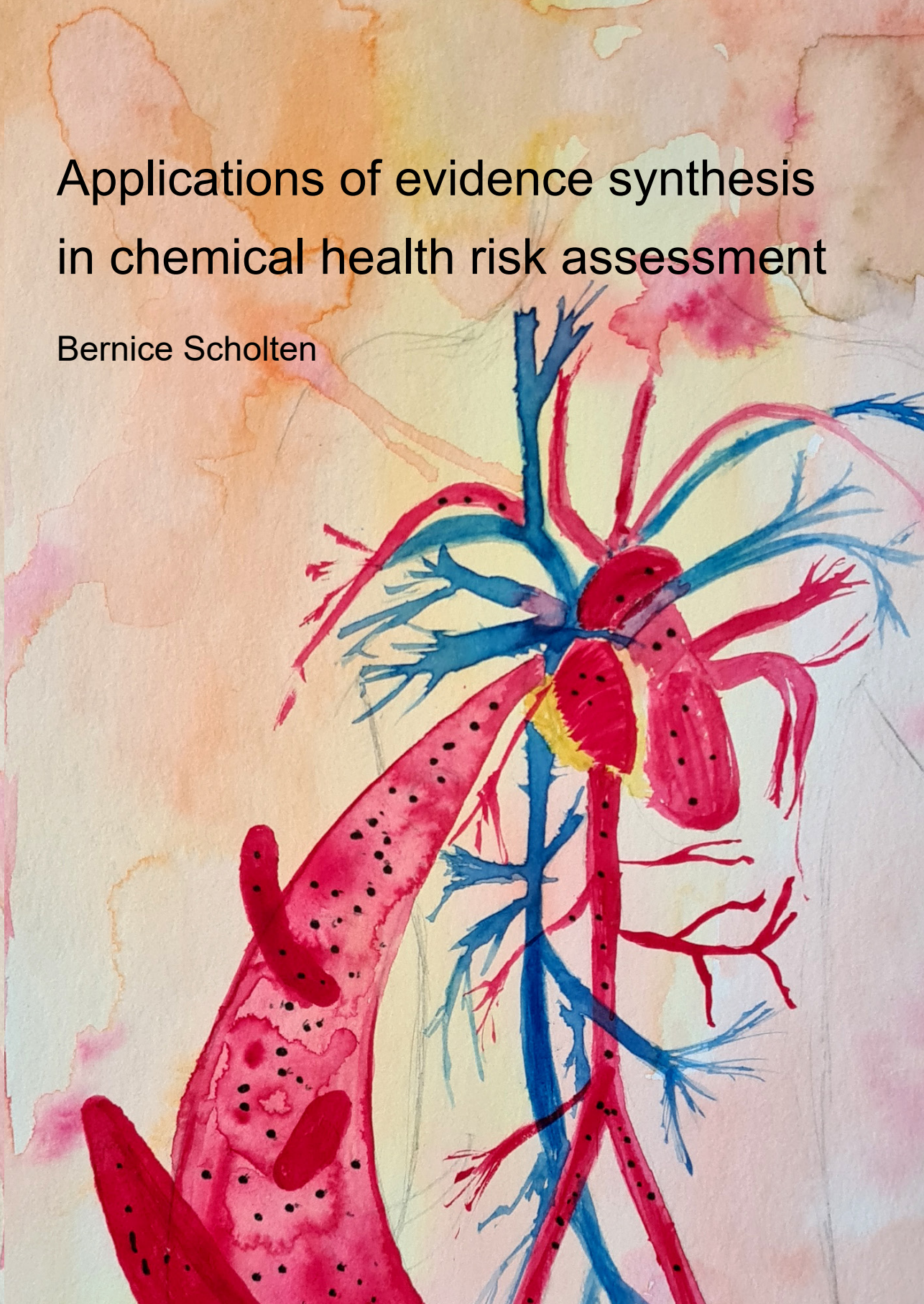
Applications of evidence synthesis in chemical health risk assessment

Bernice Scholten



Applications of evidence synthesis in chemical health risk assessment

Bernice Scholten

Bernice Jannigje Wilhelmina Scholten
Applications of evidence synthesis in chemical health risk assessment
Doctoral thesis, Utrecht University

Cover art: Janny Scholten
Content layout: Lena Schulte-Uebbing
Printing and binding by: Print Service Ede, Ede, the Netherlands.

ISBN: 978-90-393-7563-1

Applications of evidence synthesis in chemical health risk assessment

Toepassingen van bewijssynthese bij de beoordeling van chemische gezondheidsrisico's

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 27 juni 2023 des middags te 2.15 uur

door

Bernice Jannigje Wilhelmina Scholten

geboren op 20 juli 1986
te Hellevoetsluis

Promotor:
Prof. dr. R.C.H. Vermeulen

Copromotoren:
Dr. J.J. Vlaanderen
Dr. R. Stierum

Beoordelingscommissie:
Dr. K. Audouze
Prof. dr. R.J. de Boer
Prof. dr. R. Gehring
Prof. dr. ir. J. Legler
Prof. dr. F.G.M. Russel

Some studies in this thesis were financially supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 733032

Table of Contents

Chapter 1	General Introduction	1
Chapter 2	Automated Network Assembly of Mechanistic Literature for Informed Evidence Identification to Support Cancer Risk Assessment	13
Chapter 3	Biomonitoring for occupational exposure to diisocyanates: a systematic review	51
Chapter 4	A physiologically-based kinetic (PBK) model for work-related diisocyanate exposure: relevance for the design and reporting of biomonitoring studies	81
Chapter 5	A Quantitative Meta-Analysis of the Relation between Occupational Benzene Exposure and Biomarkers of Cytogenetic Damage	123
Chapter 6	Estimation of the exposure response relation between benzene and acute myeloid leukemia by combining epidemiological, human biomarker, and animal data	161
Chapter 7	General Discussion	199
Appendices	Summary	216
	Samenvatting	222
	Scientific Publications	228
	Dankwoord	229
	Curriculum Vitae	231

Chapter 1

Introduction

Chemical risk assessment

Risk assessment (of chemicals) is defined as *the use of the factual base to define the health effects of exposure of individuals or populations to hazardous materials and situations* (NRC 1983). This assessment is historically organized in several steps (Figure 1). After the risk assessment case (problem) has been formulated, the following step is to determine if the chemical of interest is causally linked to particular health effects, that is the identification of the intrinsic hazard. The hazard characterization, or dose(exposure)-response assessment, involves assessing the relation between the magnitude of exposure and the probability of health effect occurrence. In the exposure assessment step, the extent of exposure in the given population is studied. Subsequently, the hazard and exposure characterization are combined in the risk characterization step, which describes the nature and magnitude of human risk.

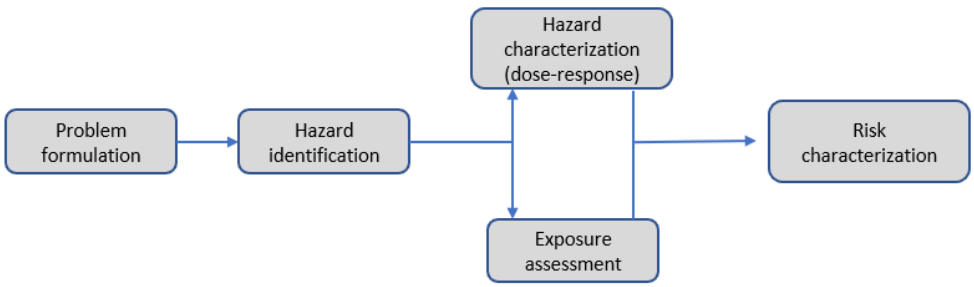


Figure 1. Chemical risk assessment paradigm (National Academies of Sciences Engineering and Medicine 2017)

Chemical risk assessment evidence streams

There are various types of evidence streams available for studying chemical induced effects, including observational epidemiology, molecular epidemiology, animal and mechanistic studies. Figure 2 displays an overview of different evidence streams of chemical risk assessment and their position in the exposure – disease continuum. The figure is partly based on the adverse outcome pathway (AOP) concept which links a molecular initiating event (MIE) to an adverse outcome (AO) via a set of key events (KE) (Ankley et al. 2010). The different evidence streams and their use for chemical risk assessment are discussed below, in order of relevance to human health outcomes.

Epidemiologic studies measure effects in the population of interest based on realistic exposures and account for population variability. However these studies often rely on less accurate exposure estimates, could suffer from potential confounding, and findings in epidemiological studies can be influenced due to reverse causation and selection bias (Hernandez et al. 2019; Loomis et al. 2018). In addition, whereas human

clinical trials are performed to assess efficacy and safety of drugs in development, for environmental or novel chemicals, it is unethical to perform controlled studies to obtain safety data in humans.

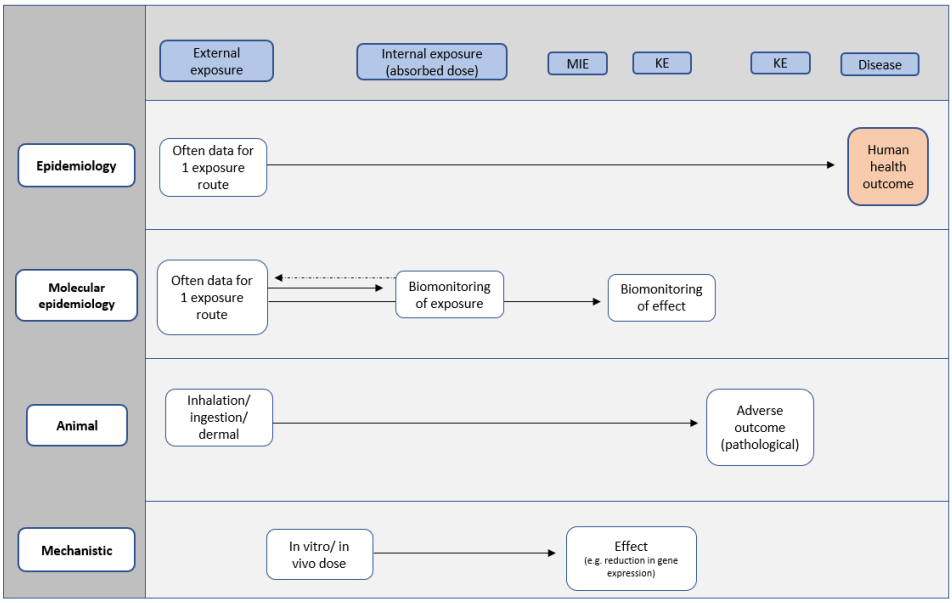


Figure 2. Overview of different evidence streams and their position in the exposure – disease continuum. The figure is partly based on the adverse outcome pathway (AOP) concept which link a molecular initiating event (MIE) to an adverse outcome (AO) via (a set of) key events (KE) (Ankley et al. 2010).

Molecular epidemiology bridges epidemiology and molecular biology research by measuring the internal dose via biomarkers of exposure, and biologic effects via biomarkers of effect. (Perera and Whyatt 1994). Like epidemiologic studies, molecular epidemiology studies may account for the differences in target population and individual susceptibility. It has been proposed that chemical risk assessment will improve if the relationships between key events along the exposure-health paradigm are established. Biomonitoring is one such tool that can link external exposure and biologically effective dose or effects (McNally et al. 2012).

Exposure biomonitoring studies may be used to refine exposure assessment by accounting for all routes of exposure leading to the absorbed dose. Biomarkers of effects give an assessment of the effect of a chemical on physiological processes and can be an indicator of early onset of pathophysiology and/or disease. However, for compounds with a short half-life, biomonitoring data mostly reflects daily variation, and, from a risk assessment perspective, it is often unclear to which extent effects recorded in biomonitoring studies (for either short-lived and the more persistent compounds) relate to the development of a clinical (chronic) response (Clewell et al. 2008). In addition the interpretation of biomonitoring data for risk assessment is not

always straightforward given there is a lack of human biomonitoring guidance values or biomonitoring equivalents (Louro et al. 2019).

Experimental animal studies are generally used for both hazard identification and hazard characterization. Animal studies typically involve fitting an empirical dose-response curve to apical responses to derive a point of departure (POD) or benchmark dose (USEPA 2005). Using pharmacokinetic information on differences between test species and humans, the POD is subsequently converted to a human-equivalent POD (Crump, Chen, and Louis 2010). There is detailed guidance for conducting and reporting animal studies (i.e. the OECD guidelines), and these studies are generally accepted by risk assessment agencies. However, it is often unclear to which extent observed effects in animals, such as increased liver weight, can be extrapolated to human health effects.

In the past decades it has been common practice to use animal toxicity studies for studying hazardous effects in humans (Knight et al. 2021). However there is an increasing demand for innovation in chemical hazard and risk assessment. Global chemical production is continuing at high rates (Persson et al. 2022), while at the same time there are ongoing efforts to reduce, refine and replace animal use in chemical toxicity testing (European Union 2010; Törnqvist et al. 2014; Madden et al. 2020). This has contributed to the development of alternative approaches, including the consideration of mechanistic assays (Coady et al. 2019; Kavlock et al. 2018; Linkov et al. 2015). Mechanistic studies include a wide range of studies that examine effects at the tissue, cellular and molecular level (Carlson et al. 2020). By studying intermediate key events, these studies can inform on the chemical mode of action or support biological plausibility (Urban et al. 2020). Mechanistic cell-based assays are less expensive to perform and require either no or less animals. However, the extrapolation of *in vitro* findings to relevant human health effects is perhaps more difficult in comparison to the other evidence streams depicted in Figure 2.

Evidence synthesis

In this thesis the focus is on combining all available evidence within and across evidence streams, i.e. evidence synthesis, to strengthen chemical risk assessment. The starting-point is that different types of evidence streams have different strengths and limitations, and they may be combined to complement each other. For example, if information on a health outcome is combined with clear measures of exposure at the biological site of action (or a surrogate for the site of action such as serum), this can strengthen causal inference based on epidemiological evidence (National Academies of Sciences Engineering and Medicine 2017).

To better formalize evidence evaluation for risk assessment within and across evidence streams, several institutes have published guidance for the consideration and integration of evidence for chemical risk assessment (i.e. NTP 2019; ANSES

2016; IARC 2019). In general these guidance approaches distinguish between evidence assembly (identify, filter and organize the evidence based on consideration of relevance and reliability), evidence weighing (assess the relevance and reliability of evidence) and evidence integration (EFSA 2017).

By rating confidence in qualitative terms such as high, sufficient, or low, a conclusion can be drawn on the consistency and strength of evidence across evidence streams. An example is the weight of evidence (WoE) concept developed by the WHO(WHO 2009). In addition the International Agency for Research on Cancer (IARC) hazard assessment considers and qualitatively integrates three types of evidence of streams (evidence of cancer in humans, animals and mechanistic studies) in reaching overall classifications of carcinogenicity (IARC 2019; Samet et al. 2020). Evidence weighting and rating is particularly useful for hazard identification but to a lesser extent for deriving a point of departure (POD) for risk characterization, i.e. hazard characterisation. The European Food Safety Authority (EFSA) provides examples of quantitative approaches for evidence integration including statistical methods and in silico tools (EFSA 2017). Examples of in silico tools are quantitative structure–activity relationships (QSARs) and physiologically based kinetic (PBK) models. QSARs aim to relate features of physico-chemical structure to a property or biological activity, so that the biological activity of new or untested chemicals can be inferred from similar molecular structures (e.g. Schultz et al. 2003). PBK models use differential equations to describe the absorption, distribution, metabolism and elimination processes that impact the fate and transport of chemicals (Paini et al. 2019). For a given exposure scenario, PBK models can predict organ-level concentration-time profiles in target organs (forward dosimetry). Alternatively, PBK models may be used to reconstruct external exposure based on internal exposure concentrations (reverse dosimetry). In silico tools are well reflected in the Integrated Testing Strategies (ITS), which aim to incorporate multiple data from various evidence streams in a more mathematically efficient and biologically informed way (Judson et al. 2013). The focus of ITS, however, is more limited on the development and combination of methods within *in vitro* cell non-animal based models. Besides in silico tools for hazard and bioavailability assessment, other quantitative approaches for evidence synthesis include statistical methods such as meta-analysis or meta-regression. Both meta-analysis and meta-regressions are regularly applied when combining either multiple animal or epidemiologic studies.

The quantitative approaches described above are mostly used to combine evidence within the same evidence stream. Quantitatively combining evidence across evidence streams is less common, in particular for regulatory purposes. The National Academies of Sciences, Engineering, and Medicine (NAS, 2017), concluded that to date insufficient attention has been given to data integration across multiple streams of evidence. Specifically, NAS recommended more methodological research and case-studies to better employ quantitative approaches in evidence integration for risk

assessment, including the use of meta-analysis, meta-regressions, and Bayesian statistics.

Aim of this thesis

Various evidence streams can all contribute to chemical risk assessment. The goal of this thesis is to further develop and test approaches to integrate different types of evidence for chemical risk assessment in a quantitative and systematic manner, both within and across evidence streams. Research questions addressed in this thesis include:

i) How can the large volume of mechanistic data be efficiently summarized?
ii) Will exposure assessment and hazard characterization benefit from the inclusion of human molecular epidemiology studies?
iii) How can data that is inherently different from each other be combined in a meaningful way?

Most work performed in this thesis is on two exemplary data rich chemicals, benzene and diisocyanates. The chemicals differ in health outcomes and kinetic behaviors. A large evidence base is available for both chemicals, including well-known biomarkers of exposure and effect. As such, benzene and diisocyanates serve as architype chemicals to evaluate how novel approaches of evidence integration can improve chemical risk assessment.

Outline of thesis

Below a description is given on the various statistical and biological informed approaches which have been used in this thesis to integrate different types of evidence, both within and across evidence streams.

Hazard identification

To be able to evaluate whether the chemical of interest is causally linked to a particular health effect, available evidence from all evidence streams needs to be organized and evaluated. This process is particularly labor intensive for mechanistic evidence because of the high volume and large diversity of study types (Guyton 2018). In Chapter 2 - *Automated network assembly of mechanistic literature for informed evidence identification to support cancer risk assessment* –an approach is presented to more efficiently and transparently identify and organize mechanistic evidence by using text mining techniques combined with molecular biological and network visualization tools.

Exposure assessment

In certain circumstances (for example when measuring external exposure is technically challenging, or when there are multiple exposure routes), human

biomonitoring can be the most reliable exposure assessment methodology as it provides an estimate of the internal exposure, i.e. the absorbed dose, of a chemical by integrating exposure from all routes (McNally et al. 2014). This also applies for diisocyanates, and therefore a range of diisocyanate biomonitoring studies have been undertaken for a variety of industrial applications, exposure scenarios and exposure biomarkers. The aim of Chapter 3 - *Biomonitoring for occupational exposure to diisocyanates: a systematic review* - was to use a systematic review approach to synthesize all available data on biomarkers of exposure, as to get a comprehensive overview of diisocyanate exposure levels.

For risk assessment purposes, exposure estimates need to be related to health based guidance values. These are often based on external exposure estimates. To be able to interpret biomonitoring data, the relationship between external and internal exposure has to be characterized, either by deriving correlation formulas, or by using PBK modeling. An advantage of using PBK models is the possibility of studying and characterizing the impact of various factors on the external-internal exposure relationship, such as working conditions (e.g. different respiration rates and exposure routes) and individual differences in physiology (McNally et al. 2012). The aim of Chapter 4 - *A physiologically-based kinetic (PBK) model for work-related diisocyanate exposure: relevance for the design and reporting of biomonitoring studies and opportunities for further improvement* – was to develop a PBK model for diisocyanates and its metabolites and provide insight into relative changes in urinary predictions upon changing exposure scenarios. Global sensitivity analyses were used to quantify uncertainty in exposure estimates, and to identify which parameters contributed most to the uncertainty in simulated biomarker levels. In addition, the model was used to calculate biological limit values, corresponding to current health based guidance values.

Hazard characterisation

Chapter 5 and 6 present innovative statistical approaches for hazard characterisation by summarizing all evidence on exposure-response relations within and across evidence streams.

Chapter 5 focused on biomarker studies because these may offer more precise information on the exact shape of the exposure-response relation, especially at low exposures. For benzene this is important because there is evidence that the exposure-response relation with cancers might be non-linear (Kim et al. 2006; Rothman et al. 1998; Vlaanderen et al. 2009). However, quantitative meta-analyses that provide summary estimates for a given change in the continuous variable of exposure are sparse for the molecular epidemiology/biomarker evidence stream. In Chapter 5 - *A quantitative meta-analysis of the relation between occupational benzene exposure and biomarkers of cytogenetic damage* - an approach is presented to quantitatively

combine all biomonitoring evidence of benzene induced genotoxicity in a meta-analysis.

Chapter 6 - *Estimation of the benzene-acute myeloid leukemia exposure-response curve by combining epidemiological, animal and human biomarker data* – explored to which extent integrating heterogenous data from different evidence streams can enhance chemical risk assessment, i.e. better characterize the exposure response relationship. This chapter focused on benzene induced acute myeloid leukemia (AML). For all available animal studies, molecular epidemiology studies, and epidemiological studies on this topic, results were extracted and summarized using Bayesian meta-regression models. Estimates of the exposure-response curve (ERC) are compared when using all available data versus using data from AML epidemiologic studies only. Moreover, an aim was to explicitly describe the steps required to harmonize data across evidence streams, including important underlying assumptions.

In Chapter 7 – *Discussion* – a critical overview is provided with reflections on how different advanced approaches for evidence synthesis may improve chemical risk assessment. First the evidence integration approaches applied for benzene and diisocyanates is considered, followed by a reflection on overarching methodological issues that emerged over these chapters.

References Introduction

Ankley, Gerald T., Richard S. Bennett, Russell J. Erickson, Dale J. Hoff, Michael W. Hornung, Rodney D. Johnson, David R. Mount, et al. 2010. “Adverse Outcome Pathways: A Conceptual Framework to Support Ecotoxicology Research and Risk Assessment.” *Environmental Toxicology and Chemistry* 29 (3): 730–41. https://doi.org/10.1002/ETC.34.

ANSES. 2016. “Evaluation Du Poids Des Preuvesal’Anses: Revue Critique de La Litterature et Recommandationsal’etape d’identification Des Dangers. Rapport d’expertise Collective.”

Carlson, Laura M., Frances A. Champagne, Deborah A. Cory-Slechta, Laura Dishaw, Elaine Faustman, William Mundy, Deborah Segal, et al. 2020. “Potential Frameworks to Support Evaluation of Mechanistic Data for Developmental Neurotoxicity Outcomes: A Symposium Report.” *Neurotoxicology and Teratology* 78 (March): 106865. https://doi.org/10.1016/J.NTT.2020.106865.

Clewell, Harvey J., Yu Mei Tan, Jerry L. Campbell, and Melvin E. Andersen. 2008. “Quantitative Interpretation of Human Biomonitoring Data.” *Toxicology and Applied Pharmacology* 231 (1): 122–33. https://doi.org/10.1016/J.TAAP.2008.04.021.

Coady, Katie, Patience Browne, Michelle Embry, Thomas Hill, Eeva Leinala, Thomas Steeger, Lidka Maślankiewicz, and Tom Hutchinson. 2019. “When Are Adverse Outcome Pathways and Associated Assays ‘Fit for Purpose’ for Regulatory Decision-Making and Management of Chemicals?” *Integrated Environmental Assessment and Management* 15 (4): 633–47. https://doi.org/10.1002/IEAM.4153.

Crump, Kenny S., Chao Chen, and Thomas A. Louis. 2010. “The Future Use of in Vitro Data in Risk Assessment to Set Human Exposure Standards: Challenging Problems and Familiar Solutions.” *Environmental Health Perspectives* 118 (10): 1350–54. https://doi.org/10.1289/EHP.1001931.

EFSA. 2017. “Guidance on the Use of the Weight of Evidence Approach in Scientific Assessments.” *EFSA Journal* 15 (8). https://doi.org/10.2903/J.EFSA.2017.4971.

European Union. 2010. “Directive 2010/63/EU of the Eurpean Parliament and of the Council of 22 September 2010 on the Protection of Animals Used for Scientific Purposes.”

Hernandez, Antonio F., Aleksandra Buha, Carolina Constantin, David R. Wallace, Dimosthenis Sarigiannis, Monica Neagu, Biljana Antonijevic, A. Wallace Hayes, Martin F. Wilks, and Aristidis Tsatsakis. 2019. “Critical Assessment and Integration of Separate Lines of Evidence for Risk Assessment of Chemical Mixtures.” *Archives of Toxicology 2019 93:10* 93 (10): 2741–57. https://doi.org/10.1007/S00204-019-02547-X.

IARC. 2019. “Preamble to the IARC Monographs, Amended January 2019.” 2019. https://monographs.iarc.who.int/wp-content/uploads/2019/07/Preamble-2019.pdf.

Judson, Richard, Robert Kavlock, Matthew Martin, David Reif, Keith Houck, Thomas Knudsen, Ann Richard, et al. 2013. “Perspectives on Validation of High-Throughput Assays Supporting 21st Century Toxicity Testing.” *ALTEX* 30 (1): 51. https://doi.org/10.14573/ALTEX.2013.1.051.

Kavlock, Robert J., Tina Bahadori, Tara S. Barton-Maclaren, Maureen R. Gwinn, Mike Rasenberg, and Russell S. Thomas. 2018. “Accelerating the Pace of Chemical Risk Assessment.” *Chemical Research in Toxicology* 31 (5): 287–90. https://doi.org/10.1021/ACS.CHEMRESTOX.7B00339/ASSET/IMAGES/MEDIUM/TX-2017-003397_0001.GIF.

Kim, Sungkyoon, Roel Vermeulen, Suramya Waidyanatha, Brent A. Johnson, Qing Lan, Nathaniel Rothman, Martyn T. Smith, et al. 2006. “Using Urinary Biomarkers to Elucidate Dose-Related Patterns of Human Benzene Metabolism.” *Carcinogenesis*. https://doi.org/10.1093/carcin/bgi297.

Knight, Derek J., Hubert Deluyker, Qasim Chaudhry, Jean Marc Vidal, and Alie de Boer. 2021. “A Call for Action on the Development and Implementation of New Methodologies for Safety Assessment of Chemical-Based Products in the EU – A Short Communication.” *Regulatory Toxicology and Pharmacology* 119 (February): 104837. https://doi.org/10.1016/J.YRTPH.2020.104837.

Linkov, Igor, Olivia Massey, Jeff Keisler, Ivan Rusyn, and Thomas Hartung. 2015. “From ‘Weight of Evidence’ to Quantitative Data Integration Using Multicriteria Decision Analysis and Bayesian Methods.” *ALTEX* 32 (1): 3. https://doi.org/10.14573/ALTEX.1412231.

Loomis, Dana, Neela Guha, Amy L Hall, and Kurt Straif. 2018. “Identifying Occupational Carcinogens: An Update from the IARC Monographs.” *Occup Environ Med* 75: 593–603. https://doi.org/10.1136/oemed-2017-104944.

Louro, Henriqueta, Milla Heinälä, Jos Bessems, Jurgen Buekers, Theo Vermeire, Marjolijn Woutersen, Jacqueline van Engelen, et al. 2019. “Human Biomonitoring in Health Risk Assessment in Europe: Current Practices and Recommendations for the Future.” *International Journal of Hygiene and Environmental Health* 222 (5): 727–37. https://doi.org/10.1016/J.IJHEH.2019.05.009.

Madden, Judith C., Steven J. Enoch, Alicia Paini, and Mark T.D. Cronin. 2020. “A Review of In Silico Tools as Alternatives to Animal Testing: Principles, Resources and Applications.” *Alternatives to Laboratory Animals : ATLA* 48 (4): 146–72. https://doi.org/10.1177/0261192920965977.

McNally, Kevin, Richard Cotton, John Cocker, Kate Jones, Mike Bartels, David Rick, Paul Price, and George Loizou. 2012. “Reconstruction of Exposure to M-Xylene from Human Biomonitoring Data Using PBPK Modelling, Bayesian Inference, and Markov Chain Monte Carlo Simulation.” *Journal of Toxicology* 2012. https://doi.org/10.1155/2012/760281.

National Academies of Sciences Engineering and Medicine. 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. National Academies Press. https://doi.org/10.17226/24635.

National Academies of Sciences Engineering, and Medicine. 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. National Academies Press. https://doi.org/10.17226/24635.

NRC. 1983. “Risk Assessment in the Federal Government: Managing the Process.” *National Academy Press*.

NTP. 2019. “Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration.” https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf.

Paini, A., J. A. Leonard, E. Joossens, J. G.M. Bessems, A. Desalegn, J. L. Dorne, J. P. Gosling, et al. 2019. “Next Generation Physiologically Based Kinetic (NG-PBK) Models in Support of Regulatory Decision Making.” *Computational Toxicology* 9 (February): 61–72. https://doi.org/10.1016/J.COMTOX.2018.11.002.

Perera, F. P., and R. M. Whyatt. 1994. “Biomarkers and Molecular Epidemiology in Mutation/Cancer Research.” *Mutation Research/Environmental Mutagenesis and*

Related Subjects 313 (2–3): 117–29. https://doi.org/10.1016/0165-1161(94)90043-4.

Persson, Linn, Bethanie M. Carney Almroth, Christopher D. Collins, Sarah Cornell, Cynthia A. de Wit, Miriam L. Diamond, Peter Fantke, et al. 2022. "Outside the Safe Operating Space of the Planetary Boundary for Novel Entities." *Environmental Science and Technology* 56 (3): 1510–21. https://doi.org/10.1021/ACS.EST.1C04158/ASSET/IMAGES/LARGE/ES1C04158_0002.JPEG.

Rothman, N., W. E. Bechtold, S. N. Yin, M. Dosemeci, G. L. Li, Y. Z. Wang, W. C. Griffith, M. T. Smith, and R. B. Hayes. 1998. "Urinary Excretion of Phenol, Catechol, Hydroquinone, and Muconic Acid by Workers Occupationally Exposed to Benzene." *Occupational and Environmental Medicine*. https://doi.org/10.1136/oem.55.10.705.

Samet, Jonathan M., Weihsueh A. Chiu, Vincent Cogliano, Jennifer Jinot, David Kriebel, Ruth M. Lunn, Frederick A. Beland, et al. 2020. "The IARC Monographs: Updated Procedures for Modern and Transparent Evidence Synthesis in Cancer Hazard Identification." *Journal of the National Cancer Institute*. https://doi.org/10.1093/jnci/djz169.

Schultz, T. Wayne, Mark T.D. Cronin, John D. Walker, and Aynur O. Aptula. 2003. "Quantitative Structure–Activity Relationships (QSARs) in Toxicology: A Historical Perspective." *Journal of Molecular Structure: THEOCHEM* 622 (1–2): 1–22. https://doi.org/10.1016/S0166-1280(02)00614-0.

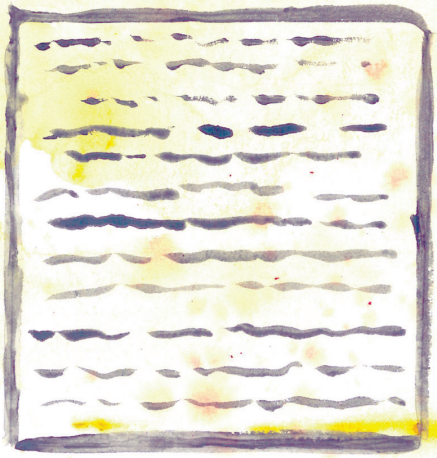
Törnqvist, E, A Annas, B Granath, E Jalkesten - PloS One, and Undefined 2014. 2014. "Strategic Focus on 3R Principles Reveals Major Reductions in the Use of Animals in Pharmaceutical Toxicity Testing." *PLoS ONE* 9 (7). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0101638.

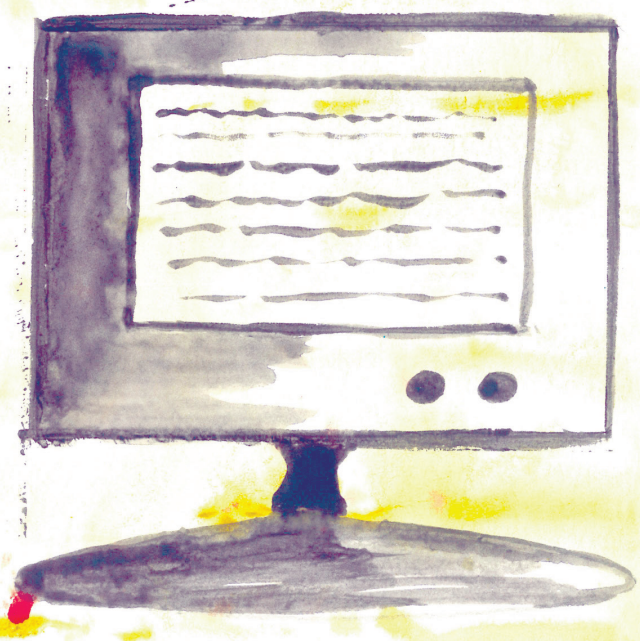
Urban, Jonathan D., Daniele S. Wikoff, Grace A. Chappell, Craig Harris, and Laurie C. Haws. 2020. "Systematic Evaluation of Mechanistic Data in Assessing in Utero Exposures to Trichloroethylene and Development of Congenital Heart Defects." *Toxicology* 436 (April): 152427. https://doi.org/10.1016/J.TOX.2020.152427.

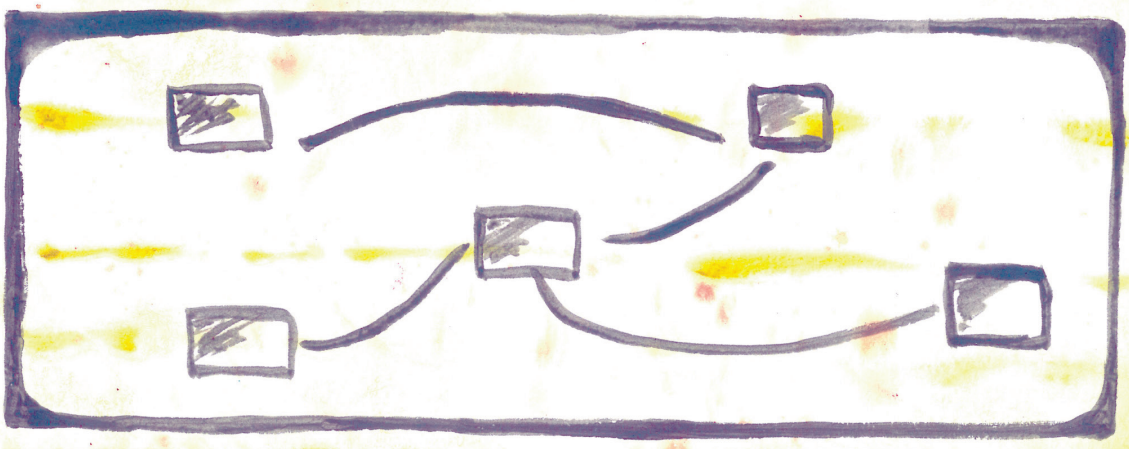
USEPA. 2005. "Guidelines for Carcinogen Risk Assessment." *Risk Assessment Forum US Environmental Protection Agency* 1 (DC EPA/630/P-03 F).

Vlaanderen, Jelle, Nathaniel Rothman, Lützen Portengen, Qing Lan, Hans Kromhout, and Roel Vermeulen. 2009. "Flexible Meta-Regression to Assess the Shape of the Benzene–Leukemia Exposure–Response Curve." *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.0901127.

WHO. 2009. "Food Safety. Project to Update the Principles and Methods for Theassessment of Chemicals in Food. Principles and Methods for the Risk Assessment of Chemicals in Food." *EHC 240*.







Chapter 2

Automated Network Assembly of Mechanistic Literature for Informed Evidence Identification to Support Cancer Risk Assessment

Bernice Scholten, Laura Guerrero Simón, Shaji Krishnan, Roel Vermeulen, Anjoeka Pronk, Benjamin M. Gyori, John A. Bachman, Jelle Vlaanderen*, Rob Stierum*

*authors contributed equally

Published in: *Environmental Health Perspectives,* 2022, 130(3), 037002.

Abstract

Background

Mechanistic data is increasingly used in hazard identification of chemicals. However, the volume of data is large, challenging the efficient identification and clustering of relevant data.

Objectives

We investigated whether evidence identification for hazard assessment can become more efficient and informed through an automated approach that combines machine reading of publications with network visualization tools.

Methods

We chose 13 chemicals that were evaluated by the International Agency for Research on Cancer (IARC) *Monographs* program incorporating the key characteristics of carcinogens (KCCs) approach. Using established literature search terms for KCCs, we retrieved and analysed literature using Integrated Network and Dynamical Reasoning Assembler (INDRA). INDRA combines large-scale literature processing with pathway databases and extracts relationships between biomolecules, bioprocesses and chemicals into statements (*e.g.*, "benzene activates DNA damage"). These statements were subsequently assembled into networks and compared with the KCC evaluation by IARC, to evaluate the informativeness of our approach.

