm not nothing.

I'll never be nothing.

I can't want to be nothing.

Apart from that, I have in me all the dreams of the world.

Extract from the poem "Tabacaria" (Tobacco Shop) by Álvaro de Campos,

alter ego of Fernando Pessoa, Portugal, 1933.

ABSTRACT

Toxicology is the discipline that investigates the possible adverse effects of chemical exposure on human, animal and environmental health. Chemical risk assessment is the process that aims to identify potentially hazardous substances and describes the probability of adverse outcomes associated with their exposure. Biological changes and adverse effects do not occur after a threshold level is surpassed, but gradually and following a sequence of linked events. Traditionally, the no-observed-adverse-effect-level (NOAEL) approach has been used to detect the highest dose at which no adverse effect was observed. However, the NOAEL approach has methodological limitations and disadvantages that have resulted in it being increasingly replaced by the scientifically more advanced benchmark dose (BMD) approach. The BMD-modelling approach is a flexible method that takes all uncertainty and variability in the data into account, providing better estimates of doses leading to the potential adverse effects. Nonetheless, there are a number of knowledge gaps that need to be addressed and a lack of consensus persists regarding certain methodological aspects of this modelling strategy.

The overall aim of this thesis was to contribute to the BMD field and expand the knowledge base by applying this approach to the areas of risk assessment and pharmaceutical development, addressing some identified challenges and discussing potential improvements. In particular, this thesis covers three topics that are interconnected, namely the choice of the Critical Effect Size (CES) (**study I** and **III** to **VI**), the analysis of multiple endpoints (**study II** to **VI**) and the assessment of chemical mixtures (**study I**, **V** and **VI**). These topics were applied to data from studies on chemicals, namely per- and polyfluoroalkyl substances (PFAS) (**study I** and **VI**), polychlorinated biphenyls (PCBs) (**study IV** and **V**) and a candidate drug in pharmaceutical development (**study II** and **III**) and the pesticide norflurazon (**study VI**).

Study I combined human and animal data in order to derive the probabilistic risk for a 10% decrease in total triiodothyronine (T_3) hormone levels, depending on residency time. The human data consisted of perfluorooctanesulfonic acid (PFOS) and perfluorohexanesulfonic acid (PFHxS) serum levels from the resident population in Ronneby, a Swedish village that was highly exposed to PFAS through contaminated drinking water. The animal data originated from a 6-month subchronic study in monkeys, exposed to PFOS once a day. This integrated probabilistic risk assessment (IPRA) analysis demonstrated that longer exposure periods were associated with a larger proportion of the population at risk, ranging from 2.1% (90% C.I. 0.4% – 13.1%) to 3.5% (90% C.I. 0.7% – 21.8%) for residents exposed to PFOS and PFHxS for at least 1 or 29 years, respectively. This risk was mostly distributed among women, and exposure duration was the

greatest source of uncertainty (60.8%). It was concluded that IPRA is an advantageous method to calculate the risk for adverse effects, in comparison to the deterministic Margin of Exposure aproach (MoE).

Study II analyzed data from three subsequential safety assessment studies performed in rats to investigate the potential toxicity of an anti-oncogenic candidate drug in pharmaceutical development. The partial least squares (PLS) modelling approach was used to detect associations between clinical signs observed during the study, a 5% body weight decrease and pathological findings noted after study termination. Piloerection, eyes half shut and slightly decreased motor activity were the signs that were most strongly associated with the pathological findings, and the models accurately predicted the injuries observed in the thymus, testes, epididymides and bone marrow. The findings indicate that an evaluation of clinical signs as an integrated toxicity evaluation has potential 3R (Replacement, Reduction and Refinement of animal use) gains, especially in terms of Refinement of animal studies. The study suggests that the PLS-modelling approach can be employed to predict pathological changes, monitor animal welfare and support the decision-making process during pre-clinical safety and toxicity assessment studies.

Study III analyzed the same data as **study II**, but applied the BMD-modelling approach instead, with a different objective, namely to describe potential relationships between the dose and the findings made in the 63 examined endpoints. The endpoints modelled included biochemistry and hematology endpoints, body weight changes, organ pathology findings and clinical observations. The resulting BMDs and BMDLs were compared to the study NOAEL (or LOAEL) and were often lower than the estimates of the NOAEL approach. A 5% change was also compared to the findings based on an adversity threshold derived from the observed and endpoint-specific magnitude of change. Additionally, the BMD-modelling was also considered to have a strong focus on the Refinement of animal studies. In summary, it was shown that modelling multiple endpoints is desirable, providing a more complete overview of the potential toxicity of a candidate drug and improving the pharmaceutical development process.

Study IV assessed the potential toxicity of PCB-156 (2,3,3′,4,4′,5-hexachlorobiphenyl) following a 90-day study in rats exposed daily through their diet. Dose-dependent toxicological effects were described, including body and organ changes but also in the assessed retinoid system endpoints. Retinoid disruption and effects in the organs of rats were demonstrated employing the BMD dose-response modelling approach, revealing that the apolar liver retinoid concentrations were the most sensitive endpoint. The retinoid system was shown to be sensitive to PCB-156 exposure, and it was suggested that its endpoints should be more often considered for chemical risk assessment purposes.

Study V employed the BMD method to calculate relative potency factors (RPFs) for seven PCBs (PCB-28, 77, 105, 118, 128, 153 and 156) and one PCB-mixture. PCB-126 was used as an index chemical, and the eight 90-day regulatory toxicity studies for the individual congeners were performed under the same conditions (the PCB-mixture study was 28 days long). The liver apolar retinoids levels and concentration, and the remaining endpoints examined, estimated greater RPFs than those calculated by the World Health Organization (WHO) in 2006 (Van den Berg et al., 2006), being suggestive of a hazard underestimation. In fact, the potency factors estimated in this study, based on the ethoxyresorufin-O-deethylase (EROD) enzymatic activity (a historically used endpoint to calculate RPFs), were the lowest in comparison to other endpoints for which RPFs were calculated. In summary, RPFs were useful to describe the potential toxicity of structurally similar compounds, expressed in units equivalent to the index chemical, and the retinoid system proved once again to be susceptible to changes following low-dose PCB exposure.

Study VI focused on the choice of CES, a matter of debate when applying the BMD method to continuous data. Currently, there is no internationally harmonized approach to choosing the CES, and five strategies were examined: the EFSA default value of 5% or 10%, the US EPA 1 SD approach, an endpoint-specific CES based on historical data, the General Theory of Effect Size (GTES) and expert judgment. All examined strategies featured advantages and limitations, and the different choices of CES led to distinct reference values when applied to five case-studies, analyzing PFAS, PCB-156 and a pesticide (norflurazon) data. Although some of these strategies delivered similar CES values, it was not always the case, and reliance on a single method to choose the CES is not recommendable. It was concluded that expert judgment is irreplaceable and that the decision-making process performed by risk assessors and managers regarding the likely threshold of adversity should be supported by BMD analysis of the data comparing different CES. This could lead to a better overview of the data package and understanding of the doses leading to different magnitudes of effects, which would lead to better motivation of the choices and decisions made.

In conclusion, this thesis demonstrates that the BMD method is a flexible modelling approach to assess the potential effects of several classes of substances, such as PFAS, PCBs and candidate drugs. Possible applications in the chemical risk assessment and pharmaceutical development areas were demonstrated. Additionally, it was shown that the BMD approach has a strong 3R potential and extracts a considerable amount of information from the data. The BMD approach is in chemical risk assessment to stay, and much like a Swiss army knife, it is a useful and multi-purpose tool that will support you in the derivation of reference values of superior quality.

POPULAR SCIENCE SUMMARY

We are all daily and continuously exposed to chemicals that can be beneficial or harmful to our health. Toxicology is the field of science that studies the potential adverse effects of chemicals on humans, animals and the environment. This potential impact depends not only on the danger posed by the chemicals but also on the amount and duration of exposure. For example, while water is not usually harmful, one might die after drinking 15 liters in one day. Risk assessment is the area of Toxicology concerned with studying the potential risk to human, animal and environmental health as a result of chemical exposure. Risk assessment is divided into four steps: hazard identification, hazard characterization, exposure assessment and risk characterization. This thesis focuses on the hazard characterization step, more specifically on dose-response assessment, which calculates the dose or concentration that will be compared to the exposure measured in the studied population in the next step. The ultimate goal of risk assessment is thus to establish reference values that are non-toxic and as safe as possible for human, animal and environmental health. Reference values can, for example, be the acceptable daily intakes found in the nutritional declaration of foodstuff.

In this thesis, knowledge is deepened in the dose-response assessment using an approach based on mathematical models, called the benchmark dose (BMD) method. Compared to other historically used modelling approaches, such as the NOAEL, the BMD method features many advantages, including taking all data into account, interpolation between doses and handling uncertainty more efficiently.

The general objective of this thesis is to investigate and exemplify how the BMD approach can be employed to lead to better risk assessment through dose-response evaluations of superior quality. It focused on three questions where the knowledge could be improved, namely the choice of CES (**study I** and **III** to **VI**), analysis of multiple endpoints (**study II** to **VI**) and the assessment of chemical mixtures (**study I**, **V** and **VI**). These three topics were applied to data from chemicals for which the potential hazard needs to be better understood, specifically per- and polyfluoroalkyl substances (PFAS) (**study I** and **VI**), polychlorinated biphenyls (PCBs) (**study IV** and **V**), a candidate drug in pharmaceutical development (**study II** and **III**) and the pesticide norflurazon **(study VI**).

In **study I**, data from a Swedish population in Ronneby, highly exposed to PFAS through drinking water, were combined with data from a study in monkeys to assess the risk of a potentially adverse 10% decrease in the total T_3 hormonal levels. It was shown that the risk for adverse effects, i.e. disrupted thyroid hormonal balance, was not negligible. In fact, the risk was mainly distributed among women and depended on residency time – longer residency times were associated with greater risk.

Study II examined three pharmaceutical development studies, where laboratory rats were exposed to an anti-cancer candidate drug, to research how clinical signs and behavioral changes might be associated with pathological changes in organ tissues observed after study termination. The PLS regression method was employed to predict the post-mortem pathological findings based on the clinical signs observed during the studies. Thus, the usefulness of the PLS method for study assessment and animal welfare monitoring was investigated, enabling potential refinement of research using animals according to the 3R-principles (Replacement, Reduction and Refinement). The most important signs observed were piloerection, eyes half shut and slightly decreased motor activity, which were the best predictors of toxicity found later in the pathological findings in the rats' organs. The models predicted well changes in the thymus, testes, epididymis, bone marrow and a 5% change in body weight.

In **study III**, the same data as for **study II** were analyzed, this time employing the BMD approach to model all 63 endpoints for which data were available. It was shown that the BMD method was useful and applicable for pharmaceutical development purposes. The dose-response analysis could be adapted to the size of the change desirable to measure and it was helpful for understanding at what doses the effects occur, which was valuable for the comparison of changes across different endpoints. In brief, the BMD approach contributes not only to the safety assessment of pharmaceutical drugs but also to the dose-setting process of future studies, focusing simultaneously on the 3R principles.

In **study IV**, the adverse effects of exposure to a toxic chemical (PCB-156) were described, based on the findings in laboratory rats exposed for 90 days through their diet. BMD-modelling showed that increasing doses of PCB-156 were associated with effects in the rats' liver, kidneys and lungs, but also changes in body weight and liver enzymatic activities. It was suggested that changes in the retinoid levels and concentrations in different organs could be sensitive biomarkers for detecting adverse effects following PCB-156 exposure.

Study V followed a similar train-of-thought as **study IV** but applied the BMD approach to eight studies on individual PCBs and one PCB mixture. The aim was to compare the potency of different compounds in relation to PCB-126, the best-studied PCB, based on changes in the retinoid system endpoints, and other endpoints such as organ and body weight changes, but also the historically important ethoxyresorufin-O-deethylase (EROD) enzymatic activity. The calculated potency factors estimated a greater toxic potential for the studied PCBs than the latest World Health Organization values, published in 2006. Thus, a potential hazard underestimation for the researched PCBs was suggested.

In **study VI**, a key measure and matter of discussion in the BMD field was analyzed – what percentage of effect from the background needs to be measured to be considered adverse. Five strategies of international importance were assessed and applied to data from five case-studies examining to illustrate the different possible outcomes, namely PFAS, PCB-156 and a pesticide (norflurazon). It was concluded that while the combination of different strategies may provide insights into the data, the judgment of experts is irreplaceable and will remain the principal driver of decision-making for risk assessment and management purposes.

In summary, this thesis shows that the use of the BMD-modelling approach for risk assessment and pharmaceutical development purposes is useful and advantageous. It extracts more information from the data, leading to better-motivated decisions and reference values for the potential adverse levels at which chemicals may affect the organisms under examination.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Alla utsätts vi dagligen och kontinuerligt för kemiska substanser som kan vara gynnsamma eller skadliga för vår hälsa. Toxikologi är vetenskapen som studerar de skadliga effekterna av kemikalier på människor, djur och miljö. Den potentiella påverkan beror på de inneboende faror som kemikalien har, men också på mängden, koncentrationen och exponeringens varaktighet. Det går till exempel att dö genom att dricka 15 liters vatten på en dag, trots att vatten är en kemikalie som normalt inte är skadlig. Riskbedömning är ett område inom toxikologi där man studerar och värderar risker med kemikalier för människa, djur och miljö. Riskbedömning kan indelas i fyra steg: faroidentifiering, farokaraktärisering, exponeringsbedömning och riskkaraktärisering. Denna avhandling fokuserar på steg två – farokaraktärisering – och mer specifikt på dos-responsanalys, i vilken man beräknar den dos eller koncentration som sedan kan jämföras med den uppmätta exponeringen. Ett viktigt mål med riskbedömning är att fastställa riktvärden som anses vara icke-toxiska och så säkra som möjlig för människors hälsa. Två exempel på riktvärden, som har betydelse i avhandlingen, är de åtgärdsgränser som Livsmedelsverket tagit fram för PFAS i dricksvatten och de tolerabla dagliga intagen som den europeiska livsmedelssäkerhetsmyndigheten (EFSA) tar fram.

I den här avhandling fördjupas kunskapen inom dos-responsbedömning med hjälp av en metod som använder matematiska modeller och kallas benchmark dos-modellering (BMD). Jämfört med traditionella metoder som att beräkna icke-effekt-nivåer (no-observed-adverse-effect-level, NOAEL), har BMD många fördelar. Bland annat tar BMD-metoden hänsyn till alla data och gör det möjligt att interpolera mellan doser. Dessutom kan metoden beskriva osäkerhet i data på ett bättre sätt.

Det övergripande syftet med denna avhandling är att utveckla och exemplifiera hur BMD-metoden kan användas för att förbättra riskbedömningar. Den fokuserade på tre övergripande frågeställningar där kunskapen behöver förbättras, nämligen valet av kritisk effekt-nivå (CES) (**studie I** och **III** till **VI**), analys av flera endpoints (**studie II** till **VI**) samt bedömning av kemiska blandningar (**studie I**, **V** och **VI**). Dessa tre frågeställningar tillämpas på data från olika kemikalier: per- och polyfluoralkylsubstanser (PFAS) (**studie I** och **VI**), polyklorerade bifenyler (PCB) (**studie IV** och **V**), en läkemedelskandidat som testas under läkemedelsutveckling (**studie II** och **III**), samt ett bekämpningsmedel (**studie VI**).

Studie I utgår från data relaterat till en population i Ronneby som exponerats för höga halter av PFAS via dricksvattnet. Information om exponeringen kombineras med resultat från en studie på apor, för att bedöma risken att drabbas av en 10% minskning av ett sköldkörtelhormon (total T_3).

Studien visar att risken för ogynnsamma effekter, det vill säga påverkan på balansen av sköldkörtelhormon, inte var försumbar och att risken var störst bland kvinnor. Högst risk hade de kvinnor som bott i området under längst tid.

I **studie II** studerads hur kliniska symtom och beteendeförändringar hos laboratorieråttor kan kopplas till patologiska förändringar i olika organ. PLS-regressionsmetoden används på data från läkemedelsutvecklingsstudier för att förutsäga patologiska skador utifrån de kliniska symptomen som observeras. Användbarheten av PLS-metoden undersöks både för toxikologisk bedömning och för övervakning av djursvälfärd. Det senare innebär en potentiell förbättring av försöksdjurens välfärd i enlighet med 3R-principerna (Replacement, Reduction and Refinement). De viktigaste kliniska symtomen som observeras är piloerektion (resning av pälshår), halvstängda ögon och något minskad motorisk aktivitet. Dessa symtom visade sig vara bra prediktorer för de patologiska fynden i råttornas organ. Modellerna förutspådde förändringar i sköldkörteln, testiklar, bitestikel, benmärg samt en 5% kroppsviktsminskning.

I **studie III** analyseras samma data som i **studie II**, denna gång med BMD-metoden för att modellera samtliga 63 endpoints för vilka data fanns tillgängliga. Studien visar att BMD-metoden är användbar för läkemedelsutveckling. Dos-responsbedömning kan anpassas till önskvärd storlek av förändringen och den kan hjälpa till att förstå vid vilka doser olika effekter inträffar. Sammanfattningsvis bidrar BMD-metoden inte bara till säkerhetsbedömningen av läkemedelskandidater utan också till att bestämma lämpliga doser för framtida studier, samtidigt som man kan förbättra djurvälfärden enligt fokuserar på 3R-principerna.

I **studie IV** beskrivs de negativa effekterna av exponering för en giftig kemikalie (PCB-156), baserat på fynden hos laboratorieråttor som exponerats i 90 dagar genom sin diet. BMD-modellering visar att ökande doser av PCB-156 är associerat till effekter i råttornas lever, njurar och lungor, men också till förändringar i kroppsvikt och leverenzymatiska aktiviteter. En slutsats är att förändringar i retinoidnivåer i olika organ kan vara känsliga markörer för att upptäcka negativa effekter efter PCB-156 exponering.

Studie V följer en liknande röd tråd som **studie IV**, men tillämpar BMD-metoden på åtta studier av individuella PCB:er samt en PCB-blandning. Syftet är att jämföra styrkan hos olika PCB:er i förhållande till en mycket välstuderad PCB – PCB-126. De effekter som ingår i studien är förändringar i retinoidsystemet, organ- och kroppsviktsförändringar, samt den historiskt ofta använda enzymaktiviteten av etoxyresorufin-O-deetylas (EROD). De beräknade potensfaktorerna visar att det finns en större toxisk potential de studerade PCB:erna än vad som framgår av Världshälsoorganisationens potensfaktorer från 2006. Således, baserat på resultat i denna studie, kan det finnas en underskattning av risken för de undersökta PCB:erna.

I **studie VI** analyseras en viktig del av BMD-metoden – vilken procentuell förändring av en effekt man bör anse vara negativ för hälsan och använda som utgångspunkt för beräkningarna? Fem strategier av internationell betydelse utvärderas och tillämpas på data från fem olika fall-studier. De kemikalier som undersöks, för att illustrera de olika möjliga resultaten, var PFAS, PCB-156 samt ett bekämpningsmedel (norflurazon). Slutsatsen är att en kombination och aktiv jämförelse av olika strategier är önskvärd och kan ge fördjupande insikter. Därför krävs expertbedömning och en sådan bör förbli en central del av riskbedömning och riskhantering.

Sammanfattningsvis visar denna avhandling att användningen av BMD-modelleringsmetoden för riskbedömning och läkemedelsutveckling är användbar och har många fördelar. Den extraherar mer information från data, vilket leder till bättre motiverade beslut och riktvärden för de potentiellt skadliga nivåer vid vilka kemikalier som undersökts kan påverka organismerna.

RESUMO CIENTÍFICO EM LINGUAGEM POPULAR

Todos nós estamos expostos diária e continuamente a compostos químicos que podem ser benéficos ou prejudiciais para a nossa saúde. A Toxicologia é a ciência que estuda os efeitos potencialmente adversos das substâncias químicas nos humanos, animais e no ambiente. Este potencial impacto depende, por sua vez, do perigo que o químico representa, mas também da quantidade, da concentração e duração da exposição ao químico. Por exemplo, a água, que é um químico normalmente não prejudicial, pode ser fatal se um humano beber 15 litros num só dia. A avaliação do risco é a área de Toxicologia que se ocupa em estudar o potencial risco para a saúde humana, animal e ambiental, devido à exposição química. A avaliação do risco dividir-se em identificação do risco, caracterização do perigo, avaliação da exposição e caracterização do risco. Esta tese foca-se na etapa de caracterização do perigo, mais especificamente na avaliação dose-resposta, que calcula a dose ou concentração que provoca efeitos adversos, que será depois comparada com a avaliação da exposição medida na população estudada, no passo denominado caracterização do risco. O objetivo final da avaliação de risco é estabelecer valores de referência não tóxicos e tão seguros quanto possível para a saúde humana, animal e ambiental. Os valores de referência podem ser, por exemplo, as encontrados na declaração nutricional dos alimentos.

Nesta tese, aprofunda-se o conhecimento na avaliação dose-resposta utilizando uma abordagem denominada benchmark dose-modelling (BMD). Comparado com outras abordagens historicamente usadas, como no-observed-adverse-effect-level (NOAEL), o método BMD apresenta muitas vantagens, incluindo levar em consideração todos os dados, interpolação entre doses e descrição da incerteza nos dados de forma mais eficiente.

O objetivo geral desta tese é investigar e exemplificar como o método BMD pode ser empregue para uma melhor avaliação do perigo que os químicos podem representar através de avaliações dose-resposta de qualidade superior. Concentrou-se em três áreas onde o conhecimento pode ser melhorado, nomeadamente a escolha do critical effect size (CES) (**estudo I** e **III** a **VI**), análise de múltiplos parâmetros (**estudo II** a **VI**) e a avaliação de misturas químicas (**estudo I**, **V** e **VI**). Estes três tópicos foram ilustrados com dados de estudos de químicos para os quais o perigo potencial precisa ser melhor compreendido, nomeadamente substâncias per- e polifluoroalquil (PFAS) (**estudo I** e **VI**), bifenilos policlorados (PCBs) (**estudo IV** e **V**), um candidato a medicamento em desenvolvimento farmacêutico (**estudo II** e **III**) e o pesticida norflurazon (**estudo VI**).

No **estudo I**, dados de uma população sueca residente em Ronneby, altamente exposta a PFAS através de água contaminada, foram combinados com dados de um estudo em macacos para avaliar o risco de uma diminuição potencialmente adversa de 10% nos níveis hormonais de T_3 total.

Foi demonstrado que o risco de efeitos adversos, ou seja, perturbação do equilíbrio hormonal da tiroide, não era insignificante. Maioritariamente mulheres estavam em riscos de efeitos adversos, risco este maior quanto mais longamente estas viveram em Ronneby.

No **estudo II** foram examinados três estudos de desenvolvimento farmacêutico, onde ratos de laboratório foram expostos a um candidato a medicamento contra o cancro, para pesquisar como sinais clínicos e alterações comportamentais podem estar associados a alterações patológicas em tecidos de órgãos observadas após o término do estudo. O método de regressão PLS foi empregue para prever alterações patológicas nos órgãos com base nos sinais clínicos observados durante os estudos. Deste modo, a utilidade do método PLS foi investigada para a avaliação de estudos e do bem-estar animal, possibilitando uma potencial melhoria dos estudos em animais, de acordo com os princípios 3R (Replacement, Reduction and Refinement). Os sinais mais importantes observados foram piloereção, olhos semicerrados e atividade motora ligeiramente diminuída, sendo estas as melhores variáveis para prever a toxicidade encontrada posteriormente nas alterações patológicas nos órgãos dos ratos. Os modelos previram bem mudanças no timo, testículos, epidídimo, medula óssea e uma alteração 5% no peso corporal.

No **estudo III**, os mesmos dados do **estudo II**, desta vez analisados empregando o método BMD para estudar os 63 parâmetros para os quais foram recolhidos dados. Foi demonstrado que o método BMD é útil e aplicável para fins de desenvolvimento farmacêutico. A análise de dose-resposta pode ser adaptada à dimensão da mudança que se deseja medir para melhor descrever a que doses os efeitos ocorrem, o que é útil para comparação de efeitos em diferentes parâmetros. Em suma, o método BMD contribui não só apenas para a avaliação da segurança dos candidatos a medicamentos durante estudos farmacêuticos, mas também para o processo de seleção das doses em estudos futuros, focando-se simultaneamente nos princípios dos 3R.

No **estudo IV**, foram descritos os efeitos adversos da exposição a um produto químico tóxico (PCB-156), com base nos resultados de um estudo em ratos de laboratório expostos por 90 dias através de dieta. O método BMD mostrou que maiores doses PCB-156 foram associadas a efeitos no fígado, rins e pulmões dos ratos, mas também alterações no peso corporal e nas atividades enzimáticas do fígado. Foi sugerido que as alterações nos níveis dos compostos retinóides em diferentes órgãos podem ser usados como marcadores sensíveis para a deteção de efeitos adversos após a exposição ao PCB-156.

O **estudo V** seguiu uma linha de pensamento semelhante ao **estudo IV**, mas aplicou o método BMD a oito estudos onde PCBs foram estados individualmente, assim como uma mistura de PCBs. O objetivo foi comparar a potência dos diferentes compostos em relação ao PCB-126, o PCB mais

bem estudado, com base em alterações nos parâmetros do sistema retinóide, assim como outros como alterações de peso corporal e de diferentes órgãos, para além da historicamente importante atividade enzimática da etoxiresorufina-O-deetilase (EROD). Os fatores de potência calculados para os PCBs estudados estimaram um potencial tóxico maior do que os últimos valores da Organização Mundial da Saúde, publicados em 2006. Assim, sugeriu-se uma subestimação do risco que os PCBs pesquisados podem representar.

No **estudo VI**, foi analisada uma questão central no método BMD – qual a percentagem de efeito, isto é, qual o tamanho da mudança que precisa ser medida para ser considerado adversa. Cinco estratégias de importância internacional foram avaliadas e aplicadas a dados de cinco casos de estudos para ilustrar os diferentes resultados possíveis, nomeadamente PFAS, PCB-156 e um pesticida (norflurazon). Concluiu-se que, embora a combinação de diferentes estratégias possa fornecer sugestões perspicazes sobre mudanças relevantes para os efeitos observados, o julgamento de especialistas é insubstituível e continuará a ser a principal maneira através da qual os avaliadores de riscos irão basear as suas decisões.

Em resumo, esta linda tese mostra que o uso do método BMD para fins de avaliação de risco e desenvolvimento farmacêutico é útil e vantajoso. Não só extrai mais informação dos dados obtidos, como leva a tomadas de decisão baseadas em melhores argumentos e num melhor entendimento dos níveis potencialmente adversos através dos quais os químicos em estudo podem afetar os organismos examinados.

RESUMEN DE DIVULGACIÓN CIENTÍFICA

Todos estamos expuestos diaria y continuamente a compuestos químicos que pueden ser beneficiosas o perjudiciales para nuestra salud. La Toxicología es un campo científico en el cual se estudian los posibles efectos adversos de los productos químicos en seres humanos, animales y el medio ambiente. Este potencial impacto depende no solo del peligro que representan los productos químicos en sí, sino también de la cantidad y la duración de la exposición a ellos. Por ejemplo, aunque el agua no suele ser dañina, beber 15 litros en un día puede ser mortal. La evaluación de riesgos es el área de la Toxicología que se ocupa del estudio del riesgo potencial para la salud humana, animal y ambiental como resultado de la exposición química. La evaluación del riesgo se divide en cuatro pasos: identificación del peligro, caracterización del peligro, evaluación de la exposición y caracterización del riesgo. Esta tesis se centra en el paso de caracterización del peligro, más específicamente en la evaluación de la dosis-respuesta, en la cual se calcula la dosis o concentración usada, en el siguiente paso, para comparación con la exposición medida en la población estudiada. El objetivo final de la evaluación de riesgos es, por lo tanto, el establecer concentraciones de compuestos químicos, usadas como valores de referencia, en las cuales no se evidencia toxicidad y por lo tanto proveen un rango de seguridad para la salud humana, animal y ambiental. Los valores de referencia pueden ser, por ejemplo, los niveles de ingestas diarias aceptables, encontrados en la declaración nutricional de alimentos.

Esta tesis profundiza el conocimiento de la evaluación dosis-respuesta utilizando un enfoque basado en modelos matemáticos, denominado método de benchmark dose (BMD). En comparación con otros enfoques de modelado utilizados históricamente, como el no-observed-adverse-effect-level (NOAEL), el método BMD provee ventajas considerables tales como la consideración de todos los datos, la interpolación entre dosis y el manejo de la incertidumbre de manera más eficiente.

El objetivo general de esta tesis es investigar y ejemplificar cómo se puede emplear el método BMD para conducir a una mejor evaluación del riesgo a través de evaluaciones de dosis-respuesta con calidad superior. La tesis se centra en tres preguntas en las cuales el conocimiento actual puede ser incrementado, estas son la elección de CES (**estudio I** y **III** a **VI**), el análisis de múltiples puntos finales (**estudio II** a **VI**) y la evaluación de mezclas químicas (**estudio I**, **V** y **VI**). Estos tres temas se aplicaron a datos de sustancias químicas cuyo peligro potencial requiere mayor comprensión, específicamente sustancias perfluoroalquiladas y polifluoroalquiladas (PFAS) (**estudios I** y **VI**), compuestos bifenilos policlorados (PCB) (**estudios IV** y **V**) y un fármaco candidato en desarrollo farmacéutico (**estudio II** y **III**).

En el **estudio I**, datos de una población sueca en Ronneby altamente expuesta a PFAS a través del agua potable se combinaron con datos de un estudio en monos para evaluar el riesgo de disminución potencialmente adversa del 10% en los niveles de la hormona T_3. Se demostró que el riesgo de efectos adversos, como por ejemplo la alteración del equilibrio hormonal de la tiroides, no fue insignificante. De hecho, el riesgo se distribuyó principalmente entre mujeres y fue dependiente del tiempo de residencia en el área ─ mayores tiempos de residencia se asociaron con mayor riesgo.

El **estudio II** examinó tres estudios de desarrollo farmacéutico en los que se expuso a ratas de laboratorio a un fármaco candidato contra el cáncer, con el fin de investigar cómo signos clínicos y cambios de comportamiento en los animales podrían estar asociados con cambios patológicos en los tejidos de los órganos observados después de la finalización del estudio. Se empleó el método de regresión partial least squares (PLS) para predecir los hallazgos patológicos post-mortem en función de los signos clínicos observados durante los estudios. Por lo tanto, se investigó la utilidad y aplicabilidad del método PLS para la evaluación del estudio y el control del bienestar animal, lo que permitió el refinamiento potencial de la investigación con animales de acuerdo con los principios de 3R (Replacement, Reduction and Refinement). Los signos más importantes observados fueron piloerección, ojos entrecerrados y actividad motora levemente disminuida, los cuales fueron los mejores predictores de toxicidad encontrados posteriormente en los hallazgos patológicos en los órganos de las ratas. Los modelos predijeron bien los cambios en el timo, los testículos, el epidídimo, la médula ósea y un cambio del 5 % en el peso corporal.

En el **estudio III**, se analizaron los mismos datos que en el **estudio II**, esta vez empleando el enfoque de BMD para modelar los 63 criterios de valoración para los que se recopilaron datos. Se demostró que el método BMD era útil y aplicable para fines de desarrollo farmacéutico. El análisis de dosis-respuesta se pudo adaptar al tamaño del cambio deseable para medir y ayudó a comprender a cuales dosis se producen los efectos, lo cual fue valioso para la comparación de cambios en diferentes criterios de valoración. En resumen, el enfoque del método BMD contribuyó no solo a la evaluación de la seguridad de los estudios farmacéuticos, sino también al proceso de establecimiento de dosis de futuros estudios, centrándose simultáneamente en los principios de las 3R.

En el **estudio IV**, se describieron los efectos adversos de la exposición a un compuesto químico tóxico (PCB-156), con base en los hallazgos en ratas de laboratorio expuestas durante 90 días a través de la dieta. El método de BMD mostró que dosis crecientes de PCB-156 estaban asociadas con efectos en el hígado, los riñones y los pulmones de las ratas, pero también con cambios en el peso corporal y las actividades enzimáticas del hígado. Se sugirió que los cambios en los niveles de retinoides y concentraciones en diferentes órganos podrían ser biomarcadores sensibles para detectar efectos adversos después de la exposición al PCB-156.

El **estudio V** siguió una línea de pensamiento similar a la del **estudio IV** pero aplicó el método BMD a ocho estudios sobre PCB individuales y una mezcla de PCB. El objetivo era comparar la potencia de diferentes compuestos en relación con el PCB-126, el PCB mejor estudiado, en función de los cambios en los criterios de valoración de retinoides pero también otros, como cambios en el peso corporal y de los órganos, pero también en el valor históricamente importante de actividad enzimática de la etoxiresorufina-O-deetilasa (EROD. Los factores de potencia calculados estimaron un potencial tóxico mayor para los PCB estudiados que los últimos valores de la Organización Mundial de la Salud, publicados en 2006. Por lo tanto, se sugirió una subestimación del peligro potencial para los PCB investigados.