Results

We found, in general, larger networks for those chemicals which the IARC has evaluated the evidence to be strong for KCC induction. Larger networks were not directly linked to publication count, given that we retrieved small networks for several chemicals with little support for KCC activation according to the IARC, despite the significant volume of literature for these specific chemicals. Additionally, interpreting networks for genotoxicity and DNA repair showed concordance with the IARC KCC evaluation.

Discussion

Our method is an automated approach to condense mechanistic literature into searchable and interpretable networks based on an *a prior* ontology. The approach is no replacement of expert evaluation, but, instead, provides an informed structure for experts to quickly identify which statements are made in which papers and how these could connect. We focused on the KCCs because these are supported by well-described search terms. The method needs to be tested in other frameworks as well to demonstrate its generalizability.

Introduction

Risk assessment of chemicals is commonly based on toxicological or epidemiological studies. Mechanistic studies can be used to complement animal or epidemiological data to inform mechanisms of toxicity, dose-response assessment and hazard identification (National Academies of Sciences Engineering and Medicine 2017). However, it is generally recognized that summarizing mechanistic data is challenging, in part because of the large diversity of study types and the high volume of available studies (EFSA 2018). At present, there is still no generally accepted procedure to structure, analyse and interpret mechanistic studies in an efficient way (Guyton et al. 2018; Wikoff et al. 2019). Further, the process of evaluating available mechanistic data, including reading manuscripts and evaluating the associated data, is labor intensive.

There is growing interest in using machine learning and other approaches to reduce the human burden in screening studies for relevance and facilitate systematic review processes (Howard et al. 2016). For example the “Sciome Workbench for Interactive computer-Facilitated Text-mining” (SWIFT)-review tool has been developed to identify and visualize whether the currently available data for a chemical of interest is rich or poor. The Table Builder and Health Assessment Workplace Collaborative (HAWC) are tools meant to share results of systematic review searches and risk of bias assessments, and include possibilities for data analyses(Shapiro et al. 2018). Other researchers have applied a bioinformatics approach to structure and analyse mechanistic data. Carvaillo et al.(2019), for example, combined text mining and systems biology by creating a tool, the AOP-helpFinder, that enriches adverse outcome pathways (AOPs). This tool could assist risk assessors in the identification of relevant associations between certain chemicals of interest and AOP components. Guha et al. (2016) combined information on chemical structure with database integration and automated text mining, and as such prioritized agents for hazard identification. Here we propose an approach for identification and prioritization of data and knowledge for use in hazard characterization of chemicals that combines text mining with network visualization tools. We apply our approach within the context of the International Agency for Research on Cancer (IARC) *Monographs* program for the evaluation of carcinogens, which evaluates mechanistic data using a well-defined framework and ontology: the 10 key characteristics of carcinogens (KCCs).

The KCCs have been recently identified by a series of IARC workshops (Smith et al. 2016). The IARC has used mechanistic data to strengthen conclusions on carcinogen classifications since 1991 (IARC 1992) but developed the 10 KCCs to create a more systematic method for the evaluation of mechanistic data to support hazard assessment for carcinogens. The KCCs comprise the properties of known human carcinogens (e.g., having genotoxic or immunosuppressive properties) and data on these characteristics can support the evidence of carcinogenicity (Smith et al. 2016*)*.

To retrieve information based on the KCCs from the scientific literature, the IARC monographs staff developed a working list of search terms for the KCCs (Guyton et al. 2018) (Table 1). In 2019 the Preamble to the IARC Monographs, which outlines procedures on scientific review and evaluation of carcinogenic hazards, was updated; the KCCs are now used as the basis for the evaluation of mechanistic data (Samet et al. 2020; IARC 2019).

Table 1. Ten key characteristics of carcinogens (KCCs) and corresponding search terms (taken from Guyton et al. 2018).

KCC	Description (label)[a]	Query	Search term[b]
1	Is electrophilic or can be metabolically activated (electrophilic)	1	"pharmacokinetics"[MeSH Terms] OR "pharmacokinetics"[Subheading] OR "absorption"[MeSH Terms] OR "distribution"[Title] OR "excretion"
2	Is genotoxic (genotoxic)	2	(“Mutation”[Mesh] OR “Cytogenetic Analysis”[Mesh] OR “Mutagens”[Mesh] OR “Oncogenes”[Mesh] OR “Genetic
3	Alters DNA repair or causes genomic instability (DNA repair)		Processes”[Mesh] OR "genomic instability"[MesH] OR chromosom* OR clastogen* OR “genetic toxicology” OR “strand break” OR “unscheduled DNA synthesis” OR “DNA damage” OR “DNA adducts” OR "SCE" OR "chromatid" OR micronucle* OR mutagen* OR "DNA repair" OR "UDS" OR "DNA fragmentation" OR "DNA cleavage")
4	Induces epigenetic alterations (epigenetics)	3	"rna"[MeSH] OR "epigenesis, genetic"[MesH] OR rna OR "rna, messenger"[MeSH] OR "rna" OR "messenger rna" OR mrna OR "histones"[MeSH] OR histones OR epigenetic OR miRNA OR methylation
5	Induces oxidative stress (oxidative stress)	4	"reactive oxygen species"[MeSH Terms] OR "reactive oxygen species"[All Fields] OR "oxygen radicals"[All Fields] OR "oxidative stress"[MeSH Terms] OR "oxidative"[All Fields] OR "oxidative stress"[All Fields] OR "free radicals"[All Fields]
6	Induces chronic inflammation	5	inflamm* or immun* or chemokine or cytokine or leukocyte or white blood cell
7	(inflammation) Is immunosuppressive (immunosuppressive)		
8	Modulates receptor-mediated effects (receptor-mediated)	6	"Hormones, Hormone Substitutes, and Hormone Antagonists"[Mesh] OR "Endocrine Disruptors"[Mesh] OR "Thyroid Hormones"[Mesh] OR "Estrogens"[Mesh] OR "Progesterone"[Mesh] OR "Receptors, Estrogen"[Mesh] OR "Receptors, Androgen"[Mesh] OR "Receptors, Progesterone"[Mesh] OR "Receptors, Thyroid Hormone"[Mesh] OR "Receptors, Aryl Hydrocarbon"[Mesh] OR "Peroxisome Proliferator-Activated Receptors"[Mesh] OR "constitutive androstane receptor"[Supplementary Concept] OR "farnesoid X-activated receptor"[Supplementary Concept] OR "liver X receptor"[Supplementary Concept] OR "Retinoid X Receptors"[Mesh]
9	Causes immortalization (immortalization).	7	"Cell Transformation, Neoplastic"[Mesh] OR "Cell Proliferation"[Mesh] OR apoptosis OR "necrosis"[MeSH] OR
10	Alters cell proliferation, cell death or nutrient supply (cell dynamics)		"DNA Replication"[Mesh] OR "Cell Cycle"[Mesh] OR brdu OR thymidine OR angiogenesis

[a] *In parenthesis, a label is provided for reference to the specific KCC in this paper.*
[b] *Truncated search terms are identified by an asterisk.*

We explored the use of the Integrated Network and Dynamical Reasoning Assembler (INDRA) (Gyori et al. 2017) coupled to the Reach natural language processing system (Valenzuela-Escárcega et al. 2018). INDRA aims to aggregate claims about causal biological and chemical mechanisms extracted by text-mining tools into a mechanistic *in silico* model. The type of *in silico* model can be defined *a priori* based on an evaluation framework. We imported the results of the literature search based on the queries by Guyton et al. (2018) into INDRA. INDRA retrieves so called "causal assertions" (i.e., statements in which an entity such as a small molecule or a protein interacts with or regulates another entity such as a protein or biological process) from the literature extracted by Reach, and performs a series of assembly steps to transform these relationships into networks. It needs to be mentioned that the term causal in the context of the computational INDRA environment does not automatically imply biological or toxicological causality, but is related to the strength of a computationally inferred value between entities (belief score (BS)).

Importantly, we do not present an approach for an automated hazard characterization. We compared the obtained networks with the IARC's evaluation to assess the informativeness of our approach to synthesize the available evidence into a predefined ontology (i.e., KCCs) and to suggest prioritization of KCCs for certain chemicals. We interpreted correspondence between the expert evaluation and our automated approach for informed evidence identification to indicate usefulness of our approach as a first step in evidence synthesis.

Methods

Figure 1 displays a comparison between our approach (Figure 1A), and the approach by IARC (Figure 1B), together with a potential application of our approach to aid in identifying and prioritizing relevant information for full-text review of included studies. We chose 13 chemicals that have been mechanistically evaluated in IARC Monographs 112 – 125 and were classified in different IARC carcinogen categories: benzene (1), pentachlorophenol (1), dichlorodiphenyltrichloroethane (DDT; 2A), hydrazine (2A), diazinon (2A), glyphosate (2A), malathion (2A), melamine (2B), parathion (2B), pyridine (2B), allyl chloride (3), β-picoline (3), and coffee (3).

Compounds were selected if they fulfilled the criteria of being evaluated by the IARC, that is, from Monograph 112 onward, based on the potential for induction of the KCCs (see "IARC evaluations for evidence of KCC activation" in the Supplemental Material), hereafter referred to as IARC evidence. After evaluating the assembled data, the IARC classified the evidence on the basis of collective expert judgment as strong, moderate, weak, or no evidence. These classifications are based on various criteria, as outlined in the IARC's Instructions for Authors (IARC 2017).

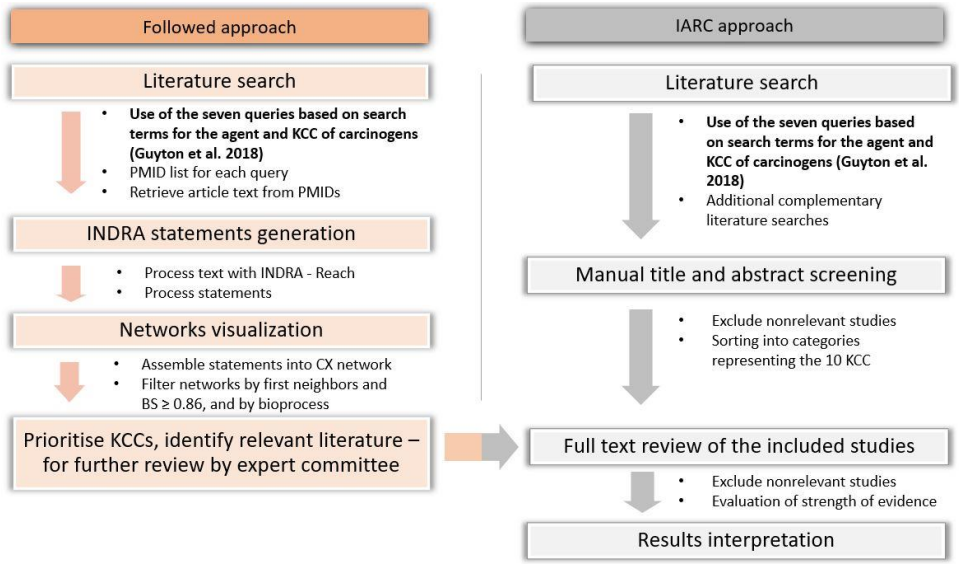


Figure 3. Two different approaches to investigate and summarize mechanistic data.(A) the followed approach and (B) the IARC approach. Note: BS, belief score; CX, a network exchange format; IARC, International Agency for Research on Cancer; INDRA, Integrated Network and Dynamical Reasoning Assembler; KCCs, key characteristics of carcinogens; PMID, PubMed identifier.

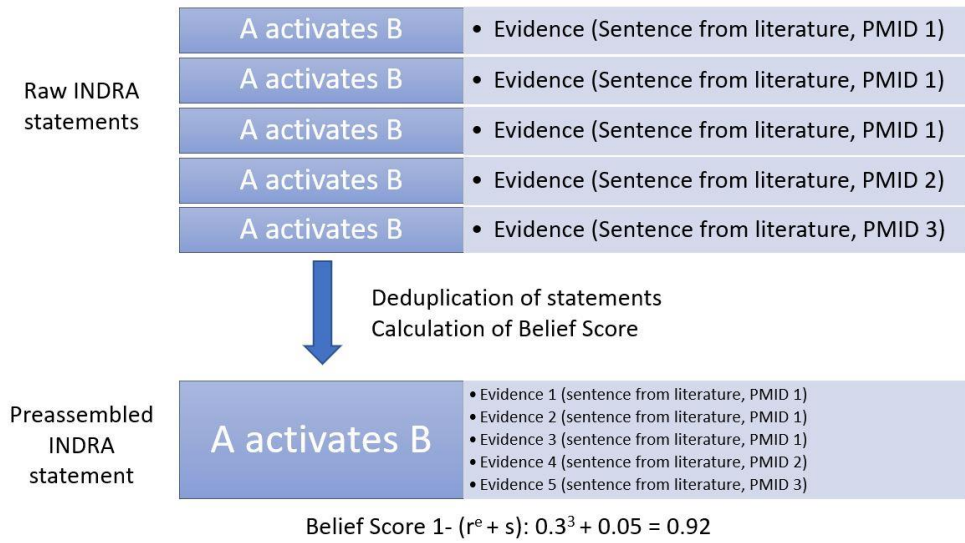


Figure 4. Integrated Network and Dynamical Reasoning Assembler (INDRA) preassembly of statements and calculation of a belief score (BS). The default preassembly function would count in this example 5 evidences (e), thus the BS would be $1 - (r^e + s) = +0.05 = 0.95$. *The function that calculates BS starts with empirical estimates of the prior random (r) and systematic error (s) rates of reading systems. In the present paper Reach was used, and its default built-in values are r=0.3 and s=0.05. With the correction we applied (referred to as deduplication of statements), evidence 1,2 and evidence 3 are counted as one, because they were retrieved from the same paper hence e =3, resulting in a BS of 0.92. Note: AIM2, absent in melanoma 2; Annexin-V-FLUOS, Annexin-V-fluorescence; Casp1, Caspase 1; DSB, double strand breaks; H2AX, phosphorylated H2AX; r, empirical estimate of the prior random rate; and s, empirical estimate of the systematic error rates; TET, ten-eleven-translocation.*

We started our approach with a literature search using the PubMed database, based on the predefined query search terms for the KCCs (Table 1; Guyton et al. 2018). Per query, the search terms were combined with the chemical name(s), as referred to by the IARC, contained within the article title and a date limitation from 1 January 1900 to 1 March 2020 in the following format: '*chemical name*[Title] AND (*Guyton search terms*) AND (1900/01/01[PDat] : 2020/03/01[PDat])'. Note that three chemicals were evaluated later (allyl chloride, pentachlorophenol, β-picoline), hence the search term for these chemicals was extended to October 2021. Each of these searches returned a list of PubMed identifiers [IDs (PMIDs)], each corresponding to an article. We used INDRA's literature retrieval module to obtain the full-text content or the abstract corresponding to each PMID returned by these searches. When available, the full-text was retrieved from PubMed Central or Elsevier [an application programming interface (API) key was used to get access to this content via the Elsevier Text and Data Mining API]. For PMIDs, for which the full-text content was not available, the abstract was retrieved.

The retrieved text content (for each of the chemicals: seven lists of article texts based on seven searches, spanning the 10 different KCCs) were then processed with Reach, an opensource natural language processing (NLP) system for the biomedical literature that is able to read and extract mechanistic descriptions of biological processes from text (Valenzuela-Escárcega et al. 2018). Reach is a type of "event extraction" system for biology that can detect and normalize information about putative interactions among biological entities and processes (Ananiadou et al. 2010). The system can recognize agents (e.g., proteins, bioprocesses, chemicals), link these to corresponding identifiers in knowledge bases [including UniProt, InterPro, Human Metabolome Database (HMDB), PubChem, and Geno Ontology (GO)], and extract events or interactions (e.g., multiple types of positive or negative regulation). To run Reach, we used the *reach* module of INDRA (specifically, the *process_text* and *process_nxml_str* functions) which provides a Python interface to running the Reach system and processing its extractions into INDRA Statements (see Gyori 2017). INDRA Statements represent a hypothetical, potentially causal influence relation between two agents (e.g., a chemical and a bioprocess). For example, for the sentence "Hydroquinone induces extensive apoptosis in the cells.", the generated INDRA Statement reads: "Activation(hydroquinone(PUBCHEM:785), apoptotic process(GO:0006915))," which represents that hydroquinone (recognized with the database identifier PUBCHEM:785) activates apoptosis (recognized with the database identifier GO:0006915). (The assignment of database identifiers to entity texts is known as named entity normalization or simply as grounding.)

We thus obtained a list of "raw" (i.e., unprocessed) INDRA Statement objects gathered from Reach after extracting relations from the content of each article's text retrieved in the previous step (Figure 2). Every raw statement contains a set of attributes that includes all information necessary to identify the given putative mechanism and its

participants being represented. Each statement also has an *evidence* attribute that contains additional provenance information and annotations, for instance text content references (i.e., PMIDs), and the specific sentence from which the statement was extracted, as well as, for instance, whether the sentence was recognized as a hypothetical statement. The evidence here and below is referred to as *technical evidence* emerging from text mining, which is not automatically and directly equal to biological/toxicological evidence in the context of hazard assessment.

After obtaining all raw statements from the retrieved text of each query, we applied several knowledge assembly steps using the *assemble_corpus* module of INDRA, with the aim of filtering, improving, deduplicating, and calculating BSs for the statements before assembling them into a network.

This assembly process consists of the following steps:

1. Filtering out hypothetical statements with the function *filter_no_hypothesis()*:
   The Reach system labels a statement as hypothesis when the evidence text for that statement contains one of the default words likely to belong to a hypothesis (e.g., test, consider, predict, speculate, suggest, theorize). This step removes all statements that have been labeled as hypotheses.
   Entity renormalization using the function *map_grounding()*:
   Entities from reading systems, such as Reach, are often incorrectly normalized (i.e., an incorrect database ID is assigned to them). INDRA integrates both expert-curated maps to improve entity normalization and machine-learned models [calling on thy Puthon package Acromine-based Disambiguation of Entities From Text context (Adeft) (Steppi, Gyori, and Bachman 2020) to choose between competing senses of ambiguous acronyms (e.g., “IR” can refer to the insulin receptor but also to ionizing radiation). INDRA also standardizes IDs [e.g., when available, it provides equivalent IDs for PubChem compounds in Chemical Entities of Biological Interest (ChEBI), Chemical Abstracts Service (CAS), ChEMBL and other databases)] and the names of agents to their standard names [e.g., HUGO Gene Nomenclature Committee (HGNC) gene symbols, GO labels]).
2. Filtering out agents without associated database identifiers with the function *filter_grounded_only()*.
3. Running preassembly with the function *run_preassembly()*:
   In the last step of assembly, statements are deduplicated (equivalent statements are merged into one statement) and the associated evidence is gathered in an evidence list (Figure 2). Subsequently, the BS are calculated by INDRA. For each INDRA Statement, the BS is a numerical value between 0 and 1, calculated as a function of the Statement’s supporting evidence. The function that calculates BS starts with empirical estimates of the prior random (r) and systematic error (s) rates of reading systems. In the present paper

Reach was used, and its default built-in values are $r = 0.3$ and $s = 0.05$. Coming from a single source, the error probability and BS are:

$$errorprobability = r^e + s,$$
$$BeliefScore = 1 - errorprobability$$

with *e* being the number of pieces of evidence for that statement. Thus, assuming an assembled statement has four pieces of evidence, the BS would be $1 - (0.3^4 + 0.05)$ = 0.94 (Figure 2). Hence, the BS is based on the amount of evidence, that is, the more evidence supporting the statement the higher the BS (Figure 2). These can be supportive, but do not directly refer to a "belief" by, for example, toxicological experts in the cumulative scientific community in the context of KCC hazard assessment.

To avoid counting repeated sentences from the same paper as distinct appearances of the same assertion, we counted sentences coming from the same paper as constituting only a single claim for the purpose of BS calculation. However, if a single paper provided evidence for different KCCs, all these evidence were taken into account. For each query, the number of PMIDs for which INDRA statements were obtained, was compared to the total number of PMIDs retrieved (expressed as percentage in between parenthesis, Table 2).

All programming steps were executed in the environment Spyder (version 3.3.6) and Python (version 3.7.4). The Python script can be found at: https://github.com/bernice493/INDRA_hazard. All information related to INDRA was retrieved from https://indra.readthedocs.io/en/latest/.

Once all the steps of this process were finished, the statements were assembled into a network using INDRA's *assemblers.cx* module. These networks were generated in CX format and visualized in Cytoscape (version 3.7.2); a bio-informatics software platform for visualizing molecular interaction networks (https://cytoscape.org/), and publicly available via National Data Exchange (Table S1). The resulting networks consist of nodes (rectangles) which represent biological entities, and edges (arrows between the nodes) which represent proposed biological or chemical interactions/mechanisms between these entities. Nodes are colored on the basis of the type of entity they represent: bioprocess (orange), chemicals (green), proteins (light blue), protein family (dark blue), and others (being nodes not classified into one of the before mentioned entities, gray). The edges can indicate different events, that is, activation, inhibition, complex formation, negative amount regulation, positive amount regulation, and post-translational modification, as implied by the underlying INDRA statements.

To reduce network complexity for further visual and statistical analysis (addressing sizes and support of the different networks), the following network filtering steps were

taken: a) the chemical of interest and its first neighbors (i.e., directly adjacent nodes) were selected and retained, b) only the nodes connected by edges with a BS ≥ 0.86 (two or more pieces of evidence) were retained (note that this is an arbitrary cut-off), and c) the nodes or group of nodes not connected to the main network (containing the chemical of interest as central node) were removed (see “Filtering networks” in the Supplemental Material). We also filtered the networks on the basis of the classes: bioprocesses and other processes, for KCC 2 (genotoxic) and KCC 3 (DNA repair), for example (see “Filtering Networks” in the Supplemental Materials). This filtering helps to focus the attention on potentially relevant biological processes that are possibly directly influenced by the chemical in question.

After creating and filtering the networks, additional quantitative descriptive network information was collected, including the number of nodes and edges. This information represents numerical values describing the network size. In addition, the sum of the BSs (sBS) obtained from all statements within each of the networks (seven networks per chemical, 10 chemicals) was calculated. This provides information on the overall abundance of claims supporting the edges contained within the network. The Wilcoxon rank test was used to compare the sBS metric with the IARC Monographs working group (i.e., IARC evidence) classifications.

Aside from serving as input to the network analyses, all original INDRA Statements are also collated in al list. After the assembly step, the statements are saved in JavaScript Object Notation (json) format. These files can be visualized with a json viewer. We used json2table.com. As such information can be retrieved on, for example, the PubMed ID and the original sentence of the paper on which the evidence is based.

Results

The number of analyzed articles for each query, the percentage of articles (i.e., PMIDs) for which INDRA Statements were produced in relation to the retrieved articles, and the percentages of PMIDs with open access, are displayed in Table 2. INDRA Statements were extracted ~ 30% of all papers retrieved based on the KCC input, indicating that a reasonable amount of input literature contained information suited to automated processing.

From the INDRA Statements, 91 networks (i.e., seven networks (KCC 1, 2/3, 4, 5, 6/7, 8, 9/10) for 13 compounds) were created (Table 3). Table 3 provides a comparison between the KCC network size (i.e., the number of nodes and edges) and network support (i.e., sBS of the edges), and the actual evaluation for KCC activation by the IARC (see “IARC evaluations for evidence of KC activation” in the Supplemental Material).

A large network represents that there is (potentially) a richer body of mechanistic literature for that chemical discussed in the context of that KCC. In general, higher sBS were observed for those compounds for which the IARC has proposed strong evidence for induction of KCCs (Figure 3). This is also corroborated by Wilcoxon statistical analysis. Across all chemicals and KCCs, INDRA derived sBS tended to be significantly higher for KCCs for which the IARC has classified the KCC evaluation as strong or moderate evidence, than for those for which the IARC evidence was rated “weak” or “no” [median (interquartile range)= 3.5 [13.7] vs. 0.9 [2.2]; p-value: 0.0003], but there was considerable overlap.

Table 2. Processed articles for each query.

KCC	Benzene (1)	Penta-chloro-phenol (1)	DDT (2A)	Diazinon (2A)	Hydrazine (2A)	Glyphosate (2A)	Malathion (2A)	Melamine (2B)	Parathion (2B)	Pyridine (2B)	Coffee (3)	β-Picoline (3)	Allyl chloride (3)
KCC1	501	139	281	60	69	68	99	71	168	370	173	8	3
Electrophilic	(30%)	(4%)	(35%)	(35%)	(26%)	(50%)	(30%)	(41%)	(33%)	(38%)	(31%)	(13%)	(0%)
	(14%)	(7%)	(9%)	(7%)	(9%)	(32%)	(15%)	(20%)	(21%)	(25%)	(21%)	(0%)	(0%)
KCC2 & 3	740	107	130	39	126	183	84	39	51	619	191	0	7
Genotoxic & DNA	(42%)	(7%)	(39%)	(38%)	(47%)	(46%)	(44%)	(46%)	(27%)	(35%)	(52%)	(0%)	(14%)
repair	(28%)	(10%)	(22%)	(38%)	(25%)	(37%)	(38%)	(36%)	(22%)	(34%)	(21%)	(0%)	(0%)
KCC4	255	61	108	29	48	99	41	19	36	173	149	6	1
Epigenetics	(41%)	(8%)	(58%)	(55%)	(19%)	(58%)	(59%)	(37%)	(33%)	(34%)	(46%)	(17%)	(0%)
	(37%)	(16%)	(47%)	(59%)	(17%)	(55%)	(54%)	(53%)	(36%)	(34%)	(32%)	(0%)	(0%)
KCC5	332	135	67	99	84	147	121	38	47	350	235	0	4
Oxidative stress	(42%)	(4%)	(60%)	(39%)	(54%)	(61%)	(38%)	(39%)	(43%)	(30%)	(47%)	(0%)	(0%)
	(37%)	(10%)	(49%)	(56%)	(39%)	(63%)	(71%)	(71%)	(49%)	(33%)	(32%)	(0%)	(0%)
KCC6 & 7	577	38	96	46	46	52	58	53	23	247	248	5	3
Inflammation &	(35%)	(11%)	(36%)	(46%)	(33%)	(62%)	(62%)	(37%)	(22%)	(38%)	(40%)	(20%)	(0%)
immune-suppressive	(26%)	(16%)	(34%)	(48%)	(30%)	(38%)	(57%)	(36%)	(26%)	(30%)	(39%)	(0%)	(0%)
KCC8	74	35	215	23	35	46	25	16	33	206	172	0	0
Receptor mediated	(33%)	(3%)	(46%)	(48%)	(20%)	(58%)	(36%)	(50%)	(30%)	(31%)	(31%)	(0%)	(0%)
	(35%)	(20%)	(40%)	(61%)	(14%)	(33%)	(52%)	(44%)	(58%)	(31%)	(26%)	(0%)	(0%)
KCC9 & 10	240	42	65	28	54	60	43	48	19	338	169	2	2
(Immortalization &	(42%)	(19%)	(55%)	(54%)	(46%)	(53%)	(37%)	(33%)	(26%)	(57%)	(34%)	(50%)	(0%)
cell dynamics)	(49%)	(12%)	(48%)	(54%)	(24%)	(47%)	(58%)	(56%)	(37%)	(40%)	(25%)	(0%)	(0%)
TOTAL unique PMIDs associated with chemical name appearing in Title and KCC queries[b]	1933	427	788	233	368	506	360	215	320	1820	1069	18	16

Note: DDT, dichlorodiphenyltrichloroethane; INDRA, Integrated Network and Dynamical Reasoning Assembler; KCCs, key characteristics of carcinogens; PMID, PubMed identifier

[a] First line within each row : total number of processed articles; second line, in parenthesis "()": the percentage (%) of full text articles in relation to the number of processed articles; third line: the percentage (%) of PMIDs for which INDRA statements were retrieved.

[b] The TOTAL number shows the unique papers emerging from the KCC-specific queries (not counting the repeated ones in more than one query).

Table 3. Network statistics for 13 chemicals in relation to IARC classifications.

Chemical	KCC 1 Electrophilic	KCC 2 & 3 Genotoxic & DNA repair	KCC 4 Epigenetics	KCC 5 Oxidative stress	KCC 6 & 7 Inflammation & immunosuppressive	KCC 8 Receptor-mediated	KCC 9 & 10 Immortalization & cell dynamics
Benzene (1), n = 1933	🌀🌀🌀	🌀🌀🌀	?	🌀🌀🌀	🌀🌀🌀	🌀🌀🌀	🌀🌀🌀
	16/19 (16.73)	43/70 (62.25)	18/21 (18.38)	17/22 (19.75)	21/23 (20.31)	5/4 (3.50)	22/32 (28.83)
Penta-chloro-phenol (1), n = 427	🌀🌀🌀	🌀🌀🌀	?	🌀🌀🌀	🌀🌀	🌀🌀🌀	🌀🌀🌀
	2/1 (0.86)	3/2 (1.51)	5/4 (2.81)	5/4 (3.44)	0	3/3 (1.95)	2/2 (1.51)
DDT (2a), n =788	?	🌀🌀	?	🌀🌀🌀	🌀🌀🌀	🌀🌀🌀	🌀🌀
	4/3 (2.64)	8/7 (6.17)	13/12 (10.52)	11/12 (10.68)	5/6 (5.43)	16/18 (15.63)	7/7 (6.35)
Diazinon (2a), n = 233	?	🌀🌀🌀	?	🌀🌀🌀	🌀	🌀	🌀
	0	3/2 (1.78)	2/1 (0.86)	12/13 (11.54)	3/2 (1.78)	2/1 (0.86)	3/2 (1.78)
Glyphosate (2a), n = 506	🌀	🌀🌀🌀	?	🌀🌀🌀	🌀	🌀	🌀
	3/2 (1.72)	10/12 (10.68)	16/17 (15.14)	25/29 (26.01)	4/3 (2.66)	3/2 (1.78)	7/7 (6.11)
Hydrazine (2a), n = 368	🌀🌀🌀	🌀🌀🌀	?	🌀🌀🌀	?	🌀	🌀🌀🌀
	0	2/1 (0.92)	2/1 (0.86)	4/3 (2.64)	2/1 (0.86)	0	0
Malathion (2a), n= 360	🌀	🌀🌀🌀	?	🌀🌀🌀	🌀🌀🌀	🌀🌀🌀	🌀🌀🌀
	2/1 (0.86)	8/9 (7.91)	7/6 (5.22)	27/48 (43.05)	16/16 (14.00)	3/3 (2.64)	15/16 (14.11)
Melamine (2b), n = 215	?	🌀	?	?	🌀🌀🌀	?	?
	2/1 (0.86)	4/3 (2.58)	2/1 (0.92)	12/13 (11.70)	3/2 (1.78)	2/1 (0.86)	9/9 (8.04)
	🌀	🌀🌀	?	🌀	🌀	🌀	🌀🌀🌀

Chemical	KCC 1 Electrophilic	KCC 2 & 3 Genotoxic & DNA repair	KCC 4 Epigenetics	KCC 5 Oxidative stress	KCC 6 & 7 Inflammation & immunosuppressive	KCC 8 Receptor-mediated	KCC 9 & 10 Immortalization & cell dynamics
Parathion (2b), n = 320	11/12 (10.38)	0	0	3/2 (1.78)	0	6/8 (6.88)	0
Pyridine (2b), n = 1820	?	🌀	?	🌀	🌀🌀	?	?
	0	3/2 (1.72)	5/5 (4.45)	0	0	3/2 (1.78)	2/1 (0.92)
Coffee (3), n = 1069	?	🌀	🌀	🌀	🌀	🌀	🌀
	2/1 (0.86)	0	2/1 (0.86)	7/8 (7.04)	12/13 (20.36)	3/6 (5.45)	0
Allyl chloride 3), n = 16	🌀🌀🌀	?	?	?	?	?	🌀
	0	0	0	0	0	0	0
β-Picoline (3), n = 18	?	🌀	?	?	?	?	?
	0	0	0	0	0	0	0

Note: DDT, dichlorodiphenyltrichloroethane; IARC, International Agency for Research on Cancer; KCCs, key characteristics of carcinogens.

[a] The symbols refer to the IARC evaluation: 🌀🌀🌀: strong evidence that KCC is induced; 🌀🌀: moderate evidence for induction of KCC. 🌀: no or weak evidence of the induction of KCC. ?: no adequate data for an evaluation to be made. It regularly occurred that the IARC evaluations differed for the various KCCs, contained within one literature query. For example, for KCC 2 the evidence could have been weak whereas for KCC3 the evidence was strong. Because the two KCCs are combined we chose to always use the stronger evidence (in this example we marked the box "strong"). To view the full networks see Table S1, where URLs for each chemical network are provided.

[b] The numbers in each cell represent number of nodes/edges and sum of belief score (sBS) of the edges between parentheses.

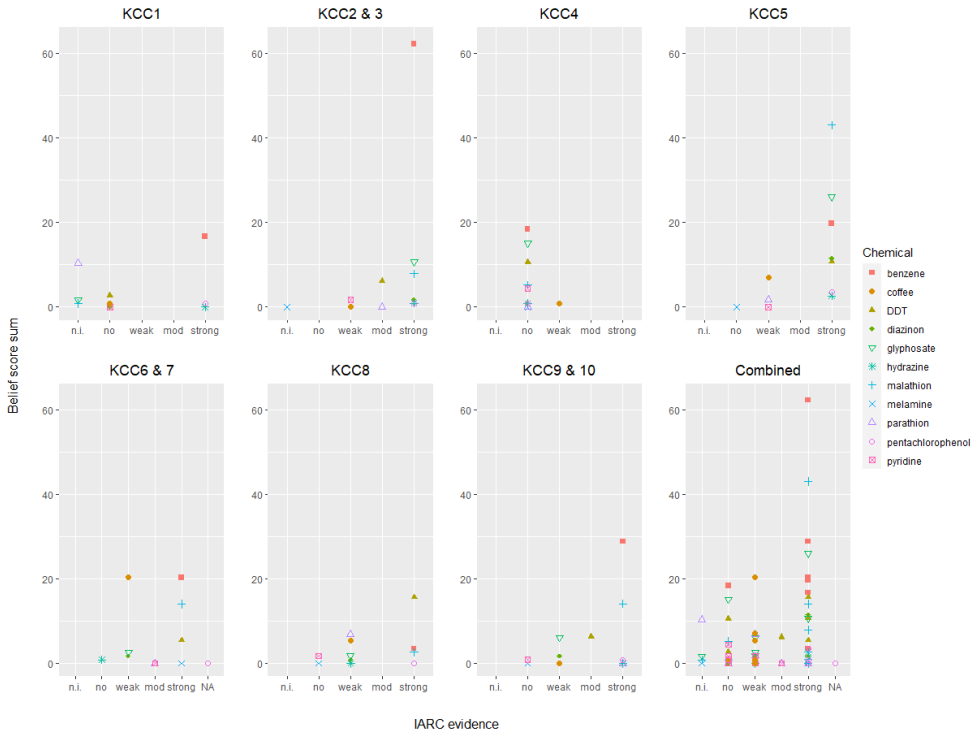


Figure 3. Relation of IARC evidence for key characteristics of cancer (KCC) result, vs. sum of the belief scores (sBS) from all filtered edges of a network. Note: IARC, International Agency for Research on Cancer; DDT, dichlorodiphenyltrichloroethane; KCC 1: electrophilic, KCC 2: genotoxic, KCC 3: DNA repair, KCC 4: epigenetics, KCC 5: oxidative stress, KCC 6: inflammation, KCC 7: immunosuppressive, KCC 8: receptor-mediated, KCC 9: immortalization, KCC 10: cell dynamics; mod, moderate; NA, not applicable; n.i., not sufficient data available for evaluation.

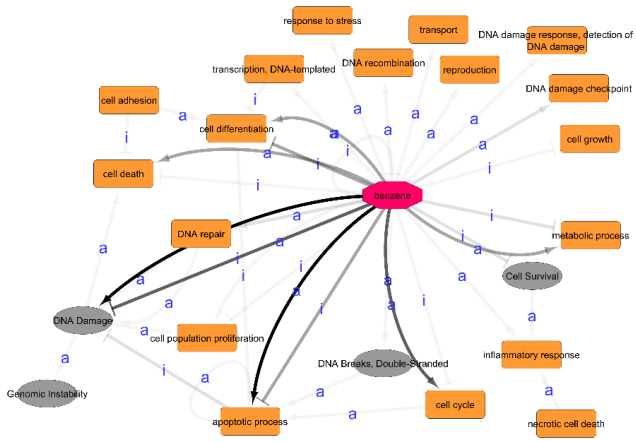


Figure 4. Benzene network for key characteristics of carcinogens (KCCs) on genotoxicity (KCC 2) and DNA repair (KCC 3) , filtered on bioprocess (orange rounded rectangle-shaped box) and other processes (gray ellipse-shaped box). A darker arrow represents a higher belief score (BS): black: BS 0.95, dark gray: BS 0.94, gray: BS 0.92, light gray: BS 0.86, transparent gray: BS 0.65. Arrows: activation, an arrow with a raised edge: inhibition. Note: a: activation, i: inhibition

From Table 3, we did notice for a number of KCCs the evidence was strong according to the IARC, but that the networks and the sBS were small. In those cases, the number of PMIDs (Table 2) was also low. Conversely it does not appear that a larger number of PMIDs, resulting from the KCCs *and* chemical specific queries, always results in larger networks, that is, although benzene has the highest number of PMIDs (1,933) and the largest networks, the networks for pyridine and coffee are considerably smaller even though these compounds have the second (1,820) and third (1,069) highest number of PMIDs, respectively. Most compounds show a network for KCC 4 (epigenetic alterations) although the IARC concluded for all compounds but coffee that there was no sufficient evidence to evaluate the induction of this KCC (Table 3).

By using the filtering option on bioprocesses and other processes, we did not limit ourselves to the first neighbors of the compound of interest, hence allowing us to investigate relations between events, also further away from the compound of interest (see “Filtering networks” in the Supplemental Material, Figure S1). An example is given for benzene (Figure 4), where benzene activates DNA damage, whereas DNA damage in turn can activate cell death. Further necrotic cell death is associated with activation of an inflammatory response. This relation (cell death activates an inflammatory response) is also described in the AOP wiki databases (ID 1776). In addition, the process of how the disruption of the cell cycle can lead to apoptotic processes (Figure 4) is described in the AOP wiki databases (ID 1712).

Figures S2 – S13 show the outcome of “Bioprocess and ‘other’ filtering” for each of the chemicals for KCC 2 (genotoxicity) and 3 (DNA repair) (originating from Query 2). We see that for all those compounds (benzene, DDT, hydrazine, diazinon, glyphosate, malathion, parathion, pentachlorophenol), for which the IARC evaluated the evidence to be moderate or strong for the activation of this specific KCC, terms (both from “Bioprocesses” and “other”) related to genotoxicity and DNA repair did appear in the networks. The terms include, for example, DNA damage and (inhibition of) DNA repair. Conversely, considering the chemicals for which the IARC evaluated the evidence to be weak or absent (melamine, pyridine, allyl chloride, β-Picoline and coffee) these two terms were not observed in the networks for pyridine and coffee, (nor for allyl chloride and β-Picoline for which no network could be created), but only in the network for melamine.

Discussion

In this work, we investigated whether evidence identification for chemical hazard assessment could be supported using an automated, computational approach. As an example, we explored the use of this approach for identifying KCCs as used in the evaluation of mechanistic evidence in the IARC Monograph program. Using text mining and network analysis approaches (i.e., INDRA), we found concordance between computationally inferred networks strength (high BS) and the IARC KCC

evaluations, especially for those compounds for which the IARC has evaluated the evidence to be strong for KCC induction. As such, our example application suggests that compounds with larger networks and higher sBS scores, could be prioritized for hazard identification, making the process of evidence identification for hazard assessment more efficient and transparent.

The output of our approach generates an inventory of available studies, as well as a categorization of data in the form of networks. These generated networks can further be filtered to retrieve information on mechanisms of action by filtering only on bioprocesses (Figure 4; Figures S2-S12). This type of visualization can be used as tool to assist in the interpretation of the literature for mechanistic evaluation of compounds within the KCC framework (i.e. informed evidence identification).

Recently Barupal et al. (2021) published a study on prioritizing cancer hazard assessments for IARC Monographs using an integrated approach of database fusion and text mining. The authors also used the KCCs as input but, unlike the investigation we conducted, Barupal et al (2021) mainly looked at publication count, as well as coverage across 34 different databases relevant to cancer, for an agent. Our approach is different in that we are not only identifying possibly relevant literature by the sheer counting of numbers of publications per chemical (Table 2), but that our approach also uses a systems biology-inspired textmining environment (i.e., INDRA) to extract data from the *individual* articles and compile these data into potentially meaningful biological networks, describing the possible relations between biomolecules and chemicals, bioprocesses in the context of KCCs (i.e. informed evidence identification). Thus, our work expands beyond the evaluation of publication density or coverage of toxicological content in databases. Importantly, we observed that the number of publications derived from the KCC specific literature queries (which is driven by general scientific interest in the chemical) proved not to be an accurate indication for potential KCC activation, at least as inferred here from automated network assembly.

Although promising, using our automated computational approach has several limitations that should be kept in mind when interpreting the results. Stringent filtering on BS, for example retaining only results with a BS ≥ 0.86, can exclude relevant results reported in a single study because only one single study could point out a relevant result that now might be discarded. For example, if we consider an unfiltered network for KCC 9, 10 for parathion (KCC9,10 – Query 7; Figure S13), we observe connections between parathion and apoptotic process, as well as between parathion and cell population proliferation. Both statements have a BS of 0.65, indicating that single studies contribute to these statements. Both processes are linked to KCC 9 and 10, and, according to the IARC, parathion indeed induces KCC 9 and 10; however, this observation would have gone unnoticed upon more stringent filtering. So for smaller networks it might thus be worthwhile to also investigate the larger, unfiltered networks. Conversely, our network analysis does not distinguish between positive and negative regulation when filtering by BS, so it can occur that a network is large but contains

processes that are actually favorable, for example, inhibition of DNA damage. The potential directionality can be further investigated by displaying inhibition vs. activation statements (an example is given for benzene in Figure S14). Last, we applied a filtering step for hypothetical statements by excluding statements that contain certain signaling words such as “suggest”. However, using the word “suggest” is sometimes preferred, particularly in human studies, to avoid the use of causal language. Hence filtering statements with reference to “suggest” can potentially exclude data from articles that use the wording “suggest” avoiding the use of causal language, and bias toward articles that inappropriately use causal language.

In other cases, the network showed potentially relevant findings, however not specifically for the KCC for which the network was originally created. An example is parathion, Query 1, which, according to the generated network, can activate cell death, modify testosterone, or inhibit acetylcholinesterase (Figure S15). Given that KCC1 is on electrophilicity the findings from Figure S15 would for example “fit” better under KCC 10 which refers to cell death.

Furthermore, we have not evaluated the selected studies’ informativeness or study quality after the filtering steps. Relevant questions, such as whether the observed mechanisms can also operate in humans, *in vitro* vs. *in vivo* models, the quality of the studies, biological significance of mechanistic end points, whether evidence is consistent within and among KCCs, for example, were not considered yet. Of course, this can be adopted in the process, that is to modify the initial PubMed query (e.g. select only human studies) but this requires experts to stratify or limit the evidence base to a-priori domains or quality assessments.

The composition of the literature query as input, in our case the search terms by Guyton et al. (2018), is quite influential when retrieving the PMIDs. This was most notable for KCC 4 (induces epigenetic alterations). For many compounds (all but coffee) we see that the IARC states that for this specific KCC there is not sufficient data available for an evaluation. However, we regularly observe large networks for KCC4 (Table 3, e.g., benzene, DDT). We discovered that this may be due to the description of Guyton’s queries for Q3/KCC 4: the query includes the terms “rna” or “rna, messenger” (because non-coding RNAs are recognized epigenetic alterations (Chappell et al. 2016)). However, this resulted, for our computational approach, mainly in the activation of events such as DNA damage, DNA damage check or cell survival. These statements do not match examples of relevant evidence according to the IARC’s instructions (IARC, 2017), which, for KCC4, should involve, for example, terms associated with DNA methylation or histone modification. When we adjust the search term for this specific query by leaving out the ‘rna’ term we see that the adjusted networks are much smaller together with a reduction in BS (Figure S16).

We noticed that the percentage of PMIDs for which we received INDRA Statements was moderate (Table 2). A search on a number of PMIDs for which we retrieved no

INDRA statements showed that some papers (mostly older ones) had no abstract or the study was non-English. For our particular case study, we retrieved full papers when open access and relied on abstracts for others. We did this to make the methodology as open as possible for use by scientists in the hazard assessment process and we conject that the most important results of a study would be made available in the abstract and as such the impact of not having full access to all papers might be limited. However, this does illustrate that although we used the same search terms as the IARC working group, the evidence base (i.e., the selected studies to either generate networks [for our approach] or to evaluate the evidence [for the IARC working group]) was not identical for both approaches. We focused specifically on the IARC and the KCCs because these provide well-defined search terms for identifying literature but we recognize other institutes (e.g. National Toxicology Program Report on Carcinogens, U.S. Environmental Protection Agency) also include mechanistic data in their hazard assessment on carcinogens, including adaptations of the KCC literature queries (NTP 2016).

Last, we did not manually annotate papers for which relations are relevant and then check which of these the reading system (in our case, Reach) can pick up. The closest relevant evaluation as to the performance of Reach was done by Glavaški and Velicki (2021), who found a good accuracy of Reach but noted the extraction performance could be improved.

Our approach does not claim to fully automate and replace manual evaluation of mechanistic literature as is done in hazard identification, such as the IARC Monographs program. Instead, it could potentially be helpful in the prioritization of chemicals in relation to KCCs for further review, that is to identify and create a network-based inventory of available studies, the content of which is to be further evaluated by an expert committee. Even though our findings are not directly generalizable outside the IARC framework, there is no reason to assume that our approach would not work well in other (noncancer) hazard identification programs using a well-defined framework for the evaluation of mechanistic data such as the KCCs. Future work should also focus on strategies to qualitatively or quantitatively assess the strength of the evidence that is provided in the mechanistic literature. This would require identifying those study characteristics that are typically used by experts to define study quality and developing approaches to systematically extract these from identified publications in an automated way.

References Chapter 2

Ananiadou, S., S. Pyysalo, J. Tsujii, and D.B. Kell. 2010. “Event Extraction for Systems Biology by Text Mining the Literature.” *Trends in Biotechnology* 28 (7): 381–90.

Barupal, D. K., M. K. Schubauer-Berigan, M. Korenjak, J. Zavadil, and K. Z. Guyton. 2021. “Prioritizing Cancer Hazard Assessments for IARC Monographs Using an Integrated Approach of Database Fusion and Text Mining.” *Environmental International* 156.

Carvaillo, Jean Charles, Robert Barouki, Xavier Coumoul, and Karine Audouze. 2019. “Linking Bisphenol S to Adverse Outcome Pathways Using a Combined Text Mining and Systems Biology Approach.” *Environmental Health Perspectives*. https://doi.org/10.1289/EHP4200.

Chappell, Grace, Igor P. Pogribny, Kathryn Z. Guyton, and Ivan Rusyn. 2016. “Epigenetic Alterations Induced by Genotoxic Occupational and Environmental Human Chemical Carcinogens: A Systematic Literature Review.” *Mutation Research - Reviews in Mutation Research*. https://doi.org/10.1016/j.mrrev.2016.03.004.

EFSA. 2017. “The Science of Combining Apples and Oranges: Joint EFSA/EBTC Scientific Colloquium on Evidence Integration in Risk Assessment.” https://doi.org/DOI: 10.2903/sp.efsa.2018.EN-1396.

Glavaški, M, and L Velicki1. 2021. “Humans and Machines in Biomedical Knowledge Curation: Hypertrophic Cardiomyopathy Molecular Mechanisms’ Representation.” *BioData Min* 14 (45). https://doi.org/10.1186/s13040-021-00279-2.

Guha, Neela, Kathryn Z. Guyton, Dana Loomis, and Dinesh Kumar Barupal. 2016. “Prioritizing Chemicals for Risk Assessment Using Chemoinformatics: Examples from the IARC Monographs on Pesticides.” *Environmental Health Perspectives*. https://doi.org/10.1289/EHP186.

Guyton, Kathryn Z., Ivan Rusyn, Weihsueh A. Chiu, Denis E. Corpet, Martin van den Berg, Matthew K. Ross, David C. Christiani, Frederick A. Beland, and Martyn T. Smith. 2018. “Application of the Key Characteristics of Carcinogens in Cancer Hazard Identification.” *Carcinogenesis*. https://doi.org/10.1093/carcin/bgy031.

Gyori, Benjamin M, John A Bachman, Kartik Subramanian, Jeremy L Muhlich, Lucian Galescu, and Peter K Sorger. 2017. “From Word Models to Executable Models of Signaling Networks Using Automated Assembly.” *Molecular Systems Biology*. https://doi.org/10.15252/msb.20177651.

Howard, B.E., J. Phillips, K. Miller, A. Tandon, M. Deepak, M.R. Shah, S. Holmgren, et al. 2016. “SWIFT-Review: A Text-Mining Workbench for Systematic Review.” *Systematic Reviews* 87 (5).

IARC. 1992. “IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 54: Occupational Exposure to Mists and Vapours from Strong Inorganic Acids; and Other Industrial Chemicals.” Lyon.

———. 2017. “Instructions to Authors for the Preparation of Drafts for IARC Monographs.” 2017. http://monographs.iarc.fr/ENG/Preamble/instructions.php.

———. 2019. “Preamble to the IARC Monographs, Amended January 2019.” 2019. https://monographs.iarc.who.int/wp-content/uploads/2019/07/Preamble-2019.pdf.

National Academies of Sciences Engineering and Medicine. 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. National Academies Press. https://doi.org/10.17226/24635.

NTP. 2016. “Standard Search Strings for Literature Database Searches: Appendix to the Draft Handbook for Preparing Report on Carcinogens Monographs.” https://ntp.niehs.nih.gov/ntp/roc/handbook/rochandbookappendix_508.pdf.

Samet, Jonathan M., Weihsueh A. Chiu, Vincent Cogliano, Jennifer Jinot, David Kriebel, Ruth M. Lunn, Frederick A. Beland, et al. 2020. “The IARC Monographs: Updated Procedures for Modern and Transparent Evidence Synthesis in Cancer Hazard Identification.” *Journal of the National Cancer Institute*. https://doi.org/10.1093/jnci/djz169.

Shapiro, A.J., S. Antoni, K.Z. Guyton, R.M. Lunn, D. Loomis, I. Rusyn, G.D. Jahnke, et al. 2018. “Software Tools to Facilitate Systematic Review Used for Cancer Hazard Identification.” *Environmental Health Perspectives2* 126 (10).

Smith, Martyn T., Kathryn Z. Guyton, Catherine F. Gibbons, Jason M. Fritz, Christopher J. Portier, Ivan Rusyn, David M. DeMarini, et al. 2016. “Key Characteristics of Carcinogens as a Basis for Organizing Data on Mechanisms of Carcinogenesis.” *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.1509912.

Steppi, Albert, Benjamin Gyori, and John Bachman. 2020. “Adeft: Acromine-Based Disambiguation of Entities from Text with Applications to the Biomedical Literature.” *Journal of Open Source Software*. https://doi.org/10.21105/joss.01708.

Valenzuela-Escarcega, M.A., G Hahn-Powell, T Hicks, and M Surdeanu. 2015. “A Domain-Independent Rule-Based Framework for Event Extraction.” https://aclanthology.org/P15-4022.pdf.

Valenzuela-Escárcega, Marco A., Özgün Babur, Gus Hahn-Powell, Dane Bell, Thomas Hicks, Enrique Noriega-Atala, Xia Wang, Mihai Surdeanu, Emek Demir, and Clayton T Morrison. 2018. “Large-Scale Automated Machine Reading Discovers New Cancer-Driving Mechanisms.” *Database*. https://doi.org/10.1093/database/bay098.

Wikoff, Daniele S., Julia E. Rager, Grace A. Chappell, Seneca Fitch, Laurie Haws, and Susan J. Borghoff. 2019. “A Framework for Systematic Evaluation and Quantitative Integration of Mechanistic Data in Assessments of Potential Human Carcinogens.” *Toxicological Sciences*. https://doi.org/10.1093/toxsci/kfy279.

Supplemental Material - Network links

Table S1. Links to NDEX networks for full (unfiltered) networks

Compound	Link to NDEX network
Melamine	http://www.ndexbio.org/#/networkset/ca837d12-e205-11ea-99da-0ac135e8bacf
Coffee	http://www.ndexbio.org/#/networkset/f0fd7ea5-e205-11ea-99da-0ac135e8bacf
Benzene	http://www.ndexbio.org/#/networkset/aa24c62c-e202-11ea-99da-0ac135e8bacf
DDT	http://www.ndexbio.org/#/networkset/832a429d-e205-11ea-99da-0ac135e8bacf
Malathion	http://www.ndexbio.org/#/networkset/bdb76d31-e205-11ea-99da-0ac135e8bacf
Glyphosate	http://www.ndexbio.org/#/networkset/acd07de0-e205-11ea-99da-0ac135e8bacf
Diazinon	http://www.ndexbio.org/#/networkset/9044727e-e205-11ea-99da-0ac135e8bacf
Hydrazine	http://www.ndexbio.org/#/networkset/9b77b9ef-e205-11ea-99da-0ac135e8bacf
Parathion	http://www.ndexbio.org/#/networkset/d4bcd423-e205-11ea-99da-0ac135e8bacf
Pyridine	http://www.ndexbio.org/#/networkset/e6f45f54-e205-11ea-99da-0ac135e8bacf

Supplemental Material - Filtering networks

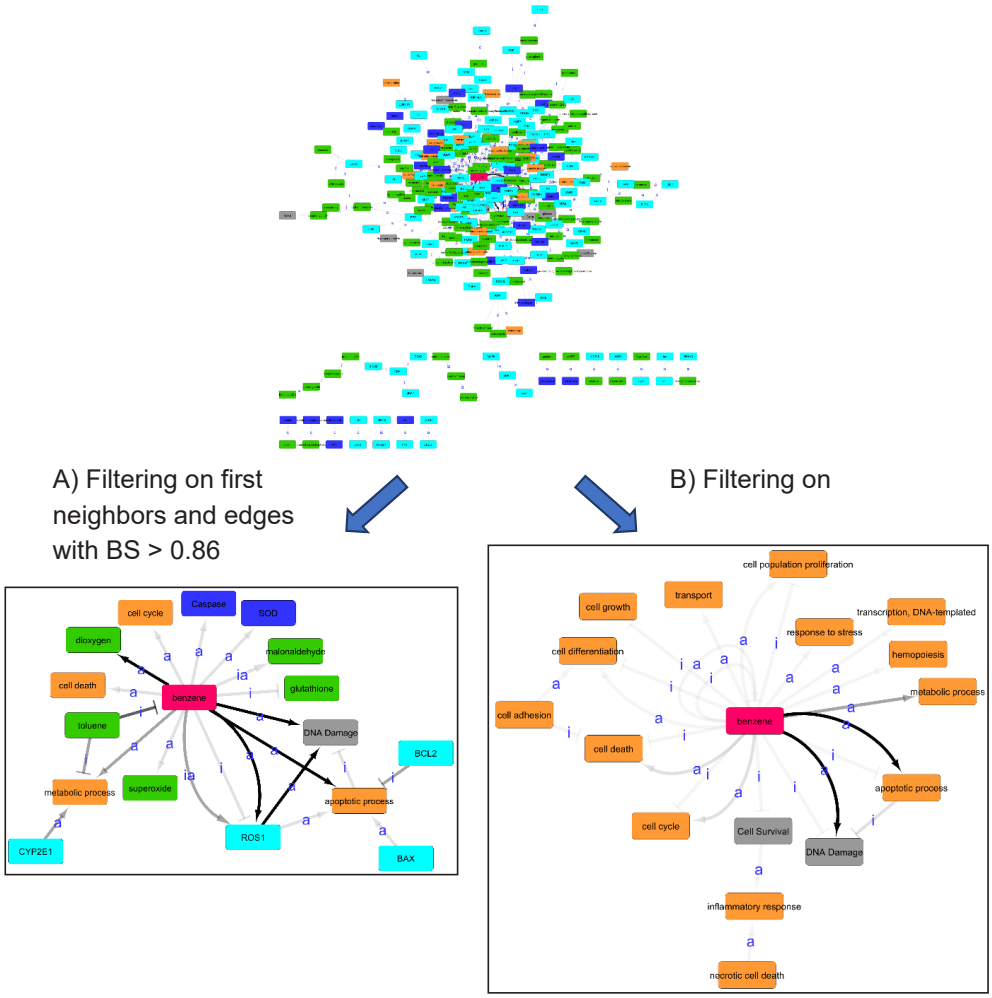


Figure S1. Different network filtering examples of benzene KC5 network. On top: full network; A: Filtering on first neighbors and edges with BS > 0.86. B:Filtering on bioprocess and other (node type).
The colors in the networks refer to the following entities: orange: bioprocess, green: chemical, light blue: protein, dark blue: protein family, grey: other.

Supplemental Material - Bioprocesses and other processes

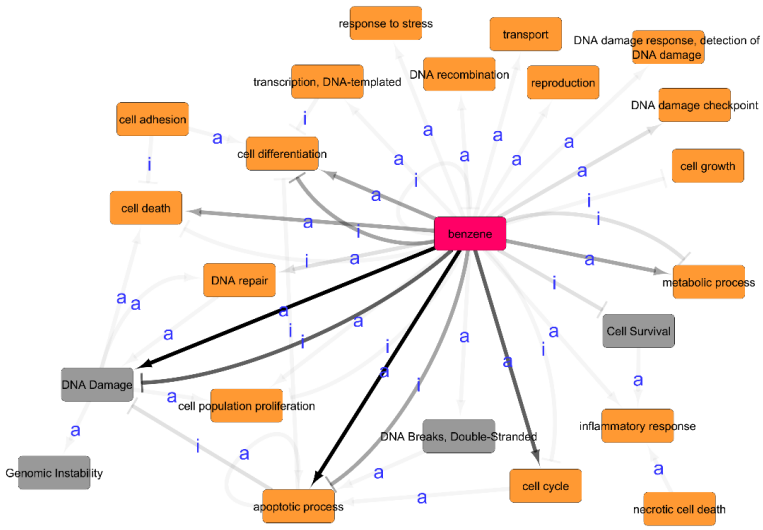


Figure S2. Network for benzene on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

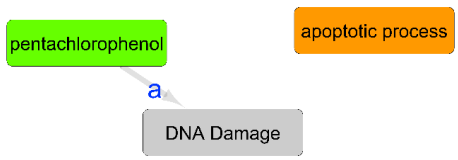


Figure S3. Network for pentachlorophenol on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event.

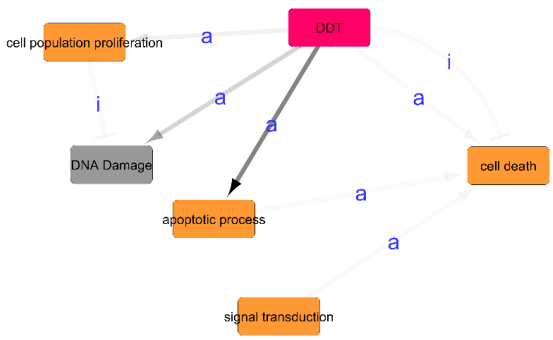


Figure S4. Network for DDT on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

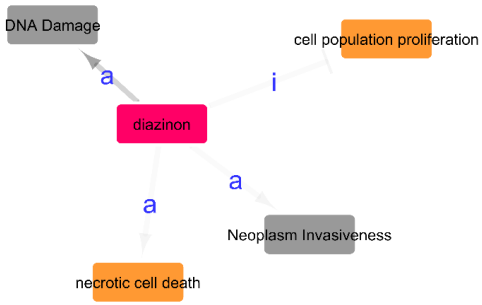


Figure S5. Network for diazinon on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

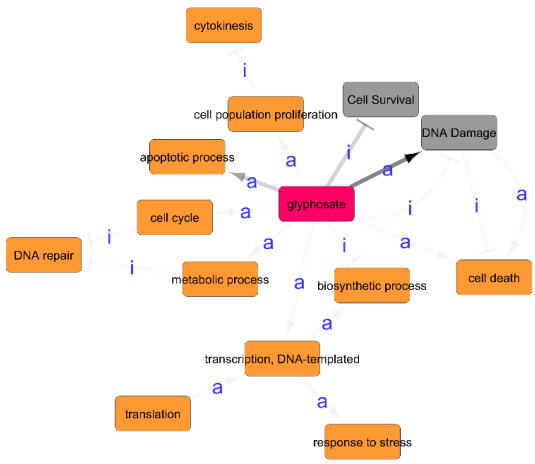


Figure S6. Network for glyphosate on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

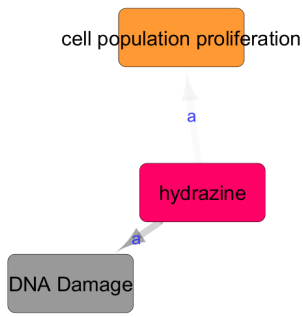


Figure S7. Network for hydrazine on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

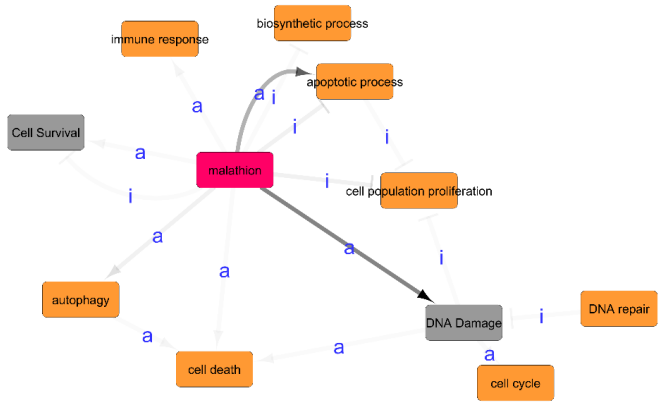


Figure S8. Network for malathion on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

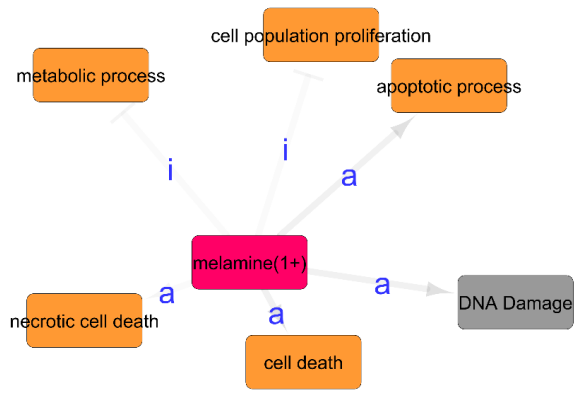


Figure S9. Network for melamine on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

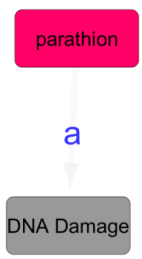


Figure S10. Network for parathion on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

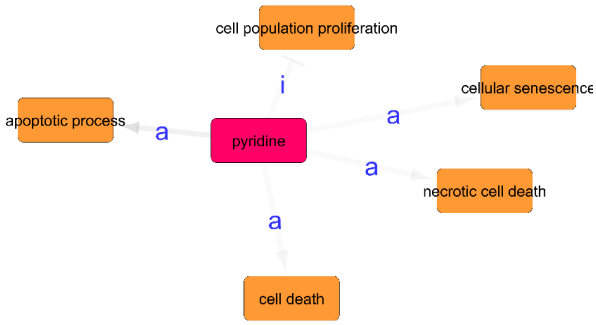


Figure S11. Network for pyridine on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

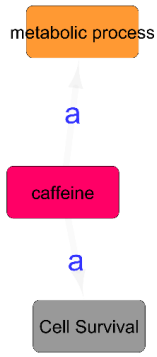


Figure S12. Network for caffeine on KC 2 (genotoxicity) and 3 (DNA repair), filtered on bioprocesses (orange box) and other processes (grey box).
a: activating event, i: inhibiting event.

Supplemental Material - Parathion Query 7

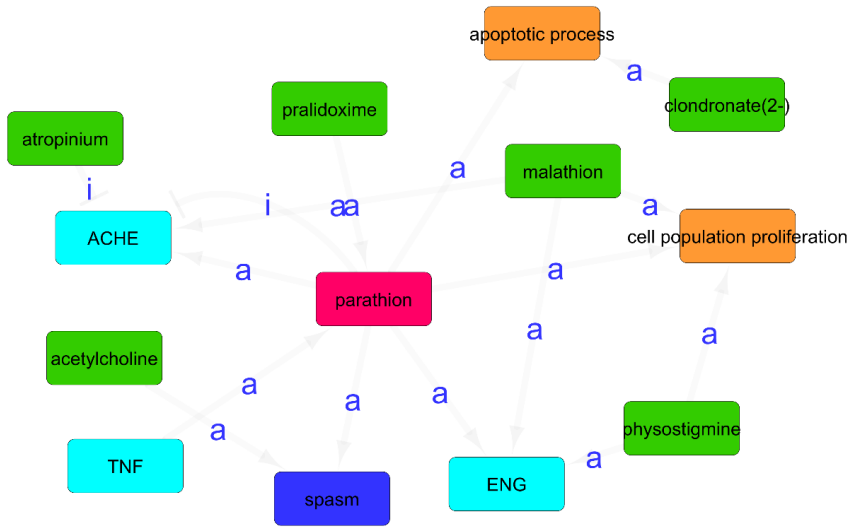


Figure S13. Parathion Query 7 – unfiltered network

Supplemental Material - Bioprocess Inhibition /Activation Graphs

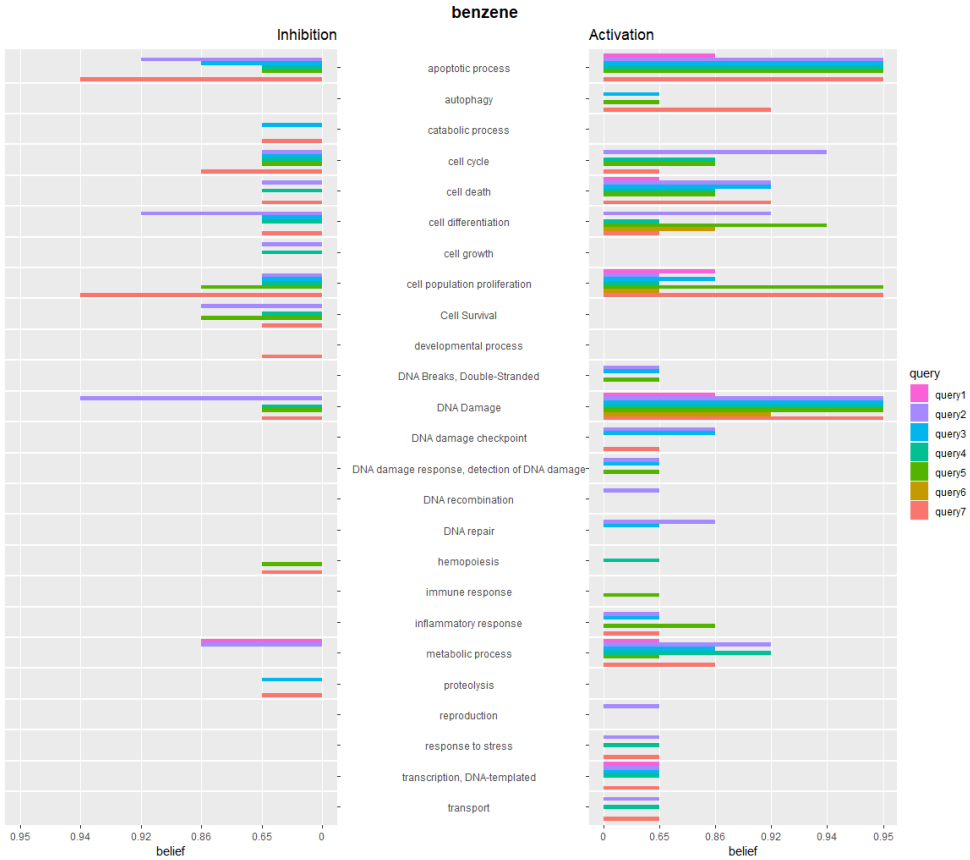


Figure S14. Figure of (in)activation of bioprocesses by benzene

Supplemental Material - Parathion – Query 1

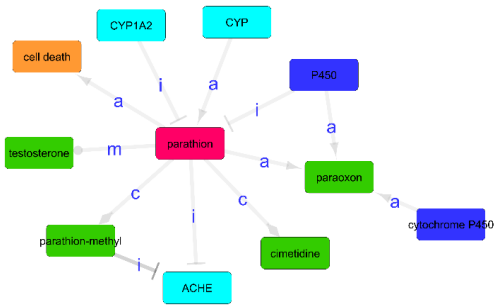


Figure S15. Unfiltered network for parathion – Query 1.

Supplemental Material - KCC 4 adjustment

Guyton query: benzene[Title] AND ("rna"[MeSH] OR "epigenesis, genetic"[MesH] OR rna OR "rna, messenger"[MeSH] OR "rna" OR "messenger rna" OR mrna OR "histones"[MeSH] OR histones OR epigenetic OR miRNA OR methylation)

Adjusted query: benzene[Title] AND ("epigenesis, genetic"[MesH] OR "histones"[MeSH] OR histones OR epigenetic OR miRNA OR methylation)

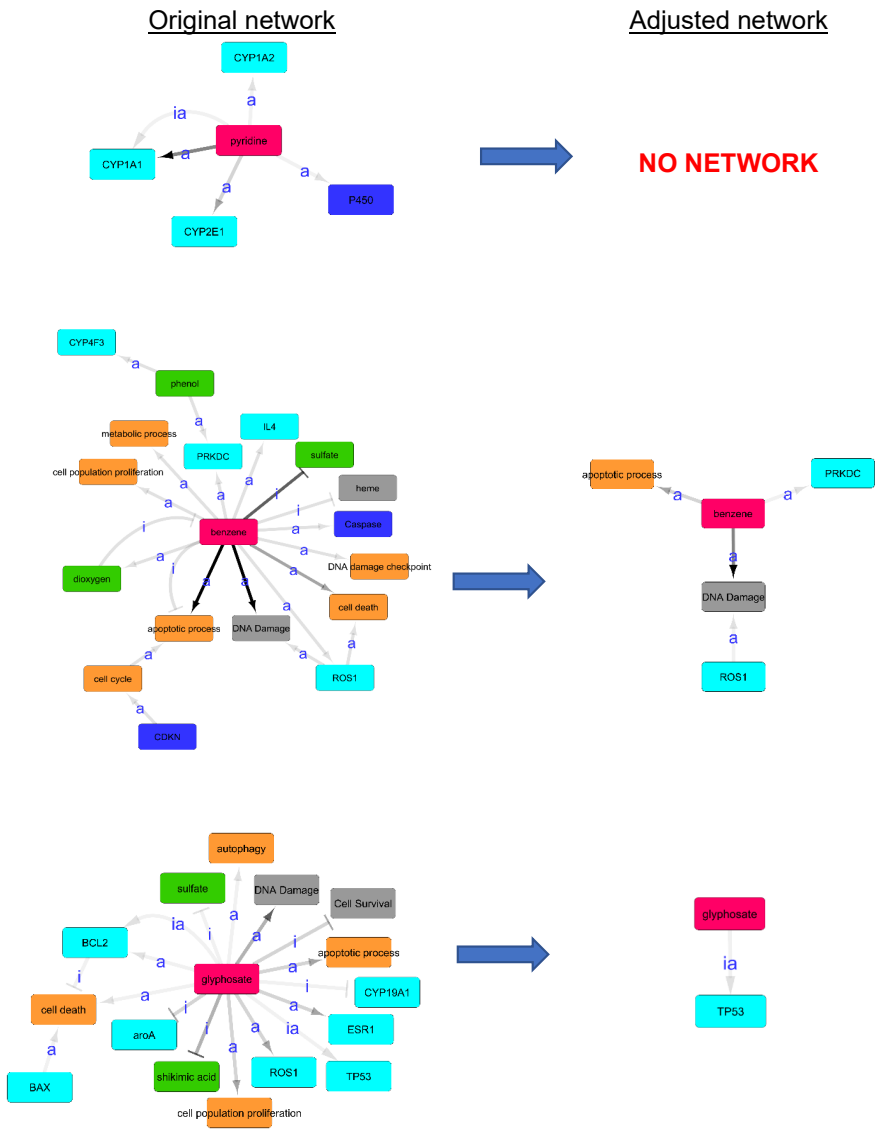


Figure S16. Three examples of how a change in the description for KC4 results in smaller networks. Left figures: networks based on original KC4 description by Guyton et al. (2018), right figures: new figure based on adjusted query description.