En el **estudio VI**, se analizó una medida clave y tema de discusión en el campo de la BMD: qué porcentaje del efecto de fondo se debe medir para que se considere adverso. Se evaluaron y aplicaron cinco estrategias de importancia internacional a los datos de cinco estudios de caso que examinaron PFAS, PCB-156 y un plaguicida (norflurazon), para ilustrar los diferentes resultados posibles. Se concluyó que, si bien la combinación de diferentes estrategias puede proporcionar información sobre los datos, el juicio de los expertos es irremplazable y seguirá siendo el principal impulsor de la toma de decisiones con fines de evaluación y gestión de riesgos.

En resumen, esta tesis muestra que el uso del método BMD para fines de evaluación de riesgos y desarrollo farmacéutico es útil y ventajoso. Este método facilita la extracción de más información de los datos, suministrando herramientas para la toma de decisiones y el establecimiento de niveles de concentraciones potencialmente adversas a las cuales productos químicos pueden afectar organismos bajo examen.

LIST OF SCIENTIFIC PAPERS

I. **Vieira Silva A**, Ringblom J, Lindh C, Scott K, Jakobsson K, Öberg M. 2020. A probabilistic approach to evaluate the risk of decreased total triiodothyronine hormone levels following chronic exposure to PFOS and PFHxS via contaminated drinking water. Environmental Health Perspectives 128(7):076001. doi.org/10.1289/ehp6654

II. **Vieira Silva A**, Norinder U, Liiv E, Platzack B, Öberg M, Törnqvist E. 2021. Associations between clinical signs and pathological findings in toxicity testing. ALTEX – Alternatives to Animal Experimentation 38(2):198-214. doi.org/10.14573/altex.2003311

III. **Vieira Silva A**, Ringblom J, Moldeus P, Törnqvist E, Öberg M. Benchmark dose-response analyses for multiple endpoints in drug safety evaluation. 2021. Toxicology and Applied Pharmacology 433:115732. doi.org/10.1016/j.taap.2021.115732

IV. **Vieira Silva A**, Chu I, Feeley M, Bergman Å, Håkansson H*, Öberg M*. 2022. Dose-dependent toxicological effects in rats following a 90-day dietary exposure to PCB-156 include retinoid disruption. Reproductive Toxicology 107:123-139. doi.org/10.1016/j.reprotox.2021.09.012

V. **Vieira Silva A**, Chu I, Håkansson H, Öberg M. Relative potency factors for polychlorinated biphenyls (PCBs) based on retinoid level changes in rats' liver. Manuscript.

VI. **Vieira Silva A***, Brunken L*, Öberg M. The choice of Critical Effect Size affects the outcome of chemical risk assessment using benchmark dose-modelling – a retrospective analysis illustrated with case-studies focusing on PFAS, PCB-156 and the norflurazon pesticide. 2022. Manuscript.

* = These authors contributed equally to this work.

Publication during the doctoral studies not included in this thesis:

Norgren K, Tuck A, **Vieira Silva A**, Burkhardt P, Öberg M, Munic Kos V. High content screening of bisphenols and their mixtures in conditions of low-intensity adipogenesis of human mesenchymal stem cells (hMSCs). 2022. Food and Chemical Toxicology 161:112842. doi.org/10.1016/j.fct.2022.112842

The publications in this thesis are referred to by their roman numerals (I – VI). The articles are reproduced with permission from the respective publishers.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

3R	Replacement, Reduction and Refinement of animal use in research and testing
ADI	Acceptable daily intake
AhR	Aryl hydrocarbon receptor
AIC	Akaike Information Criterion
AOP	Adverse outcome pathways
APROBA	Approximate probabilistic risk analysis
BMD	Benchmark dose
BMDL	Lower limit of the BMD's 90% confidence interval
BMDU	Upper limit of the BMD's 90% confidence interval
BMR	Benchmark Response
CAS	Chemical Abstracts Service
CES	Critical effect size
ECHA	European Chemicals Agency
EFSA	European Food Safety Authority
EROD	Ethoxyresorufin-O-deethylase
FAO	Food and Agriculture Organization of the United Nations
GTES	General Theory of Effect Size
IBMD	Individual benchmark dose distribution
IEXP	Individual exposure distribution
IMoE	Individual Margin of Exposure
IPCS	WHO International Programme on Chemical Safety
IPRA	Integrated probabilistic risk assessment
JECFA	Joint FAO-WHO Expert Committee on Food Additives

JMPR	Joint FAO-WHO Meeting on Pesticide Residues
LOAEL	Lowest-observable-adverse-effect level
MA	Model averaging
MOE	Margin of Exposure
NOAEL	No-observed-adverse-effect-level
OECD	Organisation for Economic Co-operation and Development
PCB	Polychlorinated biphenyl
PCB-156	2,3,3′,4,4′,5-hexachlorobiphenyl
PFAS	Per- and polyfluoroalkyl substances
PFNA	Perfluorononanoic acid
PFOA	Perfluorooctanoic acid
PFOS	Perfluorooctanesulfonic acid
PFHxS	Perfluorohexanesulfonic acid
PLS	Partial least squares
PoCE	Probability of Critical Exposure
PoD	Point-of-departure
POP	Persistent Organic Pollutant
PROD	Pentoxyresorufin-O-dealkylase
RPF	Relative potency factor
RIVM	National Institute for Public Health and the Environment of the Netherlands
SD	Standard deviation
TCDD	2,3,7,8-tetrachlorodibenzo-*p*-dioxin
UDPGT	Uridine 5'-diphospho-glucuronosyl transferase
US EPA	United States Environment Protection Agency
WHO	World Health Organization

1 INTRODUCTION

The main focus of this thesis is to contribute to the advancement of research in Toxicology in general and risk assessment in particular. It focuses on the applications and challenges faced by the benchmark dose-response modelling approach. Special attention is given to the choice of the Critical Effect Size (CES), the assessment of multiple endpoints and chemical mixtures. The substances assessed, namely per- and polyfluoroalkyl substances (PFAS), polychlorinated biphenyls (PCBs) and an anti-oncogenic candidate drug, were chosen as relevant examples for which the risk and hazard characterization are not complete and could potentially benefit from the application of the benchmark dose (BMD) approach. Table I shows the combinations between the three main topics studied and the examined substances.

Table I. The relation between the three main investigated topics (CES, multiple endpoints and chemical mixtures) and the substances examined (PFAS, PCBs and a candidate drug in pharmaceutical development).

	CES	**Multiple endpoints**	**Chemical mixtures**
PFAS	Study I and VI	Study VI	Study I and VI
PCBs	Study IV and V	Study IV and V	Study V
Pharmaceutical development	Study III	Study II and III	Not applicable

The selection of a CES, i.e. the percentage of change at which the effect is considered to be adverse and non-random/due to noise, is an essential step when employing the BMD approach for chemical risk assessment. The CES chosen needs to reflect a relevant change, from a toxicological point of view. For example, 10% might be a significant change for hemoglobin levels, but may be irrelevant when examining liver enzymatic changes. For risk assessors and BMD users, the choice of CES is still a matter of debate – the strategy chosen differs between countries, organizations and individuals. In **study I** and **III** to **V**, the reasoning behind the choice of a particular level was motivated and discussed. **Study VI** took it one step farther and performed a retrospective analysis of five case-studies, investigating how five of the most used strategies might lead to similar results – or disagree with one another. The different strategies and consequent results were discussed, demonstrating how important it is to choose the CES wisely.

Secondly, the assessment of multiple endpoints may be challenging from the mathematical and risk assessor's point-of-view, but might provide a more complete picture and, therefore, a better understanding of the ongoing changes. The common practice of choosing a single, critical

endpoint for *in vivo* study assessment was challenged in **study II** to **V**, where the BMD and partial least squares (PLS) methods were employed to analyze multiple endpoints. **Study II** used the PLS method to investigate if information from several endpoints can be combined to predict pathological findings in rat organs. **Study III** employed the BMD approach to model 63 endpoints, in order to detect possible associations and dose-dependent changes in the investigated endpoints. **Study IV** and **V** investigated if the retinoid system endpoints could be used to describe potentially toxic effects at low doses, compared to other endpoints used earlier, and compared the potency of different PCBs.

Lastly, the assessment of the exposure outcome to chemical mixtures is still challenging. Historically, chemical exposures have been studied on an individual basis – "chemical X exposure is associated with effect Y", in simplified terms. The outcomes studied are usually a result of one predominant chemical exposure. However, realistic exposure scenarios are far more complex, as we are permanently exposed to different chemicals at different concentrations and timepoints, through different exposure routes such as food, water or air. A realistic assessment of a mixture should discern the contribution of individual chemicals to the effect(s) in the organs studied. Thankfully, the BMD-modelling approach can support risk assessors by addressing these challenges. In this thesis, the issues related to chemical mixtures were addressed in **study I**, **V** and **VI**. Of particular relevance is the estimation of relative potency for structurally similar chemicals, which was the focus of **study V** (see also the paper by Norgren and colleagues, 2022).

In summary, this thesis focuses on applying the BMD and PLS methods for dose-response assessment and hazard characterization, addressing the CES, multiple endpoints and chemical mixtures assessment, rather than on the specific chemicals assessed. Both modelling strategies have a great 3R (Replacement, Reduction and Refinement) potential, enabling the refinement and eventual reduction of animal use. The following section frames the literature background to the research questions being investigated.

2 BACKGROUND

"*What is there that is not a poison?*" was the rhetorical question posed by Paracelsus in 1538, widely considered the father of Toxicology, which summarizes the quest of the field. Toxicology is the scientific discipline that studies, methodically, the possible adverse effects of chemical exposure (including diverse agents such as radiation and nanomaterials) on living organisms and ecosystems (Wexler and Hayes, 2018). In June 2015, by the 50th anniversary of the Chemical Abstracts Service (CAS, a division of the American Chemical Society), the world's 100 millionth molecule was registered. This represents ten times as many molecules as there were in the database 25 years before, with an impressive average of a new molecule being registered every 7 minutes. As of April 2022, the CAS register featured 194 million substances, and the registration rate increased to a newly registered molecule every 11 minutes. A recent study estimated that over 350 000 chemicals and mixtures are produced and used globally (Wang et al., 2020). These chemicals can, in turn be combined in an unfathomably large number of combinations. These chemicals and mixtures constitute a considerable toxicological challenge to assess, in terms of potential hazard or safety, not only due to the large numbers but also because of the lack of information about many of the substances. Adding to the inherent complexity posed by the exposure to chemical mixtures, different organisms, individuals and populations will vary in sensitivity and manifest effects differently. Therefore, the Toxicology field faces a considerable challenge: understanding the associations and causality between chemical exposure and its effects.

Humankind has resorted to modelling for millennia in order to understand the surrounding environment. The purpose of all experimental, mechanistic and mathematical models is the same – to simplify a complex real-world system into understandable terms. Models, particularly mathematical models, are and have been of fundamental importance to describe past and current phenomena. But they are also used to make predictions, for decision-making and communication purposes (Schichl, 2004). Data *per se* are not valuable if they cannot be translated into information, and information is only useful if processed and used appropriately. Therefore, it is unsurprising that mathematical models achieved an important position in different fields including politics, economics, medicine, toxicology, and physics. Currently, few decisions in these fields are taken without performing an *a priori* mathematical analysis of the available data.

Paracelsus postulated in 1538 that, "*All things are poison and nothing (is) without poison. Solely the dose determines that a thing is not a poison*", suggesting that chemicals should not lead to adverse effects under a certain exposure threshold. Substances deemed harmless, such as water, are safe only under certain exposure limits (15 liters per day can be deadly)(Radojevic et al., 2012). Determining potentially safe exposure limits is challenging and constitutes one of the fundaments

of risk assessment, the area within Toxicology that aims to study and assess the risk as a basis for decision-making. The exposure limits are then enforced to protect or prevent an effect on the population. However, considerable efforts are made to define what an acceptable effect level or exposure is. Additionally, the same exposure level may represent different risks, depending on the subpopulation group, age, sex, and underlying disease factors, for example.

The concern for adverse effects on human and environmental health following chemical exposure has driven mathematical modelling in Toxicology. Currently, different mathematical modelling approaches are employed by toxicologists in order to address the various challenges faced, and to improve their understanding of biochemical reactions, kinetics, epidemiology, and *in vitro* and *in vivo* dose-response associations. Thus, the observations made by Paracelsus five centuries ago are still relevant in the context of risk assessment today. Lastly, in order to better understand what risk assessment is, it is necessary to understand the concepts of risk and how it is managed.

2.1 RISK, RISK ASSESSMENT AND RISK MANAGEMENT

Risk, by definition, is the probability of an adverse outcome. Hazard is the source or situation from which the danger stems. For example, a great white shark is a hazard, which could be dangerous or harmless – the risk only occurs if one swims in waters populated by these animals. Risk assessment, in this case, is the evaluation of the probability of being attacked by the shark given the location, duration and depth of swimming.

In Toxicology, risk assessment is the area that combines hazard identification, dose-response assessment/hazard characterization and exposure assessment in order to characterize the risk following chemical exposure. Dose-response assessment, also denominated hazard characterization by the World Health Organization (WHO), is a crucial step in risk characterization that refers to the modelling approaches employed to study effects associated with exposure and constitutes a cornerstone of the understanding of chemical risks (WHO IPCS, 2020). Risk assessment is often employed within regulatory toxicology, and consists of the evaluation of toxicological data by researchers, the industry and authorities who regulate the use and exposure of chemicals in society (Faustman, 2018; Schwenk et al., 2002). The ultimate goal of risk assessment is to characterize the potential chemical hazard, by using tools that describe as accurately as possible, the risk inherent to chemical exposure (Aven, 2016). Risk assessment may also be protective or predictive: the former aims to prevent while the latter aims to predict adverse effects on humans and the environment (Rudén et al., 2019). In order to assess the potential outcome of chemical exposure and their impact on humans, the risk assessment process

may be divided into four steps, following an initial problem formulation (NRC, 1983; Renwick et al., 2003; WHO ICPS, 2009) (Figure 1 and Table II):

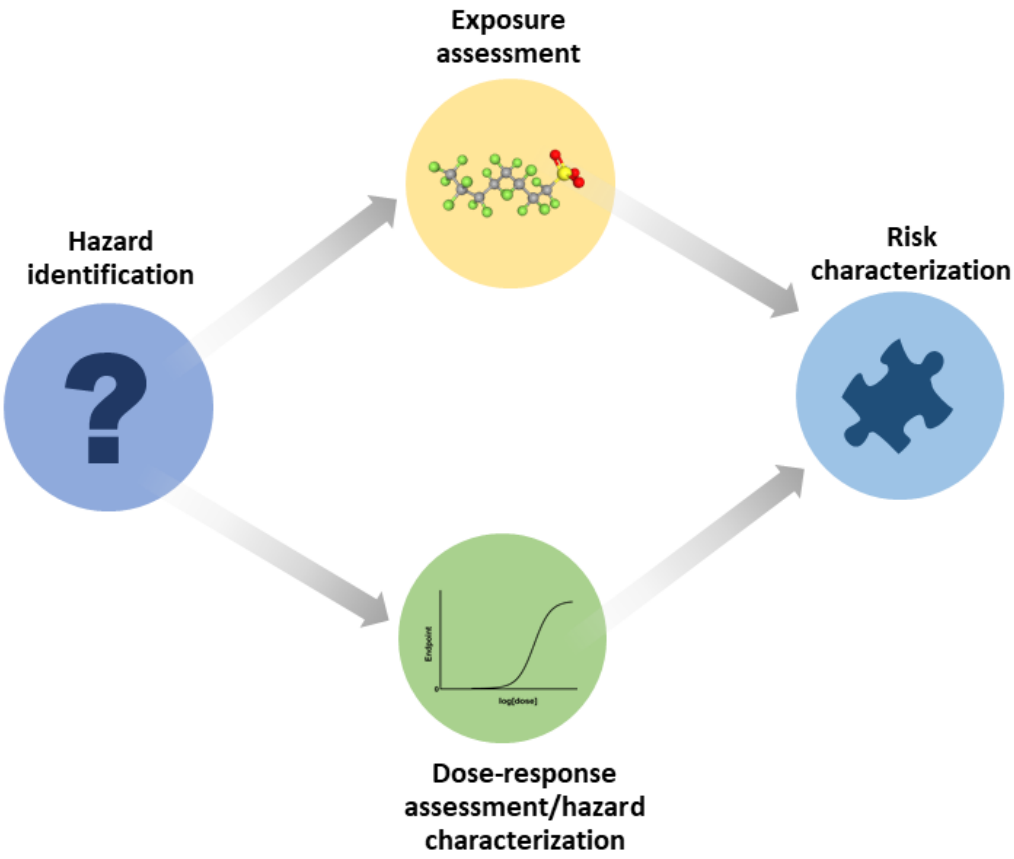

Figure 1. The risk assessment process steps: hazard identification, dose-response assessment/hazard characterization, exposure assessment and risk characterization. Inspired by the WHO IPCS *Environmental Health Criteria 239. Principles for modelling dose-response for the risk assessment of chemicals* (2009).

Table II. The risk assessment paradigm and the related problem formulation. Adapted from the WHO IPCS *Environmental Health Criteria 239. Principles for modelling dose-response for the risk assessment of chemicals* (2009) and the work by Renwick and colleagues (2003).

Step	Description	Content
Problem formulation	Frame the scope and objective of the assessment	Identification of the knowledge gap Compile and analyze prior knowledge Desired outcomes
Hazard identification	Identifies the type and nature of the adverse health effects, regardless of the dose/exposure	Human studies *in vivo* toxicology studies *in vitro* toxicology studies Structure-activity studies
Dose-response assessment/hazard characterization	Qualitative or quantitative description of inherent properties of an agent having the potential to cause adverse health effects	Selection of the critical dataset Modes/mechanisms of actionDose-response for critical effect Kinetic variability Dynamic variability
Exposure assessment	Evaluation of the exposure levels (concentration or amountof a particular agent) that a population or ecosystem is exposed to, across different sources of exposure (water, food, soil, air)	Magnitude Frequency Duration Route Extent
Risk characterization	Combination of the previous steps to describe, as accurately as possible, the risk of potentially adverse effects in the population Advice for decision-making	Probability of occurrence Severity Given population Attendant uncertainties

Problem formulation identifies which compounds need to be considered in the risk assessment. It is important to identify the properties or effects of concern and frame the knowledge gaps clearly. In **study I**, for example, the exposure duration was a factor of concern – did longer-term residents display a higher risk of adverse effects compared to those who lived in Ronneby for one year? In **study V**, the differences in the potential for disruptive effects in the retinoid system following exposure to different PCBs were investigated.

Risk characterization provides the grounds for decision-making for risk management and risk communication (Renwick et al., 2003). However, it depends heavily on the availability of data. Consequently, it is important to employ appropriate methods to extract as much information as possible from the available data. Therefore, sensitive and reliable methods for assessing a chemical dose-response association or dependency, such as the BMD-modelling and the integrated probabilistic risk assessment (IPRA) approach, are progressively gaining more importance in the risk assessment area (see section **2.4 The benchmark dose approach** and **2.4.3 Integrated Probabilistic Risk Assessment (IPRA)**). Additionally, it is more important than ever to present data in a meaningful manner so that risk assessors and researchers can model it independently from the original study authors; this means sharing summarized data in terms of arithmetical means, standard deviations (SDs) and *n* number of units, for example. This will improve the public availability of data and promote the transparency of the risk assessors' analysis and conclusions.

Risk management relates to actions taken to control the hazard(s) resulting from the risk assessment process. Risk managers often examine and debate the methodology underpinning the risk assessment report before discussing and motivating the decisions which will inform the subsequent action taken – action which might include implementing exposure limits or remediation measures. For example, lower exposure limits might be set for the general population in order to protect sensitive sub-populations such as asthmatics and children (Johansson, 2016; Larsson, 2018). Estimates of the risk, such as the margin of safety, are taken into account to estimate the resources and extension of remediation actions required. Risk managers are often decision-makers who evaluate the different possibilities and decide when, where and how to take action (Figure 2).

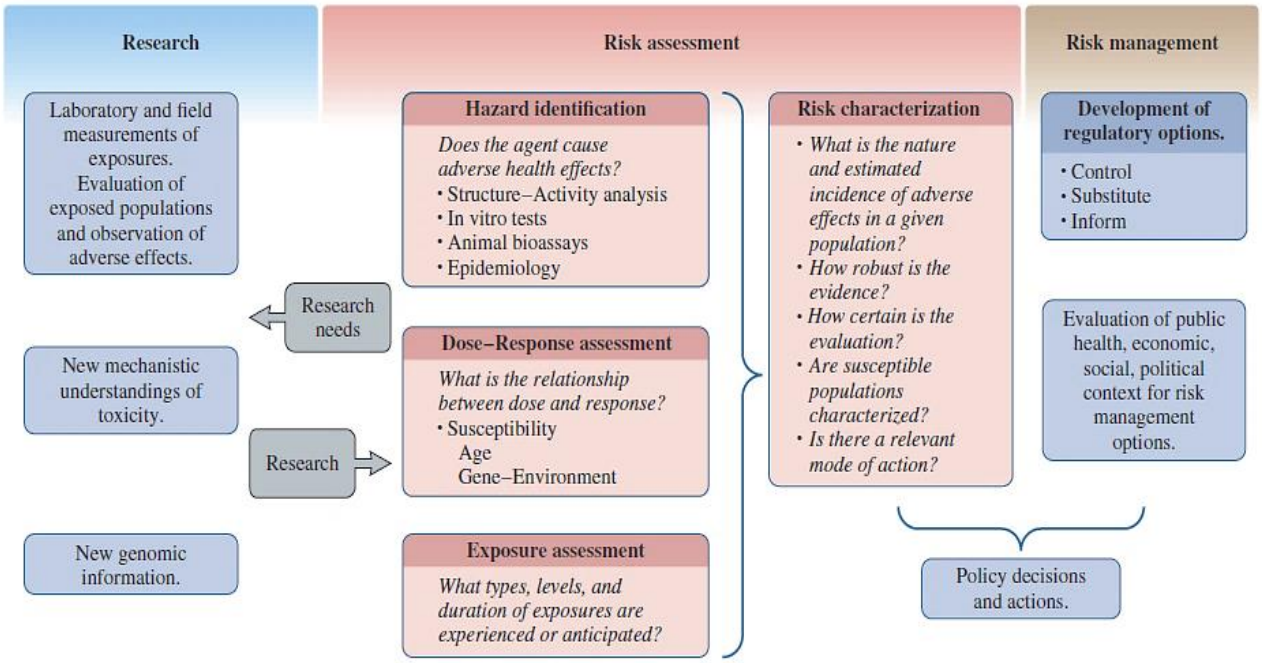

Figure 2. Research, risk assessment and risk management framework. Reproduced with permission from McGraw-Hill Education from Chapter 4. Risk Assessment, in Casarett & Doull's *Toxicology – The Basic Science of Poisons* (9th edition, 2018).

Risk managers face challenges that stem from the limitations in the risk assessment process, such as the decision-making based on no-observed-adverse-effect-level (NOAEL) or lowest-observed-adverse-effect-level (LOAEL) values, or inappropriate choice of CES when performing BMD-modelling. These may have an impact on the resulting limit values and, therefore, the whole risk assessment outcome. Therefore, it is of fundamental importance to employ the best tools in the toolbox, i.e. to analyze good quality data with appropriate methods in order to deliver the best information and estimates, to support the risk management decision – for example setting exposure limits or initiating remediation actions.

Risk communication relates to the interaction between individuals (or groups of individuals, such as the public at large) and the stakeholders (such as risk managers/assessors and authorities). It includes the flow of information to and from the stakeholders, how to articulate the risk being examined and how systemic communication leads to a chemically safer society. Expressed in terms of the shark metaphor, risk communication would be informing the hypothetical swimmer about the hazard in the water, through, for example, signs or billboards. Additional measures would include educating the population about what a shark bite entails (or exposure to a particular chemical), to prevent future harm. Risk management, in this case, could include providing a flotation device to rescue the swimmer, e.g. a boat (safest remediation option but more resource-demanding) or by a second swimmer going out to rescue the first swimmer (riskier and least safe option). Additional risk management actions could include removing the shark(s) from the water, a costly remediation measure that would affect the ecosystem, or tracking their location to inform safe bathing zones.

Whereas risk assessment and risk management are performed at a collective level, risk perception is the understanding of the probability of an effect or event occurring at an individual or subpopulation level. Risk perception may be divided into controllable/uncontrollable and observable/not observable risks, which represent different degrees of hazard (Figure 3). However, risk perception varies considerably, depending on socio-cultural and geographical factors, and it also varies depending on sex and personal disposition. For example, the residents of Siberia will comprehend the hazard that a shark bite entails on an abstract level, even if they have never swum in shark-infested waters. Usually, action is taken either preventively, to avoid the risk, or reactively, to contain it.

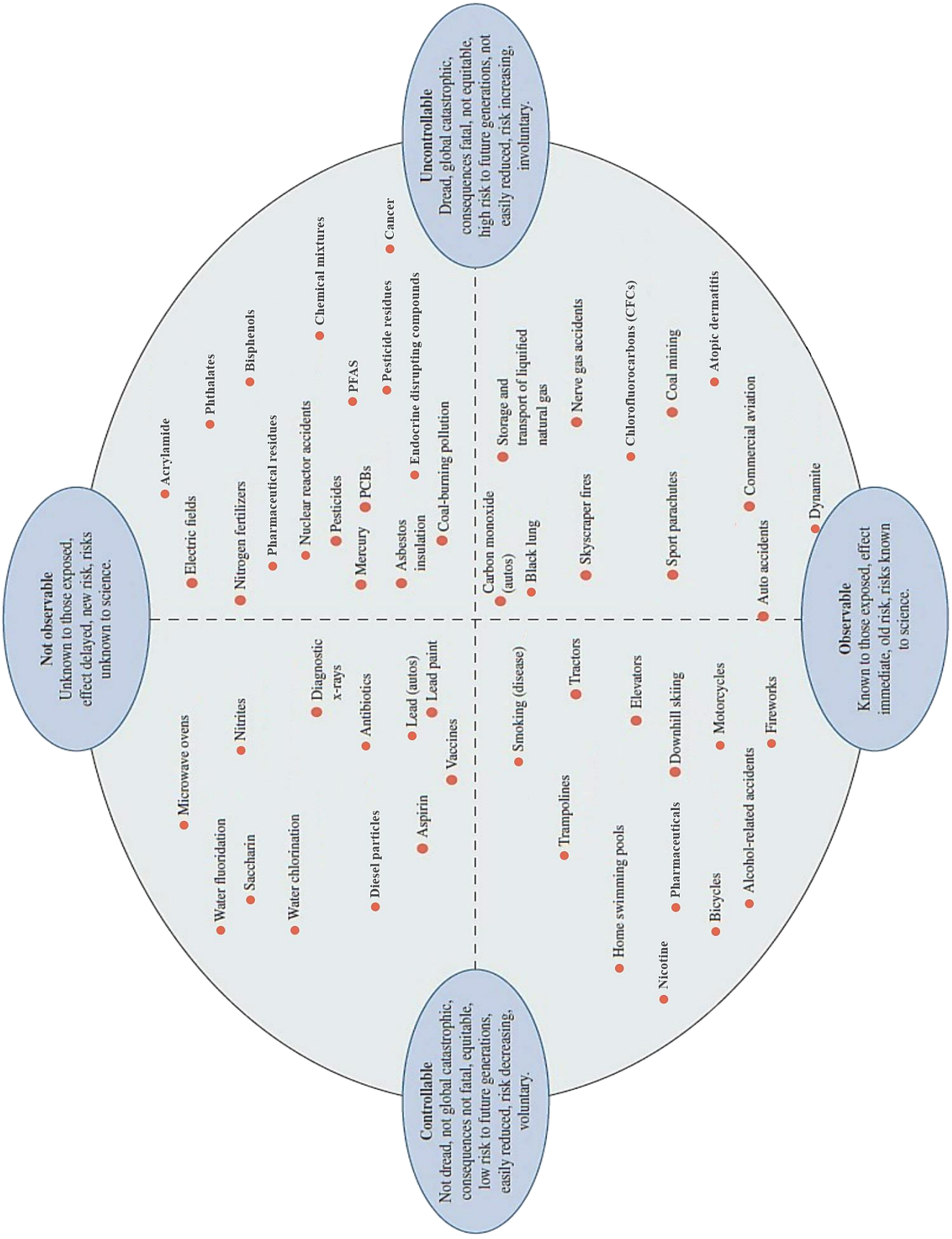

Figure 3. The distribution of risks of different agents and situations, categorized as Controllable/Uncontrollable and Observable/Not observable. Each risk requires a specific assessment in order to take relevant risk management and risk communication measures. Adapted from and reproduced with permission from McGraw Hill Education *Casarett & Doull's Toxicology – The Basic Science of Poisons*, Chapter 4. Risk Assessment (9th edition, 2018).

2.2 DOSE-RESPONSE ASSESSMENT

A key step in the risk assessment process is the dose-response assessment, which aims to describe and quantify the relationship between the chemical exposure and the associated effect (WHO IPCS, 2009). Historically, a panoply of approaches has been used, among them the half-maximal effective concentration (EC_{50}), the NOAEL and the BMD method. Most biological changes progress gradually with increasing exposure for many endpoints, which is the contrary to changes in a threshold-like manner (Slikker et al., 2004; Slob, 1999). For example, in the particular case of genotoxic carcinogenic substances, it is assumed that there is no threshold effect, i.e. a single molecule may cause an effect, although at such a low probability that it is impossible to quantify (EFSA, 2005; Slob, 1999).

The NOAEL and the BMD stand among the most commonly used approaches for dose-response assessment. The NOAEL approach estimates the highest tested dose/concentration at which an adverse effect is not observed (WHO IPCS, 2020). The BMD-modelling approach, on the other hand, calculates the lowest dose/concentration (BMDL) or the best estimate (BMD) at which an effect is expected to occur. Although their definition differs, both these approaches estimate a plausible threshold level, the so-called point-of-departure (PoD), which is the lowest dose or concentration that deviates from the normal levels of the assessed endpoint. The PoD is then used to set reference values and exposure limits for the population of interest (WHO IPCS, 2020). However, the NOAEL is a surrogate measure to the PoD, while the BMDL is a better estimate, incorporating the uncertainty in the data (Jensen et al., 2021a). The PoD is used to calculate exposure limits and reference values, such as the Acceptable or Tolerable Daily Intakes (ADI/TDIs), by the application of assessment/uncertainty factor(s) (WHO IPCS, 2020). The ADIs/TDIs are levels at which exposure may occur without adverse effect(s) expected to be observed. Moreover, the NOAEL and the BMD approaches differ methodologically in terms of underlying reasoning and might lead to different estimates. Therefore, consideration is required in determining the choice of method and desired robustness of the analysis and subsequent results, since each approach may lead to different conclusions and interpretations.

Currently, the NOAEL is the most commonly used approach among toxicologists in the risk assessment and pharmaceutical development areas. It is used by authorities and expert groups with a global impact such as the European Chemicals Agency (ECHA), European Food Safety Authority (EFSA), US Environmental Protection Agency (US EPA), the Joint FAO-WHO Expert Committee on Food Additives (JECFA) (jointly administered by the Food and Agriculture Organization of the United Nations, FAO, and WHO), the FAO-WHO Joint Meeting on Pesticide Residues (JMPR) and the Organisation for Economic Co-operation and Development (OECD). Although the tide is changing, its simplicity and familiarity have slowed the paradigm shift towards the BMD-modelling approach.

2.3 THE NO-OBSERVED-ADVERSE-EFFECT-LEVEL (NOAEL) APPROACH

The NOAEL is a historically important approach which determines the highest dose for which the response is not statistically significantly different from the control group (WHO IPCS, 2020). It typically employs a statistical test to compare groups, such as the t-test, Holm-Šídák or Dunnett's test, to compare the means of the different groups against the control, although other tests have also been used (Hothorn, 2021). The LOAEL is the lowest dose that is statistically significantly different from the control group, at which the adverse effect is expected to occur (WHO IPCS, 2020). The NOAEL approach aims to detect the highest dose tested that did not lead to adverse effects. However, the NOAEL is not a risk-free dose or concentration – it might represent the level for which adverse effects were not observed but, in some cases, might correspond to the level over which adverse effects may occur (WHO IPCS, 2020).