Supplemental Material - IARC evaluations for evidence of KC activation

Benzene

Volume 120

- KC1: Strong evidence that benzene is metabolically activated to electrophilic metabolites
- KC2: Strong evidence that benzene is genotoxic
- KC3: Strong evidence that benzene alters DNA repair or causes genomic instability
- KC4: No adequate data for an evaluation to be made
- KC5: Strong evidence that benzene induces oxidative stress
- KC6: No adequate data for an evaluation to be made
- KC7: Strong evidence that benzene is immunosuppressive
- KC8: Strong evidence that benzene modulates receptor-mediated effects
- KC9: No adequate data for an evaluation to be made
- KC10:Strong evidence that benzene alters cell proliferation, cell death, or nutrient supply

Pentachlorophenol

Volume 117

- KC1: Strong evidence that pentachlorophenol is metabolically activated to electrophilic metabolites
- KC2: The evidence for genotoxicity of pentachlorophenol is strong
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: The evidence is strong that pentachlorophenol induces oxidative stress
- KC6: There is moderate evidence that pentachlorophenol induces chronic inflammation
- KC7: There is moderate evidence that pentachlorophenol is immunosuppressive
- KC8: There is strong evidence that pentachlorophenol modulates receptor-mediated effects
- KC10: There is strong evidence that pentachlorophenol alters cell proliferation, cell death, or nutrient supply that can occur in humans

DDT

Volume 113

- KC1: No adequate data for an evaluation to be made
- KC2: The evidence for genotoxicity of DDT is moderate.
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made

- KC5: The evidence is strong that DDT induces oxidative stress, and this can occur in humans
- KC6: The evidence that DDT induces inflammation is moderate.
- KC7: The evidence is strong that DDT is immunosuppressive, and this can operate in humans.
- KC8: The evidence is strong that DDT modulates receptor-mediated effects that can operate in humans
- KC9: No adequate data for an evaluation to be made
- KC10: The evidence is moderate that DDT alters cell proliferation or cell death.

Diazinon

Volume 112

- KC1: No adequate data for an evaluation to be made
- KC2: The evidence for the genotoxicity of diazinon is strong and appears to operate in humans
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: The evidence that diazinon can induce oxidative stress is strong.
- KC6: No adequate data for an evaluation to be made
- KC7: Because of the limited available data, the evidence for immunosuppression as a mechanism of carcinogenicity for diazinon is weak.
- KC8: The evidence for receptor-mediated mechanisms in the potential carcinogenicity of diazinon is weak.
- KC9: No adequate data for an evaluation to be made
- KC10: Overall, the effects on proliferation are weak.

Glyphosate

Volume 112

- KC1: Glyphosate is not electrophilic
- KC2: Strong evidence that glyphosate causes genotoxicity
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC 5: Strong evidence that glyphosate causes oxidative stress
- KC6 and 7: Weak evidence that glyphosate may affect immune system
- KC8: Weak evidence that glyphosate induces receptor-mediated effects (potentially aromatase activity, or ligand for Ahr, PXR)
- KC9: No adequate data for an evaluation to be made
- KC 10: Weak evidence that glyphosate may affect cell proliferation or death (several studies reported cytotoxicity and cell death – the latter attributed to the apoptosis pathway)

Hydrazine

Volume 115

- KC1: Strong evidence that hydrazine is electrophilic or can be metabolically activated

- KC2: Strong evidence that hydrazine is genotoxic
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: Strong evidence that hydrazine induces oxidative stress
- KC6: No adequate data for an evaluation to be made
- KC7: No adequate data for an evaluation to be made
- KC8: Weak evidence that hydrazine modulates receptor-mediated effects
- KC9: No adequate data for an evaluation to be made
- KC10: Strong evidence that hydrazine alters cell proliferation, cell death, and nutrient supply

Malathion

Volume 112

- KC1: Malathion is not electrophilic (but its bioactive metabolite, malaoxon, can covalently modify B-esterases specifically at the active site serine residue; however, it is unknown whether electrophilicity of malaoxon plays a role in carcinogenesis)
- KC2: There is strong evidence that exposure to malathion-based pesticides is genotoxic
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: There is strong evidence that malathion induces oxidative stress
- KC6: There is strong evidence that malathion induces inflammation
- KC7: The evidence for immunosuppression as an effect of exposure to malathion is moderate
- KC8: There is strong evidence that malathion modulates receptor-mediated effects
- KC9: No adequate data for an evaluation to be made
- KC10: There is strong evidence for alteration of cell proliferation

Melamine

Volume 119

- KC1: No adequate data for an evaluation to be made
- KC2: there is evidence that melamine is not genotoxic
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: No adequate data for an evaluation to be made
- KC6: There is strong evidence that melamine induces chronic inflammation in the urinary tract.
- KC7: There is weak evidence that melamine is immunosuppressive
- KC8: No adequate data for an evaluation to be made
- KC9: No adequate data for an evaluation to be made
- KC10: No adequate data for an evaluation to be made

Parathion

Volume 112

- KC1: Parathion is not electrophilic (but its bioactive metabolite, paraoxon, can covalently modify B-esterases specifically at the active site serine residue; however, it is unknown whether the electrophilicity of paraoxon plays a role in carcinogenesis)
- KC2: Moderate evidence that parathion is genotoxic
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: Weak evidence that parathion induces oxidative stress
- KC6: Weak evidence that parathion induces chronic inflammation
- KC7: Weak evidence that parathion is immunosuppressive
- KC8: Weak evidence that parathion modulates receptor-mediated effects
- KC9: No adequate data for an evaluation to be made
- KC10: Strong evidence that parathion alters cell proliferation, cell death or nutrient supply

Pyridine

Volume 119

- KC1: No adequate data for an evaluation to be made
- KC2: Weak evidence that pyridine is genotoxic
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: Weak evidence that pyridine induces oxidative stress
- KC6: Moderate evidence that pyridine induces chronic inflammation
- KC7: No adequate data for an evaluation to be made
- KC8: No adequate data for an evaluation to be made
- KC9: No adequate data for an evaluation to be made
- KC10: No adequate data for an evaluation to be made

Coffee

Volume 116

- KC1: No adequate data for an evaluation to be made
- KC2: Weak evidence that coffee is genotoxic
- KC3: Weak evidence that coffee alters DNA repair or cause genomic instability
- KC4: Weak evidence that coffee induces epigenetic alterations
- KC5: Weak evidence that coffee induces oxidative stress
- KC6: Weak evidence that coffee induces chronic inflammation
- KC7: No adequate data for an evaluation to be made
- KC8: Weak evidence that coffee modulates receptor-mediated effects
- KC9: No adequate data for an evaluation to be made

- KC10: Weak evidence that coffee consumption alters cell proliferation / Moderate evidence that coffee consumption increases cell death through apoptosis

Allyl chloride

Volume 125

- KC1: Allyl chloride is electrophilic and alkylates DNA
- KC2: Evidence on genotoxicity is incoherent across studies
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: No adequate data for an evaluation to be made
- KC6: No adequate data for an evaluation to be made
- KC7: No adequate data for an evaluation to be made
- KC8: No adequate data for an evaluation to be made
- KC9: No adequate data for an evaluation to be made
- KC10: There is weak (suggestive) evidence that allyl chloride alters cell proliferation, cell death, or nutrient supply

β-picoline

Volume 122

- KC1: No adequate data for an evaluation to be made
- KC2: No evidence for genotoxicity
- KC3: No adequate data for an evaluation to be made
- KC4: No adequate data for an evaluation to be made
- KC5: No adequate data for an evaluation to be made
- KC6: No adequate data for an evaluation to be made
- KC7: No adequate data for an evaluation to be made
- KC8: No adequate data for an evaluation to be made
- KC9: No adequate data for an evaluation to be made
- KC10: No adequate data for an evaluation to be made

Chapter 3

Biomonitoring for occupational exposure to diisocyanates: a systematic review

Bernice Scholten, Laura Kenny, Radu-Corneliu Duca, Anjoeka Pronk, Tiina Santonen, Karen S. Galea, Miranda Loh, Katriina Huumonen, Anne Sleeuwenhoek, Matteo Creta, Lode Godderis, Kate Jones

Published in: *Annals of work exposures and health,* 2020, 64(6), 569-585.

Abstract

Diisocyanates are a group of chemicals that are widely used in occupational settings. They are known to induce various health effects, including skin- and respiratory tract sensitization resulting in allergic dermatitis and asthma. Exposure to diisocyanates has been studied in the past decades by using different types of biomonitoring markers and matrices. The aim of this review as part of the HBM4EU project was to assess: (i) which biomarkers and matrices have been used for biomonitoring diisocyanates and what are their strengths and limitations; (ii) what are (current) biomonitoring levels of the major diisocyanates (and metabolites) in workers; and (iii) to characterize potential research gaps. For this purpose we conducted a systematic literature search for the time period 2000-end 2018, thereby focussing on three types of diisocyanates which account for the vast majority of the total isocyanate market volume: hexamethylene diisocyanate (HDI), toluene diisocyanate (TDI) and 4,4'-methylenediphenyl diisocyanate (MDI). A total of 28 publications were identified which fulfilled the review inclusion criteria. The majority of these studies (93%) investigated the corresponding diamines in either urine or plasma, but adducts have also been investigated by several research groups. Studies on HDI were mostly in the motor vehicle repair industry [with urinary hexamethylene diamine results ranging from 0.03 to 146.5 µmol/mol creatinine]. For TDI, there is mostly data on foam production [results for urinary toluene diamine (TDA) ranging from ~0.01 to 97 µmol/mol creatinine] whereas the available MDI data are mainly from the polyurethane industry (results for methylenediphenyl diamine range from 0.01 to 32.7 µmol/mol creatinine). About half of the studies published were prior to 2010 hence might not reflect current workplace exposure. There is large variability within and between studies and across sectors which could be potentially explained by several factors including worker or workplace variability, short half-lives of biomarkers, and differences in sampling strategies and analytical techniques. We identified several research gaps which could further be taken into account when studying diisocyanates biomonitoring levels: (i) the development of specific biomarkers is promising (e.g. to study oligomers of HDI which have been largely neglected to date) but needs more research before they can be widely applied, (ii) since analytical methods differ between studies a more uniform approach would make comparisons between studies easier, and iii) dermal absorption seems a possible exposure route and needs to be further investigated. The use of MDI, TDI and HDI has been recently proposed to be restricted in the European Union unless specific conditions for workers' training and risk management measures apply. This review has highlighted the need for a harmonized approach to establishing a baseline against which the success of the restriction can be evaluated.

Introduction

Diisocyanates are a group of chemicals containing two isocyanate functional groups (R–N=C=O). These low molecular weight compounds first alter a human protein before becoming allergenic. As such they are further known to induce various health effects, including skin and respiratory tract sensitization resulting in allergic dermatitis and asthma (DECOS, 2018). There is also concern of potential genotoxicity and carcinogenicity of diisocyanates, with the degradation products and metabolites of 4,4'-methylenediphenyl diisocyanate (MDI) and toluene diisocyanate diisocyanate (TDI) both being classified as mutagenic and carcinogenic (ECHA, 2005; IARC, 1999; DECOS, 2018).

The two major diisocyanates in the European market are MDI (CAS 101-68-8; 100 000 – 1 000 000 tonnes per annum; ECHA, 2019a) and TDI (CAS 584-84-9 for 2,4-TDI and CAS 26471-62-5 for the mixture of 2,4-TDI/2,6-TDI; 100 000 – 1 000 000 tonnes per annum, ; ECHA, 2019b, c). A third diisocyanate with widespread use, especially in vehicle paints, is hexamethylene diisocyanate (HDI; CAS 822-06-0; 10,000 - 100,000 tonnes per annum; ECHA, 2019d). In Europe, MDI, TDI and HDI account for more than 95% of the volume of diisocyanate production (ECHA, 2017). Since there are no suitable alternatives for the majority of applications, the usage is not expected to decline in near future (ECHA, 2017). In addition to these three compounds, several oligomeric products (e.g. for HDI) and various other diisocyanates are registered in the European market [e.g. 1,5-naphthalene diisocyanate (NDI), CAS 3173-72-6; isophorone diisocyanate (IPDI), CAS 4098-71-9] and are manufactured and/or imported in smaller yet notable amounts.

Since diisocyanates are widely used in different applications in industry (including in the manufacturing of polyurethanes (PURs) and as hardeners in industrial paints, glues, varnishes and resins), occupational exposure during production and handling of these materials is a concern (McDonald et al, 2005). Workplace exposure to diisocyanates has been studied historically by using different types of biomonitoring markers and matrices. Because of the diversity of industrial usages of diisocyanates, and the variety of biomarkers, we aimed to systematically identify and report relevant occupational biomonitoring studies reporting use of diisocyanates published between 2000 and 2018, focussed on addressing the following questions:

i) Which biomarkers and matrices have been used for biomonitoring diisocyanates; what are their strengths and limitations?
ii) What are (current) biomonitoring levels of the major diisocyanates (and metabolites) in workers?
iii) What are potential research gaps with regard to studying diisocyanate biomonitoring levels?

Methods

A literature search was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology (Moher et al., 2009) from year 2000 to end 2018 in PubMed and Web of Science, using the following search terms for MDI, HDI and TDI: ("Occupational" OR "worker*") AND ("biomonitor*" OR "biomarker*" OR "urin*"). The full chemical names were also used.

A total of 161 publications were retrieved. Publications were subsequently evaluated based on the abstract (independently by two reviewers). Publications were excluded if published in a language other than English, they reported in-vitro or animal studies, or were mechanistic papers. Review papers identified from the 161 papers were screened for potential additional studies that were not retrieved via the systematic review. This evaluation process resulted in 59 publications being identified for full review by two reviewers. Upon full review, 31 studies were additionally excluded due to various criteria such as results being reported in earlier studies, no report of any biomarker results, volunteer studies or method development papers. One paper by De Palma et al. (2012), could not be retrieved. One relevant paper, by Jones et al. (2017), was missed by the predefined search criteria. Twenty-eight publications were then taken forward to the next stage of the review process.

Data from these publications were extracted by using a bespoke template that collected information on: study type, study participants, chemicals investigated, type of biomarker and matrix, measurement techniques and quality assurance. In order to compare results across studies, we standardized results as far as possible as µmol/mol creatinine for urine (converting any uncorrected values using an approximate creatinine value of 12 mmol/L (Cocker et al., 2011) and marking any approximate corrections as ~), nmol/L for plasma, and pmol/g for haemoglobin (Hb) and albumin.The publications were reviewed and ranked independently by one of two reviewers using a modified version of the LaKind scoring criteria (Table 1) (LaKind et al., 2014). The LaKind criteria were developed to assess study quality for non-persistent biomarker studies and were used here to give an indication of the overall quality of the study. This considers the specificity of the biomarkers used and the analytical techniques, the quality of the study design, sample handling and quality assurance. A sample of papers (15, ~20%) was independently scored by both reviewers, with the results being compared for quality assurance purposes. Very few instances were identified where scoring was diametrically opposed (one reviewer scored Tier 1 (highest quality), the other Tier 3 (lowest quality). Both sets of scoring for these sample papers were reviewed by a third researcher with discussion on harmonizing scoring approaches for the remaining papers. Papers were scored against eight categories with the total score potentially ranging from 8 (highest quality) to 24 (lowest quality), see Table 1.

Table 1. Adaptation of LaKind scoring criteria for isocyanates mini-review. Each paper was scored from 1 (Tier 1) to 3 (Tier 3) for each of the eight components, giving total possible scores from 8 (highest quality) to 24 (lowest quality).

Assessment component	TIER 1	TIER 2	TIER 3
Study participants	>20 occupationally exposed individuals	5-20 occupationally exposed individuals	Any other study (<5 occupationally exposed individuals, volunteers, general population)
Chemicals under investigation	HDI, TDI and/or MDI	IPDI, NDI	Any other isocyanates
Exposure biomarker and matrix	Biomarker in a specified matrix has accurate and precise quantitative relationship with external exposure, internal dose, or target dose e.g. diamines, Hb-adducts	Evidence exists for a relationship between biomarker in a specified matrix and external exposure, internal dose, or target dose but limited application e.g. other protein adducts or conjugates	Biomarker in a specified matrix is a poor surrogate (low accuracy and precision) for exposure/dose e.g. experimental biomarkers, non-specific markers such as general effect markers
Biomarker specificity	Biomarker is derived from exposure to one parent chemical.	Biomarker is derived from a limited number of parent chemicals, such as diamines.	Biomarker is derived from multiple parent chemicals with varying types of adverse endpoints.
Technique	Instrumentation that provides unambiguous identification and quantitation of the biomarker at the required sensitivity (e.g., GC–HRMS, GC–MS/MS, LC–MS/MS).	Instrumentation that allows for identification of the biomarker with a high degree of confidence and the required sensitivity (e.g., GC–MS, GC–ECD).	Instrumentation that only allows for possible quantification of the biomarker but the method has known interferants (e.g., GC–FID, spectroscopy).
Method characteristics –	Acceptable LoD		LoD above current state-of – the-art
Any specific weaknesses in study design leading to a TIER 3 score to be noted.	Samples with a known history and documented stability data or those using real-time measurements.	Stability not specifically assessed, but samples were stored appropriately and analyzed promptly.	Specific reason to query stability. E.g. samples with unknown history or known issues.
	Samples are contamination-free from time of collection to time of measurement (e.g., by use of certified analyte-free collection supplies and reference materials, and appropriate use of blanks both in the field and lab). Research includes documentation of the steps taken to provide the necessary assurance that the study data are reliable.	Study not using/documenting these procedures.	There are known contamination issues and no documentation that the issues were addressed.
Quality assurance	Study has used external QA where appropriate	Some QA used (note details)	No QA
Matrix adjustment	Study includes results for adjusted and non-adjusted concentrations if adjustment is needed. 24 h total urine collection is considered tier 1.	Study only provides results using one method (matrix-adjusted or not).	No established method for adjustment (e.g., adjustment for hair, saliva)

Results

Overview of various biomarkers available

Several biomarkers and matrices have been used to study diisocyanate exposure. Considering the very high chemical reactivity of diisocyanates molecules due to the two NCO chemical groups, the direct analysis of the parent compounds in human matrices, like urine or blood, is not possible. For this reason, the biomonitoring techniques used to monitor diisocyanates exposure investigate the presence of products of chemical degradation, such as diamines in urine, or products of metabolism, such as acetylated amines or protein adducts in urine or blood samples. Here, we will provide an overview of the available biomarkers, including reported suitability and half-life.

Amines

The majority of the papers accessed for this review (>90%) studied the corresponding diamines in either urine or plasma. This is based on analysing isocyanate derived diamines released by hydrolysis of protein adducts in plasma or urine (Cocker, 2007). The corresponding diamines for the three diisocyanates in this review are hexamethylene diamine (HDA) for HDI, toluene diamine (TDA) for TDI (both isomers) and methylene dianiline (MDA) for MDI. Since the elimination half-lives of these derived diamines in urine are relatively short (2–5 h), urine samples should be collected at the end of exposure (i.e. end of workshift for occupational exposure). Indeed, most of the papers here report post-shift samples ('spot' samples in the case of urine) although some also report pre-shift and others after a weekend. Diamines are not specific, i.e. diamines themselves can also be common industrial chemicals . In addition, where the primary exposure is not to the diisocyanate monomer (particularly HDI exposures), it is unclear whether the diamine method is measuring only monomer exposure; data suggest that it may be more than just monomer but unlikely to include all oligomers/pre-polymers (Cocker, 2011).

Protein adducts

Protein adducts, isolated from blood samples, can be divided in two groups: albumin adducts and Hb adducts (Sabbioni et al., 2007). The first group can be produced by direct reaction of diisocyanates with the protein, or through an intermediate step where diisocyanates react with glutathione, in both cases adducts of albumin are formed (Figure 1). These adducts are considered specific biomarkers of diisocyanates exposure; only molecules that contain one or more isocyanate (N=C=O) group can react directly with albumin or through an intermediate step with glutathione (Sabbioni et al., 2010). For albumin adducts, the most studied molecules are represented by MDI or acetyl-MDI reacting with lysine residues (MDI-Lys and AcMDI-Lys). It has been shown that albumin adducts, having a half-life of 20–25 days, could be used for the

determination of short to medium-term exposure to diisocyanates (Sabbioni et al., 2010).

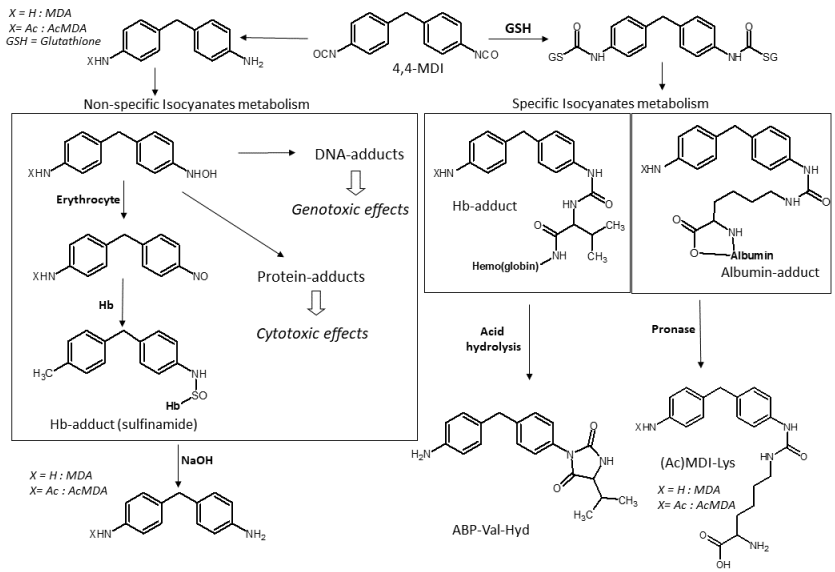


Figure 1 – General overview of the metabolic pathway of 4,4-MDI, as proposed by Gries et al. (2013); Sabbioni et al. (2010); Sabbioni (2017)

The second group of protein adducts, the Hb adducts, are formed through the reaction between diisocyanates and globin proteins. Hb adducts are also assumed to form through an intermediate step where diisocyanates react with glutathione (Gries and Leng, 2013). These Hb adducts can be measured directly (Gries and Leng, 2013) and are therefore specific, or they can be released by hydrolysis, resulting in diamines and are therefore non-specific. Hb has a lifetime of 120 days and its adducts therefore reflect longer-term exposure to diisocyanates (Flack et al. 2011).

Whilst the mechansims of sensitization are still unclear, it seems that glutathione conjugates may promote immune responses (Wisnewski et al, 2013) and therefore albumin adducts could potentially be considered biomarkers of effect as well as exposure.

Biomarker data

For amines data, the presented results focus on post-shift samples as these are most commonly reported and allow comparison between studies.

HDI

Reported biomonitoring levels

HDI is predominately used in spray paints within the motor vehicle repair (MVR) industry, see Table 2 for a summary of the studies identified. Seven studies were

identified as being from the MVR sector, two of these are from European countries (UK and Netherlands) (Jones et al., 2013; Pronk et al., 2006). The number of workers in all these studies varies substantially from 45 (Netherlands) to 995 (UK). The Netherlands study (Pronk et al., 2006) observed urinary HDA levels up to 150.2 µg/g creatinine (146.5 µmol/mol creatinine). Sampling was carried out at multiple time points throughout the day with the highest mean exposures occurring between the early afternoon and evening. Mean exposures for these time points were all ~20 µmol/mol creatinine. This is high compared with the UK study results (Jones et al., 2013) where the maximum result was ~20 µmol/mol creatinine and to an Australian study of 196 MVR workers which had almost 100% 'none detects' with only three results above the limit of quantification (LoQ) of 0.5 µmol/mol creatinine (Hu et al., 2017). Also the Netherlands may be considered high in comparison to results from the USA (Gaines et al., 2010) that reported post shift urine samples with HDA levels between <0.04 and 65.9 µg/L (~<0.03 – 47.3 µmol/mol creatinine).
Other industry sectors have not been so well studied, with only two UK studies adding to the list. These were studies at small and medium sized enterprises (SMEs) producing polyurethane elastomer (Cocker et al., 2009) and a survey of SMEs with a variety of uses (Creely et al., 2006). Results were reported as total isocyanates (HDA + TDA + MDA + IPDA urine levels) and so it was not possible to compare HDI exposure with other studies. There were a low number of detects in the polyurethane (PUR) study (Cocker et al., 2009), with 9 of the 13 positive sample results being above the UK biological monitoring guidance value of 1 µmol/mol creatinine showing potential for individual exposure in this industry (HSL, 2005).

One of the biomarkers reported for the first time was TAHI (trisaminohexyl isocyanurate), a specific metabolite to the oligomer HDI isocyanurate (Robbins et al., 2018). TAHI was detected in a third of the 111 exposed workers. Most of the data in the retrieved studies from the MVR described exposure to the HDI monomer (due to methods and standards not being available for oligomer exposure detection in urine prior to 2018) even though oligomers make up the bulk of 2-pack spray paints (Rosenberg and Savolainen, 1986; Fent et al., 2008). Pronk et al. (2006) also demonstrated significantly higher concentrations of HDI oligomers in personal inhalation samples when compared with HDI monomer. Mean values of NCO (isocyanate content) exposure for sprayers was 2.1 µg/m^3 of NCO from monomer exposure and 116.3 µg/m^3 of NCO from oligomer exposure.

Correlations

Some studies show a relationship between plasma levels of HDA and air levels of HDI, such as a US MVR study (Flack et al., 2010) here plasma HDA levels were weakly correlated (r = 0.22) with personal inhalation exposure samples. A study in the following year (Flack et al., 2011) from the same researchers showed that plasma HDA correlated best with inhalation exposure when taking cumulative exposure into account, r=0.61. This was also observed for Hb adducts, however Hb and plasma

levels were not correlated, most likely due to the different turnover times of these proteins in the body. Gaines et al. (2010) and Pronk et al. (2006) reported that dermal, as well as inhalation exposure, was a significant predictor of urinary biomarker levels.

TDI

Reported biomonitoring levels

TDI is a volatile diisocyanate used in foam blowing, glues/adhesives and lacquers, see Table 3 for a summary of the studies identified.

Table 2. Summary of HDI exposure studies for the main processes.

Sector	Study Populations (Country, no. workers)	Biomonitoring data (Expressed as range (median))	Notable correlations / comments	LaKind scoring[a]	References
MVR	USA, 15	Plasma HDA: 0.012-0.71 (0.061*)µg/L HDA-Hb: 1.3-37 (3.0*) ng/g Hb	Hb and plasma weakly associated. Air: Correlated with cumulative exposure (Hb: r^2= 0.34, $p < 0.05$; P: r^2= 0.37, $p < 0.05$) 10 x higher Hb adducts than plasma due to cumulative exposure and turnover times. Positive association between HDA-Hb adduct concentration and HDI exposure was strongest with cumulative dermal (N = 12, r2 = 0.32, p = 0.058), cumulative inhalation (N = 12, r2 = 0.35, p = 0.042), or cumulative air exposure (N = 12, r2 = 0.34, p = 0.048).	15	Flack *et al.* 2011
	USA, 46	Plasma HDA: 0.02-0.92 µg/L	Inhalation correlation r = 0.22, p = 0.026	14	Flack *et al.* 2010
	USA, 48	Urine HDA: <0.04-65.9 µg/L (0.10*) [~0.03 – 47.2 µmol/mol cr.]	Dermal and inhalation exposure found to be significant predictors of urinary biomarker levels.	12	Gaines *et al.* 2010
	USA, 15	Urine TAHI <LoD – 1.99 µg/L (means) [<LoD – 0.39 µmol/mol cr.]	TAHI reported for first time. Positive correlation between HDI isocyanurate exposure and total urine TAHI concentration (r = 0.14 with creatinine adjustment)	13	Robbins *et al.* 2018
	Australia, 196	Only 3 above LoQ (0.5 µmol/mol cr.)	Positive spray booths thought to be just as effective as negative because of high level of non-detects.	11 External QA	Hu *et al.* 2017
	UK, 995	Pre intervention: 1.34 (90%) µmol/mol cr. Intervention : 0.60 (90%) µmol/mol cr. Post intervention: 0.68 (90%) µmol/mol cr.	Participants invited to a Safety and Health Awareness Day (SHAD). Samples taken before and after show lower results after the intervention.	11	Jones *et al.* 2013
	Netherlands, 55 (10 workers from industrial paint shop)	Urine HDA: <2.9 – 146.5 µmol/mol cr. (means)	Higher levels of oligomers than monomers detected in air samples. Highest concentrations of HDA in urine seen in the afternoons and early evening. Dermal exposure was a predictor of the presence of HDA.	11 24-h samples	Pronk *et al.* 2006

Sector	Study Populations (Country, no. workers)	Biomonitoring data (Expressed as range (median))	Notable correlations / comments	LaKind scoring[a]	References
Other	UK, 71	Urine HDA: 56 < LoD, 13 > LoD. 9 > Biological Monitoring Guidance Value (BMGV). <0.5 – 10.1 (1.8) µmol/mol cr.	About 25 companies were visited that were involved in the manufacture of polyurethane products.	11	Cocker *et al.* 2009
	UK, 67	Urine HDA: Results reported as total isocyanates, HDA most common detected n=21	Low airborne concentrations, only 20% above LoQ. Mixing and pouring tasks seen as a major potential source of exposure. Biased towards good practice.	12	Creely *et al.* 2006

[a] *The lower the LaKind score the better the overall quality (possible range 8-24)*
[b] *Geometric mean (rather than median)*

Continuous foam production has been the most studied individual process [six papers from five different European Union (EU) countries], although the number of workers per study was generally small (N <30) (Table 3). Maximum observed results ranged from 3.9 µmol/mol creatinine (Swierczynska-Machura et al., 2015) to 97 µmol/mol creatinine (Geens et al., 2012).

Moulding processes have been studied in larger (n=18 – 90) worker populations than for continuous foam production (n= 4 - 26) but were only reported in two countries: the UK (Cocker et al., 2009; Keen et al., 2012) and Sweden (Sennbro et al., 2004; Tinnerberg et al., 2014). If mixed processes (which include some moulding companies but results are not reported separately) are included (Littorin et al., 2007; Sakkinen et al., 2011) then three countries are covered (UK, Sweden and Finland) and maximum results for urinary TDA range from ~3.2 (Tinnerberg et al., 2014) to ~110 µmol/mol creatinine (Sennbro et al., 2004). If the Sennbro study (2004) is excluded, the results are more comparable across studies, with maximum urine TDA levels of ~3.2 (Tinnerberg et al., 2014), >6.5 (Keen et al., 2012), only 90th percentile reported), 15.5 (Cocker et al., 2009), ~29.3 (Littorin et al., 2007) and 39 (Säkkinen et al., 2011) being reported; all results in µmol/mol creatinine.

There were very few studies looking at other uses of TDI, such as glues, spray adhesives or heat guns; these were sometimes included in “mixed” studies involving multiple sites but the results were not reported separately. Sakai et al. (2005) reported on urethane spray painting for lacquering musical instruments. This study showed a good correlation between urine TDA and airborne TDI (r>0.9 for post-shift creatinine-corrected urinary 2,6-TDA and 2,6-TDI). Results up to 19 µmol/mol creatinine were reported. One study (Rosenberg et al., 2002) examined thermal degradation processes such as cutting, welding and grinding. Airborne levels were generally less than 5% of the occupational exposure limit (OEL) and urine TDA levels were very low (all less than 1 µmol/mol creatinine).

Correlations

Except where respiratory protective equipment (RPE) use or skin contact was significant (and noted), there was generally a strong correlation (r>0.8) between urine TDA and airborne TDI (Swierczynska-Machura et al., 2015; Geens et al., 2012; Sennbro et al., 2004; Littorin et al., 2007; Sakai et al., 2005; Sakkinen et al., 2011). However, the resulting estimates of a urine TDA level from a 5 ppb TDI exposure varied significantly from 4.1 µmol/mol creatinine (Swierczynska-Machura et al., 2015); (only based on GM correlations) to ~66 µmol/mol creatinine (Sennbro et al., 2004).

Fewer data were presented on correlation with air monitoring in other studies compared to continuous foam production (Table 3) although Sennbro et al. (2004), Littorin (2007) and Sakkinen (2011) all report positive correlations (r>0.75).

Table 3. Summary of TDI exposure studies for the main processes.

Sector	Study Populations (Country, no. workers)	Biomonitoring data (expressed as range (median))	Notable correlations / comments	LaKind scoring[a]	References
Continuous foam production	Poland (n=20)	Sum-TDA (U)= <0.01 to 3.9 µmol/mol cr.	Positive for geometric mean (GM) in each group (r.0.9) *U_totTDA (µmol/mol) = 0.10777_TDI(µg/m³) + 0.2178* [5ppb TDI = 4.1 µmol/mol cr.] RPE used and observed to impact TDA(U) results (no correlation between individual urinary TDA concentrations and TDI air concentrations)	12	Swierczynska-Machura *et al.*, 2015
	UK (n=26, 13 handlers, 13 non)	Sum-TDA (U)= <~0.4 to 7 (2.21) µmol/mol cr. (handlers)	No correlation between post-shift urinary TDA concentration and airborne TDI concentrations (r=0.027). Dermal considered a significant factor – urine 20x higher for same airborne exposure.	16 Methodology not well described	Austin *et al.*, 2007
	Belgium* (n=9)	Sum-TDA (U)= 4.4 - 142.6 (18.01*) µg/L [21 samples] [~3 – ~97 (~12.3) µmol/mol cr.]	*TDA (µg/g)=0.547_TDI (µg/m³) - 1.636*, r=0.917 [5ppb TDI =19.2 µg/g, 17.8 µmol/mol cr.] Proposed measuring "increase over shift" to exclude accumulation.	12 External QA	Geens *et al.*, 2012
	Finland (n=17)	Sum-TDA (U)= <0.05 to 39 µmol/mol cr.	Good correlation between airborne TDI and urinary TDA in post shift samples (r = 0.91 and 0.86 for the two different factories studied)	13	Karaai *et al.*, 2001
	Sweden (n=6)	2,4-TDA(U) / 2,6-TDA(U) 0.5 – 5.4 / 0.2 – 4.7 µg/L [~0.3 – 3.7 / 0.14 – 3.2 µmol/mol cr.] 2,4-TDA(P) / 2,6-TDA(P) 0.1 – 14 / 0.7 – 12 µg/L	Samples taken Monday morning so results not comparable to other studies. Only reported levels 'above reference value'.	12 Creatinine and specfic gravity	Tinnerberg *et al.*, 2014
	Sweden** (n=4 in 2000, n=6 in 2005)	2,4-TDA (U) ~ 0 - 10 µmol/mol cr. 2,6-TDA (U) ~ 0 – 35 µmol/mol cr. 2,4-TDA (P) / 2,6_pTDA (µg/L) In 2000: 2.9-27.2 (7.0) / 8.2-62.1 (30.8) In 2005: 0.5-1.3 (1.0) / 2-11.8 (4.0)	Urine results only presented graphically.	14 Small study	Tinnerberg *et al.*, 2008
	Finland (n=17)	Sum-TDA(U) 0.2 – 39(4.9) µmol/mol cr. Sum-TDA(P) 0.4 – 70.8 (5.6) µg/L	Air TDI and plasma TDA correlated (r=0.91). Plasma and urine TDA well correlated (r=0.97).	13	Sakkinen *et al.*, 2011

Sector	Study Populations (Country, no. workers)	Biomonitoring data (expressed as range (median))	Notable correlations / comments	LaKind scoring[a]	References
		Sum_TDA(Hb) 0.012 – 0.33 (0.047) (nmol/g)			
	UK (n=71)	Sum-TDA(U) <0.5 - 15.5 (1.3) µmol/mol cr. 2,6-TDA(U) <0.5 - 13.2 (0.8) µmol/mol cr. 2,4-TDA (U) <0.5 – 5.6 (0.7) µmol/mol cr.	The companies visited were involved in the manufacture of PUR products	11	Cocker *et al.*, 2009
	UK (n=90)	Sum-TDA(U) (µmol/mol cr.) <0.4 – 6.5 (90%, median <LOD)	Positive association observed in 4 pairs of samples (air and urine). Air levels <LoD at 2/5 sites and only 1/11 samples >WEL (20 µg/m^3 NCO) – no further data. 446 samples analysed of which 280 were below the detection limit. Follow-up to Cocker, 2009	12	Keen *et al.*, 2012
Mixed	Sweden (n=136) (including : moulding -, continuous foam -, and flame-lamination plants)	2,6-TDA(U) <0.05 – 43.1 µg/L [~<0.03 – 29.3 µmol/mol cr.] 2,6-TDA(P) <0.05 – 62.1 µg/L	Correlation (r>0.86) with air levels for both urine and plasma for same-day samples. Same sites as Sennbro, 2004.	12	Littorin *et al.*, 2007
	Sweden (n=81)	2,4-TDA(U) / 2,6-TDA(U) / Sum-TDA(U) <0.1 – 47 (4.5) / <0.1 – 115 (3.7) / <0.1 – 162 (9.7) µg/L [~<0.07 – 32 (3.1) / - 78 (2.5) / -110 (6.6) µmol/mol cr.] 2,4-TDA(P) / 2,6-TDA(P) / Sum-TDA(P) <0.1 – 31 (7.4)/ <0.1-42 (6.1) / <0.1 – 70 (14) µg/L	High correlations between air exposure and urinary biomarker levels (ranging from 0.75-0.88) or plasma biomarker levels (ranging from 0.50-0.78). 2,6-TDA(U) (µg/L) = 2.7_TDI(ppb) + 0.02 (r=0.88) [5ppb=97 µg/L, ~66 µmol/mol cr.]	13	Sennbro *et al.*, 2004

Sector	Study Populations (Country, no. workers)	Biomonitoring data (expressed as range (median))	Notable correlations / comments	LaKind scoring[a]	References
	Japan* (n=18) (spraying urethane paints)	Individual results not reported except graphically, 2,6-TDA(U) <19 µmol/mol cr.	2,6-TDA(U) (µg/g) = 6.6_TDI (ppb) - 1.43 (r=0.91) [5ppb = 29 µmol/mol cr.]	13 Both raw & corrected data	Sakai *et al.*, 2005
	Finland* (n=6, car repair ; n=15 other PUR processes	Car repair Sum-TDA(U) <0.02–0.76 (0.23) µmol/mol cr. Other processes <0.02 – 0.17 (0.07) µmol/mol cr.	Other processes included milling and turning of PUR-coated metal cylinders, injection moulding of thermoplastic PUR, welding of district heating pipes and joint welding of PUR floor covering.	11	Rosenberg *et al.* 2002

[a] *The lower the LaKind score the better the overall quality (possible range 8-24).*
[b] *Geometric mean rather than median*
[c] *Country of origin assumed from authors' affiliation, not specifically stated in paper*

No specific albumin adducts of TDI have been measured in plasma samples, only plasma TDA was assessed after alkaline or acid hydrolysis. Sakkinen et al. (2011) analysed 2,4-TDA (6 – 270 nmol/L) and 2,6-TDA (3 – 310 nmol/L) in plasma after acid hydrolysis with sulphuric acid 3M (100° C for 16 h) and found a very good correlation (r=0.91) between the plasma total TDA and the airborne TDI. Sennbro et al. (2004), Littorin et al. (2007) and Tinnerberg et al. (2014) used basic hydrolysis with 0.3 M NaOH (24 h) in order to measure plasma concentration of TDA. Similarly to Sakkinen et al. (2011), both Sennbro (2004) and Littorin (2007) found same range levels of plasma TDA as well as good correlations with the air levels of TDI (r=0.50 for 2,4-TDA and 0.78 for 2,6-TDA; and r=0.72 for 2,4-TDA and 0.86 for 2,6-TDA, respectively).

As for albumin adducts, no direct measure of TDI Hb adducts has been performed. Sakkinen et al. (2011), after isolation of globin and acid hydrolysis with sulphuric acid 3M (100° C for 16 h), were able to quantify 2,4-TDA (13 - 120 pmol/g) and 2,6-TDA (12 - 210 pmol/g) in globin, as representative for Hb adducts, but no correlation was observed between the Hb total TDA and airborne TDI. This might be, amongst others, due to the relatively low number of detects among the occupationally exposed workers (n=4 for 2,4-TDA and n= 10 for 2,6-TDA), or Hb total TDA might be a reflection of long-term exposure unlike airborne TDI measurements.

MDI

Reported biomonitoring levels

Available studies on MDI (Table 4) are mostly from the PUR sector with five papers classified under PUR production/use (Robert et al., 2007; Sennbro et al., 2006; Cocker, 2009; Keen et al., 2012; Rosenberg et al., 2002). Exposures reported in studies from France (Robert et al., 2007) and Sweden (Sennbro et al., 2006) are in reasonable agreement; urine levels of MDA up to 33.7 µg/g creatinine (19.2 µmol/mol creatinine) in the Robert (2007) study and 78 µg/L (~32.6 µmol/mol creatinine) in the Sennbro (2006) study. Low levels were reported in the UK PUR elastomer industry with only six of the 71 workers sampled having detectable exposures (<0.5 – 0.7 µmol/mol creatinine) in one study (Cocker et al., 2009) and a reported 90th percentile of 0.5 µmol/mol creatinine in a second study (Keen et al., 2012). A study conducted in Finland (Rosenberg et al., 2002) examined exposure to thermal degradation products of PURs in a number of processes including grinding and welding in MVR, milling and turning of PUR-coated metal cylinders, injection moulding, welding and cutting heating pipes, joint welding, and heat-flexing of PUR floor covering. Exposures were low overall (0.01 – 3.1 µmol/mol creatinine) with pipe layers receiving the highest exposures.

One paper was identified from Sweden (Littorin et al., 2000) in which 150 workers were using glue containing MDI, with some of these workers using the glue heated. Maximum MDA urine levels from workers using heated glue were over five times

higher than those using non-heated glue, 9.4 µg/L (~3.9 µmol/mol creatinine) and 1.8 µg/L (~0.8 µmol/mol creatinine), respectively.

Three papers were identified from the construction industry (Sabbioni et al. 2007, 2010; Henriks-Eckerman et al., 2015). Urine MDA levels were low for both study groups; 0.017 -16.4 nmol/L (~0.001 – 1.4 µmol/mol creatinine) in a Switzerland study (Sabbioni et al., 2007) and 0.1 – 0.2 µmol/mol creatinine in a Finnish study (Henriks-Eckerman et al., 2015). Workers were reported to be involved in a range of activities, including spray foaming which could generate higher exposures, but RPE was used in the majority of cases.

In the 2007 study from Sabbioni, MDA in urine from acid hydrolysis correlated well with acetyl MDA (acMDA), released by alkaline hydrolysis, and Hb-MDA with r = 0.86 and r = 0.39 respectively ($p < 0.01$). Inhalation and dermal were both identified as routes of exposure in the study of Finnish workers (Henriks-Eckerman *et al*., 2015); this was also the only study to report dermal concentrations (88% workers <2 µg MDI/10cm^2 on hand). A French PUR study of 169 workers also noted evidence of dermal exposure (Robert et al., 2007).

Specific albumin adducts (MDI-Lys and AcMDI-Lys) have been measured by Sabbioni *et al*. (2010) in two groups of workers: construction workers (n=65) and workers of a non-specified MDI plant (n=27). The specific biomarker MDI–Lys, was found in 63% of the construction workers, in 64% of the plant workers, and in only 15% for the control group. In Sabbioni *et al*. (2017) a higher concentration (750 pmol/g) was found in workers with asthma who had recently worked with MDI compared with asthmatic workers who had not used MDI for more than 3 months (MDI-Lys 191 pmol/g). These MDI- Lys biomarkers are more specific than hydrolyzed MDA levels (as they are not confounded with MDA itself) and can be used as exposure biomarkers. Sabbioni (2007) found that these biomarkers accumulate in the body over time. Elevated levels were observed in workers samples after working with MDI over a period of 4 – 7 months.

Gries and Leng (2013) have found a higher concentration of the specific MDI biomarker ABP-Val-Hyd in exposed workers (0.15-16.2 pmol/g) compared with non-exposed workers (below LoQ). Säkkinen *et al*. (2011) measuring MDA after acid hydrolysis of globin, found the same order of magnitude between exposed and non-exposed workers; this is most likely due to the hydrolysis releasing non-specific adducts and converting them to MDA.

Correlations

Although MDI is less volatile than HDI and TDI, correlations with air levels were still observed. Sennbro *et al*. (2006) reported weak but significant correlations ($r>0.5$) with air levels for both urine ($p <0.01$) and plasma ($p < 0.05$) biomarkers. The Finnish study

from the construction industry (Henriks-Eckerman *et al*., 2015) observed that the use of RPE led to lower post shift urine levels of MDA.

Sakkinen *et al.* (2011) reported urine MDA levels of workers (n = 65) at 0.015 – 1.4 µmol/mol creatinine in post shift samples and plasma concentrations of MDA at 9 – 13 nmol/L. There were insufficient positive plasma results to investigate correlations between urine and plasma levels of MDA. Tinnerberg *et al.* (2014) reported MDA urine levels of workers (n = 21) at 0.5 – 8.4 µg/L (~0.2 – 3.5 µmol/mol creatinine) and plasma concentrations of MDA at 0.4 – 19.4 µg/L (2.0 - 98.0 nmol/L). Significant correlations ($p < 0.01$) were seen between plasma MDA and urinary MDA levels ($r = 1.000$, 0.988, and 0.986 for unadjusted, specific gravity adjusted and creatinine adjusted urine respectively) (Tinnerberg *et al*., 2014).These results cannot be directly compared to the Sakkinen study as the samples were taken on a Monday morning after a weekend of no exposure.

Discussion

Biomarkers for biomonitoring of diisocyanates

The majority of the papers (> 90%) assessed for this review studied the corresponding diamines in either urine or plasma. Where both have been studied, a good correlation ($r>0.85$) is generally seen between diamines in urine and plasma (Tinnerberg *et al*., 2014; Tinnerberg and Mattsson, 2008; Sennbro *et al*., 2003) although significant individual variation was observed and correlations could not be compared between studies due to sample collection differences or the way the data were reported. Aside from the diamine biomarkers, the quantity of identified literature published is limited. The use of adducts has been investigated by several research groups although different adducts have been used such as diisocyanate-specific Hb adducts, diamine-specific Hb adducts and diisocyanate-specific albumin adducts. One major benefit of the adduct biomarkers is the ability to study exposures over the longer-term; amines are really only representing same day or previous day exposures whereas adducts can represent exposures over weeks (albumin) and months (Hb). It is probably necessary to consider both peak and chronic exposures in the assessment of occupational asthma risk as the exact mechanisms of sensitization and exacerbation are still unclear. Development of more specific biomarkers (U-TAHI for HDI-isocyanurate, ABP-Val-Hyd for MDI, MDI-lysine for MDI) is reported by individual papers. These are yet to be used widely and there are potential issues with standards (they are not commercially available yet) so demonstrating comparability of results will be difficult initially. To date there are few data correlating adducts to diamines or airborne levels so these cannot be interpreted with respect to risk at this time. Although the LaKind scoring criteria provide useful clarity on the strengths and weaknesses of papers when considering papers as part of a review, the ‘total score’ is less useful as papers that have different strengths may have the same ‘total score’. The ‘total score’

should therefore not be the sole consideration when differentiating papers. Despite this, the best quality papers can be identified by a low score. From this review, it is clear that many papers still do not adequately present methodological details (often citing another paper with no relevant details in the present paper). Other factors, such as quality assurance and sample handling, are also often not reported although they may have in fact been done. There is therefore a need for journal reviewers to request these details more often. Given the widespread facility to submit supplementary data for papers, more reporting of both corrected and uncorrected urine data should be encouraged because there is still not always a consensus on whether particular biomarkers are best reported in one or other format. Some criteria (such as study size) do not reflect quality per se but rather the generalizability of the study results to other scenarios.

Table 4. Summary of MDI exposure studies for the main processes.

Sector	Study Populations (Country, no. workers)	Biomonitoring data (expressed as range (median))	Notable correlations / comments	LaKind scoring[a]	References
(Rigid) foam productio n	Finland, n=57	Urine: 0.015-1.4 (0.13) µmol/mol cr. Plasma: 1.8 – 2.6 µg/L 18 - 37 pmol/g (Hb)	Airborne levels (very low or not detected) and task time not associated with urinary biomarker levels.	13	Sakkinen 2011
	Sweden, n=18	Urine: 0.5-8.4 µg/L [~0.2 – 3.5 µmol/mol cr.] Plasma: 0.4-19.4 µg/L	Plasma and urinary MDA correlated after 2 days of no exposure. (P > 0.986)	12 Creatinine and specfic gravity	Tinnerberg 2014
PUR industry (Generic)	France, n=169 (19 factories)	Urine: <0.1-23.6 µg/L [<0.5 – 19.25 µmol/mol cr.]	Association with skin exposure. Elevated pre-shift levels but not cumulative. Higher MDI % in formulations not associated with higher results.	13 Both raw & corrected data	Robert 2007
	Sweden, n=18	Urine: 0.3-78 (2) µg/L [~0.13 – 32.7 (0.8) µmol/mol cr.] Plasma: 0.2-74 (0.7) µg/L	Weak but significant correlations with air. p < 0.01 (U), p < 0.05 (P)	12 Both raw & corrected data	Sennbro 2006
	UK, n=71	Urine: <0.5 - 0.7 µmol/mol cr. (only 6+ve/71 results above LoD)	Low levels of isocyanate exposure in the PUR elastomer industry.	11	Cocker 2009
	UK, n=90	Urine : 56/326 > LoD, 90% 0.5 µmol/mol cr. (median < LOD)		12	Keen 2012
	Finland, n=21	Urine : <0.01-3.1 µmol/mol cr.	Low exposures but highest levels seen in pipe layers.	11	Rosenberg 2002
PUR industry (Glue)	Sweden, n=150	Urine: <LoD-1.8 µg/L [~<LoD – 0.8 µmol/mol cr.] <LoD-9.4 (Heat) µg/L	Higher exposure levels when using heated glue.	11 External QA	Littorin 2000
Constructi on	Switzerland*, n=65	Urine: MDA 0.003- 3.2 µg/L [~0.001 – 1.3 µmol/mol cr.] Median (P90) Hb-MDA: 0 (0.177) pmol/g Hb Hb-AcMDA: 2 positive, 2.3 and 3.7 pmol/g Hb	U-MDA-tot correlates with U-AcMDA and Hb-MDA with r= 0.86 and r=0.39, respectively (p<0.01). U-AcMDA correlates with Hb-MDA with r= 0.47, (p<0.01). U-MDA correlates with Hb-MDA (r=0.38, p=0.05).	14 Methodology not well described	Sabbioni 2007

Sector	Study Populations (Country, no. workers)	Biomonitoring data (expressed as range (median))	Notable correlations / comments	LaKind scoring[a]	References
	Switzerland*, n=65	Albumin: MDI-Lys 0-899.4 fmol mg^{-1} AcMDI-Lys: 0-51.2 fmol mg^{-1}	Correlation MDI–Lys with MDA-Hb, r= 0.295 ($p<0.05$) Same workers as Sabbioni 2007 MDI–Lys levels were compared in a subgroup of construction workers (n = 19) which were analysed prior to isocyanate exposure and after 4–7 months of isocyanate exposure ; The MDI–Lys levels increased significantly (Wilcoxon sign test, $p < 0.01$).	12	Sabbioni 2010
	Finland, n=21	Urine: <0.1-0.2 µmol/mol cr. Dermal: 88% <2µg MDI/10cm^2 on hand	Effect of RPE lowering exposure seen in post-shift samples but not evening and following morning samples indicating 2 routes of exposure, dermal and inhalation.	12 Both raw & corrected data	Henriks-Eckerman 2015
Other	Switzerland* n=27 (chemical industry)	Urine: MDA 0 - 10.2 (1.7) nmol/L [~0 – 0.9 (0.142) µmol/mol cr.] Albumin: MDI-Lys 0-138 pmol/g AcMDI-Lys: 25.6 pmol/g (1 +ve)	Correlation MDI–Lys with MDA-Hb, r= 0.382 ($p<0.05$)	12	Sabbioni 2010
	Switzerland, n=73 (Urethane mould production)	Albumin: MDI-Lys=191 pmol/g (mean) (based on 4 workers with asthma who reported that their last activity with MDI was >3 months ago) MDI-Lys= 750 pmol/g (mean) (n=5)	Workers with confirmed asthma had significantly higher adduct levels than healthy worker.	14 Methodology not well described	Sabbioni 2017
	Germany* n=25	Urine: <500-124490 pmol/g creatinine [0.004 – 1.1 µmol/mol cr.] Hb: MDA <0.35 – 1.12 pmol/g ABP-Val-Hyd 0.15 – 16.2 pmol/g	No exposure assessment; measurement of Hb adducts and ABP-Val-Hyd reflect long term exposure (up to 120 days)	12	Gries 2013

[a] *The lower the LaKind score the better the overall quality (possible range 8-24).*
[b] *Country of origin assumed from authors' affiliation, not specifically stated in paper.*

Reported biomonitoring levels

With regard to our second aim: we have provided an overview of available studies on reported exposure in available diisocyanate biomonitoring studies. About half of the studies were prior to 2010 hence perhaps do not reflect current workplace exposures, especially given subsequent EU restrictions and more recent product developments (ECHA, 2010). Also we noticed that a substantial number of studies measured airborne levels below the OEL. This indicates that secondary exposures such as cutting, welding and grinding (after curing) are likely to be minimal. Exposure levels can only be considered for the urinary and plasma diamines where there is a sizeable body of data. There is large variability within and between studies and across sectors making it difficult to identify high risk sectors. The observed large variability indicates that elevated exposures are most likely due to individual (whether worker or workplace) factors rather than reflecting an industry or a process as a whole.

Austin (2007) observed that "handlers" of foam blocks had much greater urinary TDA levels (by up to 20 times) for the same airborne TDI exposure as "non-handlers". This finding points to the potential for dermal TDI uptake although co-exposure to TDA (known to be well absorbed through the skin and shown to be generated in foam blowing processes (Jones *et al*., 2017) possibly contributed to the body burden.

The short half-lives of amines and the timing of sampling in relation to the activities are probably also influential and another factor that could contribute to the observed large variability is the use of RPE; when comparing the studies on HDI in the MVR we noticed that virtually all sprayers in the UK study wore air-fed visors whereas these were not used in the Netherlands. Swierczynska-Machura *et al.*(2015) also stated that the lack of observed correlation with air levels in their study could be due to the use of RPE.

Another possibility for the observed variability is the potential issue with comparing TDI results across studies with regard to the different sample hydrolysis procedures. Generally, one of three different conditions is applied (acid hydrolysis for 1.5 hours, acid hydrolysis for 16 hours or alkaline hydrolysis for up to 24 hours). Alkaline hydrolysis has been reported to give higher results as more adducts are released. Note, e.g., that Geens *et al*. (2012) and Kääriä *et al*. (2001) used the higher release alkaline method in comparison to the other TDI biomonitoring studies (Table 3).

A couple of papers (Sakai *et al*., 2005; Sennbro *et al*., 2003) examined different hydrolysis conditions and concluded that the results from the different techniques are well correlated but that alkaline hydrolysis can result in ~50% higher results in urine and ~10% higher in plasma. Such differences are not so extreme as to negate a comparison of the different studies, processes and exposure risks so we have not 'corrected' reported results for hydrolysis method. Furthermore, although methods are well described and published by respected institutes, the lack of reported external

quality assurance means that direct equivalence of the studies cannot be assumed. Although this issue is not specific to diisocyanates, it is an area where journals could seek to improve reporting. It also highlights the value of studies such as HBM4EU (www.hbm4eu.eu) and DEMOCOPHES (Casteleyn *et al*., 2015) where significant effort has gone into ensuring that results from different studies/countries within the project are comparable.

Research gaps

We have identified a number of research gaps based on an analysis of the available data. Firstly, much of the focus of diisocyanate exposure assessment has been, naturally, inhalation exposure (given it causes occupational asthma) however several studies (Austin, 2007; Robbins *et al*., 2018; Robert *et al*., 2007) indicate the possibility of dermal absorption or the need to further investigate routes of exposure (Keen *et al*., 2012). As these studies have used the diamine biomarkers, it is not certain that any dermal uptake is from the intact diisocyanates. Obviously one of the primary advantages of using biomonitoring is that exposures when relying on RPE or where skin uptake is possible are difficult to assess by other means. Further, in general, we found that most studies showed a fair correlation between urinary amines levels and airborne measurements. However, in some cases where dermal exposure was likely or RPE was used, the correlation was weaker. This also illustrates the added value of taking biomonitoring samples.

Furthermore, the vast majority of papers discovered were studying HDI, TDI and/or MDI. There were brief mentions of NDI (Tinnerberg *et al*., 2014; Tinnerberg and Mattsson, 2008) and IPDI (Creely *et al*., 2006) however, these were not a focus of the review and so not specific search terms in the initial literature search. One of the new 'innovations' in isocyanate use has been the introduction of so-called "blocked isocyanates" into paints and coatings, where the –NCO group is chemically blocked until the reaction process (which usually takes place in an enclosed booth or oven). This is proposed to be a safer use although the diisocyanate is still released during the reaction process. No papers were found looking specifically at the exposures from blocked diisocyanates so the assumption of safer use has not been demonstrated and may warrant investigation.

Lastly, we looked into potential research gaps with regard to industrial sectors studied. For TDI, there is a reasonable amount of data on flexible foam production (Poland, Sweden, UK, Belgium, and Finland). Less data are available for the uses of TDI in rigid PUR manufacture and the use of heated glues. Littorin *et al*. (2000) suggest that heated glues may present a greater exposure risk than unheated glues.

The available MDI data are mainly from the PUR industry with just one study looking at exposure from glues and two concerned with the construction sector where there are potentially several different sources of exposure such as sprayed insulation. Paint

spraying in MVR has long been considered potentially high risk due to the very high concentrations of HDI that have been measured in spray booths (hundreds to thousands of μg/m^3, well in excess of current exposure limits). Whilst MVR has been well studied globally (including in the Netherlands and UK in Europe), exposures in other transport sectors such as aerospace, shipping and large commercial vehicles have not been widely reported, with data only available from Netherlands (Pronk *et al*., 2006). These sectors may present different exposure issues as the use of enclosed spray booths is impractical. The most relevant industrial uses of diisocyanates have been reported to be in the manufacturing of: diisocyanate compounds themselves, PUR and PUR composite materials, foam (spray foam applications), coatings and adhesives (RAC/SEAC, 2017). The direct manufacture of PUR plastic materials accounts for more than 90% of the use of diisocyanates. We did not find any European studies on manufacture of spray adhesives or coatings; hence these sectors might also be worthwhile investigating.

The use of MDI, TDI and HDI has been recently proposed to be restricted in the EU unless specific conditions for workers' training and risk management measures apply (RAC/SEAC, 2017/ 2018). The aim of the restriction is not, however, to ban the use of diisocyanates but rather to improve the control of diisocyanate use by obligatory training for good working practices and risk management. There is evidence that exposures can be well controlled within a study population e.g. 98% of 196 workers less than LoQ (Hu, 2017), >90% of 995 workers less than Great Britain guidance value of 1 μmol/mol creatinine (Jones *et al.*, 2013), both for HDI in MVR. However there are also examples of significantly elevated results of up to 100 μmol TDA/mol creatinine (Geens *et al*., 2012; Sennbro *et al*., 2004), which are well in excess of the American Conference of Governmental Industrial Hygienists guidance value (ACGIH, 2019) of 4.6 μmol/mol creatinine, meaning that health effects cannot be ruled out. This is the first time that this type of restriction has been proposed at the EU level and there might be an interest to follow-up on the effectiveness of the restriction if implemented. This review has highlighted the wide range of results that can be found within workers, workplaces and sectors. If the restriction proposal on diisocyanates does come into force, it should have an impact on the exposure to diisocyanates, but small and medium enterprises (with the cost implications of the restriction and being traditionally hard to reach) may still pose a challenge. Therefore, a follow-up on the effectivity especially in SMEs is of high interest to evaluate whether exposures overall have been reduced through improved training (the restriction proposal), as has been seen to some extent in the UK (Piney, 2015).

Conclusion

Based on a systematic review we provide a comprehensive summary on available biomarkers and matrices for diisocyanate exposure. Further, we summarized available

occupational diisocyanates biomonitoring studies published since 2000 including reported exposure levels, and discussed the studies in detail. Although biomonitoring studies have a number of advantages over external exposure measurements - for example when dermal exposure is likely or when respiratory protection is used - this review has highlighted the need for a harmonized approach to study and report biomonitoring levels; also to provide a baseline against which the success of the recently proposed restriction can be evaluated. We identified several knowledge gaps which could further aid studying diisocyanates biomonitoring levels: (i) the development of specific biomarkers is promising (e.g. to study oligomers of HDI which have been largely neglected to date) but needs more research before they can be widely applied, (ii) a more uniform approach of analytical methods would make comparisons between studies easier, and (iii) dermal absorption seems a possible exposure route and needs to be further investigated.

References Chapter 3

ACGIH (2019). TLVs and BEIs. Retrieved from https://www.acgih.org/forms/store/ProductFormPublic/2019-tlvs-and-beis

Austin, S. (2007). Biological monitoring of TDI-derived amines in polyurethane foam production. *Occupational Medicine, 57*(6), 444-448.

HSL. (2005) Biological monitoring for isocyanates. In *Analysis of urine to assess exposure to Isocyanates Guidance for workers, employers, and occupational health professionals* (pp. 5). Buxton, UK: Health Sciences Group, Health and Safety Laboratory.

Casteleyn, L., Dumez, B., Becker, K., Kolossa-Gehring, M., Den Hond, E., Schoeters, G., Aerts, D. (2015). A pilot study on the feasibility of European harmonized human biomonitoring: Strategies towards a common approach, challenges and opportunities. *Environ Res, 141*, 3-14. doi:10.1016/j.envres.2014.10.028

Cocker, J. (2007). Biological monitoring for isocyanates. In: Annals of Occupational Hygiene.

Cocker, J. (2011). Biological Monitoring for Isocyanates. *Annals of work exposures and health, 55*(2), 127-131. doi:10.1093/annhyg/meq083

Cocker, J., Cain, J., Baldwin, P., McNally, K. and Jones, K. (2009). A survey of occupational exposure to 4, 4′-methylene-bis (2-chloroaniline)(MbOCA) in the UK. *Annals of occupational hygiene, 53*(5), 499-507.

Cocker, J., Mason, H. J., Warren, N. D. and Cotton, R. J. (2011). Creatinine adjustment of biological monitoring results. *Occup Med (Lond), 61*(5), 349-353. doi:10.1093/occmed/kqr084

Creely, K. S., Hughson, G. W., Cocker, J. and Jones, K. (2006). Assessing isocyanate exposures in polyurethane industry sectors using biological and air monitoring methods. *The Annals of Occupational Hygiene, 50*(6), 609-621.

De Palma, G. P., Cortesi, I., Ghitti, R., Festa, D., Bergonzi, R. and Apostoli, P. (2012). Biological monitoring as a valid tool to assess occupational exposure to mixtures of 2, 4-: 2, 6-toluene diisocyanate. *La Medicina del lavoro, 103*(5), 361-371.

DECOS (2018). *Di- and triisocyanates. Health-based recommendation on occupational exposure limits* (No. 2018/20). Retrieved from The Hague, The Netherlands

ECHA (2005). European Union Risk Assessment Report METHYLENEDIPHENYL DIISOCYANATE (MDI) CAS No: 26447-40-5. EINECS No: 247-714-0 Retrieved from https://echa.europa.eu/documents/10162/9f8ad2fd-9b47-4eb6-9bf9-e0fc898f874d

ECHA (2010). Restrictions on the manufacture, placing on the market and use of certain dangerous substances, mixtures and articles(Entry 56). Retrieved from https://echa.europa.eu/documents/10162/46cb2ae0-ee6b-4319-8604-b5b1d4a41d53

ECHA (2017). Opinion on an Annex XV dossier proposing restrictions DIISOCYANATES ECHA/RAC/RES-O-0000001412-86-174/F Retrieved from https://echa.europa.eu/documents/10162/737bceac-35c3-77fb-ba7a-0e417a81aa4a

ECHA (2019a). Brief profile 4-methyl-m-phenylene diisocyanate (2,4-TDI). Retrieved from https://echa.europa.eu/fi/brief-profile/-/briefprofile/100.008.678

ECHA (2019b). Brief profile m-tolylidene diisocyanate Retrieved from https://echa.europa.eu/fi/brief-profile/-/briefprofile/100.043.369

ECHA (2019c). Brief profile on 4,4'-methylenediphenyl diisocyanate. Retrieved from

https://echa.europa.eu/fi/brief-profile/-/briefprofile/100.002.697

ECHA (2019d). Brief profile Hexamethylene diisocyanate Retrieved from https://echa.europa.eu/nl/substance-information/-/substanceinfo/100.011.350

Fent, K. W., Jayaraj, K., Ball, L. M. and Nylander-French, L. A. (2008). Quantitative monitoring of dermal and inhalation exposure to 1,6-hexamethylene diisocyanate monomer and oligomers. *J Environ Monit, 10*(4), 500-507. doi:10.1039/b715605g

Flack, S. L., Fent, K. W., Trelles Gaines, L. G., Thomasen, J. M., Whittaker, S., Ball, L. M., & Nylander-French, L. A. (2010). Quantitative plasma biomarker analysis in HDI exposure assessment. Annals of occupational hygiene, 54(1), 41-54.

Flack, S. L., Fent, K.W., Gaines, L.G.T., Thomasen, J.M., Whittaker, S.G., Ball, L.M., Nylander-French, L.A. (2011). Hemoglobin adducts in workers exposed to 1,6-hexamethylene diisocyanate. *Biomarkers. 16*(3):261–270. doi:10.3109/1354750X.2010.549242

Gaines, L. G. T., Fent, K. W., Flack, S. L., Thomasen, J. M., Ball, L. M., Richardson, D. B., Nylander-French, L. A. (2010). Urine 1,6-Hexamethylene Diamine (HDA) Levels Among Workers Exposed to 1,6-Hexamethylene Diisocyanate (HDI). *The Annals of Occupational Hygiene, 54*(6), 678-691. doi:10.1093/annhyg/meq041

Geens, T., Dugardin, S., Schockaert, A., De Cooman, G. and van Sprundel, M. (2012). Air exposure assessment of TDI and biological monitoring of TDA in urine in workers in polyurethane foam industry. *Occup Environ Med, 69*(2), 93-98.

Gries, W. and Leng, G. (2013). Analytical determination of specific 4, 4′-methylene diphenyl diisocyanate hemoglobin adducts in human blood. *Anal Bioanal Chem, 405*(23), 7205-7213.

Henriks-Eckerman, M.-L., Mäkelä, E. A., Laitinen, J., Ylinen, K., Suuronen, K., Vuokko, A. and Sauni, R. (2015). Role of dermal exposure in systemic intake of methylenediphenyl diisocyanate (MDI) among construction and boat building workers. *Toxicology letters, 232*(3), 595-600. doi:https://doi.org/10.1016/j.toxlet.2014.12.012

Hu, J., Cantrell, P. and Nand, A. (2017). Comprehensive biological monitoring to assess isocyanates and solvents exposure in the NSW Australia motor vehicle repair industry. *Annals of work exposures and health, 61*(8), 1015-1023.

IARC (1999). IARC MONOGRAPHS ON THE EVALUATION OF CARCINOGENIC RISKS TO HUMANS. Volume 71. Retrieved from https://monographs.iarc.fr/wp-content/uploads/2018/06/mono71.pdf

Jones, K., Cocker, J. and Piney, M. (2013). Isocyanate exposure control in motor vehicle paint spraying: evidence from biological monitoring. *The Annals of Occupational Hygiene, 57*(2), 200-209. doi:10.1093/annhyg/mes056

Jones, K., Johnson, P. D., Baldwin, P. E. J., Coldwell, M., Cooke, J., Keen, C., Cocker, J. (2017). Exposure to Diisocyanates and Their Corresponding Diamines in Seven Different Workplaces. *Annals of work exposures and health, 61*(3), 383-393. doi:10.1093/annweh/wxx006

Kääriä, K., Hirvonen, A., Norppa, H., Piirilä, P., Vainio, H. and Rosenberg, C. (2001). Exposure to 2, 4-and 2, 6-toluene diisocyanate (TDI) during production of flexible foam: determination of airborne TDI and urinary 2, 4-and 2, 6-toluenediamine (TDA). *Analyst, 126*(7), 1025-1031.

Keen, C., Coldwell, M., McNally, K., Baldwin, P., McAlinden, J. and Cocker, J. (2012). A follow up study of occupational exposure to 4, 4′-methylene-bis (2-

chloroaniline)(MbOCA) and isocyanates in polyurethane manufacture in the UK. *Toxicology letters, 213*(1), 3-8.

LaKind, J. S., Sobus, J. R., Goodman, M., Barr, D. B., Furst, P., Albertini, R. J., Weisel, C. P. (2014). A proposal for assessing study quality: Biomonitoring, Environmental Epidemiology, and Short-lived Chemicals (BEES-C) instrument. *Environ Int, 73*, 195-207. doi:10.1016/j.envint.2014.07.011

Littorin, M., Axmon, A., Broberg, K., Sennbro, C.-J. and Tinnerberg, H. (2007). Eye and airway symptoms in low occupational exposure to toluene diisocyanate. *Scandinavian journal of work, environment & health*, 280-285.

Littorin, M., Rylander, L., Skarping, G., Dalene, M., Welinder, H., Strömberg, U. and Skerfving, S. (2000). Exposure biomarkers and risk from gluing and heating of polyurethane: a cross sectional study of respiratory symptoms. *Occupational and Environmental Medicine, 57*(6), 396-405. doi:10.1136/oem.57.6.396

McDonald, J. C., Chen, Y., Zekveld, C. and Cherry, N. M. (2005). Incidence by occupation and industry of acute work related respiratory diseases in the UK, 1992-2001. *Occup Environ Med, 62*(12), 836-842. doi:10.1136/oem.2004.019489

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Piney, M. (2015). Reducing isocyanate exposure and asthma risk in motor vehicle repair. *International Journal of Workplace Health Management, 8*(4), 272-283. doi:10.1108/IJWHM-10-2014-0043

Pronk, A., Yu, F., Vlaanderen, J., Tielemans, E., Preller, L., Bobeldijk, I., Heederik, D. (2006). Dermal, inhalation, and internal exposure to 1, 6-HDI and its oligomers in car body repair shop workers and industrial spray painters. *Occupational and Environmental Medicine, 63*(9), 624-631.

RAC, SEAC (2017) Opinion on an Annex XV dossier proposing restrictions DIISOCYANATES. ECHA/RAC/RES-O-0000001412-86-174/F. Available at: https://echa.europa.eu/documents/10162/737bceac-35c3-77fb-ba7a-0e417a81aa4a. Accessed at 3/3/2020

Robbins, Z., Bodnar, W., Zhang, Z., Gold, A. and Nylander-French, L. A. (2018). Trisaminohexyl isocyanurate, a urinary biomarker of HDI isocyanurate exposure. *Journal of Chromatography B, 1076*, 117-129. doi:https://doi.org/10.1016/j.jchromb.2018.01.021

Robert, A., Ducos, P., Francin, J. and Marsan, P. (2007). Biological monitoring of workers exposed to 4, 4′-methylenediphenyl diisocyanate (MDI) in 19 French polyurethane industries. *International Archives of Occupational and Environmental Health, 80*(5), 412-422.

Rosenberg, C., Nikkilä, K., Henriks-Eckerman, M.-L., Peltonen, K. and Engström, K. (2002). Biological monitoring of aromatic diisocyanates in workers exposed to thermal degradation products of polyurethanes. *Journal of Environmental Monitoring, 4*(5), 711-716.

Rosenberg, C. and Savolainen, H. (1986). Determination in urine of diisocyanate-derived amines from occupational exposure by gas chromatography-mass fragmentography. *Analyst, 111*(9), 1069-1071.

Sabbioni, G., Dongari, N. and Kumar, A. (2010). Determination of a new biomarker in

subjects exposed to 4,4′-methylenediphenyl diisocyanate. *Biomarkers, 15*(6), 508-515. doi:10.3109/1354750X.2010.490880

Sabbioni, G., Vanimireddy, L. R., Lummus, Z. L. and Bernstein, D. I. (2017). Comparison of biological effects with albumin adducts of 4, 4′-methylenediphenyl diisocyanate in workers. *Archives of Toxicology, 91*(4), 1809-1814.

Sabbioni, G., Wesp, H., Lewalter, J. and Rumler, R. (2007). Determination of isocyanate biomarkers in construction site workers. *Biomarkers, 12*(5), 468-483. doi:10.1080/13547500701395636

Sakai, T., Morita, Y., Roh, J., Kim, H. and Kim, Y. (2005). Improvement in the GC–MS method for determining urinary toluene-diamine and its application to the biological monitoring of workers exposed to toluene-diisocyanate. *International Archives of Occupational and Environmental Health, 78*(6), 459-466. doi:10.1007/s00420-004-0571-9

Säkkinen, K., Tornaeus, J., Hesso, A., Hirvonen, A., Vainio, H., Norppa, H. and Rosenberg, C. (2011). Protein adducts as biomarkers of exposure to aromatic diisocyanates in workers manufacturing polyurethane (PUR) foam. *Journal of Environmental Monitoring, 13*(4), 957-965. doi:10.1039/C0EM00595A

Sennbro, C. J., Lindh, C., Mattsson, C., Jönsson, B. and Tinnerberg, H. (2006). Biological monitoring of exposure to 1, 5–naphthalene diisocyanate and 4, 4'-methylenediphenyl diisocyanate. *International Archives of Occupational and Environmental Health, 79*(8), 647-653.