The NOAEL approach has limitations, such as heavy dependence on sample size, doses tested and data quality (EFSA, 2017; Haber et al., 2018; OECD, 2014; Slob, 2014a, 2014b; US EPA, 2012). Moreover, it lacks uncertainty quantification around the PoD and characterization of the true effect size at the NOAEL (EFSA, 2017). Henry Louis Mencken's (1920) centenary statement that "*There is always a well-known solution to every human problem – neat, plausible, and wrong."* would be a very fitting description for the NOAEL approach. Not all NOAELs are wrong, but some are more wrong than others. Many biological changes occur progressively, not in a dose or concentration threshold-dependent manner. The NOAEL is a simple and straightforward approach which delivers plausible results. Its statistical pillars confer credibility, but the underlying tests are dependent on assumptions and considerations which do not necessarily take into account the nature of the biological endpoint being assessed. The conclusions reached by this approach are heavily dependent on the sample size, not promoting the 3Rs, with the inherent ethical consequences (EFSA, 2017; Jensen et al., 2019). For example, low sample size and study power can lead to falsely high NOAELs and, consequently, less protective exposure levels (EFSA, 2017; Ringblom et al., 2014). As the sample size and tested doses are often the limiting factors in *in vitro* or *in vivo* assessment studies, the detection of small events is dependent on the statistical test employed.

Additionally, it has been suggested that the derived *p*-value is not informative of biological relevance but only of differences between the groups assessed (Parens et al., 2017; Wasserstein and Lazar, 2016; Wasserstein et al., 2019). Edgeworth's pioneering work (1885) stated that statistical significance, in the form of a significant *p*-value, signals that there might be a difference between the groups observed and that the obtained results need further scrutiny. The American Statistical Association has discussed and demonstrated how unconstructive it is to dichotomize

reality into significant/non-significant and how, ultimately, this leads to poor decision-making (Wasserstein and Lazar, 2016; Wasserstein et al., 2019). The *p*-value is an assessment of the role of chance in the observed result, and a $p = 0.05$ does not investigate causality but the possibility that 1 out of each 20 observations could be produced by chance (Wasserstein and Lazar, 2016; Wasserstein et al., 2019). Therefore, a significant *p*-value does not imply that a biological effect or change is taking place; it merely suggests a plausible difference that needs to be further scrutinized. Within Toxicology, the relevance of this discussion was elegantly summarized by Parens and colleagues: "*The use of statistical significance for toxicity assumes that proof of toxicity is required, not of safety. Absence of proof of toxicity is surely not proof of the absence of toxicity. Given this framing, it should not be surprising that many cases where toxicological tests do not find toxicity result in the approval of substances that are later found to be toxic*." (Parens et al., 2017).

In conclusion, "*The NOAEL is a poor version of BMDL*", as Dr. Wout Slob stated during the EFSA workshop on the Guidance on the use of the benchmark dose approach (in Brussels, March 2017). Although the NOAEL might lead, on average, to similar values compared to BMDLs, it is more easily influenced by outliers, study design and uncertainty, leading to risk underestimations (Bokkers and Slob, 2005; EFSA, 2017; Faustman and Bartell, 1997; WHO IPCS, 2020). Therefore, a more scientifically sound and mathematically robust method has been slowly but steadily replacing the NOAEL – the BMD-modelling approach.

2.4 THE BENCHMARK DOSE APPROACH

The BMD-modelling approach was first introduced by Crump (1984) as an alternative to address the inherent limitations of the NOAEL method (Kimmel and Gaylor, 1988; Slob, 2014a, 2014b). This dose-response modelling is characterized by three fundamental differences, compared to the NOAEL approach. Firstly, a curve is fitted to the data, contrary to the groupwise testing performed by the NOAEL approach. Secondly, the BMD-modelling approach calculates the benchmark dose, which is the data-derived best mathematical estimate of the dose/concentration that leads to the pre-specified change in the response/endpoint analyzed (WHO IPCS, 2020). Thirdly, the BMD approach takes all uncertainty and variability in the data into account, calculating a confidence interval around the median (the BMD). The lower and upper limits of the 90% confidence interval for the estimated BMD are denominated as BMDL and BMDU, respectively. The resulting BMDL, or the BMD, is then used as a PoD to support decision-makers to determine reference values and exposure limits, such as the ADIs and TDIs (see section **2.2 Dose-response assessment**) (Figure 4).

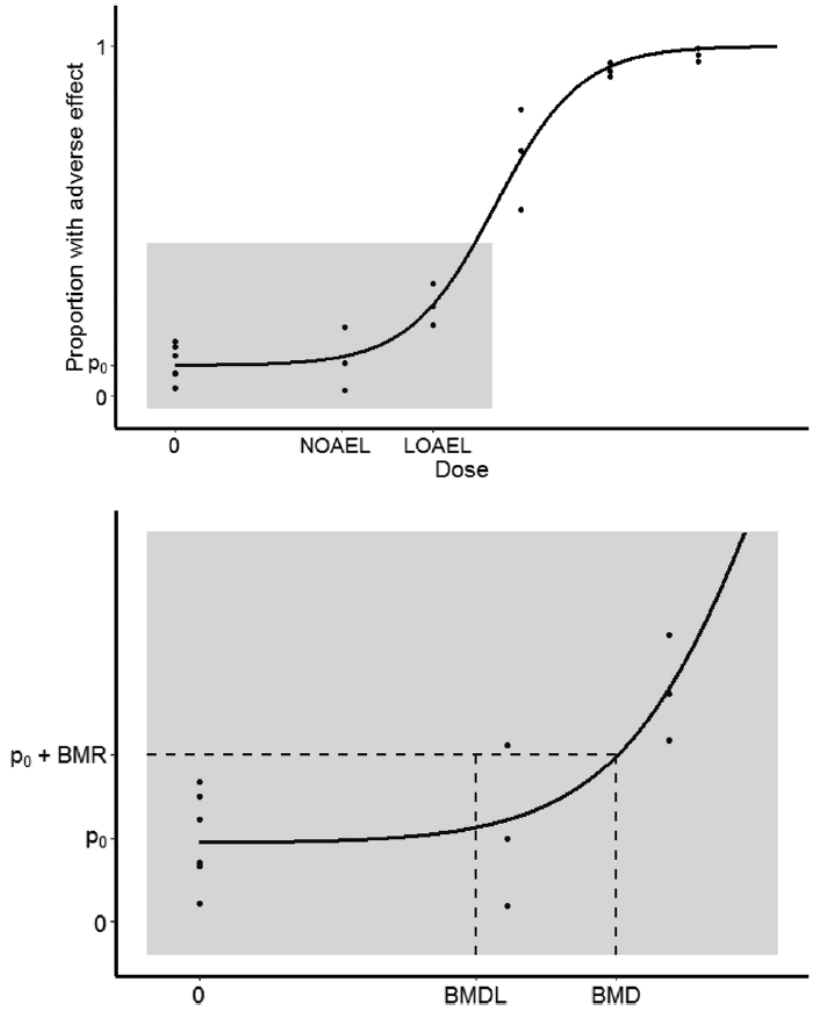

Figure 4. Benchmark dose-modelling plot, showing the resulting BMD and BMDL in comparison to the study NOAEL and LOAEL. The grey area is zoomed in the bottom plot. Reproduced with permission from the Risk Analysis journal, from the original article by Jensen and colleagues (2019).

Owing to its ability to interpolate between dose-groups, extract information from the data and compare different studies, the BMD-modelling approach has established itself as a robust tool for dose-response assessment of various classes of substances. Among others, it has been used for the risk assessment of toxic chemicals such as PFOA (perfluorooctanoic acid), TCDD (2,3,7,8-tetrachlorodibenzo-*p*-dioxin), carbon nanotubes, asbestos and glyphosate (Korchevskiy, 2021; Van den Berg et al., 2006; Zeilmaker et al., 2018).

In short, there are four main steps in the BMD approach:

1) Selection of a benchmark response (BMR), also known as Critical Effect Size (CES) for continuous data (see section **2.4.1 The choice of Critical Effect Size (CES)**);
2) Selection of a suite of models to fit the data;
3) Model fitting, comparison and selection;
4) Calculation of the BMD and the respective 90% confidence interval lower and upper limits, BMDL and BMDU.

As of April 2022, guidance was only available for the two main types of data modelled by this approach: continuous and quantal data (EFSA, 2017; US EPA, 2012; WHO IPCS, 2020). Continuous data is by far the most common type of data and it is the focus of this thesis. This type of data consists of non-negative values that are continuous and can, therefore, be divided into infinitely smaller parts. Hemoglobin concentrations, enzymatic activity or concentration and body weight are common examples of continuous variables. Many non-negative continuous variables follow a normal or a log-normal distribution, i.e. tend to have a symmetric or right (positive) skewed distribution. Additionally, it is plausible that most biological systems level off over a certain dose, through phenomena such as system saturation or excess of substrate/substance, for example. These factors need to be considered when choosing the CES, which is discussed in section **2.4.1 The choice of Critical Effect Size (CES)**.

The models fitted to continuous data are most often non-linear models, most notably the exponential (Crump, 1995; Slob, 2002) and the Hill family of models (Barton et al., 1998; Murrell et al., 1998), although many other models are available. EFSA included the inverse exponential and log-normal models in its BMD platform, available at https://r4eu.efsa.europa.eu/app/bmd (Varewyck and Verbeke, 2018), and four other models (Gamma, linearized two-stage, probit and logistic) to its draft guidance on the use of the BMD approach (EFSA, 2022). All of these four families of models are non-monotonic functions, i.e. they do not feature a bell shape but rather a sigmoidal curve, which either increases or decreases over incrementing x-axis values, featuring (or not) a ceiling/plateau level.

Model fitting, comparison and selection are important steps where disagreement persists across different guidance documents (EFSA, 2017; US EPA, 2012; WHO IPCS, 2020). Both the US EPA and EFSA recommend using software (BMDS and PROAST, respectively) that will fit different parametric frequentist models aiming to estimate the best values for the *a priori* unknown parameters *a*, *b*, *c* and *d* (EFSA, 2017; US EPA, 2012). Differences exist, however, in terms of features, available models and user interfaces (Haber et al., 2018; Jensen et al., 2019). For the model and parameter optimization procedure, the preferred method is the maximized log-likelihood method (EFSA, 2017; US EPA, 2012). This method chooses the model and respective parameters that best fit the data (EFSA, 2017; US EPA, 2012). However, a model should not have more parameters than is strictly necessary to accurately describe the data, but underparameterization (too simple models) might ultimately lead to underestimation of the risk (Ringblom et al., 2014). Although good quality data are always desirable, in cases of scarcity or poor data quality, the estimated BMDLs might be lower due to the greater uncertainty, which leads to broader confidence intervals. Lower BMDLs are conservative from a regulatory perspective, as they reduce the possibility that the established reference value is not protective; however, too low reference values might be difficult to comply with (Gelman and Greenland, 2019).

An important characteristic of the BMD approach is the ability to separate the information contained in the data from the noise, i.e. distinguish between systematic (the effect, associated with the administered dose) and other sources such as biological, system or process variation (e.g. the experimental day-to-day, between and within plate variation) (Burnham and Anderson, 2004; Ritz et al., 2013). Therefore, a criterion to compare the different models was introduced, most prominently the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which perform similarly (Burnham and Anderson, 2004). The AIC stems from the information theory field and compares the Kullback-Leibler information loss by replacing the data with a model (Akaike, 1973). It is an extension of the maximum likelihood principle and it postulates that a model should be the simplest possible to describe the data accurately, avoiding, however, the loss of too much information (Akaike, 1973). The model fit comparison using the AIC is expressed by the equation $AIC = -2\ log(L) + 2p$, with $log(L)$ constituting the log-likelihood of the model and p the number of parameters. It has been demonstrated that two models with an AIC difference under two units describe the data equally well (Burnham and Anderson, 2004). The general consensus in the field and the BMD-modelling guidelines recommends employing the AIC for model comparison and selection, as well as for model weighing in the model averaging (MA) step (EFSA, 2017, 2022; Peña et al., 2017; US EPA, 2018, 2020; WHO IPCS, 2020).

There are three main methods to calculate a confidence interval for non-linear models: the likelihood-ratio method, the bootstrap method and the delta method (Moerbeek et al., 2004). Method performance was reviewed by Moerbeek and colleagues (2004), concluding that the delta method was not appropriate for non-linear models and that the two other approaches performed similarly. The likelihood-ratio method is commonly employed for model fitting without model averaging (EFSA, 2017; EPA, 2012), but the bootstrap method, although more time-consuming, features the possibility to link to Monte Carlo simulations that depart from the original dataset (Moerbeek et al., 2004; Shao and Gift, 2014; Voet and Slob, 2007; **study I**). Varewyck and Verbeke have shown that convergence was achieved and that the results after 1000 bootstraps are comparable to those after 200 bootstraps, when estimating a 90% confidence interval (Varewyck and Verbeke, 2018). In conclusion, the bootstrap method is the preferred approach for MA, providing more confidence in the interval of possible "true" BMDs (EFSA, 2017; Voet and Slob, 2007).

2.4.1 THE CHOICE OF CRITICAL EFFECT SIZE (CES)

The Critical Effect Size (CES), a synonym to benchmark response (BMR) used when continuous data type is analyzed, is the pre-specified change in response (the dependent) variable under assessment (EFSA, 2017). It results in a BMD, the best mathematical estimate of the independent variable, usually a dose or concentration, that leads to the change specified by the BMR. The BMR is expressed differently depending on the type of data (continuous or quantal data) and the background level considered. In the hybrid approach, the BMR regards a change of response over the estimated background, and it is the recommended approach by EFSA (EFSA, 2017; Jensen et al., 2019). EFSA suggests starting with a default change of 5% (CES_5) in response for continuous data, and a 10% extra risk for quantal data, relative to the background estimate (EFSA, 2017). On the other hand, the US EPA recommends using a 1 SD change compared to the control group mean for continuous data, having the same approach as EFSA for quantal data. Different motivations on how to estimate a biologically significant change in response that is considered adverse explain these differences (EFSA, 2017; US EPA, 2012). Different results might be obtained when biological considerations are taken into account, e.g. a 5% or 20% decrease in alkaline phosphatase enzymatic activity (Figure 5). In this particular case, a 5% CES was inadequate – the change was too small to be considered a reasonable threshold for adversity, being within the variation observed in the data (Figure 5). A 20% change, on the other hand, could perhaps stand as a change in response associated with a change in enzymatic activity considered adverse (Figure 5).

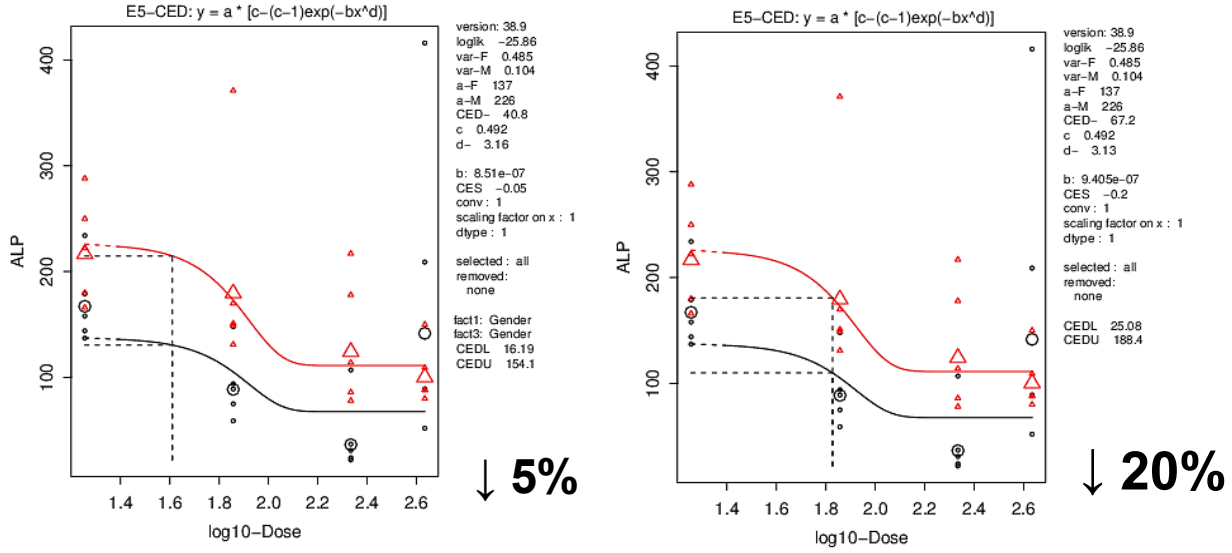

Figure 5. The benchmark dose-response modelling approach applied to alkaline phosphatase (ALP) enzymatic changes in rat. The same data was fitted with a CES of 5% (left) and 20% (right plot).

A BMR_5 for continuous data and a BMR_{10} for quantal data often lead to results comparable to the NOAEL (Allen et al., 1994; Bokkers and Slob, 2007; EFSA, 2017; Fowles et al., 1999; Sand et al., 2011). However, there is a broad consensus in the field that the BMD approach provides a more mathematically sound estimate of the critical dose leading to the effect, where uncertainty is expressed through the 90% confidence interval, contrary to the NOAEL, where the uncertainty remains unknown (EFSA, 2017; OECD, 2014; US EPA, 2012). Other strategies have been proposed for the choice of CES, most notably the point of transition in an S-shaped dose-response curve (Sand et al., 2012), intra-animal variation based on historical data (Buist et al., 2009; Dekkers et al., 2006) and a data-derived BMR approach called General Theory of Effect Size (GTES) (Slob, 2017). The BMR can also be adjusted based on expert judgment, depending on the underlying data or biological considerations, as it ultimately aims to establish the dose that leads to a biologically significant and potentially adverse change in the outcome. There is, however, room for improvement, for example in terms of choice of CES and whole study analysis with different CESs.

2.4.2 MODEL AVERAGING

Model averaging has been introduced into the benchmark dose-modelling approach to make calculations more sophisticated and address issues related to model selection and BMDL and BMDU estimation, by weighing the different models fitted to the data (Jensen et al., 2019; Wheeler and Bailer, 2008, 2009). It is currently recommended by EFSA (2017, 2022) and WHO (2020), and is under review by the US EPA (2018, 2020). Model averaging gained international relevance owing to the possibility to combine different models and reduce uncertainty (Hoeting et al., 1999; Shao and Gift, 2014; Shao and Small, 2011, 2012; Wheeler and Bailer, 2007, 2008). Model averaging has been shown to perform better than estimates based on a single parametric model (Wheeler and Bailer, 2013). In short, it combines the best-fitted models by weighing the contribution of each model based on their AIC score (Jensen et al., 2019). For example, the standard weight equals (1 / *n* number of models fitted to the data), i.e. 0.25, if four models are fitted to the data. In the indirect method, the model weight will then be adjusted based on the AIC score of the fit, scoring lower the models with a bad fit and thereby preventing these from having a significant contribution to the final model-averaged BMD and confidence interval (Aerts et al., 2020; EFSA, 2017).

When at least one model reveals a dose-dependent trend, the BMD and respective BMDL and BMDU are calculated. Model averaging-based BMDLs and BMDUs are recommended as they are more mathematically robust than simple averages or the lowest BMDL obtained (EFSA,

2017). When calculating a model-averaged point estimate, one can employ the direct or the indirect method (Aerts et al., 2020). The direct method estimates the final BMDL and BMDU based on model weighs and the BMDLs and BMDUs of each model, while the indirect method estimates an average model first and, based on it, the BMDL and BMDU are calculated (Aerts et al., 2020; EFSA, 2022). Both EFSA and US EPA have used the indirect method in their frequentist model-averaged BMD software (PROAST and BMDS) (Davis et al., 2011; EFSA, 2017; US EPA, 2020).

The features of model averaging can be employed for dose-response and exposure assessment, reflecting better data uncertainty and laying the foundation for a better-grounded risk characterization through an IPRA analysis (Voet and Slob, 2007).

2.4.3 INTEGRATED PROBABILISTIC RISK ASSESSMENT (IPRA)

In regulatory toxicology and risk assessment, decision-makers are often challenged by uncertainty and lack of knowledge, data or studies (Aven, 2016; Paparella et al., 2013). A tiered approach is often employed, where the risk assessors progress through the different layers of complexity depending on the degree of uncertainty, adversity degree and importance associated with the chemical exposure. Tier 1 refers to routine screening assessments, characterizing uncertainty qualitatively and applying default uncertainty factors between 3 and 10, for example, for intra- and interspecies extrapolations (WHO IPCS, 2018). If the exposure is likely to exceed levels of concern and there is a need to characterize and quantify the individual contributions of the sources of uncertainty to the predicted results, a Tier 2 analysis may be performed (WHO IPCS, 2018). The BMD and BMDLs can then be used in probabilistic methods such as the IPRA approach (WHO IPCS, 2018).

The IPRA approach is receiving increasing attention as it delivers better quality estimates than the simpler and traditionally used (lower tier) deterministic Margin of Exposure (MoE) approach (Voet and Slob, 2007). Integrative Quantitative Risk Assessment methods, such as the IPRA, combine dose-response with exposure assessment data, and characterize the sources of uncertainty and variability (Faustman, 2018). The novelty of IPRA lies in the iteration of the dose-response and exposure data in a probabilistic manner, generating a distribution of Individual Margins of Exposure (IMoE) (Voet and Slob, 2007). The IPRA approach has been used for risk assessment of different agents, as diverse as triclosan, nanoparticles, cadmium, mycotoxins, pesticides and acrylamide, among others (Bokkers et al., 2009; Jacobs et al., 2016; Prichystalova et al., 2017).

Additionally, the IPRA approach takes the different sources of uncertainty into account, e.g. exposure duration uncertainty and intraspecies variability, and quantifies their contribution to the final estimate (Voet and Slob, 2007; WHO IPCS, 2018). This uncertainty estimation constitutes a major advantage over traditional deterministic risk assessment strategies, like the MoE approach, which returns a single estimate where the impact of uncertainty remains unknown (Kimmel and Gaylor, 1988; Voet and Slob, 2007; WHO IPCS, 2018). Uncertainty quantification is important as it affects the quality of the estimates and provides additional information for interpretation of the results (Chatfield, 1995; Hoeting et al., 1999). The IPRA approach expresses the estimated risk through a confidence interval describing the proportion of the population at risk of having an effect due to the chemical exposure (Voet and Slob, 2007; WHO IPCS, 2018). Accordingly, BMD-modelling feeds into IPRA by providing the probabilistically estimated BMDs (or BMDLs) that will be used to calculate the individual benchmark doses (IBMD). The IBMDs, in their turn, are then used as a numerator to estimate the IMoE, defined as follows by Voet and Slob (2007):

$$IMoE = \frac{IBMD}{IEXP}$$

IEXP is defined as the individual exposure distribution and IBMD as the (probabilistically derived) individual benchmark dose distribution (Voet and Slob, 2007). From the IMoE distribution, expert judgment devises a cut-off MoE, usually MOE < 1, also known as Probability of Critical Exposure (PoCE) (Voet and Slob, 2007). In this case, the PoCE is defined as the fraction of the population for which the exposure is higher than the estimated dose leading to the adverse effect(s) (Voet and Slob, 2007). Consequently, exposure limits can be derived and reinforced to protect the population from adverse effects. For more detailed information, please refer to the World Health Organization International Programme on Chemical Safety (WHO IPCS) framework for the evaluation of uncertainty in hazard characterization (WHO IPCS, 2018).

2.4.4 MULTIPLE ENDPOINTS

Historically, most risk assessment analyses have been performed based on one critical endpoint per study. By definition, it is the endpoint considered most relevant, being sensitive to the chemical exposure and displaying potentially adverse effects for human health. It does not necessarily reflect all ongoing biological changes, but the most adverse or important effects to be avoided. It is based on the critical endpoint chosen during the hazard characterization step that the risk assessor will estimate the reference value, usually departing from a BMDL or a NOAEL value. Therefore, a critical endpoint should be sensitive to the assessed chemical, i.e. the adverse effects and changes should occur at low doses, to account and avoid effects on the remaining endpoints

at higher doses. As effects occur in a continuum and sequence of events, and not on a threshold manner, many biological changes can be detected at an early stage or at low doses (Sand et al., 2018). For example, an increased incidence of neoplasms in animals can be used as the critical endpoint for cancer risk assessment following exposure to PCBs (IARC, 2015).

Additionally, a paradigm shift occurred when computational power increased, and it became easier to compare effects across multiple endpoints. BMD-modelling can be particularly useful and speed up the analysis of whole study data through, for example, the use of the vector function *c* in the R-package PROAST. However, resistance to change is natural, and the BMD method has been criticized as the analysis of multiple endpoints may be time-consuming, and because "*different software give different answers*" (Travis et al., 2005). Nevertheless, the BMD approach is better suited to provide a holistic view of the sequence of events, demonstrating how changes can be caused by different chemical exposure levels (Sand et al., 2018). The analysis of multiple endpoints might support the choice of the critical endpoint(s), especially when one is not identified from previous experiments or the literature. Therefore, it is recommended to model all available data, in order to compare the doses at which effects occur, to obtain a better understanding of the consequences of chemical exposure – instead of relying on a single critical endpoint, which often fails to describe all ongoing biological changes, especially if it is involved in several processes simultaneously.

2.4.5 CHEMICAL MIXTURES

Many studies have shown that chemical mixtures may affect human, animal and environmental health (EC, 2012; Kienzler et al., 2016). Sometimes, mixtures being assessed are well described, but this might not always be the case for realistic scenarios, which typically comprise numerous chemicals in unknown proportions that could change over time. In short, we know little about the effects of individual chemicals when they outweigh all other exposures, but very little about the exposure outcome of chemical mixtures. This challenges how safety thresholds are established, how to define adverse effects and how information from different endpoints is combined. Ideally, a realistic assessment of a mixture should discern the contribution of individual chemicals to the effects studied, for example in terms of body weight. Humans, animals and the environment are constantly exposed to different chemicals at different concentrations and timepoints, through different exposure routes such as inhalation or via diet, making realistic exposure scenarios complex and challenging to assess (Kienzler et al., 2016). Figure 6 illustrates the complexity of exposure to chemical mixtures, for which risk assessors can assess combining the chemicals combined per exposure route (combined exposure) or considering one chemical per exposure route (aggregated exposure) (Kienzler et al., 2016).

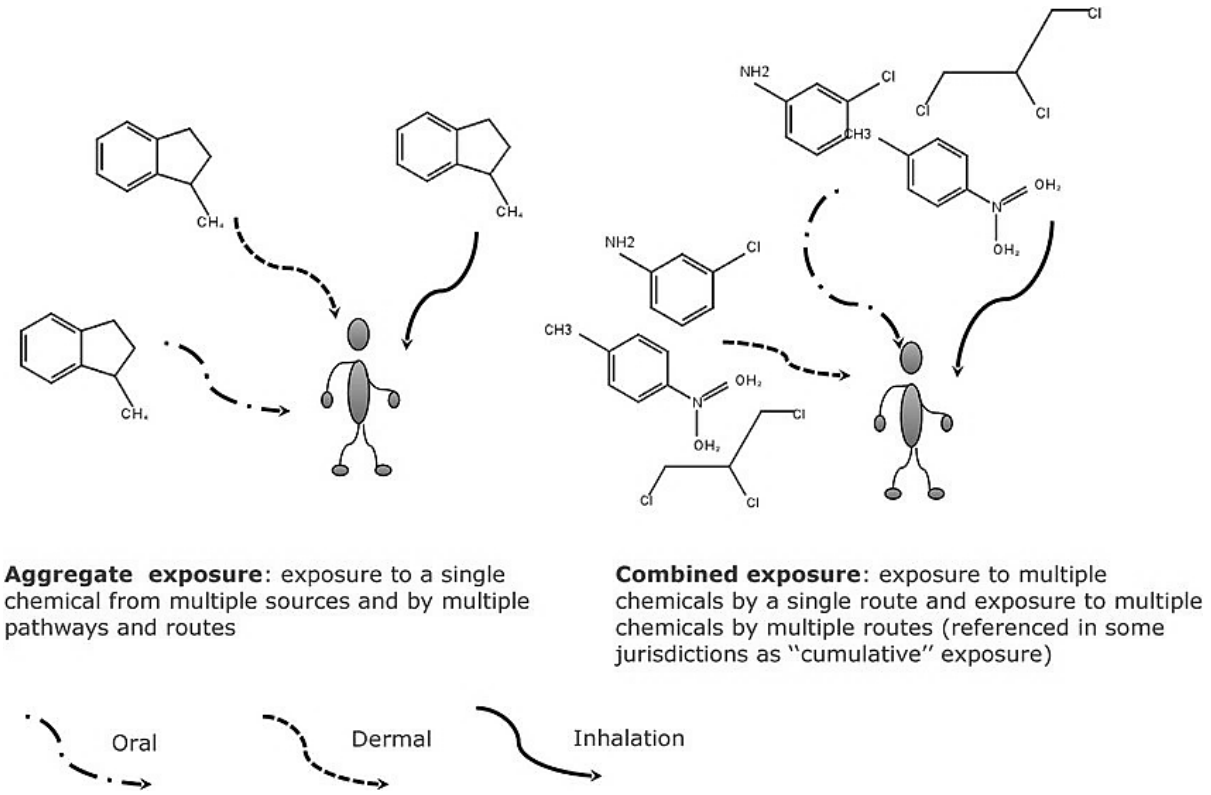

Figure 6. The combined exposure assessment can be divided into aggregate and combined exposure assessment. Aggregate exposure assessment considers "one chemical, one exposure route", while the combined exposure includes more than one chemical per exposure route. Reproduced with permission from the article by Kienzler and colleagues (2016).

One way to assess a complex chemical mixture is to characterize the individual chemicals that it consists of (EC, 2012). After the hazard characterization of the individual chemicals is described, through experimental cell- or animal models, for example, one can combine that information with exposure data in humans to characterize the risk. This information can then be used for aggregated or combined exposure assessment. However, it is still challenging to extrapolate *in vivo* and *in vitro* effects and observations to humans. There are significant differences in toxicokinetics and toxicodynamics from *in vivo* and *in vitro* models to humans, which further complicates the prediction of human effects based on non-human data. Nevertheless, one way to characterize the hazard and the contribution to the outcome of structurally similar chemicals in a mixture is to compare their effects through, for example, relative potency factors (RPFs) (EFSA, 2017). Although the relative potency factor approach is not new, it is useful for characterizing lesser-known chemicals in a mixture, departing from better-studied structurally related compounds.

2.4.6 RELATIVE POTENCY FACTORS

In order to employ the RPF approach for risk assessment of mixtures and structurally similar compounds, chemicals being assessed should have similar physicochemical, toxicological and ecotoxicological properties (EFSA, 2017; Faustman, 2018). It assumes dose addition and, therefore, that the mixture components can be expressed in terms of an index chemical-equivalents, usually well-studied substances such as TCDD or PFOA (Bosgra et al., 2009; Van den Berg et al., 2006; Zeilmaker et al., 2018). To apply the RPF approach, the substances under assessment should simultaneously fulfill three criteria, demonstrating that these are interconnected (Bosgra et al., 2009). Conversely, not fulfilling one of these premises would render the approach inapplicable (Bosgra et al., 2009). The criteria to apply the RPF approach for chemical mixture assessment are (Bosgra et al., 2009):

1) Act with a similar mode of action;
2) Parallel dose-response curves in the log-dose scale (to differ in potency only);
3) Lack of interaction between the assessed compounds.

Regarding the first criterion, evidence should demonstrate that the assessed chemicals show similar toxicity or effects, potentially acting via a similar mode of action. This qualitative appraisal is usually done through experimental evidence evaluation. Synergistic substances might interact differently with the molecular receptor/target and therefore not have a similar mode of action – even though they do not interact directly with each other. The dose addition assumption would not be fulfilled in this case, rendering the approach inapplicable (Bosgra et al., 2009).

The second criterion, the parallelism of the curves, refers to the properties of the log scale. The distance between two points on the normal (linear) scale is equivalent to adding a fixed number to the starting value. On the log scale, it is equivalent to multiplying the starting value by a fixed number. The BMD approach makes use of this log scale property and calculates the RPFs comparing the b parameter (potency) of curves proven to be parallel, expressing the results as "chemical X is 20 times more potent than chemical Z", equivalent but simpler than "chemical Z exerts 5% of the effect of chemical X". In order to draw the curves, the BMD approach performs interpolation between the doses, comparing the models via the AIC (EFSA, 2017). Thus, it is unlikely that the curves will be parallel if they do not fulfill the first criterion – a similar mode of action.

The third criterion, the lack of interaction between the compounds, may be tested by multiplying the estimated RPFs with the respective doses (at the same CES) and comparing the individual exposure with the mixed exposure experimental results. If the results are the same or similar, then

a mixed exposure does not result in synergism but in additivity, fulfilling the criterion for the lack of interaction. However, a perfect dose addition is rarely observed in real-life experiments, as the data contain non-random errors and variation.

The advantage of the RPF approach lies in its simplicity and usefulness – it is possible to assess substances for which limited information is available. It also reduces the need for *in vivo* studies, having therefore a strong 3R potential, apart from being time- and resource-saving, especially for the assessment of substances with similar modes of action, such as pesticides, dibenzo-p-dioxins, bisphenols and polychlorinated biphenyls, among others (Bosgra et al., 2009; Kienhuis et al., 2015; Müller et al., 2009; Norgren et al., 2022; US EPA, 2003; Van den Berg et al., 2006). Many international authorities suggest using the RPF approach to expedite the assessment of mixed exposures and groups of substances with a similar mode of action (EFSA, 2017; US EPA, 2003, 2012; WHO IPCS, 2020).

2.4.7 3R AND BENCHMARK DOSE-MODELLING

The BMD-modelling approach displays great potential in terms of applying the 3R principles – Replacement, Reduction and Refinement of animal use in research and testing (Slob, 2014a, 2014b; Öberg, 2010; **study III**).

The Replacement possibilities of BMD are more limited and indirect in relation to the other Rs (Reduction and Refinement), but may, for example, support the comparison and validation of alternative studies against animal studies (Slob, 2014a). Notable examples are the studies by Norgren and colleagues, that compared experimental data and modelling results to test for additivity, and Soeteman-Hernández and colleagues, who demonstrated that *in vivo* micronucleus test data can be used to predict the carcinogenic potency of chemicals (Norgren et al., 2022; Soeteman-Hernández et al., 2016).

Furthermore, Reduction can be achieved by employing the BMD-modelling approach to improve the study design by adjusting the number of dose groups and the number of animals per dose group (Kalantari et al., 2017; Ringblom et al., 2017; Shao and Small, 2012; Slob, 2014a, 2014b; Slob et al., 2005). In particular, it has been advocated that five to ten dose groups are necessary for an optimal study design, while keeping the same number of animals per study, instead of the historically used control plus three dose groups (EFSA, 2022; Kuljus et al., 2006; Shao and Small, 2012; Slob et al., 2005). Reduction of the total number of studies and number of animals used per study constitutes a substantial 3R and ethical gain. An additional Refinement gain regards the lower number of animals exposed to high doses which are often associated with potentially toxic effects (Kalantari et al., 2017; Ringblom, 2016; Ringblom et al., 2017).

In terms of Refinement, which ultimately leads to Reduction, BMD-modelling allows extracting more information from the same amount of data, compared to the NOAEL approach (Slob, 2014a, 2014b). This increased efficiency represents an important ethical gain, especially when test organisms have been sacrificed in the name of Science (Slob, 2014a, 2014b). The BMD approach models an effect or response, on the dependent variable, as a function of the dose or concentration, allowing one to study the steepness of the transition in the studied endpoint/effect and interpolate between doses (Slob, 2014a; Slob and Setzer, 2014). Both of these features are assessed by the NOAEL approach in a much more limited way (Slob, 2014a, 2014b). Additionally, BMD-modelling estimates more precise PoDs and allows the combination of data from different studies, which is not possible with the NOAEL approach. Multiple study datasets can also be used to calculate RPFs for different chemicals and test for additivity, not requiring additional animal studies (EFSA, 2017; Kienhuis et al., 2015; **study V**).

An important role in promoting the 3R principles can be played by the model averaging feature and by probabilistic approaches, as the data obtained from each study (regardless if *in vitro*, *in vivo* or human data) only reflects one possible reality. Model averaging, in particular, allows combining different models to address uncertainty and deliver a plausible range of values (BMDLs, BMDs and BMDUs). It also addresses the concern voiced by Dr. George Box that "*Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration*." (Box, 1976). The IPRA approach makes a probabilistic analysis of the dose-response and exposure data, while not requiring additional data points (see **study I** for an example). It stands as a promising strategy to extract information from the data and lead to better-grounded conclusions within the risk assessment and regulatory toxicology areas.

Travis and colleagues voiced resistance to the role that the BMD method could play in the future and put forward a number of reasons for which "*the BMD will never entirely replace the NOAEL*", questioning whether the BMD approach would not, instead, lead to a larger number of necessary animals (Travis et al., 2005). As toxicologists and risk assessors are becoming more familiar with the BMD approach and an increasing number of national and international authorities endorse it, the consensus is that the advantages largely outweigh the limitations, e.g. the knowledge threshold for its use (Muri et al., 2009; Öberg, 2010).

Lastly, other modelling or machine learning strategies, such as PLS models or random decision forest algorithms, may be combined with BMD to provide an even deeper understanding of the data. The most important 3R gains regard the improvement of animal study design and assessment, for example in terms of early detection and monitoring of signs of toxicity (see **study II** for an example). It addresses one of the greatest challenges for Refinement and chemical risk assessment,

namely the importance and contribution to the outcome of the different variables in an *in vivo* study, not only from the exposure perspective but also possible associations between the different endpoints assessed.

2.5 BENCHMARK DOSE-MODELLING IN THE PHARMACEUTICAL DEVELOPMENT AND REGULATORY CONTEXT

The growing importance of the BMD method can be illustrated by the increasing number of authorities and expert groups with expertise in Toxicology and risk assessment that have endorsed its use. Additionally, an ever-growing number of documents and reports have employed this approach for dose-response assessment. It is, therefore, natural that other areas have started using this modelling strategy, such as life sciences research, pesticides, veterinary and pharmaceutical development, due to its advantages, usefulness and wide applicability (Ecke et al., 2017; ICH, 2019; Mecklenburg, 2012; SSCS, 2018; **study III**).

Pharmaceutical development is a well-defined process, framed by regulations and guidelines established by competent authorities (EMA, 2010; FDA, 2010; OECD, 2014, 2018a). These regulations and guidance documents often require *in vivo* testing, if no suitable alternative methods are available. It is common to perform several and sequential preclinical studies, to obtain a large body of data before proceeding with resource-demanding first-in-human trials. In order to avoid unnecessary animal use and suffering, the EU decreed that all *in vivo* studies have to align with EU Directive 2010/63 with emphasis on the 3Rs (EU, 2010). In fact, significant progress has been made, e.g. in Reduction of the total number of animals required for testing and microsampling of low blood volumes (Jonsson et al., 2012; Sewell et al., 2014; Sparrow et al., 2011; Törnqvist et al., 2014), as well as Refined use of body weight loss assessment for decisions regarding the maximum tolerated dose (MTD) (Chapman et al., 2013). *In silico* models are able to predict NOAELs/LOAELs relatively well, based on quantitative relationships between the endpoints and the Simplified Molecular Input Line Entry System (SMILES) attributes (Gadaleta et al., 2021).

For pharmaceutical development purposes, the BMD-modelling approach may be used to plan and evaluate experimental data. The planning phase can be refined, addressing questions such as the placement of doses and unequal group sizes, based on expected concentrations for effects, for an improved study design (Ringblom et al., 2017; Slob 2014a, 2014b; Öberg, 2010). The evaluation phase can ascertain a better understanding of the data by describing dose-dependent effects and analyzing different levels of change. Among other types of studies, the BMD method can be useful for drug safety assessment and toxicity testing, mandatory steps

in the pharmaceutical development process. Additionally, BMD-modelling is already being used in distinct research areas such as the analysis of genomic data and key events in adverse outcome pathways (AOPs) (Chen and Shao, 2020; Simon et al., 2014; Yang et al., 2007).

In the regulatory toxicology area, the BMD-modelling approach has been mostly used for the assessment of potential dose-response relationship(s) between chemical exposure (in terms of doses or concentrations) and an effect(s). It is generally accepted as the best practice to derive the PoD and it is currently endorsed and used by many national and international regulatory agencies and expert groups such as the ECHA, EFSA, JECFA, OECD, US EPA and WHO IPCS (ECHA, 2012, 2017; EFSA, 2017; NAC/AEGL, 2001; OECD, 2014, 2018a; US EPA, 2012; WHO FAO, 2016; WHO IPCS, 2009, 2018, 2020). It is common to base an assessment on a single, critical endpoint, for which relevant toxic effects have been observed, forming the basis on which the reference value(s) will be estimated. However, as discussed in section **2.4.4 Multiple endpoints**, it might be beneficial to analyze several, if not all, endpoints for which data were collected, to obtain a more complete picture of the ongoing biological changes. Lastly, Tier 2 (uncertainty quantification) analysis, mentioned in **2.4.3 Integrated Probabilistic Risk Assessment (IPRA)**, might become more common as the BMD popularity continues to grow. Easy-to-perform probabilistic analyses, which generate dose-response and exposure distributions, might become instrumental for risk assessment of substances for which there is a limited sample size or information available, but is currently time-consuming and expertise-demanding.

2.6 PFAS

Persistent Organic Pollutants (POPs) are carbon-based chemicals that persist in the environment or living organisms and have bioaccumulative properties, i.e. tend to accumulate in organisms of higher trophic levels in the food chain. Many POPs cause adverse effects on living organisms and the environment, such as cancer and diabetes in humans, or may render water undrinkable and contaminate the soil (ATSDR, 2021; EFSA, 2020; IARC, 2015; Li et al., 2017; Li et al., 2021a; Ruzzin, 2012; Steenland and Winquist, 2021).

Many of the new chemicals in the market are anthropogenic, with useful technical properties, but information is lacking regarding exposure levels deemed safe for human health. This presents a considerable challenge for regulatory toxicologists to assess and characterize the risk associated with exposure to these chemicals (Jian et al., 2017). In order to limit the adverse effects of exposure to persisted chemicals, the Stockholm Convention categorizes POPs in Annex A (Elimination), Annex B (Restriction) or Annex C (Unintentional production). Among several substances of concern, the convention includes PFAS and PCBs. The Stockholm Convention on

Persistent Organic Pollutants stands as an example of coordinated chemical regulation, with 152 signatory countries as of April 2022 (UNEP, 2021).

PFAS are pollutants ubiquitously found in human, animal and environmental samples, known for their long half-lives and toxicity, including carcinogenic potential (ATDSR, 2021; Banzhaf et al., 2017; EFSA, 2020; Jian et al., 2017). In 2018, the OECD identified 4,730 PFAS-related chemicals on the global market (OECD, 2018b). There is considerable uncertainty regarding the detection and exact identification of these molecules, which often occur in complex mixtures of parent and derivate compounds (Sunderland et al., 2019; Sznajder-Katarzyńska et al., 2019). To date, there is no harmonized risk assessment method to evaluate cumulative risk following mixture exposure (Kienzler et al., 2016). The most studied individual PFAS chemicals are perfluorooctane sulfonic acid (PFOS) and PFOA, the former being listed under Annex B (Restriction) (decision SC-4/17, UNEP, 2017) and the latter being included in Annex A (Elimination) (decision SC-9/12, UNEP, 2012) of the Stockholm Convention on Persistent Organic Pollutants. Perfluorohexane sulfonic acid (PFHxS) and other PFAS are under evaluation, but no decision has been taken, as chemical risk assessment is a time-consuming and resource-demanding process.

It is widely known that well-studied PFAS represent only a fraction of the substances that humans are exposed to and, without class restriction, the delay between production, use and chemical regulation will only increase (Wang et al., 2017). Chemical regulation is necessary and, among other important actions, sets safe exposure levels and restricts the use of potentially harmful substances, such as PFOS and PFOA. A chemical is only restricted when its harm has been proven with sufficient evidence. However, other fluorinated molecules might be introduced to replace potentially harmful PFAS, with similar properties but unknown fates or effects, often with uncertain consequences for human health (Wang et al., 2017). Chemical replacement may sometimes lead to the so-called regrettable substitution, where the alternative is unknowingly as or more harmful than the compound intended to be replaced (Ritscher et al., 2018). Mounting evidence points out that GenX represents one such case, which was introduced to replace long-chain PFAS (Bălan et al., 2021). One way to monitor this problem and the body burden of fluorine-based chemicals is to quantify the total organofluorine compounds in human blood, and their changes over time, as targeted analysis captures only a fraction of the fluorinated entities to which humans are exposed to (Aro et al., 2021, 2022; Yeung and Mabury, 2016). As such, PFAS constitute a complex chemical regulatory challenge, requiring tools that can quickly assess the risk of chemical exposure for human health and estimate potentially safe exposure limits.

In a similar and unfortunate fashion to other countries, Sweden has discovered contaminated drinking water and sites close to airports and airfields, where PFAS-containing firefighting foams

have been used (Filipovic et al., 2015; Gyllenhammar et al., 2015; Jakobsson et al., 2014; Li et al., 2017; Stubleski et al., 2017). Recent research did not find a link between increased serum PFAS-levels and the risk of thyroid disease or inflammatory bowel disease in a Swedish subpopulation (Andersson et al., 2019; Xu et al., 2020). On the other hand, effects associated with PFAS exposure included changes in the serum lipids levels (Li et al., 2020), thyroid hormones (Li et al., 2021b), menstruation onset (Glynn et al., 2020) and body mass index changes (Gyllenhammar et al., 2018). However, PFAS exposure effects may be diffuse and do not always lead to a disease state (ATSDR, 2021; EFSA, 2020; IARC, 2015). The same PFAS exposure may result in different effects at different ages and sexes (Andersson et al., 2019; Attanasio, 2019; Dallaire et al., 2009; Eriksen et al., 2013; Grandjean et al., 2012, 2017; Knox et al., 2011; Li et al., 2020, 2021b; Lopez-Espinosa et al., 2012; Mondal et al., 2014; Shrestha et al., 2015; Wen et al., 2013). As such, human health risk assessment may benefit from probabilistic modelling strategies such as IPRA and the approximate probabilistic risk analysis (APROBA) (Bokkers et al., 2017) to deliver robust estimates of risk, as demonstrated in **study I**.

2.7 PCBs

PCBs are organic highly chlorinated chemicals that are also highly persistent and bioaccumulative (WHO, 2003). There are 209 PCBs, ranging from monochlorobiphenyl (PCB-1) to decachlorobiphenyl (PCB-209). These compounds were used in coolants, lubricants and paints owing to their useful technical properties, such as chemical stability, low flammability and vapor pressure. Despite their usefulness, their hazard was later characterized, resulting in their inclusion in Annex A (Elimination) of the Stockholm Convention on Persistent Organic Pollutants (UNEP, 2021). Nevertheless, PCBs may still impact human and environmental health owing to their long half-lives, ranging from months for the less chlorinated PCBs to several decades for the heavily chlorinated PCBs (Grandjean et al., 2008; IARC, 2015; Ritter et al., 2011; Seegal et al., 2011). The UNEP Chemicals and Waste Branch estimated that only 17% of the PCBs ever produced have been eliminated, and the remainder 83% (corresponding to approximately 14 million tonnes) are still in circulation, present either in the environment, humans or animals (UNEP, 2016).

Human exposure to PCBs may occur through, for example, the atmosphere or contaminated water and fish (Figure 7) (Judd et al., 2014; Zhu et al., 2022). However, the causality between exposure and human health effects is not easily demonstrated, as these chemicals were produced and often occurred in mixtures of various proportions. PCBs mixtures were generated in production processes that involved the addition of increasing amounts of chlorine until a target percentage by weight was achieved. Therefore, mixtures with greater amounts of chlorine will include more

highly chlorinated PCBs. The most well-known range of products featuring PCBs was the Aroclor series, produced by Monsanto Company until 1977. Despite international efforts to ban, reduce circulation and ultimately eliminate PCBs, both adults and children are still exposed through food and lactation (ATSDR, 2021; EFSA, 2018a; IARC, 2015). Therefore, adequate dose-response assessment and hazard characterization of the different PCBs is necessary. The most hazardous chemicals in this class are the 12 dioxin-like PCBs (PCB-77, 81, 105, 114, 118, 123, 126, 156, 157, 167, 169 and 189), which can activate the aryl hydrocarbon receptor (AhR) and lead to a multitude of downstream effects.

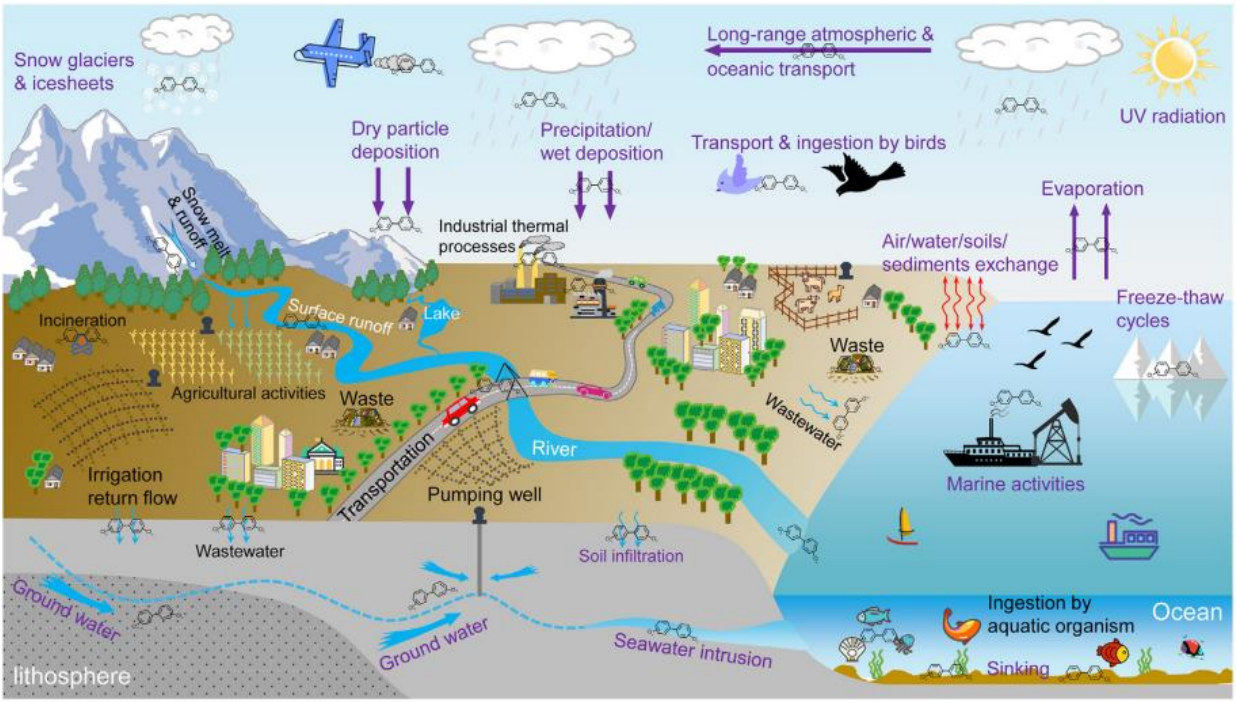

Figure 7. Schematic depiction of the PCB transportation routes, resulting in human and environmental exposure. Reproduced with permission from the article by Zhu and colleagues (2022).

Furthermore, it has been suggested that the retinoid system could be sensitive to PCB exposure (Grignard et al., 2020; Håkansson, 2020; Nilsson and Håkansson, 2002). Possible effects include liver toxicity, disrupted retinoid levels and activation of phase I and II metabolizing enzyme activities such as CYP1A1 and CYP2B (Grignard et al., 2020; Håkansson, 2020; Nilsson and Håkansson, 2002). Therefore, the toxicity of PCB-156 exposure and the relative potency of seven PCBs (PCB-28, 77, 105, 118, 128, 153 and 156), based on effects in the retinoid system, were investigated in **study IV** and **V**.

2.8 KNOWLEDGE GAP

Considering the need to use transparent and scientifically sound approaches when assessing risks, this thesis departed from the existing BMD approach to deepen the knowledge and widen the gap towards the historically employed and more limited NOAEL approach. Despite increasing evidence of the advantages and usefulness of the BMD-modelling approach for dose-response assessment in risk assessment, compared to the NOAEL approach, several knowledge gaps were identified, which need to be addressed for continuous improvement and acceptance of this method. National and international authorities such as the WHO IPCS, the US EPA and EFSA have endorsed and provided guidance on how to perform dose-response assessment using the BMD method, with the intention of deriving PoDs and estimate references values (EFSA, 2017; US EPA, 2012, 2020; WHO IPCS, 2020). This thesis explores current challenges to the BMD approach such as the choice of CES, analysis of multiple endpoints and chemical mixture risk assessment. The applications in the thesis included an IPRA analysis and the use of the BMD approach for the risk assessment of PFAS, PCBs and for pharmaceutical development purposes. More specific knowledge gaps include the following:

- Does probabilistic risk assessment using the BMD method reveal insights about the risk of health effects for the exposed population, whereas a classical deterministic evaluation might not describe them as well? (**study I**)
- How can the PLS and BMD approaches model multiple endpoints, including clinical signs, and contribute to better predictions of the outcome in subsequent studies? (**study II** and **III**)
- How can BMD-modelling be used for pharmaceutical development, specifically for the safety assessment of preclinical *in vivo* studies? (**study III**)
- How can BMD-modelling of multiple endpoints contribute to an improved understanding of toxicity of PCBs? (**study IV** and **V**)
- How do PCB-28, 77, 105, 118, 126, 128, 153 and 156 compare in potency, when employing the BMD-modelling approach to calculate RPFs based on the retinoid disruption effects observed on eight similar 90-day studies in rats? (**study V**)
- How do the different strategies for choosing the CES perform and affect the result of BMD-modelling in general and the reference values in particular? (**study IV**)

This thesis addresses these questions, aiming to contribute to the knowledge base by filling the gaps mentioned above and by improving research in the BMD field.

3 AIMS

The overall aim of this thesis is to advance the knowledge of the BMD-modelling approach by exploring a set of new applications for the toxicological and pharmaceutical development areas, and to discuss potential improvements. Additionally, it intends to address the identified knowledge gaps and strengthen the relevance and conceptual understanding of the BMD-modelling approach in the regulatory toxicology field. The main substances assessed, PFAS, PCBs and an anti-oncogenic candidate drug, were chosen as relevant chemicals for which the dose-response assessment and hazard characterization is not complete.

The specific aims for each study are as follows:

For **study I**, to use integrated probabilistic risk assessment to characterize the risk of decreased levels of total triiodothyronine (T_3) in a population which has been highly exposed to PFAS via drinking water.

For **study II**, to investigate possible associations between clinical signs, a 5% body weight change and histopathological findings observed in a series of preclinical pharmaceutical tests in rats, and to determine their potential as early markers of toxicity.

For **study III**, to examine the applicability of the BMD approach for safety assessment in preclinical studies for pharmaceutical development purposes, specifically when modelling continuous data endpoints (including biochemistry, hematology, organ pathology and clinical observations).

For **study IV**, to evaluate potential dose-dependent toxicological effects in general, and the retinoid system in specific, in rats exposed to PCB-156 via diet.

For **study V**, to compare and calculate relative potency factors for seven PCBs (PCB-28, 77, 105, 118, 128, 153 and 156) and one PCB-mixture in rats exposed via diet. The retinoid system was also studied as a potential critical endpoint for PCB exposure and RPF estimation.

For **study VI**, to investigate and compare the different strategies for the choice of CES, using case-studies comprising PFAS, PCB-156 and a pesticide. Additionally, to demonstrate the distinct possible outcomes and reference values resulting thereof.

Additionally, this thesis intends to elucidate the 3R gains from the use of the BMD and PLS modelling strategies for study assessment. Lastly, it aims to provide evidence that the BMD method is a reliable, adaptable and consistent approach for human health risk assessment, following exposure to chemicals and pharmaceuticals.

4 ANIMALS AND METHODS

This section serves as an overview of the data and methods used in this thesis. Detailed descriptions of the materials and statistical analysis can be found in the associated publications.

No animal experiments were conducted for this thesis; data from previously performed studies were reanalyzed using the BMD and PLS modelling approaches. Therefore, there was no need to perform any additional experiments, as data were available after the studies' termination. The specific animals and methods for each study were:

For **study I**, PFOS and PFHxS serum levels from 1 845 subjects from the Ronneby cohort were analyzed, who resided in Ronneby for over a year (up to December 2013). Their individual exposure levels were used to create the IEXP distributions, and the IBMD distribution departed from animal data from 44 monkeys exposed daily to PFOS for six months (Seacat et al., 2002). Data were analyzed using the R-package PROAST version 65.5 (National Institute for Public Health and the Environment, RIVM, Bilthoven, Netherlands), and the contribution of the uncertainty sources was estimated using the multiple linear regression nonnegative linear models R-package (NNLM, version 0.4.1) (Lin and Boutros, 2016). The remaining calculations were performed in R version 3.4.2 (R Development Core Team, 2017).

For **study II**, data from three pharmaceutical development studies were analyzed, which included 136 rats – 40 in study 1, 16 in study 2 and 80 in study 3. The independent variables employed to predict the outcome were all registered clinical observations, a 5% body weight change (at study termination), the toxicokinetic parameter C_{max} and the sex of the animals. The dependent variables were all of the pathological findings observed, which were found in the rat's bone marrow, epididymides, large intestines, liver, lymph node, testes and thymus. Associations were investigated employing the PLS regression method, performed using the SIMCA-P computer software (version 15, Sartorius Stedim Data Analytics AB, Umeå, Sweden).

For **study III**, the same raw data for **study II** were reanalyzed, i.e. 136 rats in three studies – 40 in study 1, 16 in study 2 and 80 in study 3. BMD-modelling was fitted to the raw data using the R-package PROAST version 69.1 (National Institute for Public Health and the Environment, RIVM, Bilthoven, Netherlands), ran on R software version 3.6.3 (R Development Core Team, 2020). The daily administered dose was the only independent variable considered. The dependent variables were the 63 endpoints for which data were available – 28 biochemistry and hematology endpoints, 22 clinical observations endpoints, 12 organs for which pathological findings were observed and, lastly, a 5% body weight change at study termination.

Study IV comprised data of 100 rats, divided into five dose groups of 10 animals per sex. The daily administered dose was the only independent variable considered. The 51 dependent variables included 19 biochemistry and hematology endpoints, three urinary ascorbic acid measurements, three liver enzymatic activities, liver uroporphyrin concentration, six endpoints regarding the apolar retinoid concentrations and total amounts in the liver, kidneys and lungs, six organ tissue concentrations, eight organ weights, total body weight gain and, finally, a 5% body weight change at study termination. BMD-modelling was performed using the R-package PROAST version 67.0 (National Institute for Public Health and the Environment, RIVM, Bilthoven, Netherlands), ran on R software version 3.6.3 (R Development Core Team, 2020).

For **study V**, data from eight 90-day *in vivo* studies and one 28-day study in rats were examined. The 90-day toxicity testing studies investigated the effects following daily dietary exposure to PCB-28, 77, 105, 118, 126, 128, 153 and 156) (Chu et al., 1994, 1995, 1996a, 1996b, 1998; Lecavalier et al., 1997; **study IV**). The 28-day study explored the possible toxic effects following a daily exposure to a PCB-mixture reflecting the congeners identified in human milk, composed (in weight %) of PCB-180 (29.0%), PCB-118 (25.7%), PCB-105 (15.0%), PCB-170 (12.1%), PCB-156 (8.4%), PCB-114 (4.0%), PCB-167 (2.51%), PCB-157 (1.83%), PCB-189 (0.62%), PCB-123 (0.4%), PCB-169 (0.18%), PCB-126 (0.15%) and PCB-77 (0.004%) (Chu et al., 2001). Each 90-day study included 100 rats, divided into five dose groups of 10 animals per sex. For the 28-day study, 35 rats were included, divided into five females per dose group. BMD-modelling was performed using the R-package PROAST version 70.0 (National Institute for Public Health and the Environment, RIVM, Bilthoven, Netherlands), ran on R software version 3.6.3 (R Development Core Team, 2020).

For **study VI**, five case-studies were chosen to illustrate the possible outcomes following the choice of different strategies to set the CES:

- For case-study 1, data from a study by Abraham and colleagues (2020) were reanalyzed. This study was chosen as it was considered the critical study in the EFSA Scientific Opinion on the *Risk to human health related to the presence of perfluoroalkyl substances in food* (EFSA, 2020). The dataset was composed of 101 one-year-olds, the independent variable was the sum of PFOA, PFOS, PFHxS and perfluorononanoic acid (PFNA) serum levels, and the dependent variable was the vaccine-induced antibodies against diphtheria;
- For case-study 2, data from a study by Grandjean and collaborators (2012) were reanalyzed. This study was chosen as it was considered the critical study in the EFSA Scientific Opinion on the *Risk to human health related to the presence of perfluorooctane sulfonic acid and perfluorooctanoic acid in food* (EFSA, 2018b). The data from 587 children (5- and 7-year-

olds) were analyzed, modelling the independent variable PFOS or PFOA serum levels and the dependent variable vaccine-induced antibodies against diphtheria;

- For case-study 3, data from a study by Seacat and colleagues (2002) were reanalyzed. This study was chosen as the critical study to analyze PFOS effects by the EFSA Scientific Panel on Contaminants in the Food Chain publication *Perfluorooctane sulfonate (PFOS), perfluorooctanoic acid (PFOA) and their salts* (EFSA, 2008). The dataset was composed of 44 Cynomolgus monkeys, the independent variable was the PFOS serum levels and the dependent variables was the serum cholesterol, HDL and total T_3 levels;
- For case-study 4, data from **study IV** were reanalyzed, employing PCB-156 levels as the dependent variable, but focusing on the rat liver apolar retinoids as the dependent variable;
- For case-study 5, data from two toxicity testing studies for norflurazon (a pesticide, CAS number 27314-13-2) were modelled. These studies were considered critical by the JMPR, based on which a reference value was established. The dataset was composed of a 1-year (Warren et al., 1990) and a 6-month (Klotzsche and Carpy, 1973) study on 32 dogs each. The independent variable was the norflurazon exposure and the dependent variable the relative and absolute liver weight.

The different strategies for setting the CES examined in **study VI** were 5% (EFSA, 2017), 1 SD (US EPA, 2012), a small effect size as defined by the GTES ($\sqrt[8]{M}$, M being the maximum response) (Slob, 2017), an endpoint-specific CES based on historical data and expert judgment. Benchmark dose modelling was performed using the R-package PROAST version 69.0 for case-study 1 and version 70.0 for case-studies 2 to 5 (National Institute for Public Health and the Environment, RIVM, Bilthoven, Netherlands), ran on R software version 3.6.3 (R Development Core Team, 2020).

5 ANIMAL USE AND ETHICAL CONSIDERATIONS

Regulatory frameworks for risk assessment and pharmaceutical development purposes may require *in vivo* testing to be performed. However, for ethical and 3R motives, no additional animals or studies should be performed other than those strictly necessary. It has been shown that Reduction can be achieved through, for example, project coordination and improved study design (Törnqvist et al., 2014). The BMD approach has also been described to have a 3R potential (Ringblom, 2016; Slob 2014a, 2014b; Öberg, 2010). In particular, both the BMD and the PLS modelling approaches depart from existing data from which they retrieve information, revealing associations between variables that could otherwise go unnoticed. Consequently, the studies in this thesis had a strong focus on the 3R and ethical gains thereof, intending to contribute to the development of Science in general and Toxicology in particular, by improving the knowledge base.

The ethical considerations for each study were:

Study I was based on previously published animal data (Seacat et al., 2002) and a human cohort managed by researchers at Lund and Gothenburg universities in Sweden. **Study I** was performed with ethical approval from the Ethics Committee at Lund University, with the permit numbers dnr 2014/267 and 2015/902. Additionally, all participants were informed about the study in which they were to participate and for which they would contribute a blood sample, after signing a consent to permit the analysis, storage and use of their data for research purposes. The animal study was performed by the industry, namely the US-based 3M company, and no mention of an ethical permit was found. However, it is stated in the study method section that "*The study was carried out in an Association for the Assessment and Accreditation of Laboratory Animal Care International (AAALAC) accredited facility and in compliance with the Animal Welfare Act regulations (9 CFR 1–4)*" and, regarding regulatory compliance, that "*All aspects of this study were conducted in accordance with the Environmental Protection Agency Good Laboratory Practice Regulations, 40 CFR 792, with the exception of the analysis performed at the Mayo Medical Laboratories.*" It was concluded that the researchers performing both studies correctly addressed the ethical aspects and that no ethical problems could arise from the re-use of these data. Additionally, **study I** demonstrated a higher risk than observed by the authors who conducted an earlier analysis employing the deterministic NOAEL approach, which questions the ethical aspects of an approach that might underestimate the risk for adverse effects on human health.

Study II and **III** analyzed data from three animal studies performed by trained technicians at the Swedish Toxicology Sciences Research Center (Swetox) facilities. These studies were approved

by the Southern Stockholm Ethical Committee for Research Animals, with the ethical permit number S7-15. The studies were performed according to Swedish animal welfare legislation L1 (SFS 1988:534) and L150 (SJVFS 2012:26 for study 1 and SJVFS 2015:24 for study 2 and 3. The research unit and the laboratory animal facility had a great 3R and ethical focus through careful study design, cage enrichment and the limited number of animals required.

Study IV featured an *in vivo* study approved by the Animal Care Committee of the Health Protection Branch, Health Canada. The animal study 1511-02-410-K341404 followed the guidelines of the Canadian Council on Animal Care (1993) and complied with the standard operating procedures of the Animal Resource Division of the Healthy Environments and Consumer Safety Branch, Health Canada. This animal study received the ethical permit number ACC90105.