Sennbro, C. J., Lindh, C. H., Tinnerberg, H., Gustavsson, C., Littorin, M., Welinder, H. and Jönsson, B. A. G. (2003). Development, validation and characterization of an analytical method for the quantification of hydrolysable urinary metabolites and plasma protein adducts of 2,4- and 2,6-toluene diisocyanate, 1,5-naphthalene diisocyanate and 4,4'-methylenediphenyl diisocyanate. *Biomarkers, 8*(3-4), 204-217. doi:10.1080/1354750031000090660

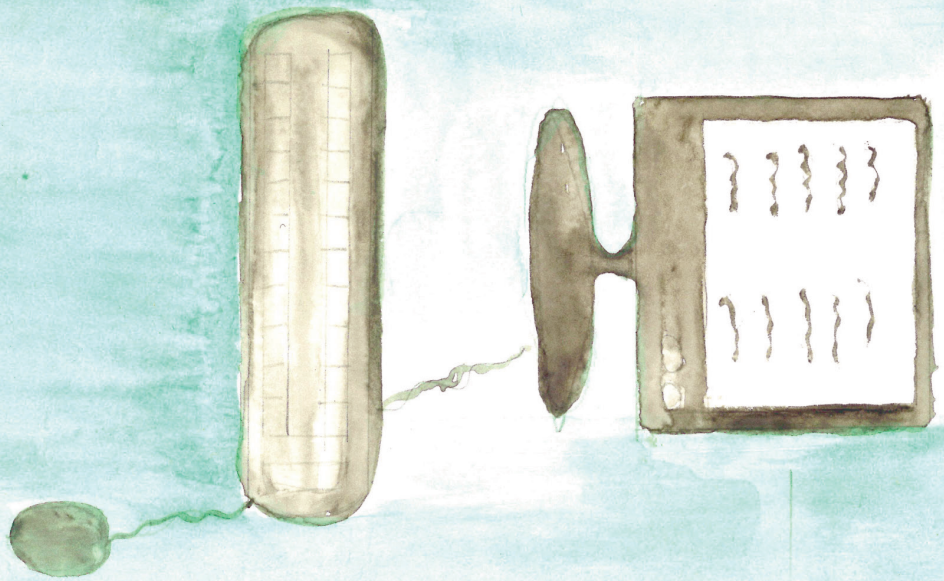
Sennbro, C. J., Lindh, C. H., Tinnerberg, H., Welinder, H., Littorin, M. and Jönsson, B. A. (2004). Biological monitoring of exposure to toluene diisocyanate. *Scandinavian journal of work, environment & health*, 371-378.

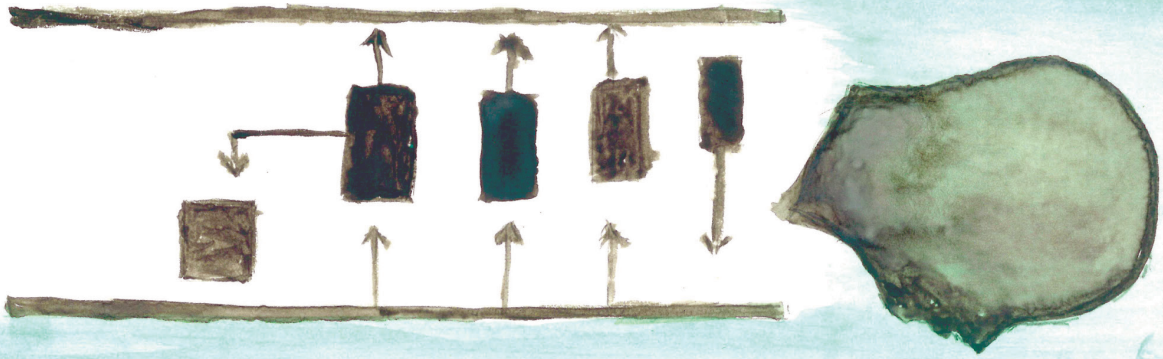
Swierczynska-Machura, D., Brzeznicki, S., Nowakowska-Swirta, E., Walusiak-Skorupa, J., Wittczak, T., Dudek, W., Palczynski, C. (2015) Occupational exposure to diisocyanates in polyurethane foam factory workers. *International Journal of Occupational Medicine and Environmental Health*. doi:10.13075/ijomeh.1896.00284

Tinnerberg, H., Broberg, K., Lindh, C. H. and Jönsson, B. A. (2014). Biomarkers of exposure in Monday morning urine samples as a long-term measure of exposure to aromatic diisocyanates. *International Archives of Occupational and Environmental Health, 87*(4), 365-372.

Tinnerberg, H. and Mattsson, C. (2008). Usage of Air Monitoring and Biomarkers of Isocyanate Exposure to Assess the Effect of a Control Intervention. *The Annals of Occupational Hygiene, 52*(3), 187-194. doi:10.1093/annhyg/men006

Wisnewski, A.V., Liu, J., Redlich, C.A. (2013). Connecting glutathione with immune responses to occupational methylene diphenyl diisocyanate exposure. *Chem Biol Interact*. *205*(1):38–45. doi:10.1016/j.cbi.2013.06.005





Chapter 4

A physiologically-based kinetic (PBK) model for work-related diisocyanate exposure: relevance for the design and reporting of biomonitoring studies

Scholten, B., Westerhout, J., Pronk, A., Stierum, R., Vlaanderen, J., Vermeulen, R., Jones, K., Santonen, T., Portengen, L.

Published in revised form in: *Environment International*, 2023, 174: 107917

Abstract

Diisocyanates are highly reactive substances and known causes of occupational asthma. Exposure occurs mainly in the occupational setting and can be assessed through biomonitoring which accounts for inhalation and dermal exposure and potential effects of protective equipment. However the interpretation of biomonitoring data can be challenging for chemicals with complex kinetic behavior and multiple exposure routes, as is the case for diisocyanates. To better understand the relation between external exposure and urinary concentrations of metabolites of diisocyanates, we developed a physiologically based kinetic (PBK) model for methylene bisphenyl isocyanate (MDI) and toluene di-isocyanate (TDI). The PBK model covers both inhalation and dermal exposure, and can be used to estimate biomarker levels after either single or chronic exposures. Key parameters such as absorption and elimination rates of diisocyanates were based on results from human controlled exposure studies. A global sensitivity analysis was performed on model predictions after assigning distributions reflecting a mixture of parameter uncertainty and population variability. Although model-based predictions of urinary concentrations of the degradation products of MDI and TDI (i.e. MDA and TDA) for longer-term exposure scenarios compared relatively well to empirical results for a limited set of biomonitoring studies in the peer-reviewed literature, validation of model predictions was difficult because of the many uncertainties regarding the precise exposure scenarios that were used. Sensitivity analyses indicated that parameters with a relatively large impact on model estimates included the fraction of diisocyanates absorbed and the proportion of diisocyanates that was assumed to bind to albumin (versus that binding to other macromolecules). We additionally investigated the effect of timing of exposure and intermittent urination, and found that both had a considerable impact on estimated urinary levels. This suggests that these factors should be taken into account when interpreting biomonitoring data and included in the standard reporting of isocyanate biomonitoring studies.

Introduction

Diisocyanates are a group of highly reactive substances with two isocyanate functional groups. Diisocyanates are potent skin and respiratory tract sensitizers and a leading cause of occupational asthma in many industrial countries (Lockey et al. 2015; Malo and Chan-Yeung 2009). Exposure occurs mostly in occupational settings where these chemicals are widely used (RAC 2020a). Measuring diisocyanate exposure is indispensable for compliance with exposure limits and accurate risk assessment. However, measuring inhalation exposure to diisocyanates is technically challenging due to their reactivity (Health Council of the Netherlands 2018). In addition, air sampling does not provide information on dermal exposure. There is evidence that dermal exposure to diisocyanates plays a role in the development of respiratory sensitization and occupational asthma (Bello et al. 2007; Redlich 2010). Air measurements, typically performed outside the respiratory protective equipment (RPE) do not take into account the efficacy of RPE that is often used (McNally et al. 2012). However diisocyanates have complex kinetic behavior which hampers interpretation of biomonitoring data. From controlled human exposure studies after single diisocyanate exposures (Skarping, Brorson, and Sangö 1991; Torbjorn Brorson, Skarping, and Sango 1991; Budnik et al. 2011) we know that diisocyanates generally have a short half-life. As such, the interpretation of biomonitoring data is influenced by recent temporal factors such as the exposure scenario on that day (i.e. moment of exposure, exposure duration and exposure route) as well as subject behavior (i.e. use of PPE, moment of urination). Biomonitoring studies in chronically exposed workers indicate possible accumulation as elevated urinary amines in samples taken before the work shift on Monday morning have been reported (Tinnerberg et al. 2014; Geens et al. 2012). It has been hypothesized that this is due to formation of albumin adducts which circulate in the blood before being degraded and excreted (Health Council of the Netherlands 2018; ATSDR 2018). This is referred to as a two phase urinary elimination pattern, with the first phase being related to the more recent exposure and a half-life over several hours, and the second much slower one probably related to the degradation of TDI and MDI conjugates (Lind et al. 1996).

Here we aimed to give quantitative insight on how urinary biomarkers of diisocyanate exposure are affected by several factors - including potential dermal exposure, accumulation in chronically exposed workers and differences between diisocyanates in elimination kinetics. For these purposes physiologically based kinetic (PBK) models can be applied as they are known to be useful tools in data analysis to interpret the relation between external and internal exposures (Hoer et al. 2020). PBK models use differential equations to estimate the concentration of a chemical or its metabolites in urine or blood. A PBK model can extrapolate across exposure scenarios (thereby reducing the need for data on specific exposure scenarios), and help understand

which parameters are most influential for kinetic and dynamic behavior of the chemical in question.

To that end we developed a PBK model for diisocyanates, focusing on two of the most commonly used isocyanates: 4,4'-methylene diphenyl diisocyanate (MDI) and toluene diisocyanate (TDI). To our knowledge there are two published PBK models available for (degradation products of) diisocyanates. Luu, Hutter, and Bushar (1998) developed a model to estimate exposure to 2,4-toluenediamine (TDA; a degradation product of TDI) as released from the degradation of the polyester urethane foam used in silicone breast implants. The other PBK model was developed to model exposure to 4,4-methylenedianiline (MDA; a degradation product of MDI) from the potting materials of hemodialyzers (Luu and Hutter 2000). Until now there is no PBK model available for diisocyanates that estimates biomarker levels in urine after chronic exposure to TDI or MDI via both inhalation and dermal exposure, and includes accumulation of metabolite levels due to the formation of plasma adducts. For deriving diisocyanate biological limit values based on external health based limit values (i.e. occupational exposure limits (OEL)), current practice is to derive correlation formulae based on occupational studies that report both air and urine measurements. We hypothesized that it would be useful to develop a PBK model to better understand and characterize how factors such as the exposure scenario and subject behavior have an effect on the presumed external – internal exposure relation.

Methods

PBK Model

Model structure

Our proposed diisocyanate PBK model (figure 1) includes the following main compartments: lungs, skin, gut, gut lumen, liver, arterial blood, venous blood, and all other ("remaining") parts of the body. Because diisocyanates are known to reside in the lungs for some time before they are taken up in the blood circulation (ATSDR 2018), we distinguish between a combined alveolar air, epithelial lining fluid and lung tissue ("interstitial") compartment and a ("vascular") compartment that is directly connected to the systemic circulation. Most diisocyanates are expected to bind rapidly to albumin and other macromolecules and we therefore distinguish between unbound ("[ub]"), albumin-bound ("[alb]"), and (other) macro-molecule-bound ("[mcr]") fractions within each of the compartments. For MDI, we additionally recognize a metabolized ("[met]") fraction.

The model allows for exposure to diisocyanates through inhalation and through skin absorption. Following exposure, transport and distribution of diisocyanates across the different compartments is assumed to be driven by blood flow rates. Elimination of diisocyanates and their conjugated forms is included in the PBK model via the urine

or feces. Urinary excretion of absorbed isocyanates is expected to take place following degradation of albumin and (at least for MDI) after hydrolyzation in the liver (Schütze et al. 1995). We accommodate the relatively long(er) retention of albumin-bound diisocyanates by modelling excretion only from the macro-molecule-bound fraction and using a relatively slow transfer (elimination) from the albumin-bound fraction into this compartment.

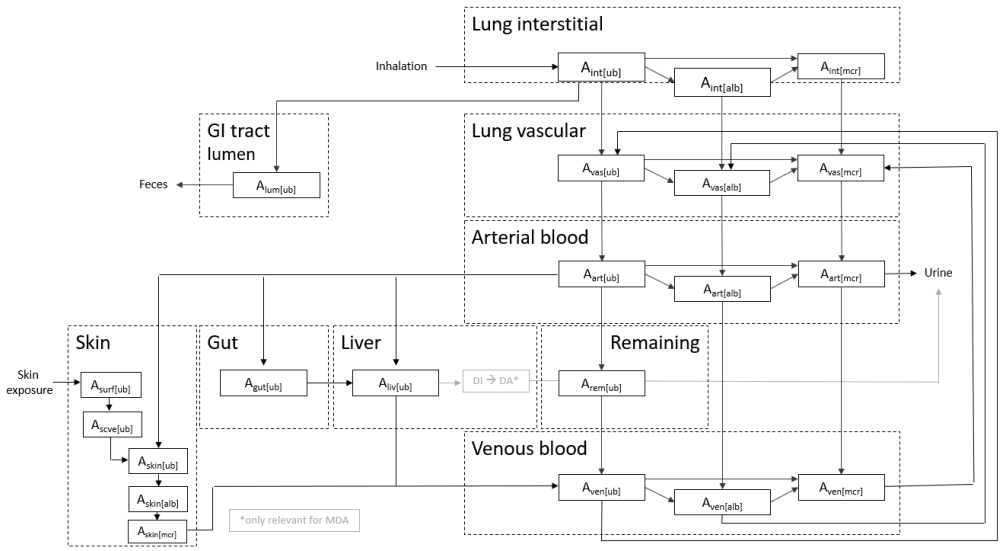


Figure 1. Schematic picture of physiological based kinetic model for diisocyanates (MDI and TDI).

Differential equations

For each of the compartments shown in figure 1, the amount (in mg) and concentration (in mg/L) of diisocyanates in each fraction (i.e. unbound, albumin-bound, or macro-molecule bound) is indicated by $A_{compartment[fraction]}$ and $C_{compartment[fraction]}$ respectively, where the concentration is calculated by dividing the amount by the compartment volume.

Lungs

The amounts of unbound, albumin-bound, and (other) macro-molecule-bound diisocyanates in the interstitial lung compartment are described by the following differential equations:

$$dt(A_{int[ub]}) = Q_{br} * F_{abs} * C_{air} - k_{alb} * A_{int[ub]} - k_{mcr} * A_{int[ub]} - k_{vas} * A_{int[ub]}$$
$$dt(A_{int[alb]}) = k_{alb} * A_{int[ub]} - k_{elim} * A_{int[alb]} - k_{vas} * A_{int[alb]}$$
$$dt(A_{int[mcr]}) = k_{mcr} * A_{int[ub]} + k_{elim} * A_{int[alb]} - k_{vas} * A_{int[mcr]}$$

where Q_{br} is the breathing rate in L/h, F_{abs} is the fraction absorbed (the non-absorbed fraction is excreted through the GI tract, see below), C_{air} is the inhalation exposure in mg/L, k_{alb} and k_{mcr} are the binding rate constants (unit h^{-1}) for binding of unbound diisocyanates to albumin and other macromolecules, k_{vas} is the transfer rate constant (unit h^{-1}) for transfer from the interstitial to the vascular lung compartment, and k_{elim} is the elimination rate constant (unit h^{-1}) for the degradation of albumin into (smaller) macromolecules.

The amounts of unbound, albumin-bound, and (other) macro-molecule-bound diisocyanates in the vascular lung compartment are described by the following differential equations:

$$dt(A_{vas[ub]}) = k_{vas} * A_{int[ub]} - k_{alb} * A_{vas[ub]} - k_{mcr} * A_{vas[ub]} + Q_c * (C_{ven[ub]} - C_{vas[ub]})$$

$$dt(A_{vas[alb]}) = k_{vas} * A_{int[ub]} + k_{alb} * A_{vas[ub]} - k_{elim} * A_{vas[ub]} + Q_c * (C_{ven[alb]} - C_{vas[alb]})$$

$$dt(A_{vas[mcr]}) = k_{vas} * A_{int[mcr]} + k_{mcr} * A_{vas[ub]} + k_{elim} * A_{vas[ub]} + Q_c * (C_{ven[mcr]} - C_{vas[mcr]})$$

where Q_c is the cardiac output (in L/h).

Dermal exposure

The amounts of (unbound) diisocyanates on the surface of the skin and stratum corneum including viable epidermis (scve; see Krüse 2006) are described by the following equations:

$$dt(A_{surf[ub]}) = SA_{derm} * D_{derm[ub]} - Q_{scve} * (C_{surf[ub]} - C_{scve[ub]}) - k_{evap} * A_{surf[ub]}$$

$$dt(A_{scve[ub]}) = Q_{scve} * (C_{surf[ub]} - C_{scve[ub]}) - K_{alb[skn]} * A_{scve[ub]} - P_{bld[scve]} * Q_{scve} * C_{scve[ub]}$$

with SA_{derm} the exposed surface area of the skin (in cm^2), D_{derm} the dose applied to the skin (in mg/cm^2/hour), CL_{scve} the fat-content adjusted stratum corneum permeability coefficient transferred into a clearance value (in L/h), k_{evap} the dermal evaporation rate (unit h^{-1}; see below), and p_{scve} the stratum corneum and viable epidermis to blood partition coefficient. The formula used to calculate the dermal evaporation rate (k_{evap}) is loosely based on (ECHA 2009) and includes the molecular weight (MW) and vapor pressure (VP, unit Pa) of the diisocyanate:

$$k_{evap} = c_{evap} \times \frac{MW \times 8.7 \times VP \times SA_{derm}}{8.314 * 303.0 * 10}$$

Based on the study by Hamada (2018), the constant c_{evap} was fixed to achieve an approximate 50% evaporation rate for MDI.

The amounts of diisocyanates in the different fractions in the skin compartment are described by the following equations:

$$dt(A_{skn[ub]}) = Q_{skn} * (C_{art[ub]} - C_{skn[ub]}) + P_{bld[scve]} * Q_{scve} * C_{scve[ub]}$$
$$dt(A_{skn[mcr]}) = -Q_{skn} * C_{skn[mcr]} + P_{bld[scve]} * Q_{scve} * Q_{scve[mcr]}$$

where Q_{skn} is the blood flow to and from the skin compartment (in L/h) and $K_{alb[skn]}$ is the binding rate constant of binding to albumin (and perhaps other macro-molecules) in the skin. Note that we assume that albumin-bound diisocyanates leave the skin compartment only after degradation of the albumin (elimination).

General circulation

The arterial and venous blood compartments are subdivided into unbound, albumin and macromolecule compartments. Because of the high reactivity of diisocyanates, we assume that no unbound diisocyanates will be available for renal clearance. In addition, we assumed that only the diisocyanates bound to other macromolecules can be excreted into urine, so that albumin-bound diisocyanates only become available for excretion after conjugation (elimination). For simplification, we have modeled renal clearance via a direct removal from arterial blood rather than including the kidney as a separate compartment in the PBK model.

The amount of diisocyanates in the different arterial blood sub-compartments is described by the following differential equations:

$$dt(A_{art[ub]}) = Q_c * (C_{vas[ub]} - C_{art[ub]}) - k_{alb} * A_{art[ub]} - k_{mcr} * A_{art[ub]}$$
$$dt(A_{art[alb]}) = Q_c * (C_{vas[alb]} - C_{art[alb]}) + k_{alb} * A_{art[ub]} - k_{elim} * A_{art[alb]}$$
$$dt(A_{art[mcr]}) = Q_c * (C_{vas[mcr]} - C_{art[mcr]}) + k_{mcr} * A_{art[ub]} + k_{elim} * A_{art[alb]} - Q_{gfr} * C_{art[mcr]}$$

where Q_{gfr} is the glomerular filtration rate (in L/h).

The amount of diisocyanates in the different venous blood sub-compartments is described by the following differential equations:

$$dt(A_{ven[ub]}) = Q_{skn} * C_{skn[ub]} + (Q_{liv} + Q_{gut}) * C_{liv[ub]} + (Q_c - Q_{skn} - Q_{liv} - Q_{gut}) * C_{rem[ub]} - Q_c * C_{ven[ub]} - k_{alb} * A_{ven[ub]} - k_{mcr} * A_{ven[ub]}$$
$$dt(A_{ven[alb]}) = Q_c * (C_{art[alb]} - C_{ven[alb]}) + k_{alb} * A_{ven[ub]} - k_{elim} * A_{ven[alb]}$$
$$dt(A_{ven[mcr]}) = Q_c * (C_{art[mcr]} - C_{ven[mcr]}) + k_{mcr} * A_{ven[ub]} + k_{elim} * A_{ven[alb]}$$

where Q_{skn}, Q_{liv}, Q_{gut}, and Q_{rem} are the blood flow rates to and from the skin, liver, gut, and remaining tissue compartments (in L/h) and C_{skn}, C_{liv}, and C_{rem} the concentrations in these compartments. Note that only the unbound fraction is assumed to leave the circulation and that blood flow from the gut is through the liver.

Faecal excretion

The fraction of diisocyanates that is inhaled but not absorbed in the lungs is transferred to the gut lumen compartment. We assume that absorption of diisocyanates from the lumen into the gut is negligible, so that all diisocyanates in the gut lumen will be excreted in the feces.

The amounts of unbound diisocyanates in the gut, gut lumen and feces are described by the following equations:

$$dt(A_{gut[ub]}) = Q_{gut} * (C_{art[ub]} - C_{gut[ub]})$$
$$dt(A_{lum[ub}) = (1 - F_{abs}) * Q_{br} * C_{air} - k_{feces} * A_{lum}$$
$$dt(A_{feces}) = k_{feces} \times A_{lum}$$

where Q_{gut} is the blood flow into the gut tissue (in L/h) and k_{feces} is the transfer rate constant (unit h^{-1}) of diisocyanates from the gut lumen to the feces.

Liver metabolism

We assume for TDI that all absorbed diisocyanates are conjugated (bound to macromolecules). This is supported by studies that describe that the majority of TDI was conjugated to macromolecules in urine after inhalation exposure (Health Council of the Netherlands 2018; ATSDR 1998). For MDI there is evidence of hydrolysis in vivo: Schütze et al. (1995) conclude that upon MDI exposure the MDI hydrolysis product 4,4'-methylenedianiline (MDA) is bioavailable in humans, based on the presence of free MDA and acetylated MDA hemoglobin adducts (AcMDA) in blood and urine. Sepai, Henschler, and Sabbioni (1995) also found free MDA and AcMDA in blood and urine after MDI exposure. We assume that after hydrolysis, metabolized ("met") diisocyanate enters the general blood circulation, after which it is excreted in the urine. Based on the study by Sepai et al. (1995) we expect for MDI that about 10% of the absorbed dose is excreted in urine as free MDA or AcMDA in urine. We therefore set the value for the hydrolization rate constant (k_{hydro}) for such that about 10% of absorbed MDI is hydrolyzed into MDA. In addition, we incorporated a parallel PBK model for the distribution and urinary elimination of the metabolites after being generated in the liver compartment.

The amounts of unbound and metabolized diisocyanates in the liver are therefore described by the following differential equations:

$$dt(A_{liv[ub]}) = Q_{liv} * C_{art[ub]} + Q_{gut} * C_{gut[ub]} - (Q_{liv} + Q_{gut}) * C_{liv[ub]} - k_{hydr} * A_{liv[ub]}$$
$$dt(A_{liv[met]}) = Q_{liv} * C_{art[met]} + Q_{gut} * C_{gut[met]} - (Q_{liv} + Q_{gut}) * C_{liv[met]} + k_{hydr} * A_{liv[ub]}$$

where k_{hydr} is the rate constant (unit h^{-1}) for hydrolysis of unbound diisocyanates.

Circulation of metabolized diisocyanates through other compartments and urinary excretion are modelled using the following differential equations:

$$dt(A_{skin[met]}) = Q_{skn} * (C_{art[met]} - C_{skin[met]})$$
$$dt(A_{gut[met]}) = Q_{gut} * (C_{art[met]} - C_{gut[met]})$$
$$dt(A_{rem[met]}) = (Q_c - Q_{skn} - Q_{liv} - Q_{gut}) * (C_{art[met]} - C_{rem[met]})$$
$$dt(A_{ven[met]}) = Q_{skn} * C_{skn[met]} + (Q_{liv} + Q_{gut}) * C_{liv[met]} + (Q_c - Q_{skn} - Q_{liv} - Q_{gut}) * C_{rem[met]} - Q_c * C_{ven[met]}$$
$$dt(A_{art[met]}) = Q_c * (C_{vas[met]} - C_{art[met]}) - Q_{gfr} * C_{art[met]}$$

Remaining compartments

The amount of diisocyanates in the remaining tissues (A_{rem}) is described by the following equation:

$$dt(A_{rem[ub]}) = (Q_c - Q_{skn} - Q_{gut} - Q_{liv}) * (C_{art[ub]} - C_{rem[ub]})$$

Urinary excretion

Usually urine samples are treated with bases or acids to hydrolyze the diisocyanate-protein or diamine-protein conjugates and acetylated diamines, resulting in the formation of free diamine (Sennbro et al. 2004). With our model we estimate the levels of conjugates in urine (which are excreted as low-molecular weight conjugates), and that are ultimately converted to amines. These are TDA and MDA, for TDI and MDI respectively.

The amount of diisocyanates excreted in urine is determined by the Q_{gfr} in L/h and the amount of diisocyanate in the macromolecule and metabolite compartments of the arterial blood, and is converted to milligrams of the amine metabolite using the ratio of molecular weights:

$$dt(A_{uri[amine]}) = \frac{MW_{amine}}{MW_{parent}} * Q_{gfr} * (C_{art[mcr]} + C_{art[met]})$$

The urinary concentration of diisocyanate amines is then calculated using the following formula for the urine production rate:

$$dt(V_{uri}) = Q_{upr}$$

Where Q_{upr} is the bodyweight specific urinary production rate (unit L/h) (Davies 1993). We assume at each void that the bladder is fully emptying.

When reporting we use the creatinine adjusted concentrations by dividing the estimated amines concentrations by the urinary creatinine concentration (parameter C_{creat}).

Model parameters

A description of the various parameters and how these are derived or calculated, can be found in Table 1. Table 2 includes an overview of the hyperparameters that are used for the calculation or derivation of the parameters from Table 1. Information from various (clinical) studies was used to determine the central estimates, ranges and uncertainty of the physiological parameters in the PBK model. Some parameter values were not readily available in the literature and were therefore calibrated using data obtained from human experimental exposure studies (Skarping, Brorson, and Sangö 1991; Brorson, Skarping, and Sango 1991; Budnik et al. 2011). For example, the estimate of the fraction absorbed ("F_{abs}") was based on experimental data suggesting that approximately 20% of diisocyanate is absorbed and excreted in urine. Another example is the half-life transfer from the lung interstitial to the lung vascular compartment (T_{vas}, hr), which is calibrated such that after about 6 hours most of TDI has been excreted, based on the studies by (Torbjörn Brorson, Skarping, and Sangö 1991; Budnik et al. 2011). MDI is excreted more slowly (Budnik et al. 2011) and therefore the T_{vas} values for TDI and MDI differ. T_{vas} is subsequently converted to a transfer rate constant (K_{vas}, unit h^{-1}) (Table 1). A more detailed description of the controlled human experimental exposure studies can be found in the Appendix (Appendix A).

Table 1. Description of parameters used for differential equations and corresponding formulas.

Parameter name	Description	Formulae
K_{vas}	Transfer rate constant for transfer from the interstitial to the vascular lung compartment (h^{-1})	$Log(2)/\ T_{vas}$
K_{alb}	Binding rate constants for binding of unbound diisocyanates to albumin (h^{-1})	$Falb * Log(2)/\ T_{albmcr}$
K_{mcr}	Binding rate constants for binding of unbound diisocyanates to macromolecules (h^{-1})	$(1 - Falb) * (\ Log(2)/\ T_{albmcr})$
K_{elim}	Elimination rate constant for the degradation of albumin into (smaller) macromolecules) (h^{-1})	$Log(2)/\ K_{elim}$
k_{hydro}	Rate constant for hydrolysis of unbound diisocyanates (h-1)	$Log(2)/\ T_{hydro}$
K_{feces}	Transfer rate constant from the gut lumen to the feces	$Log(2)/\ T_{feces}$
$K_{alb[skn]}$	Binding rate constant of binding to albumin in the skin	$Log(2)/\ T_{alb[skn]}$
CL_{scve}	Fat-content adjusted stratum corneum permeability coefficient transferred into a clearance value (L/h)	$(1-Ffat_{epi}+Ffat_{epi}*10^{logp})/(1-Ffat_{bld}+Ffat_{bld}*10^{logp}) * 10^{0.74*logp - 0.006*MW - 2.8} * SA_{derm}/1000)$
Q_{upr}	Urine production rate (L/h)	$UPR * BW$
Q_{gut}	Blood flow to the gut (L/h)	$QF_{gut} * Qc$

Q_{liv}	Blood flow to the liver (L/h)	$QF_{liv} * Qc$
Q_{skn}	Blood flow to the skin (L/h)	$QF_{skn} * Qc$

Table 2. Hyperparameters for PBK model
N: normal (distribution), T N: truncated normal (distribution)

Model parameter	Description	Default value	Uncertainty distribution	Reference
Chemical specific				
MW	Molecular weight of the parent compound	174.16 (TDI) 250.25 (MDI)	-	
MW_{met}	Molecular weight of the amine	122.17 (TDA) 198.26 (MDA)	-	
LogP	LogP	3.74 (TDI) 5.22 (MDI)	-	(PubChem 2022) (USEPA 2011)
VP	Vapor pressure (Pa)	2.7 (TDI) 4.9 x 10^{-4} (MDI)	-	(Health Council of the Netherlands, 2018)
F_{abs}	Fraction absorbed	0.2	U [0.1- 0.3]	Based on the controlled human exposure study of Budnik et al. (2011) for MDI and TDI, for TDI only the study by Brorson, Skarping, and Sangö (1991), and for MDI the animal study by Gledhill et al. (2005).
T_{albmcr}	Half-life of binding to albumin and other macromolecules (hr)	TDI: 0.167 MDI: 1.5	TDI: U [0.033 – 0.3] MDI: U [0.3 – 2.7]	
F_{alb}	Proportion bound to albumin (%)	0.2	U [0.04 – 0.36]	
T_{vas}	Half-life transfer from lung interstitial to lung vascular (hr)	TDI: 0.083 MDI: 14	TDI: U [0.017 – 0.15] MDI: U [2.8 – 25]	
T_{hydro}	Half-life hydrolysis (acetylation) in liver (hr)	TDI: 4.0 MDI: 0.033	TDI: U [0.8 – 7.2] MDI: U [0.007 – 0.06]	Based on 10% MDA in urine (Sepai 1995)
Individual				
BW	Body weight (kg)	75	T N (75, 7.5) [60 – 90]	
(volume) fractions				
VF_{bld}	Volume fraction blood	0.082	T N (0.082,0.004) [0.073 – 0.09]	The volumes of compartments were based on the values reported for adult males (Valentin and Streffer 2002)
VF_{lun}	Volume fraction lung	0.013	T N (0.013, 0.001) [0.012 – 0.015]	
VF_{skn}	Volume fraction skin	0.004	T N (0.004, 0.0002) [0.004 – 0.005]	
VF_{gut}	Volume fraction gut	0.023	T N (0.023,0.001) [0.021 – 0.025]	
VF_{liv}	Volume fraction liver	0.025	T N (0.025, 0.001) [0.022 – 0.027]	
V_{lumen}	Volume gut lumen	0.9	T N (0.9, 0.045) [0.81 – 0.99]	
F_{hem}	% hematocrit	0.45	T N (0.45, 0.045) [0.36 – 0.54]	
Blood flows				
QF_{skn}	Fractional blood flow to skin	0.175	T N (0.175, 0.009) [0.158 – 0.193]	
QF_{gut}	Fractional blood flow to gut	0.19	T N (0.19, 0.010) [0.171 – 0.209]	

Model parameter	Description	Default value	Uncertainty distribution	Reference
QF_{liv}	Fractional blood flow to liver	0.065	T N (0.065, 0.003) [0.059 – 0.072]	
Flow rates				
Q_{br}	Breathing rate (L/h)	1000	N (1000, 100) [800,1200]	(HERAG 2007)
Q_c	Cardiac output (L/h)	390	T N (390, 19.5) [351 – 430]	(Valentin and Streffer 2002)
Q_{gfr}	Glomerular filtration rate (L/h)	5.5	N (5.5, 0.7) [4,7]	(Redal-Baigorri, Rasmussen, and Heaf 2014)
T_{feces}	Half-life transfer lumen to feces (/hr)	2.3	U [2.1 – 2.5]	
Excretion				
UPR	Urine production rate (L/kg/hr)	0.00125	T N (0.00125, 0.00006) [0.00113 – 000.138]	
Ccreat	Creatinine in urine (g/L)	1.0	T N (1.0, 0.05) [0.8 – 2.0]	(Valentin and Streffer 2002)
T_{elim}	Half-life elimination from albumin (hr)	456	T N (456, 45.7) [364 – 547]	(Reed 1988)
Skin parameters				
SA_{derm}	Exposed surface area of skin (cm^2)	100	T N (100, 5) [90 – 110]	
H_{scve}	Height of skin barrier (cm)	0.012	T N (0.012, 0.001) [0.01 – 0.013]	
H_{surf}	Height of skin surface (cm)	0.010	T N (0.01, 0.0001) [0.009 – 0.011]	
Fat_{epi}	Fraction of fat in epidermis	0.02	T N (0.02, 0.001) [0.018 – 0.022]	
Fat_{bld}	Fraction of fat in blood	0.07	T N (0.07, 0.004) [0.063 – 0.077]	
$T_{alb[skn]}$	Half-life of binding to albumin in skin (hr)	1/100	U [0.033 – 0.3]	

Exposure scenarios

Short-term and chronic inhalation exposure

As several model parameters were based on information gathered from human experimental exposure studies, we first confirmed whether our model was able to mimic the time course of urinary excretion in these studies. Most of these studies used single, relatively short-term, exposures, and we therefore refer to these as “short-term exposure” scenarios. However, as our main interest was in modeling urinary excretion after occupational exposure, our main results are based on a scenario where exposure occurred continuously for 8 hours a day (from 9:00 to 17:00) on 5 days a week, for a total of 8 weeks. We refer to this scenario as the “chronic exposure” scenario. For this scenario, workers were assumed to urinate, at 8:00, 13:00, 18:00, and 23:00. Reported urinary concentrations relate to the bladder content just prior to the urination at 18:00.

Contribution of dermal exposure

For combined inhalation and dermal exposure, we used an exposure scenario that was based on an observational study by Henriks-Eckerman *et al.* (2015). The authors provided information on MDI exposure levels for both inhalation and dermal exposure for workers in the construction and boat building industry. Cumulative dermal exposure was assessed using a strip-tape technique and was estimated to average between 0.16 and 1.38 µg/10 cm^2, depending on the exact site (on the lower arm) that was sampled, while inhalation exposure was estimated to be between 0.08 and 0.8 µg/m^3. Tasks involving MDI exposure in this study typically lasted between 15-60 minutes, with each task repeated 2-20 times per shift. Based on this information, we designed our dermal exposure scenario to consist of continuous exposure to 0.015 µg MDI/cm^2/hour for 4 hours, equivalent to a cumulative exposure of 0.6 µg/10 cm^2. To investigate the (potential) impact of dermal exposures on total excreted urinary diisocyanate levels, we compared the output from our model for dermal exposure either with or without concurrent inhalation exposure to 0.4 µg/m^3 MDI.

Model comparison with published literature

To compare model predictions for chronic exposures with those reported in the literature we searched for reports from biomonitoring studies that included estimates of both external exposure and urinary concentrations in workers that were exposed to diisocyanates for a prolonged time-period. We collected information on urinary concentrations of diisocyanates, sampling times (i.e. pre-shift or post shift), potential dermal exposure, use of RPE, and task-based and/or time-weighted-average (TWA) exposures. Exposures below the limit of detection (LOD) were replaced by half the reported LOD. In case of task-based estimates of inhalation exposure only, we estimated TWA exposure under the assumption that there was no exposure outside the task.

For TDI we extracted five studies that reported post-shift urinary measurements (Carl J. Sennbro et al. 2004; Maître et al. 1993; Geens et al. 2012; Austin 2007; Swierczyńska-Machura et al. 2015), and two studies that reported pre-shift urinary measurements (Geens et al. 2012; Tinnerberg et al. 2014) (Appendix C). For MDI we extracted three studies that reported post-shift urinary measurements (C. J. Sennbro et al. 2006; Henriks-Eckerman et al. 2015; A. Bello et al. 2019), and one study that reported pre-shift urinary measurements (Tinnerberg et al. 2014) (Appendix D). An overview of estimated exposure levels and urinary diisocyanate concentrations for these studies is provided in the appendix (tables S1and S2).