Study V and the calculation of RPFs, based on the comparison of substances, do not require ethical approval. It is important, though, that these are calculated transparently, so that no ethical issues regarding their use may arise. All animal studies for which **study V** analyzed data were performed under the ethical permit number ACC90105, under the same conditions and standard operating procedures described for **study IV**. The most important ethical gain from **study V** regards the RPFs estimation, which may contribute to the Reduction and Replacement of additional animal studies, which would otherwise be carried out to illustrate the potential toxicity of these PCBs. Thus, **study V** has a strong ethical focus, making use of publicly available data.

Study VI researches the different results obtained by five strategies for choosing the CES, and discusses their potential impact. The question of how adverse an effect needs to be was raised, as too adverse effects are related to ethical aspects of toxicity testing. As greater CESs relate to more adverse effects in the organisms examined, it should be scaled to fall within the tested dose range, where potential changes were observed. However, too small CESs might be within the biological variation of the endpoint and, therefore, be deemed inadequate. It has been shown that it is fundamental importance to balance the number of animals required, the doses and the adversity of expected effects, to diminish animal suffering and improve study outcomes (Ringblom, 2016). Thus, good quality data and data sharing are important, reflecting a great ethical awareness from the researchers or scientists. For **study VI**, publicly available data were collected from terminated studies, which received an ethical permit and followed animal guidelines. Accordingly, no ethical permit was necessary for the performance of this study, which only reanalyzed data.

6 RESULTS AND DISCUSSION

Benchmark dose-modelling approach has matured over the past 35 years, gaining global importance and being employed for the dose-response assessment of diverse chemicals such as food contaminants, pesticides, and carcinogens. Despite its wide acceptance, some issues persist, such as the lack of standardization regarding the suite of models to fit the data, the comparison of model fit, the model averaging methodology and the selection of CES for continuous data (Haber et al., 2018; Jensen et al., 2019). This thesis addresses some challenges which arise from the application of the BMD approach, namely:

- the choice of CES, i.e. the analysis of the different strategies to estimate a threshold for adverse effects, discussed in sections **6.1 The choice of Critical Effect Size (CES)** and **6.5 PFAS**;
- the analysis of multiple endpoints and subsequent benefits compared to the traditional focus on a single, critical endpoint, reviewed in sections **6.2 Multiple endpoints** and **6.4 Pharmaceutical development, safety assessment and dose-response modelling**;
- the assessment of the contribution of individual chemicals in a mixture to the outcome, discussed in sections **6.3 Chemical mixtures** and **6.6 PCBs**.

The analyzed substances, namely PFAS, PCBs and a candidate drug in pharmaceutical development, were chosen as examples of challenging chemicals to assess.

6.1 THE CHOICE OF CRITICAL EFFECT SIZE (CES)

Currently, there are several strategies for choosing the CES, which aim to detect and set the threshold for adversity at an appropriate level. The most common strategies for setting the CES are at 5% from the background (EFSA, 2017), 1 SD from the control group (US EPA, 2012), endpoint-specific based on historical data (Buist et al., 2009; Dekkers et al., 2006), the General Theory of Effect Size (Slob, 2017) or expert judgment. Their level of complexity varies, as well as their ease-of-use in practice. Some approaches might return similar values, in cases of good quality data with little variability, but this is not always the case. Data modelled by BMD may have various degrees of uncertainty. Therefore, it is of fundamental importance to choose a CES that is the most adequate for the assessed chemical, endpoint and analyzed dataset.

In **studies I**, **III** and **IV**, the CES was either based on the EFSA default value or expert judgment. The BMD guidance by EFSA recognizes that expert judgment is necessary when the default value is not appropriate (EFSA, 2017). It is often the case when analyzing data with great variation or significant changes, such as enzymatic activities. However, in some cases, for example body and

organ weights, a 5% change might be adverse. Biological variation might come into play, for example when analyzing hormones or biomarkers that follow a circadian or seasonal cycle, which could add to the total variation.

Study I suggested that a 10% decrease in total T_3 levels could be an appropriate CES and threshold for adversity, whereas 5% might have been in the noise region, and 15% would measure an effect that is overly adverse in relation to the chemical and assessed endpoint. The risk of adverse effects could, thus, be overestimated if a 5% CES was employed and underestimated if a 15% CES was chosen. Furthermore, for the CES of 10%, the risk estimates depended on the residency time, with a median of 2.1% and 3.5% for people residing in Ronneby for at least 1 or 29 years, respectively, exposed to PFOS and PFHxS through contaminated drinking water.

Study III compared the resulting BMDLs and BMDs following a choice of 5% as a default value to the GTES-based CES. Results show that the GTES calculated CESs which could represent better estimates of changes considered adverse. The GTES-derived CESs were often greater than 5%, suggesting that the latter could measure changes too small to be considered adverse. However, it was not always the case – when the magnitude of change was small, the GTES-based CESs were ≤ 5%. It was concluded that while the GTES is helpful in establishing dependent CESs on the magnitude of change, 5% is a useful starting value, especially when comparing effects and endpoints across different studies. In comparison to the results obtained by the NOAEL approach, 5% and the GTES-based CES often lead to lower BMDLs and BMDs than the studies' NOAEL or LOAEL (see Table 10 in **study III**).

Similarly, **study VI** explored how the different CES may lead to disparate conclusions. To this end, data from different classes of substances were analyzed with five different CESs. The GTES-based CES was often similar to the percent change equivalent to 1 SD, which were > 15% in four out of the five case-studies examined. The obtained results were also compared to the values chosen based on expert judgment, which was concluded to be irreplaceable and a cornerstone in risk assessment, combining both biological and statistical considerations. No single strategy seemed to fit all datasets, but a comparison of the different CESs returned by the various approaches was useful to be considered for critical appraisal. For example, an expert judgment-based CES can be substantiated by a similar value estimated by the GTES, in cases where other approaches seem to suggest inadequate CESs. This was the case, for example, of case-study 1, for which the adversity threshold following a combined exposure to PFOS, PFOA, PFHxS and PFNA associated to a decreased diphtheria antibody titer in 101 one-year-old healthy children (Abraham et al., 2020) was examined. An adequate level of adversity seemed to be around 10%, while the endpoint-specific CES based on historical data (58.9%) was deemed too adverse.

In conclusion, the results of this thesis highlight that a harmonized strategy for choosing the CES is necessary. It is unlikely that a single option will fit all possible cases and datasets. Therefore, it is important to model the data with the most suitable CES, i.e. the one for which the best arguments stand for. Different possibilities might be tested based on a unified strategy to approach the data. For example, a 5% and the GTES-based CES might be used as a first step to get insight into the data. However, if a single value needs to be chosen, it is likely that expert judgment-based CES, supported by critical appraisal of the data and biological and statistical considerations, would be the best choice.

6.2 MULTIPLE ENDPOINTS

In vitro and *in vivo* studies often include several endpoints as means to measure and understand the ongoing biological changes of the investigated system. Different endpoints are often biologically related, as steps of a common process, similarly to key events in an AOP (Sand et al., 2018; Simon et al., 2014; Vinken et al., 2017). As such, the understanding of a change in an endpoint or biomarker may provide insight into possible related changes. For example, a 10% decrease in total T_3 levels might indicate thyroid metabolism disruption, especially in connection with altered levels of other thyroid hormones. However, a single and isolated measurement of total T_3 constitutes insufficient information to detect any effect attributed to chemical exposure. It requires additional information regarding the exposure, such as the chemical substance(s), exposure duration and route. Thus, the historical focus on a single, critical endpoint, might have led risk assessors to establish reference values based on the assessment of only a part of the bigger picture. The assessment of multiple endpoints is therefore necessary to gain a better understanding of all ongoing changes in an organism following chemical exposure.

Study III, **IV** and **V** challenge the traditional choice of a single critical endpoint to use as a basis for the conclusions of an assessment. In particular, **study III** highlighted the benefit of the analysis of multiple endpoints in the context of pharmaceutical development. **Study IV** and **V**, on the other hand, analyzed several endpoints which were relevant for the risk assessment of PCBs, addressing the topic and the choice of endpoint on which WHO based their estimation of RPFs for dioxin-like PCBs (Van den Berg et al., 2006).

Study III analyzed 63 endpoints for which data were collected during three independent *in vivo* studies using the same candidate drug. A total of 59 dose-response relationships were observed in at least one of the studies, of which 16 endpoints displayed changes in all three studies (Figure 1 in **study III**). Out of the 63 endpoints analyzed, 15 showed changes in one study only (Figure 1 in **study III**). The changes in the endpoints were clustered in a Spearman correlation matrix

(Figure 6 in **study III**). The importance of modelling all available endpoints, especially in a pharmaceutical development context, was highlighted. The estimated BMDs and BMDLs for several endpoints were lower than the NOAEL/LOAEL of each study and provided information that might have otherwise gone unnoticed. This additional information comes at no extra cost other than time, but the newly gained knowledge might provide important insight into the data analysis. Potential gains include a better understanding of each study which translates into better planning and dose-setting of the subsequent studies. Additional information may be invaluable as it might determine the course of action, especially in the pharmaceutical development process where unpredictable events might be observed during *in vivo* studies.

Similarly, **study IV** demonstrated that analyzing all endpoints might unveil unexpected findings, leading to serendipitous observations and findings which ultimately contribute to the advancement of Science. One such novel finding in this study was that hepatic retinoid levels and concentration changes returned lower BMDLs than other endpoints classically used for PCB risk assessment, such as EROD activity. The lower confidence limits (BMDLs) for a 5% decrease in apolar liver retinoid concentrations were 0.0008 ppm on average, corresponding to a daily dose of 0.06 μg PCB-156 per kg body weight. For a 100% induction of EROD activity, the corresponding BMDL was 0.163 ppm, corresponding to a daily dose of 12.3 μg PCB-156 per kg body weight. **Study IV** showed that PCB-156 can activate the AhR, through a marker for CYP1A1 (the EROD activity), but also that the retinoid system endpoints might be more sensitive to PCB exposure than other endpoints of historical importance. However, the BMDLs and the reference values derived based on them, are highly dependent on the choice of CES. Therefore, the CES should be carefully chosen and grounded on biological and toxicological considerations. In sum, modelling all endpoints might unveil unexpected findings that are later proven useful for chemical risk assessment.

Similarly, **study V** analyzed several endpoints to compare the toxicity of seven PCBs (PCB-28, PCB-77, PCB-105, PCB-118, PCB-128, PCB-153 and PCB-156) to that of PCB-126. It was found that the hepatic retinoid levels and concentration endpoints-based relative potency factors were greater than the latest WHO estimates (Van den Berg et al., 2006) (see section **6.3 Chemical mixtures**). The importance of these findings relates to a potential underestimation of the toxicity of these PCBs for human health. **Study V** challenged the premise that PCBs need to elicit AhR-mediated biochemical responses to establish an RPF (Van den Berg et al., 2006), suggesting that other endpoints might be considered. It was shown that the retinoid system could be sensitive to PCB exposure and it was discussed that changes in the liver's retinoid levels and concentrations could be included as critical endpoints for PCB risk assessment. In conclusion, the historically-

researched AhR-mediated toxicity might only explain part of the picture, which may need to be completed with other pieces of the puzzle which can provide important additional information.

This doctoral thesis demonstrated that the assessment of multiple endpoints when employing the BMD approach might add value to the evaluation being performed, compared to the grounding conclusions on the effects observed on a single endpoint. It might equip the risk or safety assessor with additional information and detect effects and changes that would otherwise go unnoticed. For this reason, it is strongly recommended to perform BMD-modelling in all or at least the majority of the endpoints for which data were collected and to compare the resulting BMDLs and BMDs. A selection of the most important endpoints might then be performed, ideally in rank order of relevance and plausibility for the analyzed effects (Sand et al., 2018). Lastly, applying the BMD approach in an automated or batch manner might be advantageous, using one or two strategies to choose the CES, as the traditional endpoint-by-endpoint analysis is time-consuming. The potential gains could outweigh the time invested, as a greater understanding of the data translates into better decision-making.

6.3 CHEMICAL MIXTURES

We know a little about the effects following exposure to individual chemicals, when their toxicity outweighs all other exposures and effects, but very little regarding exposure outcomes of chemical mixtures. Mixtures are composed of two or more compounds, with a known or unknown composition, and these chemicals may have different potencies depending on the assessed endpoints and effects. For example, a chemical in a small quantity might be relatively more toxic than other more abundant compounds in the same mixture, which might have a lower potency. The BMD approach can improve the assessment of chemical mixtures through, for example, dose-response modelling of the individual compounds or the mixture as a whole (sum of doses or concentrations). However, assessing realistic mixed exposure scenarios might be complex and difficult. Thus, many risk assessors often choose to focus on the effects of individual chemical exposure or to combine the effects of one chemical per exposure route (the aggregate exposure approach). Yet, to understand the contribution of the individual chemicals in a mixture to the outcome, one must be aware that the combined effects might be greater than the sum of the parts. Thus, attention is being given to methods and strategies which may be useful to identify, monitor, regulate and manage chemical mixtures and their potential adverse effects, such as the BMD-modelling approach (EC, 2012; Kienzler et al., 2016).

This section focuses on the assessment of chemical mixtures composed of PFASs (**study I** and **VI**) and PCBs (**study V**) using the BMD-modelling approach. These groups of compounds have been extensively studied, but their hazard characterization is not complete yet. It was hypothesized if the BMD method could contribute to an improved risk assessment of these classes of substances, for which exposure most often occurs as a mixture.

In **study I**, the potency of PFOS and PFHxS was estimated based on equipotent toxicity on a molar basis, i.e. 0.8 for the latter due to its shorter carbon chain length. This simple approach was based on the structural similarity of these two compounds, and the lack of PFAS RPFs published in the scientific literature. This value was somewhat similar to the National Institute for Public Health and the Environment of the Netherlands (RIVM) estimate of 0.3 for PFHxS RPF, which employed BMD-modelling to analyze relative liver weight changes in male rats (Zeilmaker et al., 2018). Nevertheless, further research is necessary to bring forward additional RPFs for PFAS, if possible based on a range of endpoints and adequate CESs.

In **study VI**, no RPFs were calculated for the PFAS-group of substances assessed in case-study 1, as the EFSA assessment (2020) derived a reference value for the mixture and not for the individual compounds. In this case, only the sum of the exposure to individual compounds was important and not the particular proportions in the mixture. This simple but powerful approach might be useful for chemicals for which the toxicity is well-established, such as PFAS, but presents some limitations. It is known that PFAS differ significantly in terms of half-lives and ADME (absorption, distribution, metabolism and excretion), which affects their potency. However, it is acceptable to disregard their potency in order to hasten the process of regulating them as a group, as it is more important to address the hazard represented by the chemical class than the individual components in a mixture (Cousins et al., 2020; Wang et al., 2017).

Relative potency factors stand as an intuitive and robust way, based on the BMD-modelling approach, to compare chemicals with similar structures and modes of action which are likely to result in a comparable exposure outcome (Bosgra et al., 2009). The usefulness of RPFs for chemical risk assessment has been shown for diverse classes of substances such as pesticides, skin sensitizers, bisphenols, PFAS and heavy metals (Bosgra et al., 2009; Kienhuis et al., 2015; Norgren et al., 2022; US EPA, 2003; Wang et al., 2021; Zeilmaker et al., 2018). RPFs can also be used to predict exposure outcomes in comparison to experimental data (Norgren et al., 2022). When calculating RPFs, results are usually expressed in equivalent units in relation to an index chemical, which is usually a well-studied chemical for which exposure effects are known. For PCB risk assessment, the index chemical traditionally used has been TCDD (Van den Berg et al., 2006) but, for **study V**, PCB-126 was chosen, standing as the most well-known and hazardous

PCB for which a study was performed, in similar conditions to the other examined PCBs (Chu et al., 1994, 1995, 1996a, 1996b, 1998, 2001; Lecavalier et al., 1997; **study IV**).

Study V estimated RPFs based for the first time on eight 90-day studies testing seven individual PCB congeners (PCB-28, PCB-77, PCB-105, PCB-118, PCB-128, PCB-153 and PCB-156), and one 28-day study investigating the effects of a PCB-mixture, derived in relation to PCB-126 (the index chemical and the eighth PCB study). Nine endpoints were analyzed to estimate RPFs, of which seven detected dose-dependent relationships for all PCBs studied. An additional analysis was made, separating the datasets by sex, hypothesizing different sensitivities based on earlier reported sexual differences for retinoid levels and CYP1A1 enzymes (Al-Azemi et al., 2009; Penaloza et al., 2014). The potency order based on the averaged RPFs for both sexes and all endpoints displaying a dose-response relationship were: PCB-156 (0.02) > PCB-118 (0.004) > PCB-105 (0.001) > PCB-153 (0.0008) > PCB-77 (0.0006) > PCB-28 (0.0003) ≈ PCB-128 (0.0003) (Table III). The absolute and relative liver weight endpoints generated the highest RPFs, higher than the retinoid changes, and the lowest RPFs were based on the liver EROD activity. The separate analysis by sex suggested that the differences were endpoint-specific, and no clear pattern was observed for all endpoints. Table III shows the RPFs estimated in **study V**, for selected endpoints, compared to WHO's toxic equivalency factors (TEFs) (Van den Berg et al., 2006).

Table III. Comparison between the toxic equivalency factors (TEFs) established by the WHO (Van den Berg et al., 2006) and study V. The relative potency factors (RPFs) estimated in study V were multiplied by a factor of 10, as PCB-126 was used as index chemical and has a TEF of 0.1 compared to TCDD (the index chemical used by WHO).

	PCB-28	**PCB-77**	**PCB-105**	**PCB-118**	**PCB-128**	**PCB-153**	**PCB-156**	**PCB-mixture**
WHO (Van den Berg et al., 2006)	NA	0.0001	0.00003	0.00003	NA	NA	0.00003	NA
Study V, liver retinoid levels	1×10^{-5}	0.01	0.005	1×10^{-5}	0.0004	0.001	0.05	1×10^{-5}
Study V, liver retinoid concentrations	0.0002	0.006	0.005	0.001	0.0005	0.002	0.04	0.02
Study V, averaged RPFs	0.0003	0.0006	0.001	0.004	0.0003	0.0008	0.02	0.01

Study V has shown that the retinoid system's endpoints are sensitive to PCB exposure and can detect adverse effects at low doses. It was therefore deemed adequate that RPFs and endpoints of the retinoid system be considered for future risk assessment of PCBs, for which exposure occurs as a mixture. The findings in rats are translatable to humans, as the retinoid system is preserved.

Additionally, all of the estimated RPFs, using the BMD-modelling approach, were greater than the latest TEFs (the RPF-equivalent term for dioxin-like compounds) established by the WHO, suggesting an underestimation of the risk posed by these substances. However, this approach assumes additivity, while synergism could represent a greater hazard, which cannot be disregarded. Thus, an update of the WHO TEFs values was suggested, which would also include the research and literature published after 2006.

In conclusion, the BMD method was appropriate to assess chemicals for which exposure occurs as a mixture, namely PFAS and PCBs. It was shown that not only can the BMD approach model the individual chemicals separately, but it can also combine several compounds in a single dose-response analysis, deriving RPFs to compare the potency of the different constituents.

6.4 PHARMACEUTICAL DEVELOPMENT, SAFETY ASSESSMENT AND DOSE-RESPONSE MODELLING

The pharmaceutical development of new drugs is a process that is divided into well-established steps, starting with drug discovery until commercialization and post-market surveillance. The pre-clinical phase is intended to study the candidate drug's safety, toxicity, pharmacokinetics, and ADME (absorption, distribution, metabolism and excretion), before the first-in-human clinical trial (Butler et al., 2017). For the development of small molecules, including anti-oncogenic drugs, which follow the International Conference on Harmonization guidelines, studies on two species are generally expected, most often rat and dog (Butler et al., 2017; ICH, 2009a, 2009b; Prior et al., 2020). The BMD approach is not systemically used in this process, although its 3R potential and ethical focus have been demonstrated, for example in terms of dose groups of unequal size (Ringblom, 2016; Ringblom et al., 2017; Slob, 2014a, 2014b). There is, however, room for improvement in terms of usage of clinical observations and multiple endpoints, for example. Clinical observations and body weight loss are often used to assess the animal's general condition and set the humane endpoint (OECD, 2000). JECFA suggested that relevant clinical observations displaying a dose-response relationship may be used as critical endpoints in order to establish a NOAEL (WHO FAO, 2016). However, they are rarely used to this end or as informative endpoints of toxicity, and have never been employed for predictive modelling.

The findings of **study II** show that the multivariate data analysis regression method PLS (Wold, 1975) could be used to predict organ injuries based on clinical signs and a 5% body weight loss. In fact, predictions with a balanced accuracy over 80% were described for pathological findings in the thymus, testes, epididymides and bone marrow. The clinical signs with the highest predictive power were piloerection, eyes half shut and slightly decreased motor activity, and 5%

body weight loss. Furthermore, merging data from all three subsequent studies resulted in better predictions than chronologically ordered datasets, i.e. predicting study 3 based on study 1 and 2. These findings suggest that the safety assessment and preclinical studies during pharmaceutical development may be improved, in terms of Refinement and Reduction of animal use, if predictive modelling strategies are employed. Predictions should be based on a 5% body weight loss and clinical signs associated with pathological findings in the organs of interest. Altogether, these improvements would strengthen the evidence on which decision-makers base their judgment for toxicity assessment, study design, dose-setting and candidate drug selection.

Study III applied the BMD-modelling approach to three *in vivo* studies performed for safety assessment purposes and demonstrated how BMD-modelling might be advantageous for the pharmaceutical development process. All endpoints for which data were collected were analyzed, using a CES of 5% of GTES-based. The results revealed that 59 out of the 63 endpoints analyzed displayed a dose-response relationship in at least one of the three sequential studies, and that these endpoints were often positively or negatively associated (Figure 6 in **study III**). Additionally, the estimated BMDLs and BMDs for the pathological findings in the rat organs were often below the studies NOAEL or LOAEL. Another relevant finding in **study III** was that clinical signs could be modelled using the BMD approach, after normalization to the control group. Lastly, the BMD method helped describe differences in sensitivity between different sexes. In sum, **study III** suggests a) using BMD-modelling to model safety assessment study results and b) modelling all endpoints for which data were collected. A more complete picture of the data package could thus be obtained, largely benefiting the team assessing the candidate drug – more information can be extracted from the data, contributing to a better understanding of the observations and to the outcome of the drug safety evaluation.

Together, the findings of **study II and III** demonstrate that there is room for improvement in the pharmaceutical development field. **Study II** shows how employing modelling approaches from other fields may be useful for obtaining new insights into the data. **Study III**, the first article employing the BMD-modelling approach for safety assessment for pharmaceutical development purposes, could be groundbreaking for the field. The pharmaceutical development field may benefit significantly from the employment of the BMD and PLS-modelling approaches and analysis of all endpoints for which data were collected, to obtain a better understanding of associated effects and the biological processes affected by the candidate drug. These modelling strategies might lead not only to Refinement of the *in vivo* studies but ultimately to a future Reduction, if fewer experiments need to be performed owing to the improved quality of early studies.

6.5 PFAS

PFAS stand as one of the chemical regulation challenges of the century. Their ubiquity, persistence and bioaccumulation would not be as problematic if they were inert and not biologically active as initially believed. Apart from their carcinogenic potential, PFAS affect the immune system and disrupt the endocrine, lipid and thyroid metabolism, among other disruptive effects (ATDSR, 2021; Boesen et al., 2020; Coperchini et al., 2020; EFSA, 2020; Li et al., 2021b; Steenland and Winquist, 2021). The most relevant human exposure sources to PFAS are drinking water, food, dust and consumer products, including cosmetics (Sznajder-Katarzyńska et al., 2021). However, contaminated drinking water may outweigh other exposure sources (De Silva et al., 2021; Gyllenhammar et al., 2019).

In 2013, a population living in Ronneby (Blekinge county, Sweden) was discovered to be accidentally exposed to high levels of PFAS, mainly PFOS and PFHxS, via contaminated drinking water (Li et al., 2017; Xu et al., 2021). Women in this population have been described to have an increased risk for skin melanoma, kidney and bladder cancer, as well as thyroid hormone disruption (Li et al., 2021b). **Study I** performed an integrated probabilistic risk assessment, an analysis that combines the individual serum PFOS and PFHxS concentrations with animal data that represent the putative exposure outcome. Earlier research on Cynomolgus monkeys (*Macaca fascicularis*) demonstrated decreased total and free T_3 levels following daily exposure to PFOS for six months (Seacat et al., 2002). The results show that the Ronneby inhabitants with internal PFOS and PFHxS serum levels associated with a 10% decrease in total T_3 were mainly women, and that longer exposure increased the risk of adverse outcome. In fact, the median probabilities of critical exposure were 0.08% (C.I. 0.02% – 2.9%) and 3.9% (C.I. 0.8% – 21.6%) for men and women, respectively, who lived in Ronneby for ≥ 1 year (n = 1 845). For those residents that lived for ≥ 29 years (n = 506), the probability of critical exposure was 0.13% (C.I. 0.03% – 5.8%) and 6.2% (C.I. 1.2% – 34.7%) for men and women, respectively. PFOS alone accounted for half of these estimates. Additionally, the analysis of the sources of uncertainty demonstrated that duration was the biggest contributor (60.8%), followed by intraspecies (17.8%) and interspecies (11.4%) toxicodynamics. As was shown in **study I**, IPRA stands as a higher tier approach that can deliver more precise estimates of risk for adverse effects. In comparison, the Swedish National Food Agency (Livsmedelsverket) estimated that 4 000 ng PFOS/L in the drinking water of the Ronneby residents accounted for 76% of TDI (based on a TDI of 150 ng PFOS kg/b.w./day, according to the EFSA assessment in 2008) (EFSA, 2008; Livsmedelsverket, 2014). The number of residents likely to present adverse effects was difficult to appreciate, and IPRA presents the potential for improving risk assessment and supporting decision-making. However, it is time-consuming and requires some degree of expertise in its application. In order to facilitate probabilistic analysis,

the WHO IPCS released an Excel-based approximate probabilistic hazard characterization tool – the latest version, 1.00_v0.13, is made available by the developers at www.researchgate.net/publication/334164164_APROBA_PLUS-V100_v013_TEMPLATE (Bokkers et al., 2017). It eases the analysis by providing default parameter values, but the user must insert the reported exposure and expert opinion-based limits. In sum, both IPRA and APROBA stand as useful alternatives to deterministic exposure estimates, when data on the exposure outcome are missing or are difficult to obtain (Bokkers et al., 2017; Chiu et al., 2018a; Voet and Slob, 2007).

However, not only PFOS and PFHxS represent a threat to human health. Other PFAS substances, such as PFOA and PFNA, have been shown to have immunotoxic, thyroid and lipid metabolism disruptive potential, among other effects (EFSA, 2020). Based on similar accumulation properties, toxicokinetic profiles and long half-lives, EFSA has decided to group the four PFAS specimens that currently represent the bulk of human burden: PFOA, PFNA, PFHxS and PFOS. **Study VI** case-studies 1 to 3 offer a historical overview of PFAS reference values, showing that the TDIs established by EFSA decreased from 150 ng PFOS/kg b.w./day (EFSA, 2008) to 0.63 ng of PFOS, PFOA, PFHxS and PFNA/kg b.w./day (EFSA, 2020). In case-study 3, it was shown that BMD-modelling could have estimated a TDI of 17.7 ng PFOS/kg b.w./day, based on the exponential model-BMDL and a CES of 5% for HDL decrease, departing from the same critical study that EFSA chose in 2008. Case-studies 1 and 2 have also demonstrated the tight relationship between the choice of CES and the resulting reference value, and how this interplay can affect the outcome of risk assessment – a hazard might be mischaracterized if the CES is not appropriate.

In conclusion, **study I** and **VI** brought additional evidence on the potential effects of PFAS, demonstrating how published data could be remodelled, which might lead to new findings and conclusions, impacting and updating the risk assessment of these substances. On the other hand, PFAS risk management measures have been mainly reactive, for example through remediation actions which can be costly and slow (Franke et al., 2019; Mudumbi et al., 2017). However, risk managers might be proactive, contributing to a severe exposure decrease if a group classification and phase-out are performed, but these processes are both slow and costly (Cousins et al., 2019; Kwiatkowski et al., 2020). Furthermore, the concept of essential use has illustrated that these chemicals should be restricted to absolutely necessary uses, when no safer alternatives are available (Cousins et al., 2019). Truly green alternatives should be fluorine-, chlorine- and bromine-free, as these elements can lead to bioaccumulation. A chemically safe world entails substances that fulfill their intended use and degrade naturally outside of it, and are not toxic to human and environmental health.

6.6 PCBs

PCBs are persistent substances that are toxic for human health, standing as another serious chemical challenge faced by humankind that ultimately resulted in a global ban in 2004 following their inclusion in the Stockholm Convention. The similar physiochemical and toxic properties of PCBs chemicals warranted a group class classification, standing now under Annex A (Elimination) of the Stockholm Convention on Persistent Organic Pollutants (UNEP, 2021). This class elimination is time- and resource-saving, hastening the process of their phase-out. However, despite the efforts made towards their elimination, the long half-lives and persistence of these chemicals explain why they can still be detected today in the chemical body burden of humans (Bignert et al., 2020; Clair et al., 2018; Raffetti et al., 2018; Ritter et al., 2011). Moreover, research has not elucidated all possible effects associated with PCB exposure (ATSDR, 2000; IARC, 2015) – in particular, the role of the retinoid system and its disruption by PCB exposure (Grignard et al., 2020).

Study IV demonstrated some of the adverse effects following a daily PCB-156 dietary exposure in a subchronic study in rats. A dose-dependency relationship was found for decreased final body weight, body weight gain and thymus weight, as well as for decreased apolar retinoid levels and concentrations in the liver and lungs and increased in the kidneys. The BMDL for a 5% decrease in apolar liver retinoid levels and concentrations were 0.008 and 0.0008 pm, respectively. The latter was about 13 times lower than the study LOAEL, established at the lowest dose tested (0.01 ppm). Additionally, dose-dependent increases in the hepatic enzymatic activities of EROD, pentoxyresorufin-O-dealkylase (PROD) and uridine 5'-diphospho-glucuronosyl transferase (UDPGT) were observed, as well as body and organ weight changes, among other effects. EROD activity is a marker for CYP1A1 and retinoic acid synthesis, while PROD is for CYP2B and retinoic acid degradation, and UDPGT for its elimination. Table IV summarizes the BMDLs estimates for a selection of the endpoints of **study IV**.

Table IV. Benchmark dose lower limits (BMDLs) for selected endpoints, following a 90-day study in which male and female rats were exposed to PCB-156 via diet.

Endpoint	CES	BMDL for males (ppm)	BMDL for females (ppm)
Body weight			
Body weight gain (week 1−13), g	− 5%	0.2	0.1
Final body weight (week 13), g	− 5%	0.4	0.4
Organ weights (total and relative)			
Liver, g	+ 5%	0.7	0.7
Liver, %	+ 5%	1.8	0.2
Kidney, g		NDR	NDR
Kidney, %	+ 5%	0.001	0.001
Lung, g		NDR	NDR
Lung, %	+ 5%	0.04	0.6
Thymus, g	− 5%	0.03	0.02
Thymus, %	− 5%	0.02	0.01
Liver enzyme activities, nmol/mg protein/min.			
EROD, ethoxyresorufin-O-deethylase	+ 100%	0.1	0.1
PROD, pentoxyresorufin-O-dealkylase	+ 100%	1.8	0.3
UDPGT, uridine 5'-diphospho-glucuronosyl transferase	+ 100%	0.6	0.2
Uroporphyrin, mg/g of liver	+ 5%	0.04	0.06
Apolar retinoids[a]			
Liver concentration, µg/g	− 5%	0.0009	0.0007
Total amount in the liver, mg	− 5%	0.008	0.008
Kidney concentration, µg/g	+ 5%	0.2	0.2
Total amount in the kidneys, g	+ 5%	0.1	0.2
Lung concentration, µg/g	− 5%	0.00009	0.000003
Total lung amount, µg	− 5%	0.00002	0.0002

NDR = No dose-response detected. [a]Apolar retinoid levels measured as free retinol and retinyl esters combined.

In conclusion, PCB-156 was found to be highly toxic in rats, and changes were associated with exposure at very low doses. The apolar retinoids changed at significantly lower doses than the lowest tested dose, migrating from the liver and lungs to the kidneys. Additionally, the retinoic acid synthesis, elimination and degradation markers were elevated, indicating that the retinoid system was disrupted and could therefore be considered sensitive to PCB-156 exposure. In summary, **study IV** illustrated the disruptive effects of PCB-156 in the retinoid system of the rat, which have not been reported before. This system is highly preserved and the findings are likely to be translatable to humans.

Case-study 4 in **study VI** examined the same raw data related to PCB-156 exposure as for **study IV**, but applied the BMD method to examine the changes in the retinoid system in the rat from a different perspective – how can the different strategies to set the CES affect the estimated reference values, and how does it impact the assessment outcome? The reference values were slightly higher when using individual data compared to the use of summary data (mean, SD and *n* number of animals). Male and female rats had different CESs, based on the 1 SD approach, 19.6% and 14.1%, respectively. Based on the results of this case-study, it was hypothesized if different CESs should be considered for different sexes or if one CES can be considered adequate for both sexes, to derive protective reference values for the human population, which is known for its heterogeneity. The derived reference values ranged between 0.05 and 3.24 µg PCB-156/kg b.w./day, depending on how the CES is chosen and which kind of data (individual or summary) is modelled. In conclusion, it was shown that the PCB-156 study data were appropriate not only to assess the potential retinoid disruption in the rat but also to illustrate the tight relationship between the choice of CES and the subsequent reference values obtained (which was discussed in section **6.1 The choice of Critical Effect Size (CES)**).

In summary, the BMD applied in **study IV** and **VI** brought additional evidence to the toxic effects of PCBs and highlighted the sex differences and the importance of including the retinoid system in the risk assessment of these chemicals. Although the retinoid system has been overlooked for a long time, these novel findings suggest that this system is sensitive to PCB exposure, advocating for the employment of the BMD-modelling approach on multiple endpoints to support and complement the risk assessment of these chemicals.

7 CONCLUSIONS

This thesis demonstrated how the BMD approach can be employed to assess the potential effects of chemical exposure, having focused on three main challenges where knowledge could be improved: the choice of CES, analysis of multiple endpoints and chemical mixture assessment. The use of the BMD approach would most likely be enhanced if these challenges were addressed in the guidance documents. Furthermore, it was shown that the BMD-modelling approach features flexibility and several methodological advantages that make it stand as a serious replacement to the more-established NOAEL approach, much like the evolution from the bicycle to a motorbike – the principle is the same, balancing yourself in motion between two wheels while holding onto a handlebar – but the instrument is more advanced, allowing you to go farther, even if at the cost of more resources. The added consumption of resources is acceptable given the better output, i.e. greater quality of results and derived reference values.

Additionally, this thesis showed the possible applications of the BMD approach in the chemical risk assessment and pharmaceutical development areas, through studies evaluating an array of relevant substances such as PFAS, PCBs and an anti-oncogenic candidate drug. Various challenges faced by the BMD approach were examined, namely the assessment of chemical mixtures, multiple endpoints and the choice of CES.

For the assessment of chemical mixtures, **study I** and **V** showed that the BMD method can support the weighing of the different components and for calculating relative potency factors. Employing the BMD approach is advantageous in cases of exposure to chemical mixtures, to characterize the contribution to the outcome of the different chemicals and to detect additive or synergistic effects (see also the paper by Norgren and colleagues, 2022).

The assessment of multiple endpoints is challenging but might be eased by the use of the BMD and PLS approaches. **Study II** showed that information from several endpoints, including clinical signs, can be combined to predict the exposure outcomes. **Study III** and **IV** challenged the established practice within the BMD field of choosing a single, critical endpoint. In turn, modelling all endpoints might unveil unexpected findings that could change the outcome of the study assessment, as a greater understanding of all changes and their meaning is achieved.

Regarding the choice of the CES, different strategies were compared and their results were examined in **study VI**. It was concluded that, despite the usefulness of other strategies to set the CES, expert judgment will continue to be the centerpiece of chemical risk assessment. Additionally, the BMD approach provides great insight into the data package, supporting the decision-making process and calculation of protective reference values.

Lastly, it was shown that the BMD approach has a strong 3R potential in terms of Refinement and Reduction, extracting considerable information from the data. Undoubtedly, the BMD approach is in chemical risk assessment to stay – much like the Swiss army knife – a useful and multi-purpose tool that allows one to derive reference value estimates of superior quality.

8 FUTURE PERSPECTIVES

The BMD approach has come a long way since the founding paper by Dr. Kenny Crump in 1984. Its development has been continuous and steady, benefiting from the increasing computational power made available. However, there are areas where the use of the BMD dose-response modelling approach may still grow, for example, in the cosmetic and pharmaceutical industries, owing to its applicability to different types of data and its ability to estimate RPFs for structurally similar compounds. Other types of data for which the BMD-modelling approach is potentially useful are epidemiological, nutrition, event-time and pest management data (Budtz-Jørgensen et al., 2001; Jensen et al., 2021a, 2021b; Whitney and Ryan, 2013).

From a methodological perspective, the addition of a fourth variable in the BMD-analysis would be beneficial. For example, when using dose as the independent variable, effect as the dependent variable and the different PCB congeners as the covariate, it would be helpful to simultaneously analyze the data using a fourth variable such as the sex of animals. The result would then be BMDs and their associated BMDLs and BMDUs for PCB-126-females, PCB-126-males and for both sexes analyzed together, and the same for PCB-156 data, when performing a simultaneous analysis. Sometimes, a dose-response relationship might be driven by the most sensitive sex, or the magnitude of the effects may be masked by a less sensitive group, potentially decreasing some of the association's strength in a joint analysis. Additionally, in a simultaneous analysis of several compounds, it is possible not to detect a dose-response relationship for a single chemical if no discrimination is made between males and females (the problem of the lack of a fourth dimension), which would otherwise only be found in a new analysis when the dataset is split by sex. This can be time-consuming or go unnoticed by the risk assessor, and therefore the inclusion of a fourth dimension (a second covariate) in BMD-modelling is desirable.

Additionally, EFSA conducted a public consultation (PC-0135) for its *Guidance on the use of benchmark dose approach in risk assessment* (EFSA, 2022). The guidance document is expected to be published in final form in late 2022 or early 2023. The most important remarks are a change from the frequentist to the Bayesian paradigm, the model averaging being reiterated as the method to calculate the confidence interval around the BMD (using 1000 bootstraps), an expansion of the suite of models to fit the data, and an appreciation of the uncertainty around the BMD estimated by BMD/BMDL and BMDU/BMDL ratios (EFSA, 2022). No major changes were made regarding the choice of CES and the role of expert judgment. However, it does not clearly endorse the RPF and IPRA approaches, which would be desirable. Although the new BMD guidance does not call for a reevaluation of earlier assessments based on the previous document, it is likely to shape how national and international authorities and

organizations perform risk assessment of chemical substances and mixtures. It consists, however, of one step in the right direction, as the guidance was reviewed to align better with the principles for dose-response assessment published by the WHO IPCS (2020).

From a practical point-of-view, the PROAST package for R needs an automatic saving function for the plots and console output obtained during a BMD-analysis, especially when several endpoints are modelled sequentially. It could also benefit from an automatic naming function, whilst saving the files in the working direction. Currently, it is time-consuming to perform an analysis of several endpoints, saving all the files and plots manually. Therefore, a batch analysis function is desirable to improve the user-friendliness of this R-package. Moreover, the latest version of the BMD-modelling R-package PROAST, version 70.3 (as of April 2022) (National Institute for Public Health and the Environment, RIVM, Bilthoven, Netherlands), does not integrate a calculation or suggestion of 1 SD from the control group as the CES, nor other approaches to select the CES other than percentages. Lastly, the latest version also performs model averaging on endpoints for which the null model, i.e. no dose-response relationship, was deemed the best fit. It is counter-intuitive and may be difficult to interpret by less experienced users.

Although EFSA and RIVM have progressed significantly in terms of public availability and user-friendliness of their online BMD platforms, the knowledge threshold for usability is still reasonably high. Education is key to promoting high-quality analysis, as the BMD analysis might significantly impact the derivation of reference values. The user interface could be simplified and better help could be provided, so that the user is aware of the (right) choices and possibilities for the specific dataset being analyzed. Furthermore, more advanced features should be available for the online version of the PROAST package (https://proastweb.rivm.nl/), such as the menu version (with its additional features) and calculation of RPFs (unavailable in version 70.1, as of April 2022). However, the web version of PROAST is more user-friendly than the US EPA Excel-based BMDS and Jensen and colleagues (2020) *bmd* package for R software environment, both lacking online versions.

Additionally, greater dissemination of the BMD approach is necessary. Teach BMD to all scientists, regulators, risk assessors and Toxicology students! Show them how useful and dose-response relationship can estimate the doses leading to the effect or change in response that they are assessing! Show them how flexible and valuable this approach is! Education and communication of the BMD-modelling approach are lagging behind its development – it is very helpful, but it is still not used on a daily basis by risk assessors, and this needs to change for safer future risk assessments. Dr. Kan Shao, from Indiana University, regularly hosts live and

free webinars that educate the (potential) users about current and coming features of the Bayesian BMD (BBMD) web-based platform (Shao and Shapiro, 2018). These events are attended by hundreds of participants, answers are given to doubts which might persist and to gather user feedback for further development of the platform. It is a great example of interaction between the users (of all backgrounds and knowledge levels) and the developers, paving the way forward for the development of the platform, based on education and user feedback.

Regarding the chemical risk assessment of POPs such as PFAS, the time from risk assessment to decision-making and action needs to be shortened. To this end, machine learning methods may be used, to predict toxicological properties based on physiochemical descriptors, and assist the data analysis and assessment performed by the risk assessors. Additionally, "omics" methods might be employed in a regulatory (and harmonized) way, in order to avoid regrettable substitution and to support the conclusions drawn by the risk assessors (Martins et al., 2019). Cases of regrettable substitution, when a chemical is replaced with another just as hazardous (or potentially worse), have been observed, and a safer chemical use could be promoted by concepts such as essential use and classification of PFAS as a group (Cousins et al., 2019; Kwiatkowski et al., 2020). PFAS will ultimately be phased out, but safer alternatives need to be identified and implemented in the meantime.

Pharmaceutical development is an ever-evolving field that benefits from being receptive to new and advantageous approaches and methods. The contribution of the BMD approach to preclinical and safety assessment studies was elucidated in this thesis and further research in this context is required. Possibilities include improved study design and unequal group sizes, more frequent use of BMD for analysis of *in vitro* and genomics data and to monitor animal welfare during *in vivo* studies (through real-time modelling of clinical signs and body weight loss). Additionally, retrospective analysis of *in vivo* and *in vitro* studies in pharmaceutical and toxicological databases, such as the US National Toxicology Program (NTP) (https://ntp.niehs.nih.gov/whatwestudy/testpgm/index.html), might provide a) a sufficiently extensive training and testing set for developing automated BMD batch and multiple endpoints analysis and b) insight into the animal-to-human dose extrapolation and limit value setting, which are still challenging.

Regarding the potential 3R gains and improvement of *in vivo* studies, Science and the regulatory toxicology field would benefit from the preregistration of animal studies, in platforms such as https://preclinicaltrials.eu/ and https://www.animalstudyregistry.org/, the latter managed by the German Centre for the Protection of Laboratory Animals (Bf3R) at the German Federal Institute for Risk Assessment (BfR). Preregistration of the study design, methodology and statistics

promotes transparency and reproducibility, potentially reducing duplicate/parallel studies, biases and misconducts such as selective outcome reporting and HARKing (hypothesis after results). Additionally, it would benefit actors conducting *in vivo* research by better study planning, feedback from peers and eventual data sharing.

For the hazard identification of chemicals, one could employ text-mining techniques to, for example, ease the process of analysis of the literature. Text-mining has successfully been used for the risk assessment of PCBs (Ali et al., 2016) and polycyclic aromatic hydrocarbons (PAHs) (Ali et al., 2021), among other substances, but not for PFAS. One easily accessible tool is the Cancer Risk Assessment using the CRAB3 tool (available at https://crab3.lionproject.net/). CRAB3 is a user-friendly tool that divides the information retrieved in scientific evidence, mode of action and toxicokinetics, and uses machine learning algorithms to colour code the text into background, objectives, methods, results and conclusion section.

Furthermore, the regulatory toxicology field and risk assessment process are evolving continuously and could draw advantage of new techniques and methods that could address the challenges these areas face. The findings made employing the PLS models require further studies to investigate if they can be used for pharmaceutical development and risk assessment purposes. Additionally, other computational toxicology and machine learning methods may be used to obtain a deeper understanding of the data, using both supervised or unsupervised methods. Unsupervised machine learning methods can detect unforeseen patterns in the data, but are less reliable and should therefore be used with caution. Supervised machine learning methods, such as PLS and random forest trees, examine a pre-labelled dataset with an *a priori* hypothesis being tested, offering the possibility to be used predictively (Eriksson et al., 2013; Smith, 2018). Other methods, such as the principal component analysis (PCA) and Bayesian kernel machine regression (BKMR) method, could address recurrent problems in Toxicology, such as collinearity quantification, e.g. addressing mixtures and weighing the contribution of individual chemicals (Bobb et al., 2015; Chiu et al., 2018b). Validation and, most importantly, method acceptance are slow processes, and the recommendation of their use by international chemical regulation guidelines can take decades to be included (Brandon et al., 2013; Muri et al., 2009). However, these new computational methods and approaches could represent improvements with a great 3R potential, especially regarding the Reduction possibilities.

Dr. John Tukey wrote in 1986 that "*The data may not contain the answer. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data. The data may not even contain an appearance of an answer*" (Tukey, 1986). In Toxicology, it reminds us that not all datasets are informative and that special

care should be put into planning before the experiments/studies are carried out. Additional data may be required if conclusions are of little confidence or in the light of new data. This ties into Parens and colleagues' words that "*Absence of proof of toxicity is surely not proof of the absence of toxicity. Given this framing, it should not be surprising that many cases where toxicological tests do not find toxicity result in the approval of substances that are later found to be toxic.*" (Parens et al., 2017). No statistical testing can replace scientific reasoning, but dose-response modelling approaches such as the BMD method can support the data analysis and decision-making process. However, caution and critical thinking are always required when evaluating modelling results. Mathematical models are blind to the reality they are modelled to and a dose-response relationship or association between two variables does not always imply a correlation. Additionally, the BMD is philosophically different from the NOAEL approach, estimating a more accurate reference value rather than a surrogate, communicating uncertainty through a confidence interval. The BMD approach is a tool that promotes the estimation of reference values of greater quality, but expert judgment is required and will continue to constitute an irreplaceable centerpiece in Toxicology.

9 ACKNOWLEDGMENTS

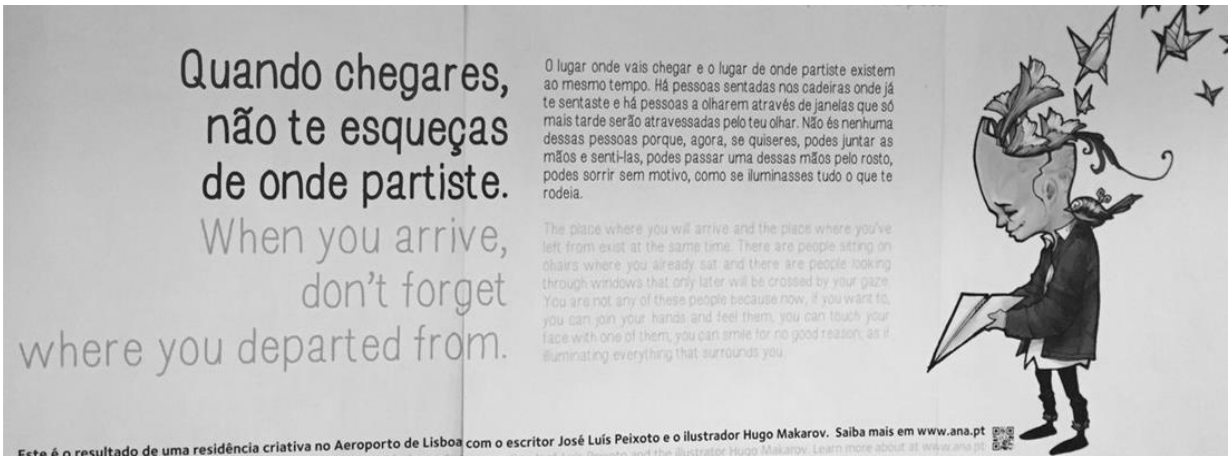

Figure 8. Art residency in Lisbon airport, featuring text by José Luís Peixoto (a great Portuguese writer) and illustration by Hugo Makarov. Photo: Antero Vieira da Silva.

This image, painted on a wall of Lisbon airport, epitomizes the journey behind this thesis. While not having forgotten my point-of-departure, I am indebted to a great number of people who have supported me or shared good moments throughout the years, making the PhD journey much more pleasurable. Being impossible to mention everyone, I would like to summarize by thanking the key people.

First of all, **Mattias Öberg**, the man without whom none of this would have been possible. More than the main supervisor, you were also a friend and your energy and enthusiasm will never be forgotten. A true *handledare*, Swedish word for supervisor but etymologically meaning "leading by the hand", precisely what happened as I took my first steps doing research independently. You didn't carry me, you led me where I am today, in scientific and academic terms, and thanks to you I have had the chance to grow personally and professionally. Thank you for believing in my potential since I was "just another" Toxicology Master student. Our years together were very good, thanks to the wisdom and supervision you have provided, but also to the independence that I have truly appreciated. It let me grow while you were engaged in contributions to working groups that I so many times lost count of, such as IMM's RUT, the EUROTOX Executive Committee, the Swedish Society of Toxicology ERT Committee, the Riskzonen podcast and the Toxicolour blog, to name a few.

Elin Törnqvist, a formidable woman who leads by example. The energy, dedication and tirelessness you have always shown to your projects are genuinely inspiring. The challenges that you have faced and overcome over the years are undoubtedly greater than the average. Despite the obstacles in your way, you have never looked sad, down or tired for a single day. All of those who have worked with you will never forget the mark you leave, and that is regardless of the many

tasks that you have successfully juggled over the years. I believe that the 3R-area would become poorer if you choose to leave it, but the Swedish National Veterinary Institute is richer with your contribution.

Annika Hanberg, a strong woman whose multitasking skills I admire. Wise and disciplined, her contribution to the Toxicology field is immense, both through the education of future toxicologists and in the risk assessment area. Her professional and personal contributions have made her be awarded the Mäster 2019 by the Karolinska Institutet's student union, and that speaks for itself – how her contribution goes beyond the borders of the Master's program. All alumni from the Toxicology Master program around the globe can vouch for her kindness, knowledge and intelligence, which leaves a mark on everyone's heart.

Anton Pohanka, a role model of a mentor. I am happy about all the meetings and fruitful discussions we have had during the past years. The Pharmacology Laboratory at the Karolinska University Hospital is fortunate to have you as their head! Your dedication to work, patients and greater causes are truly inspiring.

Joakim Ringblom, a mathematical mind and a wise young man. Your knowledge about BMD is vast and I thank you for everything that you have taught me throughout the years we have worked together. You were instrumental in laying the ground for my PhD and I remember asking a lot of silly questions, but you never grew tired of teaching me, telling me that there are no dumb questions and explaining as many times as necessary.

Helen Håkansson, for the passion and dedication that you have shown. Beyond your contribution to the PCB and retinoid system areas, and other substances that I am not aware of, you personify "enjoy what you do so you don't have to work another day in your life".

Ih Chu, although we have never met personally, I have great respect for the work you have done, coordinating the planning, performance and reporting of many studies on PCBs and other toxicants, which laid the ground for **study IV** and **V**. Thank you!

To my fellow co-authors, **Vesna Munic Kos**, **Kristina Jakobsson**, **Christian Lindh**, **Elin Liiv**, **Mark Feeley** and **Peter Moldeus**, for everything that I have learned while working with you. Big thanks too to **Ulf Norinder**, for the fun discussions and meetings!

As we all know, for great Science great leadership and management is required. For that, we can count on **Lena Palmberg**, a fantastic head of unit and the departmental director of doctoral studies, all of that aside from her own research. Not only is Lena Palmberg a great woman in words, but also in actions – careful, thoughtful and attentive, she makes sure that each person around her is looked after, no exceptions made.

Speaking of Unit and Unity, I would like to thank all of the past and present members of the Unit of Integrative Toxicology, including but not limited to **Andrea Cediel**, **Anneli Julander**, **Ann-Katrin Sjödén**, **Aishwarya (Ashu) Mishra**, **Ashesh Chakraborty**, **Axel Nordström**, **Bengt Sjögren**, **Carola Lidén**, **Carolina Vogs Elin Engdahl**, **Elina Drakvik**, **Emma Wincent**, **Gunnar Johanson**, **Jie Ji**, **Joëlle Rüegg**, **Johnny Lorentzen**, **Jolinde Kettelarij**, **Klara Midander**, **Lena Ernstgård**, **Libe Vilela**, **Linda Schenk**, **Michela Di Criscio**, **Mizanur Rahman**, **Polina Lizunkova**, **Pooria Sabouri & Zahra Dehghani**, **Selina Rinaldi**, **Shanzina Sompa**, **Swapna Upadhyay**, **Tshepo Moto**, **Urban Svedberg** and **Tania Thimraj**. It was fun and enriching to work with you! **Marialuisa Baldi** and **Marcela Quirós Lépiz**, for the good energy you brought while sharing the office – it felt emptier in the room after you left. A colleague of many years is **Koustav Ganguly**, who has shared a realistic and unromantic version of life as a researcher. Without being discouraging, your insightful comments have always given me a lot of food for thought!

Exciting was also the arrival of **Lars Brunken**, to who I like to call my brother-in-science. It brought a whole new light to my life at the office. Our friendship dates back from Swetox times, in 2016, and you have returned to Stockholm as a fellow PhD-student. I am amazed by your kindness as a person, and I hope you know that I have great hopes for you! I am sure we will continue to be friends and orbiting close to each other in the future.

There are also cool people outside the Integrative Toxicology Unit – **Ahmed Sabah**, **Anda Gliga**, **Anna Zettergren**, **Emelie Helte**, **Klara Gustin Mossegård**, **Paolo Frumento**, **Sarah McCarrick** and **Tessa Schillemans**, to name a few. Thank you for everything!

At the beginning of this PhD, there was a fantastic hub called Swetox – the Swedish Toxicology Sciences Research Center. It takes a visionary man to plan and manage such a complex organization, a unique cooperation between 11 Swedish universities, gathered around the Toxicology discipline. That man was Professor Emeritus **Åke Bergman**, whose observations are always clever. I would also like to thank Swetox's previous workers **Anne Vonk**, **Astrud Tuck**, **Erica Selin**, **Ernesto Alfaro-Moreno**, **Heike Hellmond**, **Ian Cotgrave**, **Jonas Förare**, **Maritha Marcusson-Ståhl**, **Pauliina Damdimopoulou** and **Richelle Duque Björvang**, among other colleagues who contributed to a unique working place. Swetox was something special! Additionally, the people at the General and Inhalation Toxicology at Swetox who were instrumental in the performance with great quality of **study II** and **III**: **Björn Platzack**, **Camilla Bengtsson** and **Marie Eriksson**. **Charlotte (Lotta) Nilsson**, for the great humor and energy, amazing working ability, and welcoming me into the Swedish Toxicology Society (SFT) board, which was a very fun trip! **Johan Lindberg** and **Marcus Sjödin**, for the fun months working with the pharmaceuticals in wastewater analysis!

The people I have worked with while at the Swedish Toxicology Society (SFT) board were very inspiring, including **Janne Olsson**, **Petra Thulin** and **Hanna Karlsson**, among many others. Your contribution to Toxicology in Sweden was tireless and I have learned a lot about ideal work.

But not everything was about work. There was A LOT of running, about 13 000 km between May 2018 and May 2022. Many, many hours limping out there, from beautiful sunny days with -10°C to others with 28°C! Running, the simplest and most natural physical activity, offered simultaneously great mental relaxation and a physical challenge. A lot of good ideas and thoughts came to me during solitary runs. Additionally, I must address a special thanks to Stockholm's running community, especially those friends and buddies with whom I have trained and participated in competitions throughout the years, for all of those smiles, high-fives and fist bumps. I have made friends and carry many people in my heart, many from the **Hässelby SK** and **Fredrikshof FIF** running clubs, but also outside of these clubs and abroad. I have witnessed the sacrifices people make for the joy that running brings to their life. It is difficult not to name an endless list of people, but among the key individuals, **Conny Sundh** and **Fredrik Uhrbom** stand out, the coaches that I have had and who keep things going, sacrificing a lot of their limited lifetime to others. To **Ricardo Queirós** and **Toni Gomes**, my first ever running mates, for introducing me to the world of half-marathons and endurance running. Not only but most importantly, I have learned and experienced the following, running with all of these people throughout the years:

- That endurance is about the small steps we gave achieved, and all of the sub-goals, not about thinking about the long run from the beginning;
- You have to run your own race/PhD, don't compare yourself to others because there will be always someone who is faster/better/stronger – but also slower or softer, and you should cheer up those people as well, because one day you might be like them and everyone is fighting their own battle;
- Be patient and believe in yourself – go slower if you need to, try again tomorrow if you didn't succeed, but always reach the goal you have established for yourself;
- There will be bad days, there will be cold days and days where you will be tired – but always be optimistic, and smile, because as surely as bad days will pass, good days will come, too;
- The worst experiences make the best stories, but good experiences make great stories too;
- Medals don't matter; what counts are the memories;
- Build your mental strength through running, and use running as a template for other challenges in life;
- Always celebrate your victories with your closest ones, including the small ones along the way (see the first point in this bullet list).

Running is a great metaphor for life (and a PhD), and there is a cartoon by Guy Downes that sums it up quite well:

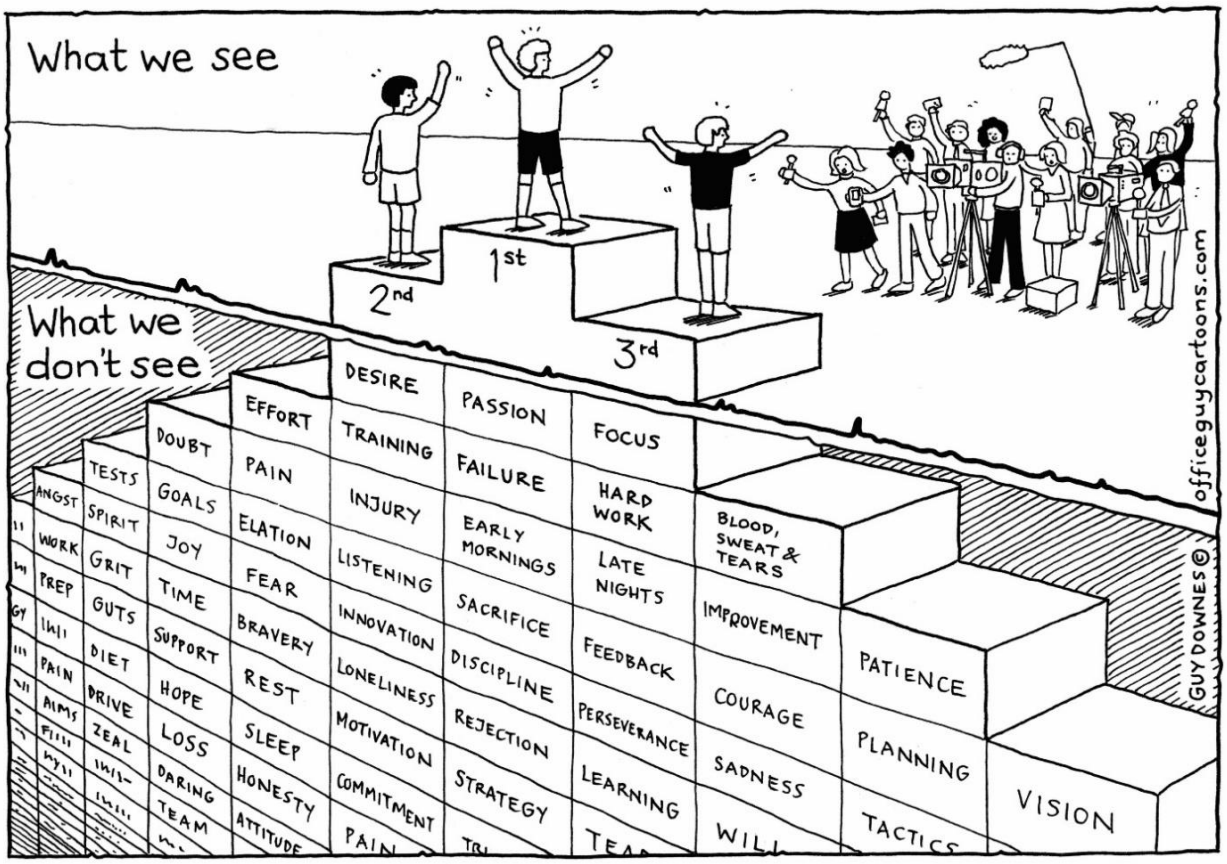

Figure 9. What we see versus what we don't see, or everything behind visible victories. Cartoon "Podium finish" by Guy Downes, Office Guy Cartoons. Reproduced with permission from the author.

Running made me meet **Joakim Svederman**, not only a great runner, but also a friend and kind-hearted soul. Your kindness and lack of judgment towards others show what a great person you are.

Filipa Sampaio, where to start? We go a long way back and I truly treasure our friendship. We have been through thick and thin throughout the years, ups and downs, and I am sure the years to come will only make our friendship grow stronger.

Drew Black, a long-time friend. Among the first people I met in Sweden, our friendship was immediate and has lasted through the years. I hope we will continue our life path together!

Daniel Portugal, **Louise Gustafson & Marco Heidrich**, **Miguel Barroso** and **Camilla Linder**, good old friends. Joyful and always in a good mood, our moments have always been fun! I am looking forward to spend more quality time with you.

Luísa Teixeira, **Couto family**, **Joelle Black**, **Michal Frankowski**, **Tiago Lopes**, **André "Macarrão" Couto**, **Vasco Câmara** and my friends and family in the Azores, Lisbon, Stockholm and Romania, for the great moments and lessons that I have learned with you.

The Portuguese gang in Stockholm, proudly represented by **Andreia Bernardes**, **Cristina Silva**, **Cláudia Sequeira**, **Joana Gil Ribeiro & Rafael Fernandes**, **João Melim**, **Luís Rocha**, **Tiago Braga** and respective partners. The years have passed but the friendship remains, and I hope we will meet more often!

Inspiring was also a lot of the people that I have met during my stay Jägargatan, a student accommodation from Karolinska Institutet during my Erasmus exchange as a bachelor's student. I met a lot of PhD-students that fueled my ambition to pursue a PhD in a relevant area. Among those people, I have become friends with **Thibaud Richard**, **Shahul Ahmeed**, **Peter Barret** and **Aga Sowińska**, among others.

Recia Atkins and **Johan Rossouw**, for the friendship and numerous good moments spent together. Your insightful comments and humor have always made me think a lot. A special thanks to **Recia** for providing many edits and comments that uplifted the language quality of my work.

And speaking of uplifts: if there is someone who has always been there since the times that we were just ambitious Master students and the dream of a PhD was still an untold and silent goal, that is my wife **Adina**. Companion from day one, she is truly my unsung hero, not only of this work but also of my life. A true angel and a wise advisor, she has been by my side and shared the ride like no one else. Always there with a positive comment or a hug when I was tired, helping me in every possible way possible to succeed. Nothing can be written or said to express my love and my most heartfelt "thank you" than a sincere, long and silent hug that only two people who love each other can give. I am looking forward to our life together, to enjoy all the small and big moments and to overcome all bumps along the way. A life that has been blessed with the much-awaited little **Gabriel Vieira da Silva** 💔💔💔

To my family in Romania, for the unconditional love and welcoming – **Gheorghe** and **Stela Tenie**, but also **Mihai & Raluca Terente**, **Andrei & Teodora Ionescu**, and all the lovely kids: **Vasile**, **Ecaterina**, **Nicolae**, **Radu**, **Ioan**, **Constantin**, **Mina**, **Maria** and **Alexandru**. My trips to Romania were and will always be better because of you. **Florina** and **Silviu Mocanu**, my wedding godparents and my expanded family, for always welcoming me to your house and letting us "live the life" – good food, good company and unforgettable moments. We had a lot of fun and I hope to continue making new memories together! **Irina**, **Maria**, **Ana** and **Teodora Mocanu**, for the past and future memories, that I will always carry in my heart. Watching you grow has been quite an experience and I am looking forward to witnessing you turn into fine young ladies!

Speaking of love, **Henrique Silva**, my little brother, I saw you growing, day by day, since you were a swollen newborn until you turned into the fine wise young man you are today, with a sharp intellect and an enormous heart! The world needs more people like you! You, too, are someone that I have an indescribable love for, and I am super happy to have you as my brother.

Vanessa Silva Pinheiro, with whom I share the unforgettable childhood memories that we will carry for the rest of our lives. We share our upbringing, knowing how special it was to share it together, and will continue the walk of life for the many years to come. To **João**, **Mateus** and **Olívia Pinheiro**, for bringing joy to life and lighting up the world around you! :D

Ricardo and **Lurdes Silva**, my parents, for the permanent support and unconditional love. They raised and formatted me to be a good citizen and an honest worker, setting the example and taking no shortcuts, not forgetting the core value of family. They, who from day one, taught me to pursue and work hard to achieve my dreams and goals, and to invest all the time and energy necessary in what I believed was important. "Patience and love", two precious virtues, as my mom often said. They who taught me to smile at all times, no matter the difficulties or how your heart feels. Aqueles pilares da Vida tão inabaláveis que me farão sentir um vazio indescritível no dia em que já não estiverem aí – os meus olhos até se marejam de lágrimas só de imaginar que esse dia pode chegar. Nunca conseguirei colocar em palavras o amor que sinto por vós, mas que espero que sintam nos abraços que trocamos quando nos encontramos e estamos juntos. O meu amor por vós é indescritível.