Sensitivity analysis

Where applicable, we report 95% uncertainty intervals (95%UI) that reflect a mix of true parameter uncertainty and between-worker variability in morphological parameters. Based on the input parameter distributions shown in Table 1, we

conducted a global sensitivity analysis (GSA) using an extended Fourier Amplitude Sensitivity Test (eFAST) (McNally 2011). The scenario used for this analysis was the "long-term exposure" scenario (consisting of 8 weeks exposure for 8 hours a day) with the focus on post-shift urinary concentrations on the last working-day at the end of this period.

Most biomonitoring studies provide little information on the timing of task-based exposures relative to the time of urine sample collection or on the frequency and timing of any intermediate urination, which could both contribute to increased exposure misclassification. We therefore investigated the sensitivity of estimated post-shift urinary diisocyanate concentrations for task-based exposures by varying the timing of exposure and the frequency and timing of intermediate urination moments.

Software

The PBK model was programmed in MCSim (version 6.2) and the R package PKSensi (version 1.2.3) was used to perform the global sensitivity analysis.

Results

Urinary biomarker levels following short-term exposure

Estimated urinary concentrations of TDA following a single inhalation exposure to 40 µg/m^3 of TDI for 4 hours are shown in the appendix (Figure S1, Appendix B). Excretion of urinary TDA peaked between 4 to 6 hours after exposure started. Most of the absorbed dose was eliminated within 10 hours after the end of exposure. This pattern is qualitatively similar to that described for human volunteers in the studies by (Skarping, Brorson, and Sangö 1991; Torbjorn Brorson, Skarping, and Sango 1991; Budnik et al. 2011). In those controlled experimental studies, the peak of excretion occurred roughly 2 to 6 hours after exposure had started, with most of the TDI excreted in urine within 24 hours from the end of exposure. The estimated urinary concentration of MDA after a single inhalation exposure to 40 µg/m3 of MDI for 4 hours is also shown in the appendix (Figure S2, Appendix B). Excretion of MDA peaked approximately 10 hours after the start of exposure, with excretion continuing long after exposure ceased. This pattern is similar to that described for human volunteers exposed to MDI in the study by Budnik et al. (2011), where the authors reported that excretion peaked approximately 14 hours after exposure and elimination was not complete after 24 hours.

Urinary biomarker levels following chronic inhalation exposure

TDI

The evolution of urinary concentrations of TDA over an 8-week period after repeated exposure to 40 µg/m3 TDI for 8 hours per day and for five days a week is provided in figure 2a. Levels of urinary TDA after the weekend on Monday morning before the shift

were estimated to be approximately 1.5 µg TDA/gr creatinine and cumulation of TDI seemed to level off after about 6 weeks.

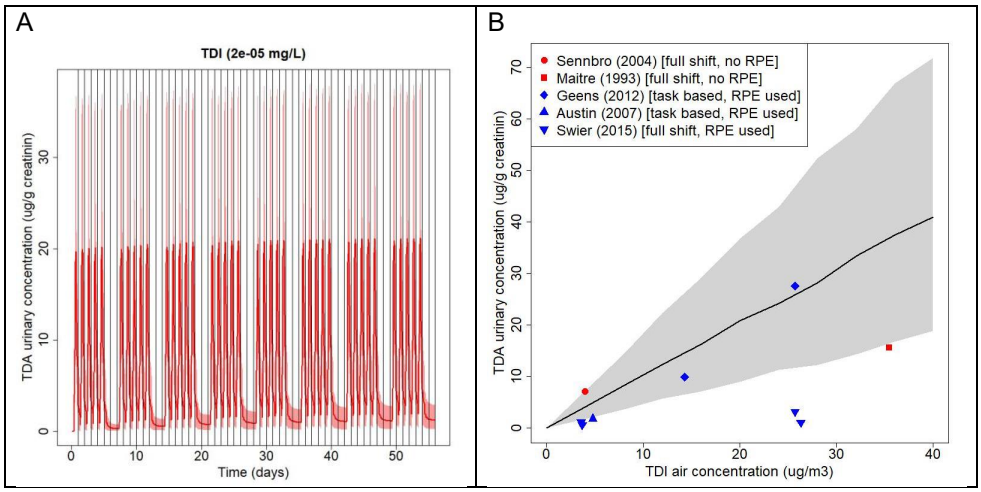


Figure 2. Predicted urinary TDA concentrations (in µg/gr creatinine) following repeated occupational exposure. A) Time evolution after exposure to 20 µg/m3 TDI. B) Exposure-response curve for urinary concentrations at the post-shift urination on the last working day. Exposure was assumed to occur on 5 days a week, for 8 hours a day (from 9:00 to 17:00), for a period of 8 weeks, with daily urination at 8:00, 13:00, 18:00, and 23:00. The red (A) and grey (B) ribbons show the 95%UI. Panel B additionally shows results from published occupational studies, with the legend indicating whether the study reported use of respiratory protective equipment (REP) by at least some of the workers.

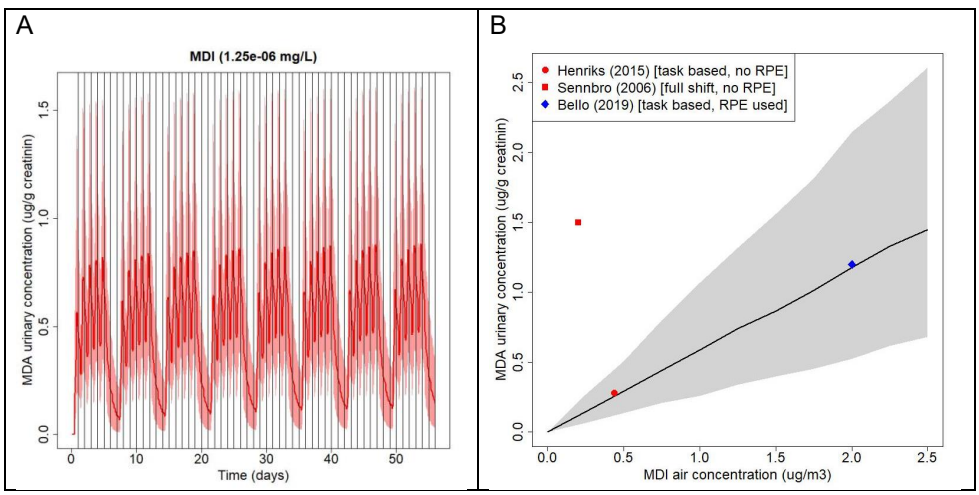


Figure 3. Predicted urinary MDA concentrations (in µg/gr creatinine) following repeated occupational exposure. A) Time evolution after exposure to 1.25 µg/m3 MDI. B) Exposure-response curve for urinary concentrations at the final post-shift urination on the last working day. Exposure was assumed to occur on 5 days a week, for 8 hours a day (from 9:00 to 17:00), for a period of 8 weeks, with daily urination at 8:00, 13:00, 18:00, and 23:00. The red (A) and grey (B) ribbons show the 95%UI. Panel B additionally shows results from published occupational studies, with the legend indicating whether the study reported use of respiratory protective equipment (REP) by at least some of the workers.

The exposure-response relation between air concentrations of TDI and estimated post-shift TDA urinary concentrations on the last workday of the 8 weeks exposure period is shown in figure 2b. There was a linear relation between air exposure and urinary concentrations, but with significant uncertainty in estimated TDA levels that included a nearly 5-fold difference between upper and lower boundaries of the uncertainty interval. Results from published occupational studies are included in the plot, to allow informal comparison. Model predictions were higher than reported urinary concentrations for almost all studies, note however that reported values were based on average exposures and average urinary concentrations for groups of workers and that respiratory protective equipment had also been used in some of the studies. A more detailed comparison is provided in the Discussion.

MDI

An overview of the time evolution of urinary concentrations of MDA over an 8-week period after repeated exposure to 1.25 µg/m^3 MDI for 8 hours per day and for five days a week is provided in figure 3a. Levels of urinary MDA after the weekend on Monday morning before the shift and after the shift on Friday afternoon in week 8 were estimated to be approximately 0.13 and 0.71 µg MDA/gr creatinine respectively, and (limited) accumulation of MDI seemed to occur even after 8 weeks.

The exposure-response relation between air concentrations of MDI and estimated post-shift MDA urinary concentrations on the last workday of the 8 weeks exposure period is shown in figure 3b. As for TDI, results from occupational studies were included for informal comparison. From the results it appears that model predictions are comparable to results from the Henriks *et al.* (2015) and Bello *et al.* (2019) studies, but not those from the Sennbro *et al.* (2006) study. A more detailed comparison is provided in the Discussion.

Impact of (short-term) dermal exposure to MDI

Our combined inhalation and dermal exposure scenario was based on results reported in the study by Henriks-Eckerman *et al.* (2015) and consisted of exposure to MDI in the air at a concentration of 0.4 µg/m^3 and a (dermal) deposition rate of 0.015 µg /cm^2/hour for 4 hours.

When including both inhalation and dermal exposure (Figure S3a, Appendix D), MDA urinary levels were estimated to be 0.03 and 0.14 µg/gr creatinine, directly after exposure and 4 hours later, respectively. Without inhalation exposure, MDA urinary levels were estimated to be 0.02 and 0.06 µg/gr creatinine, directly after exposure and 4 hours later, respectively (Figure S3b, Appendix D). The relative contribution of dermal exposure to total urinary excretion within 72 hours after exposure was approximately 25% and 4.5% of the total dermal dose had been excreted within 72 hours.

Parameter sensitivity for a fixed exposure scenario

We performed a global sensitivity analysis to assess which parameters contributed most to the uncertainty in model predictions for the “chronic” inhalation exposure scenario. For both TDI and MDI, the uncertainty in “fraction absorbed” appeared to have the largest impact on overall uncertainty, accounting for approximately 25% of total uncertainty (Figure 4 and 5). Other important contributors to overall uncertainty in predictions were the binding rate of TDI to albumin and other macro molecules and the body weight of subjects (which is often not recorded). For MDI, uncertainty in the rate of transfer for macro-molecule bound MDI from the “interstitial” to the “vascular” lung compartment was also important.

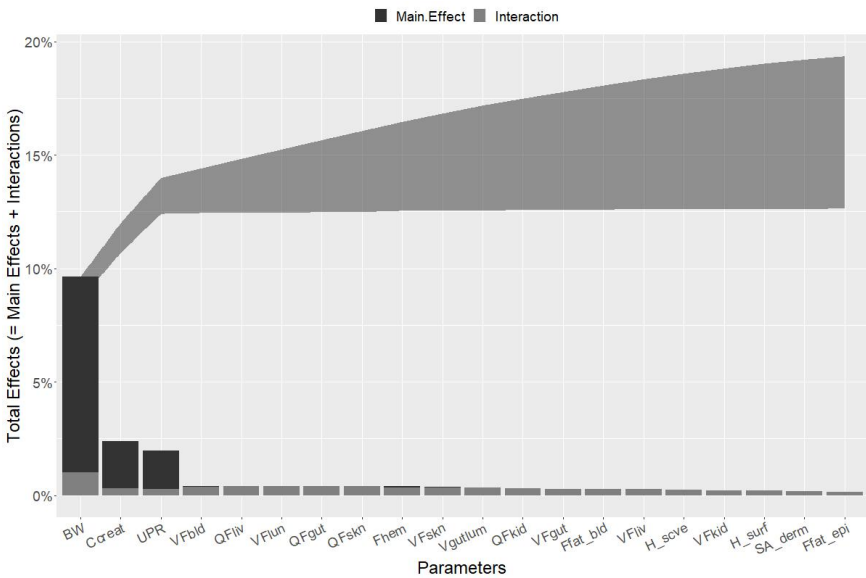


Figure 4. Lowry plot of results from the eFAST sensitivitiy analysis for predicted urinary concentrations of TDA (in µg/gr creatinine) after exposure to 20 µg/m³ TDI at the post-shift urination on the last working day. Exposure was assumed to occur on 5 days a week, for 8 hours a day (from 9:00 to 17:00), for a period of 8 weeks, with daily urination at 8:00, 13:00, 18:00, and 23:00. The total effect is comprised of the main effects (black bar) and possible interactions with other parameters (grey bar) given as the proportion of variance in the output parameter. The grey ribbon indicates the cumulative proportion of variance explained due to main effects and parameter interactions.

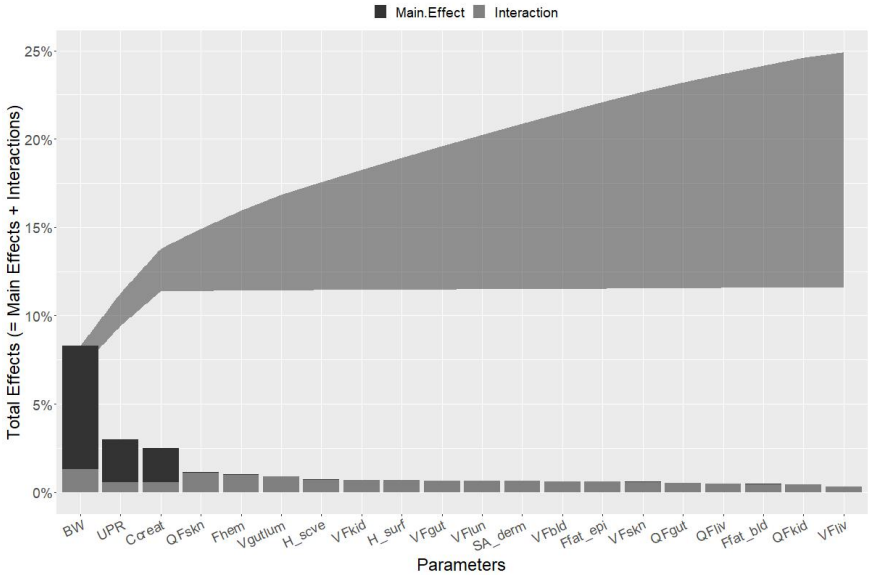


Figure 5. Lowry plot of results from the eFAST sensitivity analysis for predicted urinary concentrations of MDA (in µg/gr creatinine) after exposure to 1.25 µg/m^3 MDI at the post-shift urination on the last working day. Exposure was assumed to occur on 5 days a week, for 8 hours a day (from 9:00 to 17:00), for a period of 8 weeks, with daily urination at 8:00, 13:00, 18:00, and 23:00. The total effect is comprised of the main effects (black bar) and possible interactions with other parameters (grey bar) given as the proportion of variance in the output parameter. The grey ribbon indicates the cumulative proportion of variance explained due to main effects and parameter interactions.

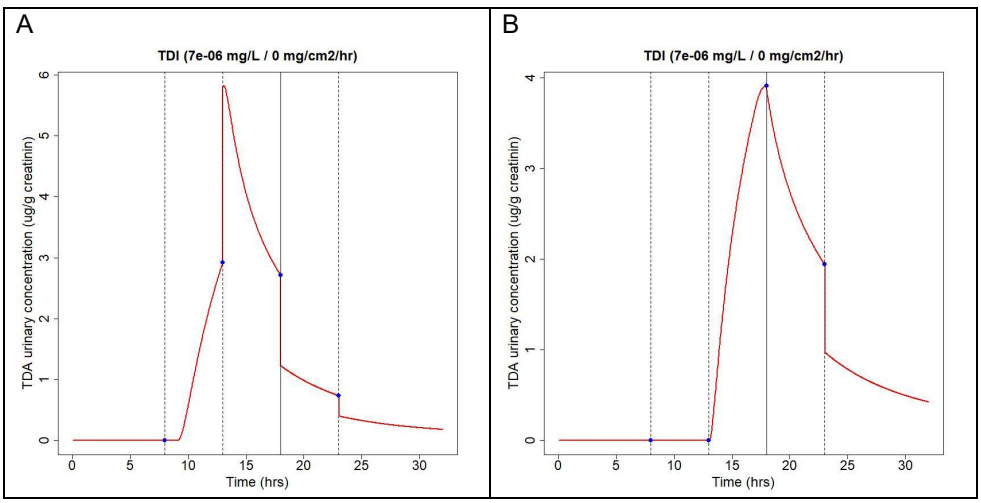


Figure 6. Effect of the timing of short-term exposure to TDI (7 µg/m3) on predicted urinary concentrations. For the early exposure scenario (A) exposure is from 9:00 to 13:00, with urination at 8:00, 13:00, 17:00, and 23:00. For the late exposure scenario (B) exposure is from 13:00 to 17:00, with the same urination moments. Urination moments are indicated by dotted lines with predicted urinary concentrations indicated by a blue dot.

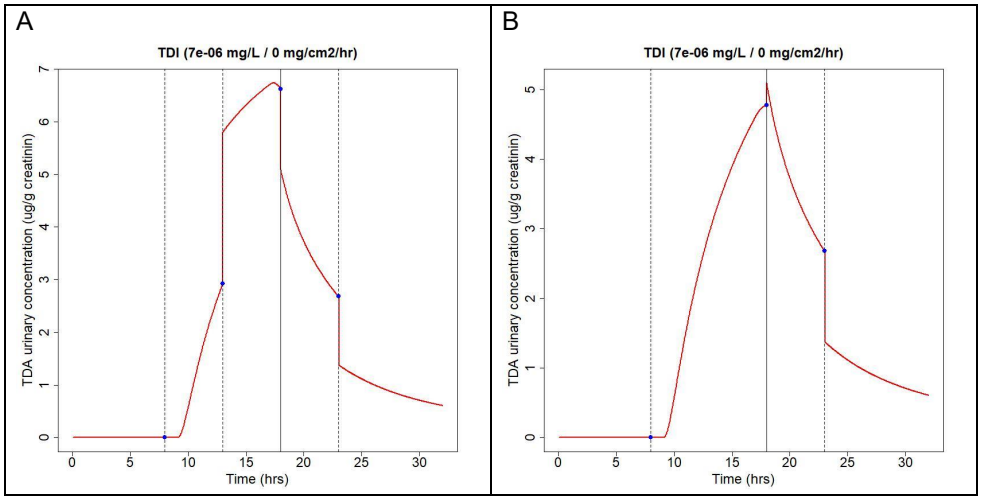


Figure 7. Effect of the frequency of urination on predicted urinary concentrations after 8 hours of exposure to TDI (7 µg/m3). For the frequent urination scenario (A) urination was at 8:00, 13:00, 17:00, and 23:00, with exposure from 9:00 to 17:00. For the infrequent urination scenario (B) there was no urination at 13:00. Urination moments are indicated by dotted lines with predicted urinary concentrations indicated by a blue dot.

Temporal factors in exposure and biomarker collection

The impact of differences in the timing of inhalation exposure relative to that for biomarker collection on estimated urinary concentrations is shown in Figure 6. Results were based on a short-term exposure scenario in which workers were exposed to 7 µg TDI/m^3 for 4 hours, but where exposure either started in the early morning at 9:00 ("early") or in the afternoon at 13:00 ("late"). Urination was at 11:00, 15:00, and 17:00 hours for both scenarios. Estimated TDA concentrations in urine obtained at the last urination moment were approximately 2.7 µg/gr creatinine for the "early" scenario (figure 6a), and 1.9 µg/gr creatinine for the "late" scenario (Figure 6b).

The impact of differences in the frequency and timing of urination is illustrated in figure 7. In these scenarios workers were exposed to 7 µg TDI/m^3 for 8 hours, but urination either at 11:00, 15:00, and 17:00 hours ("frequent"), or only at 17:00 ("infrequent"). Estimated TDA concentrations in urine obtained at the last urination moment were lower in the "infrequent" urination scenario (4.8 µg/gr creatinine; figure 7b) than in the "frequent" urination scenario (6.6 µg/gr creatinine , figure 7a).

Figure S4 in the appendix (Appendix F) shows results for the same exposure scenarios for MDI. Estimated MDA biomarker concentrations in post-shift samples were considerably higher for the "early" than the "late" scenario and also much higher for the "frequent" than the "infrequent" urination scenario illustrating that the effect of different exposure and urination scenarios on biomarker concentrations in post-shift samples critically depends on the kinetics of the isocyanates involved.

Discussion

Main findings

In this study we developed a PBK model for MDI and TDI; more specifically we modelled the renal excretion of MDA and TDA as degradation products. Global sensitivity analysis was used to assess which parameters contributed most to the uncertainty in simulated biomarker levels. Highly influential parameters included the fraction of diisocyanate absorbed, a subject's body weight, the binding rate of diisocyanates to albumin and other macromolecules, and (for MDI only) the transfer rate of macro-molecule bound MDI from the interstitial to the vascular lung compartments. We performed additional sensitivity analyses to investigate the importance of the timing of exposure and frequency and timing of urination and found that these had considerable impact on modelled urinary biomarker levels.

Model comparison

Inhalation

To be able to validate our model we compared model predictions with published data (Figure 4b and 5b). This is not straightforward because PBK models are generally compared to experimental animal studies with clear exposure regimes, whereas we aimed to estimate exposure in chronically exposed human workers for a chemical with both a fast and accumulating excretion pattern. The studies we extracted differed in terms of measuring exposure (task based versus TWA), the use of RPE (no use, very strict use, or worn inconsistently) and even in sample hydrolysis procedures. All these factors could impact urinary amine levels. For example, alkaline hydrolysis has been reported to give higher results as more adducts are released in comparison to acid hydrolysis (as reviewed in Scholten et al. 2020). The use of RPE could arguably result in lower urinary amine concentrations in comparison to our model predictions, unless the RPE was not very effective or not worn throughout the working period. And converting a task exposure to a TWA is only reliable if there was no exposure during the rest of the working period. We could not specify one specific parameter that was predictive of urinary measurements. For example, the Geens et al. (2012) study reported nearly the same urinary concentrations as we modelled (Figure 4b) but the workers in that study did wear RPE (which could have resulted in lower levels), and reported a task based estimate (which we converted to 8 hours exposure) and used the higher release alkaline method in comparison to the other TDI biomonitoring studies. Another difficulty is that levels of predicted urinary levels after longer exposures are difficult to compare because it is unclear whether exposure levels as measured on study days were representative for the entire period.

We further estimated accumulation due to binding to protein adducts by modelling urinary results on Monday morning after a weekend with no exposure. Accumulation

was about 1.5 µg/gr creatinine for TDA (based on chronic exposure to 20 µg/m^3 TDI), and 0.13 µg/gr creatinine for MDA (based on chronic exposure to 1.25 µg/m^3) (Figure 4a and 5a). A couple of studies measured pre-shift levels. For TDI, Monday morning urinary concentrations were reported to be 4.5 µg/gr creatinine (with Friday post-shift air measurements of 39.45 µg/m^3) (Geens et al. 2012), and between 0.15 and 4 µg/gr creatinine with air measurements between 0.52 and 4.1 µg/m^3 for 2,4-TDI (Tinnerberg, 2014). For MDI, Tinnerberg et al. (2014) reported 0.52 µg MDA/gr creatinine (and air sample measurements between 0.04 and 9.7 ug 4,4'-MDI/m^3). Based on these studies our model estimates seem to be in the same range as the published data, considering the various exposure scenarios.

In general we can conclude that published results are reasonably comparable, although systematically lower, to our model predictions (Figure 4b and 5b), but biomonitoring studies for diisocyanates could be more comparable if the studies were more harmonized and reported all relevant information, including the moment of exposure when it concerned a task based exposure measurement. Recently a protocol for measuring and reporting on diisocyanates was published, which is helpful in harmonizing data collecting and reporting (Jones et al. 2022).

Dermal

The dermal compartment of our model is difficult to evaluate. Henriks-Eckerman et al. (2014) reported average urinary concentrations of about 0.30 µg/gr creatinine for workers that did not use either gloves or RPE. We estimated levels of 0.03 and 0.14 µg/gr creatinine, directly after exposure and 4 hours later, when considering the same inhalation and dermal exposure as reported by Henriks-Eckerman et al. (2014). There is evidence that most skin absorbed diisocyanate is held in a skin "reservoir", which could result in subsequent slower desorption into the circulation (ATSDR 1998). With our model we estimate a relative contribution of dermal exposure to total urinary excretion (within 72 hours after exposure) of approximately 25%. Furthermore we predicted that, from the applied dose, 4.5% of the total dermal dose had been excreted within 72 hours. This seems much higher in comparison to a controlled human exposure study by Hamada et al. (2018), who dermally exposed 4 volunteers for 8 hours to MDI. They found very low amounts in urine and plasma of (about 0.01 to 0.2% of the applied dose). The authors hypothesize that MDI is rapidly absorbed but remain bound to the upper layers of the skin. Hoffmann et al. (2010) studied dermal absorption in rats and also found very low dermal absorption of MDI: at or below 1% of the applied dose. It has been argued that dermal exposure could lead to exposure via hand-to-mouth behavior. Currently our model assumes that there is no uptake in the gut after oral exposure, i.e. no systemic availability. To our knowledge there is no data available on absorption from the gut for MDI. For TDI, a study in orally exposed male rates found that 81% of the administered 14C- 2,4-TDI was recovered in feces, and only 8% in urine (Gledhill et al. 2005).

Strengths and weaknesses of model, and possibilities

The strengths of our model include the potential to model both inhalation and dermal exposure, and to account for longer exposure histories. The biologically interpretable parameters can be validated in mechanistical studies. Weaknesses of our model include the fact that for many parameters there is uncertainty regarding the exact value, as also reflected in the wide parameter distribution range (e.g. confidence interval in Figure 4b, 5b). In addition model validation was difficult because often information on the precise exposure situation was missing, for example on the moment of exposure when it concerned a TWA, or to which extent PPE was used throughout the day. For model calibration it would be helpful to have an individual dataset available with both external exposure estimates (quantitative air and dermal measures) including data on the exposure scenario (i.e. moment and duration of exposure) and individual behavior and information (i.e. moment of urination, body weight), together with well described biomonitoring data. This information could then be included in the model and as such reduce uncertainty of our common assumptions on e.g. body weight distribution or moment of urination. In addition if urinary values are measured over an extended period of time, even during/after a period of non-exposure, this provides information on the accumulation and binding to albumin, but also on the fraction absorbed. With a fully validated and calibrated model we have more certainty with regards to extrapolating across dose levels and exposure scenarios, and characterizing potential contributions of RPE and/or dermal exposure. In addition we could apply exposure reconstruction with more precision, when only urinary data is available. Lastly our model could be used for informing biological limit values (BLV). The BLV for both TDI and MDI are based on correlation formulas between air sampling and urinary data (Hartwig and MAK Commission 2021; DFG 2007). These correlations formulas were used to estimate urinary amine levels upon exposure to available MAKs. We compared the current available BLVs for diisocyanates to our model predictions. The Deutsche Forschungsgemeinschaft (DFG) derived a value of 6.13 µg TDA/gr creatinine based on the MAK value of 7 µg/m^3 and proposed a BLV of 5 µg TDA/gr creatinine (Hartwig and MAK Commission 2021). With our model we predicted urinary values of 8.7 µg TDA/gr creatinine, when modelling the last post-shift urine sample, so comparable to the value based on correlation formulae. The DFG also used correlation formulae for MDI in air and MDA in urine and proposed a value of 10 µg MDA/gr creatinine based on exposure to MDI at the MAK value of 0.05 mg/m^3. With our model we predicted MDA urinary values of 22 µg/gr creatinine, post-shift.

Recently the European Risk Assessment Committee (RAC) concluded that occupational diisocyanate exposure >0.67 µg/m^3 NCO during an entire working life was likely to result in a more than 5% excess risk of developing bronchial hyperresponsiveness (BHR) (RAC 2020b). For TDI and MDI our model predicts urinary levels of 1 and 1.4 µg/gr creatinine for TDI and MDI respectively, at this critical

exposure level. A purely health based BLV should then accordingly also be lower than the current values of 5 µg TDA/gr creatinine and 10 µg MDA/gr creatinine.

References Chapter 4

ATSDR. 1998. “Toxicological Profile for Methylenedianiline. U.S. Dept. Health & Human Services.”

———. 2018. “Toxicological Profile for Toluene Diisocyanate and Methylenediphenyl Diisocyanate.”

Austin, Shaun. 2007. “Biological Monitoring of TDI-Derived Amines in Polyurethane Foam Production.” *Occupational Medicine*. https://doi.org/10.1093/occmed/kqm085.

Bello, Anila, Yalong Xue, Rebecca Gore, Susan Woskie, and Dhimiter Bello. 2019. “Assessment and Control of Exposures to Polymeric Methylene Diphenyl Diisocyanate (PMDI) in Spray Polyurethane Foam Applicators.” *International Journal of Hygiene and Environmental Health* 222 (5): 804–15. https://doi.org/10.1016/J.IJHEH.2019.04.014.

Bello, Dhimiter, Christina A. Herrick, Thomas J. Smith, Susan R. Woskie, Robert P. Streicher, Mark R. Cullen, Youcheng Liu, and Carrie A. Redlich. 2007. “Skin Exposure to Isocyanates: Reasons for Concern.” *Environmental Health Perspectives* 115 (3): 328–35. https://doi.org/10.1289/EHP.9557.

Brorson, Torbjorn, Gunnar Skarping, and Carsten Sango. 1991. “Biological Monitoring of Isocyanates and Related Amines.” *International Archives of Occupational and Environmental Health*. https://doi.org/10.1007/bf00386374.

Brorson, Torbjörn, Gunnar Skarping, and Carsten Sangö. 1991. “Biological Monitoring of Isocyanates and Related Amines.” *International Archives of Occupational and Environmental Health 1991 63:4* 63 (4): 253–59. https://doi.org/10.1007/BF00386374.

Budnik, Lygia T., Dennis Nowak, Rolf Merget, Catherine Lemiere, and Xaver Baur. 2011. “Elimination Kinetics of Diisocyanates after Specific Inhalative Challenges in Humans: Mass Spectrometry Analysis, as a Basis for Biomonitoring Strategies.” *Journal of Occupational Medicine and Toxicology*. https://doi.org/10.1186/1745-6673-6-9.

DFG. 2007. “Addendum to 4,4’-Methylene Diphenyl Diisocyanate (Inhalable Fraction)[BAT Value Documentation, 2007].”

ECHA. 2009. “Guidance on Information Requirements and Chemical Safety Assessment.” https://echa.europa.eu/documents/10162/23047722/r14_update_version_2_rev00_en.pdf/c5c1a1e9-1251-42a6-813d-29cb8ee8279b.

Geens, Tom, Sandra Dugardin, Ann Schockaert, Geert De Cooman, and Marc Van Sprundel. 2012. “Air Exposure Assessment of TDI and Biological Monitoring of TDA in Urine in Workers in Polyurethane Foam Industry.” *Occupational and Environmental Medicine* 69 (2): 93–98. https://doi.org/10.1136/OEM.2011.064840.

Gledhill, A., A. Wake, P. Hext, E. Leibold, and R. Shiotsuka. 2005. “Absorption, Distribution, Metabolism and Excretion of an Inhalation Dose of [14C] 4,4′-Methylenediphenyl Diisocyanate in the Male Rat.” *Xenobiotica*. https://doi.org/10.1080/00498250500057591.

Hartwig, A, and MAK Commission. 2021. “The MAK Collection for Occupational Toluylendiisocyanate MAK-Begründung, Nachtrag.” *Health and Safety* 6 (2). https://doi.org/10.34865/mb58484ismd6_2ad.

Health Council of the Netherlands. 2018. “Di-and Triisocyanates Health-Based Recommendation on Occupational Exposure Limits.” The Hague.

Henriks-Eckerman, Maj Len, Erja A. Mäkelä, Juha Laitinen, Katriina Ylinen, Katri Suuronen,

Aki Vuokko, and Riitta Sauni. 2015. “Role of Dermal Exposure in Systemic Intake of Methylenediphenyl Diisocyanate (MDI) among Construction and Boat Building Workers.” *Toxicology Letters* 232 (3): 595–600. https://doi.org/10.1016/J.TOXLET.2014.12.012.

HERAG. 2007. “Health Risk Assessment Guidance for Metals. Assessment of Occupational Inhalation Exposure and Systematic Inhalation Absorption.” https://www.ebrc.de/downloads/HERAG_FS_02_August_07.pdf.

Hoer, D, HA Barton, A Paini, M Bartels, I Brandall, J Domoradzki, and J. Fisher. 2020. “Predicting Nonlinear Relationships between External and Internal Concentrations with Physiologically Based Pharmacokinetic Modeling.” *Toxicology and Applied Pharmacology* 440: 115922. https://www.sciencedirect.com/science/article/pii/S0041008X22000679.

Hoffmann, H. D., E. Leibold, C. Ehnes, E. Fabian, R. Landsiedel, A. Gamer, and A. Poole. 2010. “Dermal Uptake and Excretion of 14C-Toluene Diisocyante (TDI) and 14C-Methylene Diphenyl Diisocyanate (MDI) in Male Rats. Clinical Signs and Histopathology Following Dermal Exposure of Male Rats to TDI.” *Toxicology Letters* 199 (3): 364–71. https://doi.org/10.1016/J.TOXLET.2010.09.021.

Hozo, Stela Pudar, Benjamin Djulbegovic, and Iztok Hozo. 2005. “Estimating the Mean and Variance from the Median, Range, and the Size of a Sample.” *BMC Medical Research Methodology* 5 (1): 1–10. https://doi.org/10.1186/1471-2288-5-13/TABLES/3.

Jones, Kate, Karen S Galea, Bernice Scholten, Marika Loikala, Simo P Porras, Radia Bousoumah, Sophie Ndaw, et al. 2022. “HBM4EU Diisocyanates Study—Research Protocol for a Collaborative European Human Biological Monitoring Study on Occupational Exposure.” *International Journal of Environmental Research and Public Health 2022, Vol. 19, Page 8811* 19 (14): 8811. https://doi.org/10.3390/IJERPH19148811.

Lind, Pernilla, Marianne Dalene, Gunnar Skarping, and Lars Hagmar. 1996. “Toxicokinetics of 2,4- and 2,6-Toluenediamine in Hydrolysed Urine and Plasma after Occupational Exposure to 2,4- and 2,6- Toluene Diisocyanate.” *Occupational and Environmental Medicine* 53 (2): 94–99. https://doi.org/10.1136/OEM.53.2.94.

Lockey, James E., Carrie A. Redlich, Robert Streicher, Andrea Pfahles-Hutchens, Pertti J. Hakkinen, Gary L. Ellison, Philip Harber, et al. 2015. “Isocyanates and Human Health: Multi-Stakeholder Information Needs andresearch Priorities.” *Journal of Occupational and Environmental Medicine / American College of Occupational and Environmental Medicine* 57 (1): 44. https://doi.org/10.1097/JOM.0000000000000278.

Luu, H Do, and J Hutter. 2000. “Pharmacokinetic Modeling of 4,4’-methylenedianiline Released from Reused Polyurethane Dialyzer Potting Materials.” *Journal of Biomedical Materials Research: An Official Journal of The Society for Biomaterials, The Japanese Society for Biomaterials, and The Australian Society for Biomaterials and the Korean Society for Biomaterials* 53(3)276-2. https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-4636(2000)53:3%3C276::AID-JBM13%3E3.0.CO;2-E.

Luu, Hoan My Do, Joseph C. Hutter, and Harry F. Bushar. 1998. “A Physiologically Based Pharmacokinetic Model for 2,4-Toluenediamine Leached from Polyurethane Foam-Covered Breast Implants.” *Environmental Health Perspectives* 106 (7): 393–400. https://doi.org/10.1289/EHP.98106393.

Maître, A., M. Berode, A. Perdrix, S. Romazini, and H. Savolainen. 1993. “Biological Monitoring of Occupational Exposure to Toluene Diisocyanate.” *International Archives*

of Occupational and Environmental Health 1993 65:2 65 (2): 97–100. https://doi.org/10.1007/BF00405726.

Malo, Jean Luc, and Moira Chan-Yeung. 2009. “Agents Causing Occupational Asthma.” *Journal of Allergy and Clinical Immunology* 123 (3): 545–50. https://doi.org/10.1016/J.JACI.2008.09.010.

McNally, Kevin, Richard Cotton, John Cocker, Kate Jones, Mike Bartels, David Rick, Paul Price, and George Loizou. 2012. “Reconstruction of Exposure to M-Xylene from Human Biomonitoring Data Using PBPK Modelling, Bayesian Inference, and Markov Chain Monte Carlo Simulation.” *Journal of Toxicology* 2012. https://doi.org/10.1155/2012/760281.

PubChem. 2022. “2,6-Diisocyanatotoluene.” 2022. https://pubchem.ncbi.nlm.nih.gov/compound/2%2C6-toluene diisocyanate.

RAC. 2020a. “Opinion on Scientific Evaluation of Occupational Exposure Limits for Diisocyanates.”

———. 2020b. “Opinion on Scientific Evaluation of Occupational Exposure Limits for Diisocyanates.”

Redal-Baigorri, Belén, Knud Rasmussen, and James Goya Heaf. 2014. “Indexing Glomerular Filtration Rate to Body Surface Area: Clinical Consequences.” *Journal of Clinical Laboratory Analysis* 28 (2): 83. https://doi.org/10.1002/JCLA.21648.