I sincerely thank everyone who contributed to the four years of my PhD and those who indirectly added value to this great experience of doctoral studies at Karolinska Institutet. I apologize in advance if I failed to mention someone, but be sure that I take you all in my heart and carry you in my memories. My most genuine "thank you"!

10 REFERENCES

Abraham K, Mielke H, Fromme H, Völkel W, Menzel J, Peiser M, Zepp F, Willich SN, Weikert C. 2020. Internal exposure to perfluoroalkyl substances (PFASs) and biological markers in 101 healthy 1-year-old children: associations between levels of perfluorooctanoic acid (PFOA) and vaccine response. Archives of Toxicology 94(6):2131-2147. https://doi.org/10.1007/s00204-020-02715-4

Akaike H. 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60(2):255-265. https://doi.org/10.1093/biomet/60.2.255

Al-Azemi MK, Omu AE, Fatinikun T, Mannazhath N, Abraham S. 2009. Factors contributing to gender differences in serum retinol and alpha-tocopherol in infertile couples. Reproductive Biomedicine Online 19(4):583-590. https://doi.org/10.1016/j.rbmo.2009.05.005

Ali I, Guo Y, Silins I, Högberg J, Stenius U, Korhonen A. 2016. Grouping chemicals for health risk assessment: A text mining-based case study of polychlorinated biphenyls (PCBs). Toxicology Letters 241:32-37. https://doi.org/10.1016/j.toxlet.2015.11.003

Ali I, Dreij K, Baker S, Högberg J, Korhonen A, Stenius U. 2021. Application of Text Mining in Risk Assessment of Chemical Mixtures: A Case Study of Polycyclic Aromatic Hydrocarbons (PAHs). Environmental Health Perspectives 129(6):067008. https://doi.org/10.1289/ehp6702

Allen BC, Kavlock RJ, Kimmel CA, Faustman EM. 1994. Dose-response assessment for developmental toxicity: II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. Fundamental and Applied Toxicology 23(4):487-495. https://doi.org/10.1006/faat.1994.1133

Andersson EM, Scott K, Xu Y, Li Y, Olsson DS, Fletcher T, Jakobsson, K. 2019. High exposure to perfluorinated compounds in drinking water and thyroid disease. A cohort study from Ronneby, Sweden. Environmental Research 176:108540. https://doi.org/10.1016/j.envres.2019.108540

Aro R, Eriksson U, Kärrman A, Yeung LWY. 2021. Organofluorine Mass Balance Analysis of Whole Blood Samples in Relation to Gender and Age. Environmental Science & Technology 55(19):13142-13151. https://doi.org/10.1021/acs.est.1c04031

Aro R, Eriksson U, Kärrman A, Jakobsson K, Yeung LWY. 2022. Extractable organofluorine analysis: A way to screen for elevated per- and polyfluoroalkyl substance contamination in humans? Environment International 159:107035. https://doi.org/10.1016/j.envint.2021.107035

ATSDR, United States Agency for Toxic Substances and Disease Registry. 2000. Toxicological profile for polychlorinated biphenyls (PCBs). Department of Health and Human Services, Public Health Service, Georgia, USA. https://www.atsdr.cdc.gov/toxprofiles/tp17.pdf [accessed 31/03/2022]

ATSDR, United States Agency for Toxic Substances and Disease Registry. 2021. Toxicological profile for perfluoroalkyls. Department of Health and Human Services, Public Health Service, Georgia, USA. https://www.atsdr.cdc.gov/ToxProfiles/tp200.pdf [accessed 31/03/2022]

Attanasio R. 2019. Sex differences in the association between perfluoroalkyl acids and liver function in US adolescents: Analyses of NHANES 2013-2016. Environmental Pollution 254:113061. https://doi.org/10.1016/j.envpol.2019.113061

Aven T. 2016. Risk assessment and risk management: Review of recent advances on their foundation. European Journal of Operational Research 253:1-13. https://doi.org/10.1016/j.ejor.2015.12.023

Bălan SA, Mathrani VC, Guo DF, Algazi AM. 2021. Regulating PFAS as a Chemical Class under the California Safer Consumer Products Program. Environmental Health Perspectives 129(2):25001. https://doi.org/10.1289/ehp7431

Banzhaf S, Filipovic M, Lewis J, Sparrenbom CJ, Barthel R. 2017. A review of contamination of surface-, ground-, and drinking water in Sweden by perfluoroalkyl and polyfluoroalkyl substances (PFASs). Ambio 46(3):335-346. https://doi.org/10.1007/s13280-016-0848-8

Bignert A, Aune M, Fridén U, Gyllenhammar I, Lignell S, Nyberg E, Glynn A. 2020. Temporal trends of Swedish environmental and human milk concentrations of dioxins, furans and dioxin-like PCBs, with forecasts to 2040. Pp 1-53. https://urn.kb.se/resolve?urn=urn:nbn:se:naturvardsverket:diva-8426 [accessed 31/03/2022]

Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, Godleski JJ, Coull BA. 2015. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics 16(3):493-508. https://doi.org/10.1093/biostatistics/kxu058

Boesen SAH, Long M, Wielsoe M, Mustieles V, Fernandez MF, Bonefeld-Jørgensen EC. 2020. Exposure to Perflouroalkyl acids and foetal and maternal thyroid status: a review. Environmental Health 19(1):107. https://doi.org/10.1186/s12940-020-00647-1

Bokkers BGH and Slob W. 2005. A comparison of ratio distributions based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. Toxicological Sciences 85(2):1033-1040. https://doi.org/10.1093/toxsci/kfi144

Bokkers BGH and Slob W. 2007. Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. Critical Reviews in Toxicology 37(5):355-373. https://doi.org/10.1080/10408440701249224

Bokkers BGH, Bakker M, Boon P, Bos P, Bosgra S, van der Heijden GWAM, Janer G, Slob W, van der Voet V. 2009. The practicability of the integrated probabilistic risk assessment (IPRA) approach for substances in food. RIVM report 320121001. https://www.rivm.nl/bibliotheek/rapporten/320121001.pdf [accessed 31/03/2022]

Bokkers BGH, Mengelers MJ, Bakker MI, Chiu WA, Slob W. 2017. APROBA-Plus: A probabilistic tool to evaluate and express uncertainty in hazard characterization and exposure assessment of substances. Food and Chemical Toxicology 110:408-417. https://doi.org/10.1016/j.fct.2017.10.038

Bosgra S, van der Voet H, Boon PE, Slob W. 2009. An integrated probabilistic framework for cumulative risk assessment of common mechanism chemicals in food: an example with

organophosphorus pesticides. Regulatory Toxicology and Pharmacology 54(2):124-133. https://doi.org/10.1016/j.yrtph.2009.03.004

Box GEP. 1976. Science and Statistics. Journal of the American Statistical Association 71(356):791-799. https://doi.org/10.1080/01621459.1976.10480949

Brandon EF, Bulder AS, van Engelen JG, Mahieu CM, Mennes WC, Pronk ME, Rietveld AG, van de Ven BM, Ten Voorde SE, Wolterink G, Slob W, Zeilmaker MJ, Bessems JG. 2013. Does EU legislation allow the use of the benchmark dose (BMD) approach for risk assessment? Regulatory Toxicology and Pharmacology 67(2):182-188. https://doi.org/10.1016/j.yrtph.2013.07.005

Budtz-Jørgensen E, Keiding N, Grandjean PJB. 2001. Benchmark dose calculation from epidemiological data. Biometrics 57(3):698-706. https://doi.org/10.1111/j.0006-341X.2001.00698.x

Buist H, von Bölcsházy GF, Dammann M, Telman J, Rennen M. 2009. Derivation of the minimal magnitude of the Critical Effect Size for continuous toxicological parameters from within-animal variation in control group data. Regulatory Toxicology and Pharmacology 55(2):139-150. https://doi.org/10.1016/j.yrtph.2009.06.009

Burnham KP and Anderson DR. 2004. Multimodel inference – understanding AIC and BIC in model selection. Sociological Methods & Research 33(2):261-304. https://doi.org/10.1177/0049124104268644

Butler LD, Guzzie-Peck P, Hartke J, Bogdanffy MS, Will Y, Diaz D, Mortimer-Cassen E, Derzi M, Greene N, DeGeorge JJ. 2017. Current nonclinical testing paradigms in support of safe clinical trials: An IQ Consortium DruSafe perspective. Regulatory Toxicology and Pharmacology 87(sup3):S1-S15. https://doi.org/10.1016/j.yrtph.2017.05.009

Canadian Council on Animal Care. 1993. Guide to the care and use of experimental animals (2nd ed., Vol. 1). ISBN 0-919087-18-3. https://www.ccac.ca/Documents/Standards/Guidelines/Experimental_Animals_Vol1.pdf [accessed 31/03/2022]

Chapman K, Sewell F, Allais L, Delongeas JL, Donald E, Festag M, Kervyn S, Ockert D, Nogues V, Palmer H, Popovic M, Roosen W, Schoenmakers A, Somers K, Stark C, Stei P, Robinson S. 2013. A global pharmaceutical company initiative: An evidence-based approach to define the upper limit of body weight loss in short term toxicity studies. Regulatory Toxicology and Pharmacology 67(1):27-38. https://doi.org/10.1016/j.yrtph.2013.04.003

Chatfield C. 1995. Model uncertainty, data mining and statistical-inference. Journal of the Royal Statistical Society Series A 158(3):419-466. https://doi.org/10.2307/2983440

Chen Q and Shao K. 2020. A Novel Dose-Response Assessment Framework with Quantitatively Integrated MOA Information. Abstract No. 1523. The Toxicologist, Supplement to Toxicological Sciences 174(1):130. https://www.toxicology.org/pubs/docs/Tox/2020Tox.pdf [accessed 31/03/2022]

Chiu YH, Bellavia A, James-Todd T, Correia KF, Valeri L, Messerlian C, Ford JB, Mínguez-Alarcón L, Calafat AM, Hauser R, Williams PL. 2018a. Evaluating effects of prenatal exposure to phthalate mixtures on birth weight: A comparison of three statistical approaches. Environmental International 113:231-239. https://doi.org/10.1016/j.envint.2018.02.005

Chiu WA, Axelrad DA, Dalaijamts C, Dockins C, Shao K, Shapiro AJ, Paoli G. 2018b. Beyond the RfD: Broad Application of a Probabilistic Approach to Improve Chemical Dose-Response Assessments for Noncancer Effects. Environmental Health Perspectives 126(6):067009. https://doi.org/10.1289/ehp3368

Chu I, Villeneuve DC, Yagminas A, LeCavalier P, Poon R, Feeley M, Kennedy SW, Seegal RF, Håkansson H, Ahlborg UG, Valli VE. 1994. Subchronic toxicity of 3,3',4,4',5-pentachlorobiphenyl in the rat. I. Clinical, biochemical, hematological, and histopathological changes. Fundamental and Applied Toxicology 22(3):457-468.

Chu I, Villeneuve DC, Yagminas A, Lecavalier P, Håkansson H, Ahlborg UG, Valli VE, Kennedy SW, Bergman Å, Seegal RF, Feeley M. 1995. Toxicity of PCB 77 (3,3',4,4'-tetrachlorobiphenyl) and PCB 118 (2,3',4,4',5-pentachlorobiphenyl) in the rat following subchronic dietary exposure. Fundamental and Applied Toxicology 26(2):282-292.

Chu I, Villeneuve DC, Yagminas A, Lecavalier P, Poon R, Feeley M, Kennedy SW, Seegal RF, Håkansson H, Ahlborg UG, Valli VE, Bergman Å. 1996a. Toxicity of 2,2',4,4',5,5'-hexachlorobiphenyl in rats: Effects following 90-day oral exposure. Journal of Applied Toxicology 16(2):121-128.

Chu I, Villeneuve DC, Yagminas A, Lecavalier P, Poon R, Håkansson H, Ahlborg UG, Valli VE, Kennedy SW, Bergman Å, Seegal RF, Feeley M. 1996b. Toxicity of 2,4,4'-trichlorobiphenyl in rats following 90-day dietary exposure. Journal of Toxicology and Environmental Health 49(3):301-318.

Chu I, Poon R, Yagminas A, Lecavalier P, Håkansson H, Valli VE, Kennedy SW, Bergman Å, Seegal RF, Feeley M. 1998. Subchronic toxicity of PCB 105 (2,3,3',4,4'-pentachlorobiphenyl) in rats. Journal of Applied Toxicology 18(4):285-292.

Chu I, Lecavalier P, Håkansson H, Yagminas A, Valli VE, Poon P, Feeley M. 2001. Mixture effects of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin and polychlorinated biphenyl congeners in rats. Chemosphere 43(4):807-814. https://doi.org/10.1016/S0045-6535(00)00437-9

Clair HB, Pinkston CM, Rai SN, Pavuk M, Dutton ND, Brock GN, Prough RA, Falkner KC, McClain CJ, Cave MC. 2018. Liver Disease in a Residential Cohort With Elevated Polychlorinated Biphenyl Exposures. Toxicological Sciences 164(1):39-49. https://doi.org/10.1093/toxsci/kfy076

Coperchini F, Croce L, Ricci G, Magri F, Rotondi M, Imbriani M, Chiovato L. 2020. Thyroid Disrupting Effects of Old and New Generation PFAS. Frontiers in Endocrinology 11:612320. https://doi.org/10.3389/fendo.2020.612320

Cousins IT, Goldenman G, Herzke D, Lohmann R, Miller M, Ng CA, Patton S, Scheringer M, Trier X, Vierke L, Wang Z, DeWitt JC. 2019. The concept of essential use for determining when uses of PFASs can be phased out. Environmental Science: Processes & Impacts 21(11):1803-1815. https://doi.org/10.1039/c9em00163h

Cousins IT, DeWitt JC, Glüge J, Goldenman G, Herzke D, Lohmann R, Miller M, Ng CA, Scheringer M, Wang Z. 2020. High persistence of PFAS sufficient for their management as a chemical class. Environmental Science: Processes & Impacts 22(12):2307-2312. https://doi.org/10.1039/D0EM00355G

Crump KS. 1984. A new method for determining allowable daily intakes. Fundamental and Applied Toxicology 4(5):854-871. https://doi.org/10.1016/0272-0590(84)90107-6

Crump KS. 1995. Calculation of benchmark doses from continuous data. Risk Analysis 15(1):79-89. https://doi.org/10.1111/j.1539-6924.1995.tb00095.x

Dallaire R, Dewailly E, Pereg D, Dery S, Ayotte P. 2009. Thyroid function and plasma concentrations of polyhalogenated compounds in Inuit adults. Environmental Health Perspectives 117(9):1380-1386. https://doi.org/10.1289/ehp.0900633

Davis JA, Gift JS, Zhao QJ. 2011. Introduction to benchmark dose methods and US EPA's benchmark dose software (BMDS) version 2.1.1. Toxicology and Applied Pharmacology 254(2):181-191. https://doi.org/10.1016/j.taap.2010.10.016

Dekkers S, Telman J, Rennen MA, Appel MJ, De Heer C. 2006. Within-animal variation as an indication of the minimal magnitude of the critical effect size for continuous toxicological parameters applicable in the benchmark dose approach. Risk Analysis 26(4):867-880. https://doi.org/10.1111/j.1539-6924.2006.00784.x

De Silva AO, Armitage JM, Bruton TA, Dassuncao C, Heiger-Bernays W, Hu, XC, Kärrman A, Kelly B, Ng C, Robuck A, Sun M, Webster TF, Sunderland EM. 2021. PFAS Exposure Pathways for Humans and Wildlife: A Synthesis of Current Knowledge and Key Gaps in Understanding. Environmental Toxicology and Chemistry 40(3):631-657. https://doi.org/10.1002/etc.4935

EC, European Commission ─ Directorate-General for Health and Consumers. 2012. Toxicity and assessment of chemical mixtures. https://doi.org/10.2772/21444

ECHA, European Chemicals Agency. 2012. Guidance on information requirements and chemical safety assessment, Chapter R.8: Characterisation of dose [concentration]–response for human health. European Chemicals Agency, Helsinki, Finland. https://echa.europa.eu/documents/10162/17224/information_requirements_r8_en.pdf/e153243a-03f0-44c5-8808-88af66223258?t=1353935239897 [accessed 31/03/2022]

ECHA, European Chemicals Agency. 2017. Guidance on the biocidal products regulation – volume III – human health assessment & evaluation (parts B+C). Vol.3 (version 4). https://doi.org/10.2823/143042

Ecke F, Singh NJ, Arnemo JM, Bignert A, Helander B, Berglund ÅMM, Borg H, Bröjer C, Holm K, Lanzone M, Miller T, Nordström A, Räikkönen J, Rodushkin I, Ågren E, Hörnfeldt B.

2017. Sublethal lead exposure alters movement behavior in free-ranging golden eagles. Environmental Science & Technology 51(10):5729-5736. https://doi.org/10.1021/acs.est.6b06024

Edgeworth FY. 1885. Methods of statistics. Journal of the Statistical Society of London, Jubilee Volume:181-217. www.jstor.org/stable/25163974 [accessed 31/03/2022]

EFSA, European Food Safety Authority. 2005. Opinion of the Scientific Committee on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic. EFSA Journal 282:1-31. https://doi.org/10.2903/j.efsa.2005.282

EFSA, European Food Safety Authority. 2008. Perfluorooctane sulfonate (PFOS), perfluorooctanoic acid (PFOA) and their salts. Scientific Opinion of the Panel on Contaminants in the Food chain. EFSA Journal 653:1-131. https://doi.org/10.2903/j.efsa.2008.653

EFSA, European Food Safety Authority. 2017. Update: Use of the benchmark dose approach in risk assessment. EFSA Journal 15(1):4658. https://doi.org/10.2903/j.efsa.2017.4658

EFSA, European Food Safety Authority. 2018a. Risk for animal and human health related to the presence of dioxins and dioxin-like PCBs in feed and food. EFSA Journal 16(11):1-331. https://doi.org/ 10.2903/j.efsa.2018.5333

EFSA, European Food Safety Authority. 2018b. Risk to human health related to the presence of perfluorooctane sulfonic acid and perfluorooctanoic acid in food. EFSA Journal 16(12):e05194. https://doi.org/10.2903/j.efsa.2018.5194

EFSA, European Food Safety Authority. 2020. Risk to human health related to the presence of perfluoroalkyl substances in food. EFSA Journal 18(9):e06223. https://doi.org/10.2903/j.efsa.2020.6223

EFSA, European Food Safety Authority. 2022. Public Consultation PC-0135 on the Updated Scientific Committee Guidance on the use of benchmark dose approach in risk assessment. https://connect.efsa.europa.eu/RM/s/publicconsultation2/a0l7U0000011Q1H/pc0135 [accessed 31/03/2022]

EMA, European Medicines Agency. 2010. Guideline on repeated dose toxicity. CPMP/SWP/1042/99 Rev 1. https://www.ema.europa.eu/documents/scientific-520 guideline/guideline-repeated-dose-toxicity-revision-1_en.pdf [accessed 31/03/2022]

Eriksen KT, Raaschou-Nielsen O, McLaughlin JK, Lipworth L, Tjønneland A, Overvad K, Sørensen M. 2013. Association between plasma PFOA and PFOS levels and total cholesterol in a middle-aged Danish population. PloS One 8(2):e56969. https://doi.org/10.1371/journal.pone.0056969

Eriksson L, Byrne T, Johansson E, Trygg J, Vikström C. 2013. *Multi-and megavariate data analysis: basic principles and applications* (3rd ed). Umetrics Academy. ISBN 9197373052. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A152566 [accessed 31/03/2022]

EU, European Union. 2010. Directive 2010/63/EU of the European Parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union 53. https://doi.org/10.3000/17252555.L_2010.276.eng

Faustman EM and Bartell SM. 1997. Review of noncancer risk assessment: Applications of benchmark dose methods. Human and Ecological Risk Assessment 3(5):893-920. https://doi.org/10.1080/10807039709383733

Faustman EM. 2018. Chapter 4. Risk Assessment. In Klaasen CD (Ed.), *Casarett & Doull's Toxicology: The Basic Science of Poisons* (9th edition, p. 127-155). McGraw Hill. ISBN 978-1-25-986375-2.

FDA, United States Food and Drug Administration. 2010. Guidance for industry – M3 (R2) nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. Rev 1. US Food and Drug Administration. Maryland, USA. https://www.fda.gov/downloads/drugs/guidances/ucm073246.pdf [accessed 31/03/2022]

Filipovic M, Woldegiorgis A, Norström K, Bibi M, Lindberg M, Österås AH. 2015. Historical usage of aqueous film forming foam: A case study of the widespread distribution of perfluoroalkyl acids from a military airport to groundwater, lakes, soils and fish. Chemosphere 129:39-45. https://doi.org/10.1016/j.chemosphere.2014.09.005

Fowles JR, Alexeeff GV, Dodge D. 1999. The use of benchmark dose methodology with acute inhalation lethality data. Regulatory Toxicology and Pharmacology 29(3):262-278. https://doi.org/10.1006/rtph.1999.1295

Franke V, McCleaf P, Lindegren K, Ahrens L. 2019. Efficient removal of per- and polyfluoroalkyl substances (PFASs) in drinking water treatment: nanofiltration combined with active carbon or anion exchange. Environmental Science: Water Research & Technology 5(11):1836-1843. https://doi.org/10.1039/c9ew00286c

Gadaleta D, Marzo M, Toropov A, Toropova A, Lavado GJ, Escher SE, Dorne JLCM, Benfenati E. 2021. Integrated In Silico Models for the Prediction of No-Observed-(Adverse)-Effect Levels and Lowest-Observed-(Adverse)-Effect Levels in Rats for Sub-chronic Repeated-Dose Toxicity. Chemical Research in Toxicology 34(2):247-257. https://doi.org/10.1021/acs.chemrestox.0c00176

Gelman A and Greenland S. 2019. Are confidence intervals better termed "uncertainty intervals"? BMJ 366:l5381. https://doi.org/10.1136/bmj.l5381

Glynn A, Kotova N, Dahlgren E, Lindh C, Jakobsson K, Gyllenhammar I, Lignell S, Nälsén C. 2020. Determinants of serum concentrations of perfluoroalkyl acids (PFAAs) in school children and the contribution of low-level PFAA-contaminated drinking water. Environmental Science: Processes & Impacts 22(4):930-944. https://doi.org/10.1039/c9em00497a

Grandjean P, Budtz-Jørgensen E, Barr DB, Needham LL, Weihe P, Heinzow B. 2008. Elimination Half-Lives of Polychlorinated Biphenyl Congeners in Children. Environmental Science & Technology 42(18):6991-6996. https://doi.org/10.1021/es800778q

Grandjean P, Andersen EW, Budtz-Jørgensen E, Nielsen F, Mølbak K, Weihe P, Heilmann C. 2012. Serum vaccine antibody concentrations in children exposed to perfluorinated compounds. Journal of the American Medical Association 307(4):391-397. https://doi.org/10.1001/jama.2011.2034

Grandjean P, Heilmann C, Weihe P, Nielsen F, Mogensen UB, Budtz-Jorgensen E. 2017. Serum vaccine antibody concentrations in adolescents exposed to perfluorinated compounds. Environmental Health Perspectives 125(7):077018. https://doi.org/10.1289/ehp275

Grignard E, Håkansson H, Munn S. 2020. Regulatory needs and activities to address the retinoid system in the context of endocrine disruption: The European viewpoint. Reproductive Toxicology 93:250-258. https://doi.org/10.1016/j.reprotox.2020.03.002

Gyllenhammar I, Berger U, Sundström M, McCleaf P, Euren K, Eriksson S, Ahlgren S, Lignell S, Aune M, Kotova N, Glynn A. 2015. Influence of contaminated drinking water on perfluoroalkyl acid levels in human serum – a case study from Uppsala, Sweden. Environmental Research 140:673-683. https://doi.org/10.1016/j.envres.2015.05.019

Gyllenhammar I, Diderholm B, Gustafsson J, Berger U, Ridefelt P, Benskin JP, Lignell S, Lampa E, Glynn A. 2018. Perfluoroalkyl acid levels in first-time mothers in relation to offspring weight gain and growth. Environment International 111:191-199. https://doi.org/10.1016/j.envint.2017.12.002

Gyllenhammar I, Benskin JP, Sandblom O, Berger U, Ahrens L, Lignell S, Wiberg K, Glynn A. 2019. Perfluoroalkyl Acids (PFAAs) in Children's Serum and Contribution from PFAA-Contaminated Drinking Water. Environmental Science & Technology 53(19):11447-11457. https://doi.org/10.1021/acs.est.9b01746

Haber LT, Dourson ML, Allen BC, Hertzberg RC, Parker A, Vincent MJ, Maier A, Boobis AR. 2018. Benchmark dose (BMD) modeling: Current practice, issues, and challenges. Critical Reviews in Toxicology 48(5):387-415. https://doi.org/10.1080/10408444.2018.1430121

Hoeting JA, Madigan D, Raftery AE, Volinsky CT. 1999. Bayesian model averaging: A tutorial. Statistical Science 14(4):382-401. https://doi.org/10.1214/ss/1009212519

Hothorn LA. 2021. A statistical method for estimating the no-observed-adverse-event-level. arXiv preprint 2101.00427. https://arxiv.org/pdf/2101.00427.pdf

Håkansson H. 2020. Chapter 10. Vitamin A and the retinoid system–from nutrition to endocrine disruption. In Mantovani A and Fucic A (Eds.), *Challenges in Endocrine Disruptor Toxicology and Risk Assessment* (Issues in Toxicology No. 42, p. 268-280), The Royal Society of Chemistry. https://doi.org/10.1039/9781839160738

IARC, International Agency for Research on Cancer. 2015. Polychlorinated Biphenyls and Polybrominated Biphenyls. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 107. ISBN 978-92-832-0173-1. https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-

Humans/Polychlorinated-Biphenyls-And-Polybrominated-Biphenyls-2015 [accessed 31/03/2022]

ICH, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 2009a. Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals M3 (R2). https://database.ich.org/sites/default/files/M3_R2__Gu ideline.pdf [accessed 31/03/2022]

ICH, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 2009b. Nonclinical evaluation for anticancer pharmaceuticals S9. https://database.ich.org/sites/default/files/S9_Guideline.pdf [accessed 31/03/2022]

ICH, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 2019. Q3D (R1) Guideline for elemental impurities. https://database.ich.org/sites/default/files/Q3D-R1EWG_Document_Step4_Guideline_2019_0322.pdf [accessed 31/03/2022]

Jacobs R, Meesters JA, Ter Braak CJ, van de Meent D, van der Voet H. 2016. Combining exposure and effect modeling into an integrated probabilistic environmental risk assessment for nanoparticles. Environmental Toxicology and Chemistry 35(12):2958-2967. https://doi.org/10.1002/etc.3476

Jakobsson K, Kronholm Diab K, Lindh C, Persson B, Jönsson B. 2014. Exponering för perfluoerade ämnen (PFAS) i dricksvatten i Ronneby kommun. AMM Rapport 8:2014. In Swedish. https://sodrasjukvardsregionen.se/download/exponering-for-perfluorerade-amnen-pfas-i-dricksvatten-i-ronneby-kommun/ [accessed 31/03/2022]

Jensen SM, Kluxen FM, Ritz C. 2019. A review of recent advances in benchmark dose methodology. Risk Analysis 39(10):2295-2315. https://doi.org/10.1111/risa.13324

Jensen SM, Kluxen FM, Streibig JC, Cedergreen N, Ritz C. 2020. *bmd*: an R package for benchmark dose estimation. PeerJ 8:e10557. https://doi.org/10.7717/peerj.10557

Jensen SM, Cedergreen N, Kluxen FM, Ritz C. 2021a. A Nonmechanistic Parametric Modeling Approach for Benchmark Dose Estimation of Event-Time Data. Risk Analysis 41(11):2081-2093. https://doi.org/10.1111/risa.13708

Jensen SM, Kluxen FM, Ritz C. 2021b. Benchmark dose modelling in regulatory ecotoxicology, a potential tool in pest management. Pest Management Science. https://doi.org/10.1002/ps.6759

Jian JM, Guo Y, Zeng L, Liang-Ying L, Lu X, Wang F, Zeng EY. 2017. Global distribution of perfluorochemicals (PFCs) in potential human exposure source-a review. Environmental International 108:51-62. https://doi.org/10.1016/j.envint.2017.07.024

Johansson M. 2016. Asthmatics as a susceptible population in health risk assessment of airborne chemicals. [Doctoral thesis, Karolinska Institutet]. Karolinska Institutet Open Archive. https://openarchive.ki.se/xmlui/handle/10616/44970

Jonsson O, Villar RP, Nilsson LB, Norsten-Höög C, Brogren J, Eriksson M, Königsson K, Samuelsson A. 2012. Capillary microsampling of 25 µl blood for the determination of toxicokinetic parameters in regulatory studies in animals. Bioanalysis 4(6):661-674. https://doi.org/10.4155/bio.12.25

Judd N, Griffith WC, Faustman EM. 2004. Contribution of PCB exposure from fish consumption to total dioxin-like dietary exposure. Regulatory Toxicology and Pharmacology 40(2):125-135. https://doi.org/10.1016/j.yrtph.2004.06.004

Kalantari F, Ringblom J, Sand S, Öberg M. 2017. Influence of distribution of animals between dose groups on estimated benchmark dose and animal distress for quantal responses. Risk Analysis 37(9):1716-1728. https://doi.org/10.1111/risa.12741

Kienhuis AS, Slob W, Gremmer ER, Vermeulen JP, Ezendam J. 2015. A dose-response modeling approach shows that effects from mixture exposure to the skin sensitizers isoeugenol and cinnamal are in line with dose addition and not with synergism. Toxicological Sciences 147(1):68-74. https://doi.org/10.1093/toxsci/kfv109

Kienzler A, Bopp SK, van der Linden S, Berggren E, Worth A. 2016. Regulatory assessment of chemical mixtures: Requirements, current approaches and future perspectives. Regulatory Toxicology and Pharmacology 80:321-334. https://doi.org/10.1016/j.yrtph.2016.05.020

Kimmel CA and Gaylor DW. 1988. Issues in qualitative and quantitative risk analysis for developmental toxicology. Risk Analysis 8(1):15-20. https://doi.org/10.1111/j.1539-6924.1988.tb01149.x

Klotzsche C and Carpy S. 1973. Six month feeding study in dogs. Unpublished report no. FB73008. Sandoz Ltd, Basle, Switzerland. Submitted to WHO by Tessenderlo Kerley Inc. Phoenix, AZ, USA.

Knox SS, Jackson T, Frisbee SJ, Javins B, Ducatman AM. 2011. Perfluorocarbon exposure, gender and thyroid function in the C8 health project. Journal of Toxicological Sciences 36(4):403-410. https://doi.org/10.2131/jts.36.403

Korchevskiy A. 2021. Using benchmark dose modeling for the quantitative risk assessment: Carbon nanotubes, asbestos, glyphosate. Journal of Applied Toxicology 41(1):148-160. https://doi.org/10.1002/jat.4063

Kuljus K, von Rosen D, Sand S, Victorin K. 2006. Comparing experimental designs for benchmark dose calculations for continuous endpoints. Risk Analysis 26(4):1031-1043. https://doi.org/10.1111/j.1539-6924.2006.00798.x

Kwiatkowski CF, Andrews DQ, Birnbaum LS, Bruton TA, DeWitt JC, Knappe DRU, Maffini MV, Miller MF, Pelch KE, Reade A, Soehl A, Trier X, Venier M, Wagner CC, Wang Z, Blum A. 2020. Scientific Basis for Managing PFAS as a Chemical Class. Environmental Science & Technology Letters 7(8):532-543. https://doi.org/10.1021/acs.estlett.0c00255

Larsson K. 2018. Characterizing chemical exposure: focus on children's environment. [Doctoral thesis, Karolinska Institutet]. Karolinska Institutet Open Archive. https://openarchive.ki.se/xmlui/handle/10616/46299

Lecavalier P, Chu I, Yagminas A, Villeneuve DC, Poon R, Feeley M, Håkansson H, Ahlborg UG, Valli VE, Bergman Å, Seegal RF, Kennedy SW. 1997. Subchronic toxicity of 2,2',3,3',4,4'-hexachlorobiphenyl in rats. Journal of Toxicology and Environmental Health 51(3):265-277. https://doi.org/10.1080/00984109708984026

Li H, Hammarstrand S, Midberg B, Xu Y, Li Y, Olsson DS, Fletcher T, Jakobsson K, Andersson EM. 2021a. Cancer incidence in a Swedish cohort with high exposure to perfluoroalkyl substances in drinking water. Environmental Research 204(part C):112217. https://doi.org/10.1016/j.envres.2021.112217

Li Y, Fletcher T, Mucs D, Scott K, Lindh CH, Tallving P, Jakobsson K. 2017. Half-lives of PFOS, PFHxS and PFOA after end of exposure to contaminated drinking water. Occupational Environmental Medicine 75(1):46-51. https://doi.org/10.1136/oemed-2017-104651

Li Y, Barregård L, Xu Y, Scott K, Pineda D, Lindh CH, Jakobsson K, Fletcher T. 2020. Associations between perfluoroalkyl substances and serum lipids in a Swedish adult population with contaminated drinking water. Environmental Health 19(1):33. https://doi.org/10.1186/s12940-020-00588-9

Li Y, Xu Y, Fletcher T, Scott K, Nielsen C, Pineda D, Lindh CH, Olsson DS, Andersson EM, Jakobsson K. 2021b. Associations between perfluoroalkyl substances and thyroid hormones after high exposure through drinking water. Environmental Research 194:110647. https://doi.org/10.1016/j.envres.2020.110647

Lin X and Boutros PC. 2016. NNLM: fast and versatile non-negative matrix factorization. R package version 0.4.1. https://cran.r-project.org/package=NNLM [accessed 31/03/2022]

Livsmedelsverket (Swedish National Food Agency). 2014. Kompletterande bedömning av risker samt förslag till hanteringsåtgärder gällande perfluorerade alkylsyror (PFAA) i Ronnebys dricksvatten. In Swedish. https://www.livsmedelsverket.se/globalassets/livsmedel-innehall/oonskade-amnen/pfaa/kompletterande_bedomning_ronneby.pdf [accessed 31/03/2022]

Lopez-Espinosa MJ, Mondal D, Armstrong B, Bloom MS, Fletcher T. 2012. Thyroid function and perfluoroalkyl acids in children living near a chemical plant. Environmental Health Perspectives 120(7):1036-1041. https://doi.org/10.1289/ehp.1104370

Martins C, Dreij K, Costa PM. 2019. The State-of-the Art of Environmental Toxicogenomics: Challenges and Perspectives of "Omics" Approaches Directed to Toxicant Mixtures. International Journal of Environmental Research and Public Health 16(23):4718. https://doi.org/10.3390/ijerph16234718

Mecklenburg L. 2012. Benchmark dose modelling for hazard analysis in pharmaceutical risk assessment. Toxicology Letters 211:S77. https://doi.org/10.1016/j.toxlet.2012.03.296

Mencken HL. 1920. Prejudices: Second series. IV. *The divine afflatus*. CreateSpace Independent Publishing Platform. ISBN 9781541327047.

Moerbeek M, Piersma AH, Slob W. 2004. A comparison of three methods for calculating confidence intervals for the benchmark dose. Risk Analysis 24(1):31-40. https://doi.org/10.1111/j.0272-4332.2004.00409.x

Mondal D, Weldon RH, Armstrong BG, Gibson LJ, Lopez-Espinosa MJ, Shin HM, Fletcher T. 2014. Breastfeeding: A potential excretion route for mothers and implications for infant exposure to perfluoroalkyl acids. Environmental Health Perspectives 122(2):187-192. https://doi.org/10.1289/ehp.1306613

Mudumbi JBN, Ntwampe SKO, Matsha T, Mekuto L, Itoba-Tombo EF. 2017. Recent developments in polyfluoroalkyl compounds research: a focus on human/environmental health impact, suggested substitutes and removal strategies. Environmental Monitoring and Assessment 189(8):402. https://doi.org/10.1007/s10661-017-6084-2

Müller AK, Bosgra S, Boon PE, van der Voet H, Nielsen E, Ladefoged O. 2009. Probabilistic cumulative risk assessment of anti-androgenic pesticides in food. Food and Chemical Toxicology 47(12):2951-2962. https://doi.org/10.1016/j.fct.2009.07.039

Muri SD, Schlatter JR, Brüschweiler BJ. 2009. The benchmark dose approach in food risk assessment: Is it applicable and worthwhile? Food and Chemical Toxicology 47(12):2906-2925. https://doi.org/10.1016/j.fct.2009.03.035

NAC/AEGL, National Research Council, Committee on Toxicology, Subcommittee on Acute Exposure Guideline Levels. 2001. *Standing Operating Procedures for Developing Acute Exposure Guideline Levels for Hazardous Chemicals*. National Academies Press, Washington, USA. ISBN 978-0-309-17086-4. https://doi.org/10.17226/10122

Nilsson CB and Håkansson H. 2002. The retinoid signalling system-a target in dioxin toxicity. Critical Reviews in Toxicology 32(3):211-232. https://doi.org/10.1080/20024091064228

Norgren K, Tuck A, Vieira Silva A, Burkhardt P, Öberg M, Munic Kos V. High content screening of bisphenols and their mixtures in conditions of low intensity adipogenesis of human mesenchymal stem cells (hMSCs). 2022. Food and Chemical Toxicology 161:112842. https://doi.org/10.1016/j.fct.2022.112842

NRC, United States National Research Council. 1983. Risk assessment in the Federal Government: Managing the process. National Academies Press, Washington, USA. https://doi.org/10.17226/366

Nyström J, Benskin JP, Plassmann M, Sandblom O, Glynn A, Lampa E, Gyllenhammar I, Moraeus L, Lignell S. 2022. Demographic, life-style and physiological determinants of serum per- and polyfluoroalkyl substance (PFAS) concentrations in a national cross-sectional survey of Swedish adolescents. Environmental Research 208:112674. https://doi.org/10.1016/j.envres.2022.112674

OECD, Organisation for Economic Co-operation Development. 2000. Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. ENV/JM/MONO(2000)7. https://doi.org/10.1787/9789264078376

OECD, Organisation for Economic Co-operation Development. 2014. Guidance document 116 on the conduct and design of chronic toxicity and carcinogenicity studies, supporting test guidelines 451, 452 and 453. ENV/JM/MONO(2011)47. https://doi.org/10.1787/9789264221475-en

OECD, Organisation for Economic Co-operation Development. 2018a. Test No. 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents. https://doi.org/10.1787/9789264070707-en

OECD, Organisation for Economic Co-operation Development. 2018b. Toward a new comprehensive global database of per-and polyfluoroalkyl substances (PFASs): Summary report on updating the OECD 2007 list of per-and polyfluoroalkyl substances (PFASs). ENV/JM/MONO(2018)7. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV-JM-MONO(2018)7&doclanguage=en [accessed 31/03/2022]

Paparella M, Daneshian M, Hornek-Gausterer R, Kinzl M, Mauritz I, Mühlegger S. 2013. Food for thought… Uncertainty of testing methods – what do we (want to) know? ALTEX – Alternatives to Animal Experimentation 30(2):131. https://doi.org/10.14573/altex.2013.2.131

Paracelsus (pseudonym of Philippus Aureolus Theophrastus Bombastus von Hohenheim). 1538. Die dritte defension wegen des schreibens der neuen rezepte. In German. http://www.zeno.org/Philosophie/M/Paracelsus/Septem+Defensiones [accessed 31/03/2022]

Parens R, Nijhout HF, Morales A, Xavier Costa F, Bar-Yam Y. 2017. A Possible Link Between Pyriproxyfen and Microcephaly. PLoS Currents 9:ecurrents.outbreaks.5afb0bfb8cf31d9a4baba7b19b4edbac. https://doi.org/10.1371/currents.outbreaks.5afb0bfb8cf31d9a4baba7b19b4edbac

Peña EA, Wu W, Piegorsch W, West RW, Na L. 2017. Model Selection and Estimation with Quantal-Response Data in Benchmark Risk Assessment. Risk Analysis 37(4):716-732. https://doi.org/10.1111/risa.12644

Penaloza CG, Estevez B, Han DM, Norouzi M, Lockshin RA, Zakeri Z. 2014. Sex-dependent regulation of cytochrome P450 family members CYP1A1, CYP2E1, and CYP7B1 by methylation of DNA. The FASEB Journal 28(2):966-977. https://doi.org/10.1096/fj.13-233320

Prichystalova R, Fini J-B, Trasande L, Bellanger M, Demeneix B, Maxim L. 2017. Comparison of methods for calculating the health costs of endocrine disrupters: A case study on triclosan. Environmental Health 16(1):1-15. https://doi.org/10.1186/s12940-017-0265-x

Prior H, Haworth R, Labram B, Roberts R, Wolfreys A, Sewell F. 2020. Justification for species selection for pharmaceutical toxicity studies. Toxicology Research 9(6):758-770. https://doi.org/10.1093/toxres/tfaa081

R Core Team. 2017. R version 3.4.2. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

R Core Team. 2020. R version 3.6.3. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Radojevic N, Bjelogrlic B, Aleksic V, Rancic N, Samardzic M, Petkovic S, Savic S. 2012. Forensic aspects of water intoxication: Four case reports and review of relevant literature. Forensic Science International 220(1):1-5. https://doi.org/10.1016/j.forsciint.2012.01.021

Raffetti E, Donato F, Speziani F, Scarcella C, Gaia A, Magoni M. 2018. Polychlorinated biphenyls (PCBs) exposure and cardiovascular, endocrine and metabolic diseases: A population-based cohort study in a North Italian highly polluted area. Environmental International 120:215-222. https://doi.org/10.1016/j.envint.2018.08.022

Renwick AG, Barlow SM, Hertz-Picciotto I, Boobis AR, Dybing E, Edler L, Eisenbrand G, Greig JB, Kleiner J, Lambe J, Müller DJG, Smith MR, Tritscher A, Tuijtelaars S, Brandt PAD, Walker R, Kroes R. 2003. Risk characterisation of chemicals in food and diet. Food and Chemical Toxicology 41(9):1211-1271. https://doi.org/10.1016/S0278-6915(03)00064-4

Ringblom J, Johanson G, Öberg M. 2014. Current modeling practice may lead to falsely high benchmark dose estimates. Regulatory Toxicology and Pharmacology 69(2):171-177. https://doi.org/10.1016/j.yrtph.2014.03.004

Ringblom J. 2016. Worth weighing for: Studies on benchmark dose analysis in relation to animal ethics in toxicity testing. [Doctoral thesis, Karolinska Institutet]. Karolinska Institutet Open Archive. https://openarchive.ki.se/xmlui/handle/10616/45331

Ringblom J, Kalantari F, Johanson G, Öberg M. 2017. Influence of distribution of animals between dose groups on estimated benchmark dose and animal welfare for continuous effects. Risk Analysis 38(6):1143-1153. https://doi.org/10.1111/risa.12929

Ritscher A, Wang Z, Scheringer M, Boucher JM, Ahrens L, Berger U, Bintein S, Bopp SK, Borg D, Buser AM, Cousins I, DeWitt J, Fletcher T, Green C, Herzke D, Higgins C, Huang J, Hung H, Knepper T, Lau CS, Leinala E, Lindstrom AB, Liu J, Miller M, Ohno K, Perkola N, Shi Y, Smastuen Haug L, Trier X, Valsecchi S, van der Jagt K, Vierke L. 2018. Zurich statement on future actions on per- and polyfluoroalkyl substances (PFASs). Environmental Health Perspectives 126(8):84502. https://doi.org/10.1289/ehp4158

Ritter R, Scheringer M, MacLeod M, Moeckel C, Jones KC, Hungerbühler K. 2011. Intrinsic human elimination half-lives of polychlorinated biphenyls derived from the temporal evolution of cross-sectional biomonitoring data from the United Kingdom. Environmental Health Perspectives 119(2):225-231. https://doi.org/10.1289/ehp.1002211

Ritz C, Gerhard D, Hothorn LA. 2013. A Unified Framework for Benchmark Dose Estimation Applied to Mixed Models and Model Averaging. Statistics in Biopharmaceutical Research 5(1):79-90. https://doi.org/10.1080/19466315.2012.757559

Rudén C, Backhaus T, Bergman P, Faust M, Molander L, Slunge D. 2019. Future chemical risk management – Accounting for combination effects and assessing chemicals in groups. SOU 2019:45. Swedish Government Official Reports (SOU). https://www.government.se/4adb1a/contentassets/ee36b3e49c354bb8967f3a33a2d5ca50/future-chemical-risk-management---accounting-for-combination-effects-and-assessing-chemicals-in-groups-sou-201945 [accessed 31/03/2022]

Ruzzin J. 2012. Public health concern behind the exposure to persistent organic pollutants and the risk of metabolic diseases. BMC Public Health 12:298. https://doi.org/10.1186/1471-2458-12-298

Sand S, Portier CJ, Krewski D. 2011. A signal-to-noise crossover dose as the point of departure for health risk assessment. Environmental Health Perspectives 119(12):1766-1774. https://doi.org/10.1289/ehp.1003327

Sand S, Ringblom J, Håkansson H, Öberg M. 2012. The point of transition on the dose-effect curve as a reference point in the evaluation of in vitro toxicity data. Journal of Applied Toxicology 32(10):843-849. https://doi.org/10.1002/jat.2757

Sand S, Lindqvist R, Von Rosen D, Ilbäck N-G. 2018. Dose-Related Severity Sequence, and Risk-Based Integration, of Chemically Induced Health Effects. Toxicological Sciences 165(1):74-89. https://doi.org/10.1093/toxsci/kfy124

Schichl H. 2004. Models and the history of modeling. Modeling Languages in Mathematical Optimization 88:25-36. https://doi.org/10.1007/978-1-4613-0215-5_2

Schwenk M, Werner M, Younes M. 2002. Regulatory toxicology: objectives and tasks defined by the working group of the German society of experimental and clinical pharmacology and toxicology. Toxicology Letters 126(3):145-153. https://doi.org/10.1016/s0378-4274(01)00457-x

Seacat AM, Thomford PJ, Hansen KJ, Olsen GW, Case MT, Butenhoff JL. 2002. Subchronic toxicity studies on perfluorooctanesulfonate potassium salt in cynomolgus monkeys. Toxicological Sciences 68(1):249-264. https://doi.org/10.1093/toxsci/68.1.249

Seegal RF, Fitzgerald EF, Hills EA, Wolff MS, Haase RF, Todd AC, Parsons P, Molho ES, Higgins DS, Factor SA, Marek KL, Seibyl JP, Jennings DL, McCaffrey RJ. 2011. Estimating the half-lives of PCB congeners in former capacitor workers measured over a 28-year interval. Journal of Exposure Science & Environmental Epidemiology 21(3):234-246. https://doi.org/10.1038/jes.2010.3

Sewell F, Chapman K, Baldrick P, Brewster D, Broadmeadow A, Brown P, Burns-Naas LA, Clarke J, Constan A, Couch J, Czupalla O, Danks A, DeGeorge J, de Haan L, Hettinger K, Hill M, Festag M, Jacobs A, Jacobson-Kram D, Kopytek S, Lorenz H, Moesgaard SG, Moore E, Pasanen M, Perry R, Ragan I, Robinson S, Schmitt PM, Short B, Lima BS, Smith D, Sparrow S, van Bekkum Y, Jones D. 2014. Recommendations from a global cross-company data sharing initiative on the incorporation of recovery phase animals in safety assessment studies to support first-in-human clinical trials. Regulatory Toxicology and Pharmacology 70(1):413-429. https://doi.org/10.1016/j.yrtph.2014.07.018

SFS, Svensk författningssamling. 1988. L1 Djurskyddslag. 1988:534. In Swedish. https://rkrattsbaser.gov.se/sfst?bet=1988:534 [accessed 31/03/2022]

Shao K and Gift JS. 2014. Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. Risk Analysis 34(1):101-120. https://doi.org/10.1111/risa.12078

Shao K and Small MJ. 2011. Potential uncertainty reduction in model-averaged benchmark dose estimates informed by an additional dose study. Risk Analysis 31(10):1561-1575. https://doi.org/10.1111/j.1539-6924.2011.01595.x

Shao K and Small MJ. 2012. Statistical Evaluation of Toxicological Experimental Design for Bayesian Model Averaged Benchmark Dose Estimation with Dichotomous Data. Human and Ecological Risk Assessment: An International Journal 18(5):1096-1119. https://doi.org/10.1080/10807039.2012.707938

Shrestha S, Bloom MS, Yucel R, Seegal RF, Wu Q, Kannan K, Rej R, Fitzgerald EF. 2015. Perfluoroalkyl substances and thyroid function in older adults. Environment International 75:206-214. https://doi.org/10.1016/j.envint.2014.11.018

Simon TW, Simons SS, Preston RJ, Boobis AR, Cohen SM, Doerrer NG, Fenner-Crisp PA, McMullin TS, McQueen CA, Rowlands JC. 2014. The use of mode of action information in risk assessment: Quantitative key events/dose-response framework for modeling the dose-response for key events. Critical Reviews in Toxicology 44(sup3):17-43. https://doi.org/10.3109/10408444.2014.931925

SJVFS, Statens Jordbruksverkets författningssamling. 2012. L150 Statens Jordbruksverks Föreskrifter och allmänna råd om försöksdjur. SJVFS 2012:26. ISSN 1102-0970. In Swedish. https://www.jordbruksverket.se/download/18.3c1967aa13afeea1eb880002406/2012-026.pdf [accessed 31/03/2022]

SJVFS, Statens Jordbruksverkets författningssamling. 2015. L150 Statens Jordbruksverks Föreskrifter och allmänna råd om försöksdjur. SJVFS 2015:24. ISSN 1102-0970. In Swedish. https://www.jordbruksverket.se/download/18.7c05177614ea46421eb82598/1437380698357/2015-024.pdf [accessed 31/03/2022]

Slikker W, Andersen ME, Bogdanffy MS, Bus JS, Cohen SD, Conolly RB, David RM, Doerrer NG, Dorman DC, Gaylor DW, Hattis D, Rogers JM, Woodrow Setzer R, Swenberg JA, Wallace K. 2004. Dose-dependent transitions in mechanisms of toxicity. Toxicology and Applied Pharmacology 201(3):203-225. https://doi.org/10.1016/j.taap.2004.06.019

Slob W. 1999. Thresholds in toxicology and risk assessment. International Journal of Toxicology 18(4):259-268. https://doi.org/10.1080/109158199225413

Slob W. 2002. Dose-response modeling of continuous endpoints. Toxicological Sciences 66(2):298-312. https://doi.org/10.1093/toxsci/66.2.298

Slob W. 2014a. Benchmark dose and the three Rs. Part I. Getting more information from the same number of animals. Critical Reviews in Toxicology 44(7):557-567. https://doi.org/10.3109/10408444.2014.925423

Slob W. 2014b. Benchmark dose and the three Rs. Part II. Consequences for study design and animal use. Critical Reviews in Toxicology 44(7):568-580. https://doi.org/10.3109/10408444.2014.925424

Slob W. 2017. A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints. Critical Reviews in Toxicology 47(4):342-351. https://doi.org/10.1080/10408444.2016.1241756

Slob W, Moerbeek M, Rauniomaa E, Piersma AH. 2005. A statistical evaluation of toxicity study designs for the estimation of the benchmark dose in continuous endpoints. Toxicology Sciences 84(1):167-185. https://doi.org/10.1093/toxsci/kfi004

Slob W and Setzer RW. 2014. Shape and steepness of toxicological dose-response relationships of continuous endpoints. Critical Reviews in Toxicology 44(7):270-297. https://doi.org/10.3109/10408444.2013.853726

Smith PF. 2018. On the application of multivariate statistical and data mining analyses to data in neuroscience. The Journal of Undergraduate Neuroscience Education 16(2):R20-R32. PMC6057756

Soeteman-Hernández LG, Johnson GE, Slob W. 2016. Estimating the carcinogenic potency of chemicals from the in vivo micronucleus test. Mutagenesis 31(3):347-358. https://doi.org/10.1093/mutage/gev043

Sparrow SS, Robinson S, Bolam S, Bruce C, Danks A, Everett D, Fulcher S, Hill RE, Palmer H, Scott EW. 2011. Opportunities to minimise animal use in pharmaceutical regulatory general toxicology: a cross-company review. Regulatory Toxicology and Pharmacology 61(2):222-229. https://doi.org/10.1016/j.yrtph.2011.08.001

SSCS, Scientific Committee on Consumer Safety (managed by the Directorate-General for Health and Consumer Protection of the European Commission). 2018. The SSCS notes of guidance for the testing of cosmetic ingredients and their safety evaluation – 10th revision. SCCS/1602/18. https://doi.org/10.2875/77673

Steenland K and Winquist A. 2021. PFAS and cancer, a scoping review of the epidemiologic evidence. Environmental Research 194:110690. https://doi.org/10.1016/j.envres.2020.110690

Stubleski J, Salihovic S, Lind PM, Lind L, Dunder L, McCleaf P, Eurén K, Ahrens L, Svartengren M, van Bavel B, Kärrman A. 2017. The effect of drinking water contaminated with perfluoroalkyl substances on a 10-year longitudinal trend of plasma levels in an elderly Uppsala cohort. Environmental Research 159:95-102. https://doi.org/10.1016/j.envres.2017.07.050

Sunderland EM, Hu XC, Dassuncao C, Tokranov AK, Wagner CC, Allen JG. 2019. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. Journal of Exposure Science & Environmental Epidemiology 29(2):131-147. https://doi.org/10.1038/s41370-018-0094-1

Sznajder-Katarzyńska K, Surma M, Cieślik I. 2019. A review of perfluoroalkyl acids (PFASs) in terms of sources, applications, human exposure, dietary intake, toxicity, legal regulation, and

methods of determination. Journal of Chemistry 2019:1-20. https://doi.org/10.1155/2019/2717528

TemaNord. 2020. Retinoids in Mammalian Reproduction, with an Initial Scoping Effort to Identify Regulatory Methods. TemaNord 2020:507. Nordic Council of Ministers. ISBN 978-92-893-6530-7. https://doi.org/10.6027/temanord2020-507

Travis K, Pate I, Welsh ZK. 2005. The role of the benchmark dose in a regulatory context. Regulatory Toxicology and Pharmacology 43(3):280-291. https://doi.org/10.1016/j.yrtph.2005.07.003

Tukey JW. 1986. Sunset Salvo. The American Statistician 40(1):72–76. https://doi.org/10.2307/2683137

Törnqvist E, Annas A, Granath B, Jalkesten E, Cotgreave I, Öberg M. 2014. Strategic focus on 3R principles reveals major reductions in the use of animals in pharmaceutical toxicity testing. PLoS One 9(7):e101638. https://doi.org/10.1371/journal.pone.0101638

UNEP, United Nations Environment Programme. 2012. SC-9/12: Listing of perfluorooctanoic acid (PFOA), its salts and PFOA-related compounds. https://chm.pops.int/TheConvention/ThePOPs/ListingofPOPs [accessed 31/03/2022]

UNEP, United Nations Environment Programme Chemicals and Waste Branch. 2016. Consolidated Assessment of Efforts Made Toward the Elimination of Polychlorinated Biphenyls. https://wedocs.unep.org/20.500.11822/31249 [accessed 31/03/2022]

UNEP, United Nations Environment Programme. 2017. SC-4/17: Listing of perfluorooctane sulfonic acid, its salts and perfluorooctane sulfonyl fluoride. https://chm.pops.int/Portals/0/download.aspx?d=UNEP-POPS-COP.4-SC-4-17.English.pdf [accessed 31/03/2022]

UNEP, United Nations Environment Programme. 2021. All POPs listed in the Stockholm Convention. https://chm.pops.int/TheConvention/ThePOPs/AllPOPs/tabid/2509/Default.aspx [accessed 31/03/2022]

US EPA, United States Environmental Protection Agency. 2003. Developing Relative Potency Factors For Pesticide Mixtures: Biostatistical Analyses Of Joint Dose-Response. EPA/600/R-32/052. U.S. Environmental Protection Agency, Washington, USA. https://cfpub.epa.gov/ncea/risk/hhra/recordisplay.cfm?deid=66273 [accessed 31/03/2022]

US EPA, United States Environmental Protection Agency. 2012. Benchmark dose technical guidance. EPA/100/R-12/001. Washington, USA. https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf [accessed 31/03/2022]

US EPA, United States Environmental Protection Agency. 2018. Bayesian Models and Model Averaging Software (External Review Draft). Washington, USA. https://cfpub.epa.gov/ncea/bmds/recordisplay.cfm?deid=343986 [accessed 31/03/2022]

US EPA, United States Environmental Protection Agency. 2020. Benchmark Dose Software (BMDS) Version 3.2, User Guide. EPA/600/R-20/216. Washington, USA. https://www.epa.gov/bmds/benchmark-dose-software-bmds-32-user-guide-readme [accessed 31/03/2022]

Van den Berg M, Birnbaum LS, Denison M, De Vito M, Farland W, Feeley M, Fiedler H, Håkansson H, Hanberg A, Haws L, Rose M, Safe S, Schrenk D, Tohyama C, Tritscher A, Tuomisto J, Tysklind M, Walker N, Peterson RE. 2006. The 2005 World Health Organization Reevaluation of Human and Mammalian Toxic Equivalency Factors for Dioxins and Dioxin-Like Compounds. Toxicological Sciences 93(2):223-241. https://doi.org/10.1093/toxsci/kfl055

Varewyck M and Verbeke T. 2018. Software for benchmark dose modelling. EFSA Supporting Publications 14(2):1170E. https://doi.org/10.2903/sp.efsa.2017.EN-1170

Vinken M, Knapen D, Vergauwen L, Hengstler JG, Angrish M, Whelan M. 2017. Adverse outcome pathways: a concise introduction for toxicologists. Archives of Toxicology 91(11):3697-3707. https://doi.org/10.1007/s00204-017-2020-z

Voet H and Slob W. 2007. Integration of probabilistic exposure assessment and probabilistic hazard characterization. Risk Analysis 27(2):351-371. https://doi.org/10.1111/j.1539-6924.2007.00887.x

Wang H, Mao WF, Jiang DG, Liu SJ, Zhang L. 2021. Cumulative risk assessment of exposure to heavy metals through aquatic products in China. Biomedical and Environmental Sciences 34(8):606-615. https://doi.org/10.3967/bes2021.084

Wang Z, DeWitt JC, Higgins CP, Cousins IT. 2017. A never-ending story of per- and polyfluoroalkyl substances (PFASs)? Environmental Science & Technology 51(5):2508-2518. https://doi.org/10.1021/acs.est.6b04806

Wang Z, Walker GW, Muir DCG, Nagatani-Yoshida K. 2020. Toward a global understanding of chemical pollution: A first comprehensive analysis of national and regional chemical inventories. Environmental Science & Technology 4(5):2575-2584. https://doi.org/10.1021/acs.est.9b06379

Warren SFP, Müller F, Hopley JP, Karapally J, Hertl P. 1990. SAN 9789 H (norflurazon). A one year feeding study in beagle dogs to evaluate chronic toxicity. Unpublished report no. AGRO DOK CBK I.7399/90. Project 422-D, Sandoz Agro Ltd, Central Development Basel Toxicology, Basel, Switzerland. Submitted to WHO by Tessenderlo Kerley Inc. Phoenix, AZ, USA.

Wasserstein RL and Lazar NA. 2016. ASA statement on statistical significance and p-values. The American Statistician 70(2):131-133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein RL, Schirm AL, Lazar NA. 2019. Moving to a world beyond "$p < 0.05$". The American Statistician 73(sup1):1-19. https://doi.org/10.1080/00031305.2019.1583913

Wen LL, Lin LY, Su TC, Chen PC, Lin CY. 2013. Association between serum perfluorinated chemicals and thyroid function in US. Adults: The national health and nutrition examination survey 2007-2010. The Journal of Clinical Endocrinology & Metabolism 98(9):e1456-1464. https://doi.org/10.1210/jc.2013-1282

Wexler P and Hayes AN. 2018. Chapter 1. The Evolving Journey of Toxicology: A Historical Glimpse. In Klaasen CD (Ed.), *Casarett & Doull's Toxicology: The Basic Science of Poisons* (9th edition, p. 3-23). McGraw Hill. ISBN 978-1-25-986375-2.

Wheeler MW and Bailer AJ. 2007. Properties of model-averaged BMDLs: A study of model averaging in dichotomous response risk estimation. Risk Analysis 27(3):659-670. https://doi.org/10.1111/j.1539-6924.2007.00920.x

Wheeler MW and Bailer AJ. 2008. Model averaging software for dichotomous dose response risk estimation. Journal of Statistical Software 26(5):15. https://doi.org/10.18637/jss.v026.i05

Wheeler MW and Bailer AJ. 2009. Comparing model averaging with other model selection strategies for benchmark dose estimation. Environmental and Ecological Statistics 16(1):37-51. https://doi.org/10.1007/s10651-007-0071-7

Wheeler MW and Bailer AJ. 2013. An empirical comparison of low-dose extrapolation from points of departure (PoD) compared to extrapolations based upon methods that account for model uncertainty. Regulatory Toxicology and Pharmacology 67(1):75-82. https://doi.org/10.1016/j.yrtph.2013.06.006

Whitney M and Ryan L. 2013. Uncertainty due to low-dose extrapolation: modified BMD methodology for epidemiological data. Environmetrics 24(5):289-297. https://doi.org/10.1002/env.2217

WHO, World Health Organization. 2003. Polychlorinated biphenyls: human health aspects. Concise international chemical assessment document 55. World Health Organization, Geneva. ISBN 92-4-153055-3. https://apps.who.int/iris/bitstream/handle/10665/42640/9241530553.pdf [accessed 31/03/2022]

WHO FAO, World Health Organization & United Nations Food Agriculture Organization Joint Expert Committee on Food Additives Meeting. 2016. Eightieth report of the joint FAO/WHO expert committee on food additives: Evaluation of certain food additives and contaminants. ISBN: 978-92-4-069540-5. https://apps.who.int/iris/handle/10665/204410 [accessed 31/03/2022]

WHO IPCS, World Health Organization's International Programme on Chemical Safety. 2009. Environmental Health Criteria 239. Principles for modelling dose-response for the risk assessment of chemicals. ISBN 978-92-4-157239-2. https://apps.who.int/iris/bitstream/handle/10665/43940/9789241572392_eng.pdf [accessed 31/03/2022]

WHO IPCS, World Health Organization's International Programme on Chemical Safety. 2018. Guidance document on evaluating and expressing uncertainty in hazard characterization. ISBN 978-92-4-151354-8. https://inchem.org/documents/harmproj/harmproj/harmproj11.pdf [accessed 31/03/2022]

WHO IPCS, World Health Organization's International Programme on Chemical Safety. 2020. Chapter 5. Dose–response assessment and derivation of health-based guidance values. Environmental Health Criteria 240. Principles and methods for the risk assessment of

chemicals in food. ISBN 978-92-4-157240-8. https://www.who.int/docs/default-source/food-safety/publications/chapter5-dose-response.pdf?sfvrsn=32edc2c6_5 [accessed 31/03/2022]

Wold H. 1975. Soft modelling by latent variables: The non-linear iterative partial least squares (NIPALS) approach. Journal of Applied Probability 12(supl1):117-142. https://doi.org/10.1017/S0021900200047604

Xu Y, Li Y, Scott K, Lindh CH, Jakobsson K, Fletcher T, Ohlsson B, Andersson EM. 2020. Inflammatory bowel disease and biomarkers of gut inflammation and permeability in a community with high exposure to perfluoroalkyl substances through drinking water. Environmental Research 181:108923. https://doi.org/10.1016/j.envres.2019.108923

Xu Y, Nielsen C, Li Y, Hammarstrand S, Andersson EM, Li H, Olsson DS, Engström K, Pineda D, Lindh CH, Fletcher T, Jakobsson K. 2021. Serum perfluoroalkyl substances in residents following long-term drinking water contamination from firefighting foam in Ronneby, Sweden. Environment International 147:106333. https://doi.org/10.1016/j.envint.2020.106333

Yang L, Allen BC, Thomas RS. 2007. BMDExpress: A software tool for the benchmark dose analyses of genomic data. BMC Genomics 8(1):387. https://doi.org/10.1186/1471-2164-8-387

Yeung LWY and Mabury SA. 2016. Are humans exposed to increasing amounts of unidentified organofluorine? Environmental Chemistry 13(1):102-110. https://doi.org/10.1071/en15041

Zeilmaker M, Fragki S, Verbruggen E, Bokkers BGH, Lijzen J. 2018. Mixture exposure to PFAS: A Relative Potency Factor approach. RIVM Report 2018-0070. https://rivm.openrepository.com/handle/10029/622164 [accessed 31/03/2022]

Zhu M, Yuan Y, Yin H, Guo Z, Wei X, Qi X, Liu H, Dang Z. 2022. Environmental contamination and human exposure of polychlorinated biphenyls (PCBs) in China: A review. Science of the Total Environment 805:150270. https://doi.org/10.1016/j.scitotenv.2021.150270

Öberg M. 2010. Benchmark dose approaches in chemical health risk assessment in relation to number and distress of laboratory animals. Regulatory Toxicology and Pharmacology 58(3):451-454. https://doi.org/10.1016/j.yrtph.2010.08.015