Redlich, Carrie A. 2010. “Skin Exposure and Asthma Is There a Connection?” *Proceedings of the American Thoracic Society* 7 (2): 134–37. https://doi.org/10.1513/PATS.201002-025RM.

Scholten, B, L Kenny, R Duca, A Pronk, T Santonen, K Galea, M Loh, et al. 2020. “Biomonitoring for Occupational Exposure to Diisocyanates: A Systematic Review.” *Annals of Work Exposures and Health*.

Schütze, D., O. Sepai, J. Lewalter, L. Miksche, D. Henschler, and G. Sabbioni. 1995. “Biomonitoring of Workers Exposed to 4,4’-Methylenedianiline or 4,4’-Methylenediphenyl Diisocyanate.” *Carcinogenesis* 16 (3): 573–82. https://doi.org/10.1093/CARCIN/16.3.573.

Sennbro, C. J., C. H. Lindh, C. Mattsson, B. A.G. Jönsson, and H. Tinnerberg. 2006. “Biological Monitoring of Exposure to 1,5-Naphthalene Diisocyanate and 4,4′-Methylenediphenyl Diisocyanate.” *International Archives of Occupational and Environmental Health* 79 (8): 647–53. https://doi.org/10.1007/S00420-006-0096-5/FIGURES/2.

Sennbro, Carl J., Christian H. Lindh, Anders Östin, Hans Welinder, Bo A.G. Jönsson, and Håkan Tinnerberg. 2004. “A Survey of Airborne Isocyanate Exposure in 13 Swedish Polyurethane Industries.” *The Annals of Occupational Hygiene* 48 (5): 405–14. https://doi.org/10.1093/ANNHYG/MEH034.

Sepai, Ovnair, Dietrich Henschler, and Gabriele Sabbioni. 1995. “Albumin Adducts, Hemoglobin Adducts and Urinary Metabolites in Workers Exposed to 4,4′-Methylenediphenyl Diisocyanate.” *Carcinogenesis* 16 (10): 2583–87. https://doi.org/10.1093/CARCIN/16.10.2583.

Skarping, Gunnar, Torhjörn Brorson, and Carsten Sangö. 1991. “Biological Monitoring of Isocyanates and Related Amines - III. Test Chamber Exposure of Humans to Toluene Diisocyanate.” *International Archives of Occupational and Environmental Health*. https://doi.org/10.1007/BF00379069.

Swierczyńska-Machura, Dominika, Sławomir Brzeźnicki, Ewa Nowakowska-Swirta, and Jolanta Walusiak-Skorupa. 2015. “Occupational Exposure to Diisocyanates in Polyurethane Foam Factory Workers Good Practice in Occupational Health Services: Recommendations for Prophylactic Examinations in Patients with Diabetes Mellitus View Project Prevention of Hypertension and Diabetes in Occupational Prophylactic Care View Project.” *Article in International Journal of Occupational Medicine and Environmental Health*. https://doi.org/10.13075/ijomeh.1896.00284.

Tinnerberg, Håkan, Karin Broberg, Christian H. Lindh, and Bo A.G. Jönsson. 2014. “Biomarkers of Exposure in Monday Morning Urine Samples as a Long-Term Measure of Exposure to Aromatic Diisocyanates.” *International Archives of Occupational and Environmental Health* 87 (4): 365–72. https://doi.org/10.1007/S00420-013-0872-Y/FIGURES/3.

USEPA. 2011. “Estimation Program Interface (EPI) Suite.” 2011. https://www.epa.gov/oppt/exposure/pubs/episuitedl.htm.

Valentin, J., and Christian Streffer. 2002. “Basic Anatomical and Physiological Data for Use in Radiological Protection: Reference Values - ICRP Publication 89.” *Annals of the ICRP* 32 (3–4): 1–277. https://doi.org/10.1016/S0146-6453(03)00002-2.

Supplemental Material - Background information on model parameters

Several controlled human exposure studies were used to inform on model parameters. Briefly about these studies: in the study by Brorson, Skarping, and Sango (1991) two healthy male volunteers were exposed to diisocyanates in a test chamber. Both were occupational hygienists and had occasional exposure to diisocyanates. The volunteers remained at physical rest during the exposure, and were exposed for four hours to various TDI concentrations. All urine was repeatedly collected at 2-h intervals for up to 14 or 16 h after the start of exposure and then at an 8-h interval during the following night. In the study by Skarping, Brorson, and Sangö (1991), five healthy male volunteers were exposed to TDI for 7.5 hours. Urine was collected at 2 hour intervals up to 15 hours after exposure had started, and then at 8 and 4 hour intervals. Budnik et al. (2011) exposed workers to several diisocyanates, including MDI and TDI, for 2 hours, with urine collections at 3, 6, 9, 12 and 20 hours.

In the study by Brorson, Skarping, and Sango (1991) the cumulated amount of 2,4-TDA excreted in the urine over 24 h was ca 15% - 19% of the estimated inhaled dose of 2,4-TDI, and that of 2,6-TDA was ca 17% - 23% of the estimated inhaled dose of 2,6-TDI. Brorson et al. report that the uptake of TDAs by the blood continues to increase slightly post exposure. In the study by Skarping, Brorson, and Sangö (1991), five men were exposed to TDI for 7.5 hours. The cumulative amount of 2,4-TDA excreted in urine within 28 h ranged from 8% to 14% of the estimated dose of 2,4-TDI, and the cumulative amount of 2,6-TDA in urine ranged from 14% to 18% of the estimated 2,6-TDI dose. For MDI there is only one kinetic study in animals (Gledhill et al. 2008). This inhalation study in rats measured that, after 168 hours exposures to MDI, approximately 32% of the inhaled dose was systemically available (Gledhill et al. 2008).

Budnik et al. exposed volunteers to different diisocyanates and found, for TDI, that the major excretion peak was at 4.1 h and 4.8 h (for 2,6- and 2,4-TDA), with a half-life of about 6 hours. Further Budnik et al. report that the majority of the TDA appeared to be eliminated after 24 hours. For MDI it was evident that MDI is excreted more slowly (i.e. there is a peak in excretion after 14 hours and after 24 hours there is still a significant amount of MDI excreted in urine) when compared to TDI. This is also consistent with findings from different animal studies (Timchalk, Smith, and Bartels 1994; Gledhill et al. 2008). Kääriä et al. (2001) monitored MDI exposure and corresponding MDA urinary levels for two consecutive days in a worker population and reported that due to the long half-life of urinary MDA exposure of the previous day is reflected in the slightly higher concentrations in the second day's morning samples compared with the noontime samples. The reason for the slower elimination kinetics of MDI in comparison to TDI is not entirely clear. MDI is a larger chemical with a higher molecular weight. Gledhill et al. (2008) state that "*the reactivity of MDI is such that it may not be directly absorbed into systemic circulation from the lung, rather, it is highly likely that it reacts*

initially with the high glutathione content of the lung to form glutathione conjugates which act as a carrier". Evidence suggests that MDI has a lower reactivity to albumin when compared to TDI (Hettick and Siegel 2011). Therefore, the difference in elimination between MDI and TDI could possibly be explained by a difference in the reactivity towards albumin and other macromolecules at the site of exposure. We included this information in our model by making a distinction between MDI and TDI for the half-life transfer from non-blood to blood in lungs (i.e. we assume MDI transfers more slowly from the lung compartment to the systemic blood circulation) and the half-life binding to albumin/other macromolecules (i.e. we assume MDI binds less quickly to albumin than TDI).

Supplemental Material - Single exposure scenario

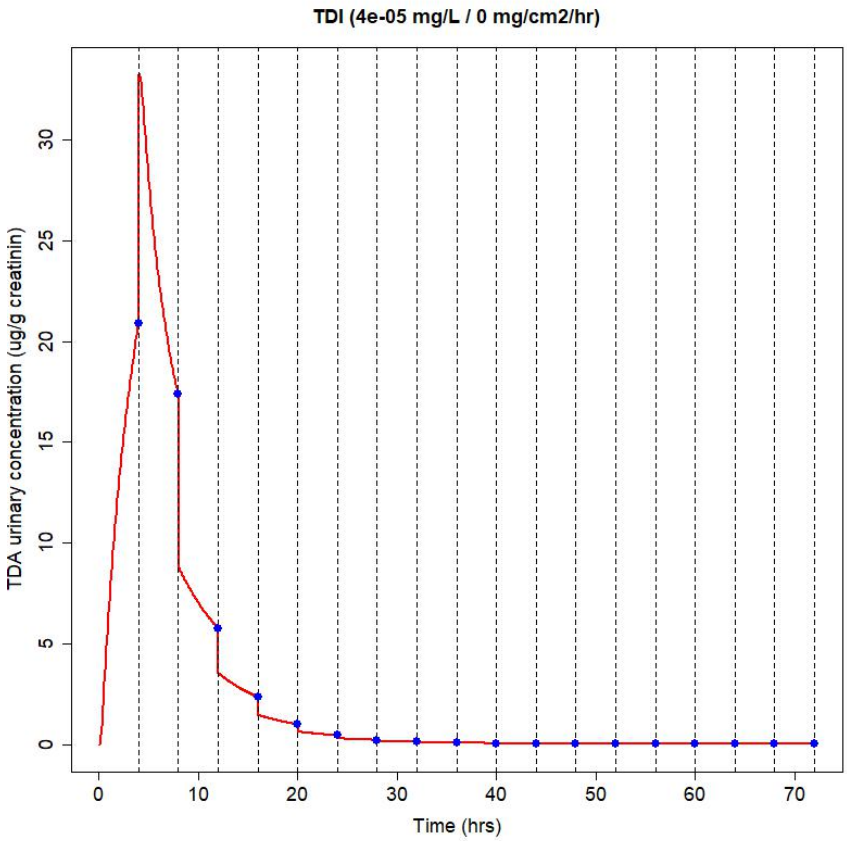


Figure S1. Estimated urinary concentrations of TDA (in µg/g creatinine) after inhalatory exposure to 40 µg/m3 of TDI for 4 hours, with urination every 4 hours. Urination moments are indicated by dotted lines with predicted urinary concentrations indicated by a blue dot.

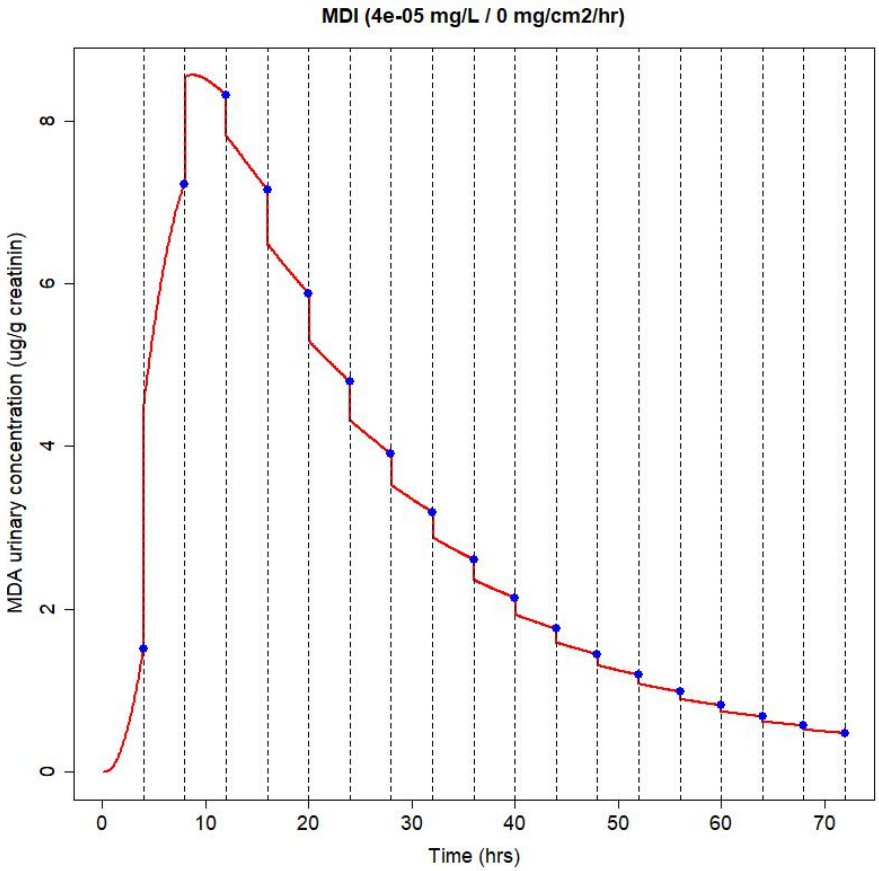


Figure S2. Estimated urinary concentrations of MDA (in µg/g creatinine) after inhalatory exposure to 40 µg/m3 of MDI for 4 hours, with urination every 4 hours. Urination moments are indicated by dotted lines with predicted urinary concentrations indicated by a blue dot.

Supplemental Material - TDI studies

Table S1. Overview available, published TDI studies with both external and internal exposure data available.

Country Sector N workers	Method of air sampling Task based or 8-hr average	Exposure duration Working history	Information on RPE	Other information	Sample collection specification	Exposure (µg/m³)	Urinary concentration [1]	Reference
Belgium PU foam production plant N = 9	Personal sampling Task based (AM of 313 min)	No information	Worn inconsistently		Hydrolysed overnight in 0.5 M of sodium hydroxide (alkaline)	39.45 (AM) 21.94 39.45 (AM) 21.94	37.46 ug/L (Friday post-shift) 13.47 ug/L (Monday post-shift) 10.95 ug/L (Friday pre-shift) 6.17 ug/L (Monday pre-shift)	(Geens et al. 2012)
UK Polyurethane foam block production N = 13	26 personal samples and 4 static. Task based (mean sampling period was 409 min)	Mean duration of potential exposure: 15 years	RPE was used for short periods when exposure was considered to be potentially high	Sampling started on the second day of the working week uTDA detected after shift in 10/13 samples	No information	5.6 (converted from NCO group equivalent – 2.7 ug/m3 NCO)	1.71 (recalculated from 2.21 umol/mol creatinine based on three below the LOD)	(Austin 2007)
Sweden Molding – ,flame lamination-, low-or non heating process plants,	Personal sampling 8 hr average	No information, age workers: 36 (median)	Air sampling paused when RPE was used	Urine collected last 4 hours of shift	Hydrolyzed for 24 hrs in sodium hydroxide en then extracted with toluene	4 (median) (<LOQ – 44 ug/m3)	7.1 ug/g creatinine (median) (range: <0.07 – 119.34 ug/gr creatinine)	(Carl J. Sennbro et al. 2004)

[1] *Urinary values are converted to µg/gr creatinine by assuming a creatinine value of 12 mmol/L. MW creatinine: 113.12 g/mol. MW TDA: 122.17 g/mol.*

Country Sector N workers	Method of air sampling Task based or 8-hr average	Exposure duration Working history	Information on RPE	Other information	Sample collection specification	Exposure (µg/m³)	Urinary concentration [1]	Reference
n=81 Poland Polyurethane foam factory N = 30	Personal air sampling 8 hr-average (time was not less than 75% of 8 hr-shift)	About 10 years	Was worn occasionally	No correlation between air and urine samples, but authors claim this could be due to RPE use.	Citric acid	3.7 (AM) 3.6 (AM) 25.7 (AM) (AM)	0.6 (AM) umol/mol creatinine (n= 10) 1.1 (AM) umol/mol creatinine (n= 3) 3.0 (AM) umol/mol creatinine (n= 2) 1.0 (AM) umol/mol creatinine (n=5)	(Swierczyńska-Machura et al. 2015)
Not clear (France or Switzerland) Production of flexible foam N = 9	Personal sampling 8 hr average	No information (workers age: 35 – 45 yrs)	No ventilation system, workers did not use PPE	Good correlation between airborne TDI and urinary TDA sampled after shift	Hydrochloric acid hydrolysis	35.39 (calculated average of nine workers)	15.7 (SD: 8.3) (ug/gr creatinine)	(Maître et al. 1993)
Sweden N = 24	Pre-shift levels			I choose the results from plant A because there was no mention of RPE use for this sector		Between 0.52 and 4.1 ug/m3 for 2,4-TDI, 0.21 and 1.5 ug/m3 for 2,6-TDI	Levels between 0.5 and 5.4 ng/ml = 0.37 and 4 ug/gr creatinine	(Tinnerberg et al. 2014)

Supplemental Material - MDI studies

Table S2. MDI studies for validation. Urinary concentrations are measured in post-shift samples unless indicated otherwise

Country Sector N workers	Method of air sampling Task based or 8-hr average	Exposure duration/ Working history	Information on RPE	Other information	Sample collection specification	Exposure (µg/m³)	Urinary concentration	Reference
USA N= 21	Personal and stationary sampling (median personal sampling time: 70 min)		PPE used by most workers (full- and half face respirators)		Acid hydrolysis (sulfuric acid)	GM: 13.8 (GSD: 4.8)	GM: 0.7 (GSD: 3.0) umol MDA/mol cr.	(Bello et al. 2019)
Sweden n=18 Moulding processes	8-hr average (median time for sampling 6.7 hr)	No information	When respiratory PPE were used the air sampling was paused, thus assuming zero exposure during periods when PPE was used.	Workers were (potentially) also exposed to NDI	hydrolysed for 24 hr in 0.3 M NaOH	0.2 ug/m3 (median) Range: 0.03 – 7.8 ug/m3	2 ug/L (median) Range: 0.4 - 38	(Sennbro et al. 2006)
Boat building industry, Construction work, production of prefabricated units n=24	Inhalation exposure measured in hood during exposing work For inhalation exposure: the sampling lasted around 15 – 60 min for each batch, and each batch was repeated 2 – 20 times per shift		RPE was used in some occasions (we extracted data for the group not using respiratory protection)	Dermal exposure measured by using tape-stripping method		Inhalation: 0.08 - 0.8 ug/m3 (without RPE) (average would then be ~ 0.44 ug/m3) the use of a powered hood reduced the inhalable concentration of MDI by 60% The medians of the amounts of MDI on the skin of the	Range of 0.1– 0.2 umol/mol creatinine (Based on a figure, it is 0.17 when no RPE is used)	(Henriks-Eckerman et al. 2015)

Country Sector N workers	Method of air sampling Task based or 8-hr average	Exposure duration/ Working history	Information on RPE	Other information	Sample collection specification	Exposure (µg/m³)	Urinary concentration	Reference
						workers using chemical protective gloves were below 0.1 ug/10 cm2.		
Sweden N = 24	Pre-shift levels					Between 0.04 and 9.7 ug 4,4'-MDI/m3 (plant C)	0.7 ng/ml	(Tinnerberg et al. 2014)

Supplemental Material - Dermal exposure to MDI

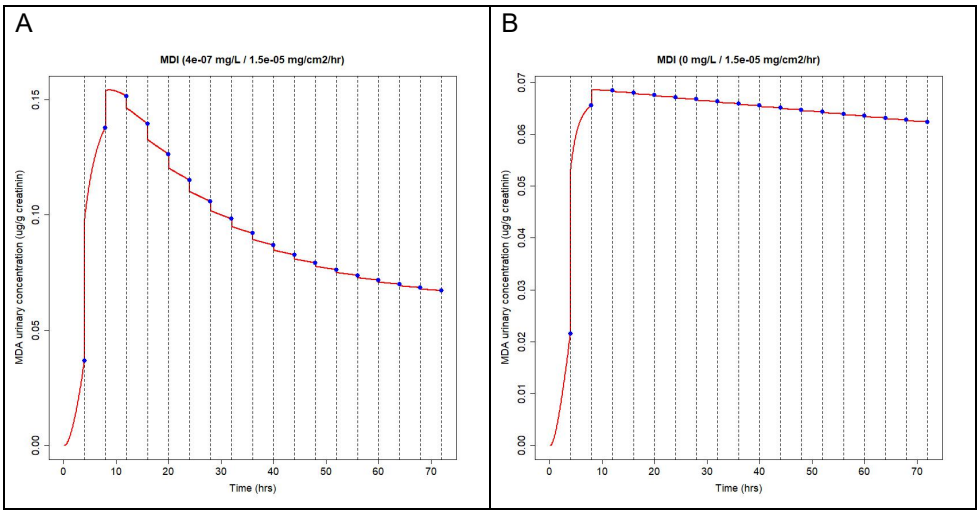


Figure S3: Predicted urinary concentrations of MDA (in µg/g creatinine) after 4 hours of both dermal (0.015 ug MDI/cm2/hour) and inhalatory (0.4 ug MDI/m3) exposure to MDI (A), or after dermal exposure only (B).

Supplemental Material - Temporal factors in exposure, analyses for MDI

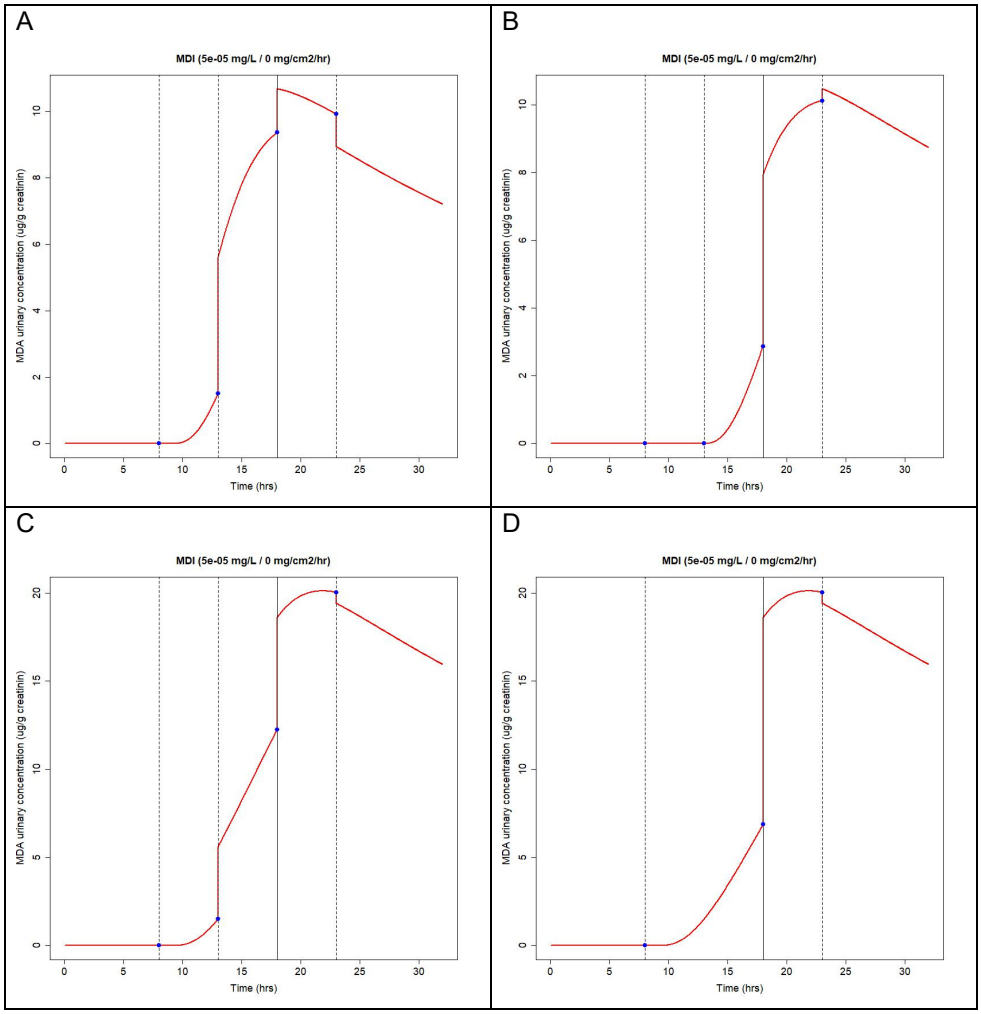


Figure S4. For the early exposure scenario (A), exposure is from 9:00-13:00 and urination at 8:00, 13:00, 18:00, and 23:00. For the late exposure scenario (B), exposure is from 13:00-17:00 and urination at 8:00, 13:00, 18:00, and 23:00. For the frequent urination scenario (C), exposure is from 9:00-17:00 and urination is at 8:00, 13:00, 18:00, and 23:00. For the infrequent urination scenario (D) , exposure is from 9:00-17:00 and urination is at 8:00, 18:00, and 23:00.

References Supplemental Material

Austin, Shaun. 2007. “Biological Monitoring of TDI-Derived Amines in Polyurethane Foam Production.” *Occupational Medicine*. https://doi.org/10.1093/occmed/kqm085.

Bello, Anila, Yalong Xue, Rebecca Gore, Susan Woskie, and Dhimiter Bello. 2019. “Assessment and Control of Exposures to Polymeric Methylene Diphenyl Diisocyanate (PMDI) in Spray Polyurethane Foam Applicators.” *International Journal of Hygiene and Environmental Health* 222 (5): 804–15. https://doi.org/10.1016/J.IJHEH.2019.04.014.

Brorson, Torbjorn, Gunnar Skarping, and Carsten Sango. 1991. “Biological Monitoring of Isocyanates and Related Amines.” *International Archives of Occupational and Environmental Health*. https://doi.org/10.1007/bf00386374.

Budnik, Lygia T., Dennis Nowak, Rolf Merget, Catherine Lemiere, and Xaver Baur. 2011. “Elimination Kinetics of Diisocyanates after Specific Inhalative Challenges in Humans: Mass Spectrometry Analysis, as a Basis for Biomonitoring Strategies.” *Journal of Occupational Medicine and Toxicology*. https://doi.org/10.1186/1745-6673-6-9.

Geens, Tom, Sandra Dugardin, Ann Schockaert, Geert De Cooman, and Marc Van Sprundel. 2012. “Air Exposure Assessment of TDI and Biological Monitoring of TDA in Urine in Workers in Polyurethane Foam Industry.” *Occupational and Environmental Medicine* 69 (2): 93–98. https://doi.org/10.1136/OEM.2011.064840.

Gledhill, A., A. Wake, P. Hext, E. Leibold, and R. Shiotsuka. 2008. “Absorption, Distribution, Metabolism and Excretion of an Inhalation Dose of [14C] 4,4’-Methylenediphenyl Diisocyanate in the Male Rat.” *Http://Dx.Doi.Org/10.1080/00498250500057591* 35 (3): 273–92. https://doi.org/10.1080/00498250500057591.

Henriks-Eckerman, Maj Len, Erja A. Mäkelä, Juha Laitinen, Katriina Ylinen, Katri Suuronen, Aki Vuokko, and Riitta Sauni. 2015. “Role of Dermal Exposure in Systemic Intake of Methylenediphenyl Diisocyanate (MDI) among Construction and Boat Building Workers.” *Toxicology Letters* 232 (3): 595–600. https://doi.org/10.1016/J.TOXLET.2014.12.012.

Hettick, J, and P.D Siegel. 2011. “Determination of the Toluene Diisocyanate Binding Sites on Human Serum Albumin by Tandem Mass Spectrometry.” *Analytical Biochemistry* 414.2: 232–38.

Kääriä, K, A Hirvonen, H Norppa, P Piirilä, H Vainio, and C Rosenberg. 2001. “Exposure to 2,4- and 2,6-Toluene Diisocyanate (TDI) during Production of Flexible Foam: Determination of Airborne TDI and Urinary 2,4- and 2,6-Toluenediamine (TDA).” *Analyst* 126 (7): 1.

Maître, A., M. Berode, A. Perdrix, S. Romazini, and H. Savolainen. 1993. “Biological Monitoring of Occupational Exposure to Toluene Diisocyanate.” *International Archives of Occupational and Environmental Health 1993 65:2* 65 (2): 97–100. https://doi.org/10.1007/BF00405726.

Sennbro, C. J., C. H. Lindh, C. Mattsson, B. A.G. Jönsson, and H. Tinnerberg. 2006. “Biological Monitoring of Exposure to 1,5-Naphthalene Diisocyanate and 4,4′-Methylenediphenyl Diisocyanate.” *International Archives of Occupational and Environmental Health* 79 (8): 647–53. https://doi.org/10.1007/S00420-006-0096-5/FIGURES/2.

Sennbro, Carl J., Christian H. Lindh, Anders Östin, Hans Welinder, Bo A.G. Jönsson, and Håkan Tinnerberg. 2004. “A Survey of Airborne Isocyanate Exposure in 13 Swedish Polyurethane Industries.” *The Annals of Occupational Hygiene* 48 (5): 405–14.

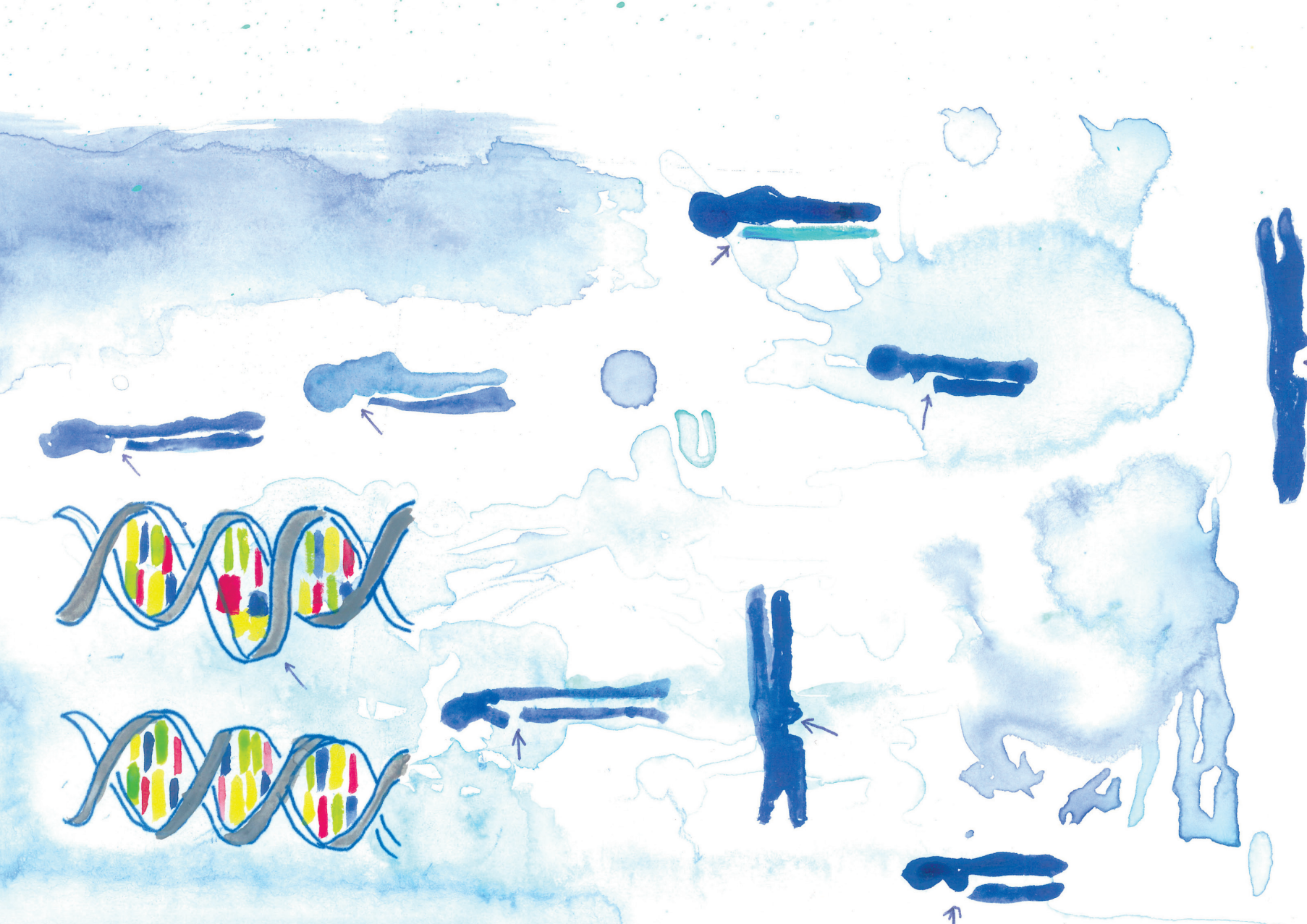
https://doi.org/10.1093/ANNHYG/MEH034.

Skarping, Gunnar, Torhjörn Brorson, and Carsten Sangö. 1991. “Biological Monitoring of Isocyanates and Related Amines.” *International Archives of Occupational and Environmental Health 1991 63:2* 63 (2): 83–88. https://doi.org/10.1007/BF00379069.

Swierczyńska-Machura, Dominika, Sławomir Brzeźnicki, Ewa Nowakowska-Swirta, and Jolanta Walusiak-Skorupa. 2015. “Occupational Exposure to Diisocyanates in Polyurethane Foam Factory Workers Good Practice in Occupational Health Services: Recommendations for Prophylactic Examinations in Patients with Diabetes Mellitus View Project Prevention of Hypertension and Diabetes in Occupational Prophylactic Care View Project.” *Article in International Journal of Occupational Medicine and Environmental Health*. https://doi.org/10.13075/ijomeh.1896.00284.

Timchalk, C., F. A. Smith, and M. J. Bartels. 1994. “Route-Dependent Comparative Metabolism of [14C]Toluene 2,4-Diisocyanate and [14C]Toluene 2,4-Diamine in Fischer 344 Rats.” *Toxicology and Applied Pharmacology*. https://doi.org/10.1006/taap.1994.1022.

Tinnerberg, Håkan, Karin Broberg, Christian H. Lindh, and Bo A.G. Jönsson. 2014. “Biomarkers of Exposure in Monday Morning Urine Samples as a Long-Term Measure of Exposure to Aromatic Diisocyanates.” *International Archives of Occupational and Environmental Health* 87 (4): 365–72. https://doi.org/10.1007/S00420-013-0872-Y/FIGURES/3.



Chapter 5

A Quantitative Meta-Analysis of the Relation between Occupational Benzene Exposure and Biomarkers of Cytogenetic Damage

Bernice Scholten, Jelle Vlaanderen, Rob Stierum, Lützen Portengen, Nat Rothman, Qing Lan, Anjoeka Pronk, Roel Vermeulen

Published in: *Environmental Health Perspectives,* 2020, 128 (8)

Abstract

Background

The genotoxicity of benzene has been investigated in dozens of biomonitoring studies, mainly by studying (classical) chromosomal aberrations (CA) or micronuclei (MN) as markers of DNA damage. Both have been shown to be predictive of future cancer risk in cohort studies and could, therefore, potentially be used for risk assessment of genotoxicity-mediated cancers.

Objectives

We sought to estimate an exposure-response curve (ERC) and quantify between-study heterogeneity using all available quantitative evidence on the cytogenetic effects of benzene exposure on CAs and MN respectively.

Methods

We carried out a systematic literature review and summarized all available data of sufficient quality using meta-analyses. We assessed the heterogeneity in slope estimates between studies and conducted additional sensitivity analyses to assess how various study characteristics impacted the estimated ERC.

Results

Sixteen CA (1,356 individuals) and 13 MN studies (2,097 individuals) were found to be eligible for inclusion in a meta-analysis. Studies where benzene was the primary genotoxic exposure and that had adequate assessment of both exposure and outcomes were used for the primary analysis. Estimated slope estimates were an increase of 0.27% CA [(95%CI: 0.08%, 0.47%); based on the results from 4 studies] and 0.27% MN [(95%CI: -0.23%, 0.76%); based on the results from 7 studies] per parts-per-million benzene exposure. We observed considerable between-study heterogeneity for both endpoints (I^2>90%).

Discussion

Our study provides a systematic, transparent, and quantitative summary of the literature describing the strong association between benzene exposure and accepted markers of genotoxicity in humans. The derived consensus slope can be used as a best estimate of the quantitative relationship between real-life benzene exposure and genetic damage in future risk assessment. We also quantitate the large between-study heterogeneity that exists in this literature, a factor which is crucial for the interpretation of single-study or consensus slopes.

Introduction

Benzene is a well-known environmental contaminant that was classified as a human carcinogen (IARC group 1) in 1979 (IARC 1979), which was recently reconfirmed (IARC 2018). Benzene and its metabolites may cause leukemia via genotoxic effects on the pluripotent hematopoietic stem cells, resulting in chromosomal changes (IARC 2009). A substantial number of cross-sectional studies have demonstrated that benzene exposure is associated with the occurrence of (classical) chromosomal aberrations (CA) and micronuclei (MN) that are considered early markers of genotoxicity (see reviews by Zhang, Eastmond, and Smith 2002; Mchale, Zhang, and Smith 2012). CA is often defined as the appearance of missing, extra, or irregular portions of chromosomal DNA (NHGRI 2016), while a micronucleus is considered to be the small nucleus that forms whenever a chromosome fragment or whole chromosome is not incorporated into the daughter nuclei after cell division (Fenech 2002). Information on intermediate endpoints is increasingly used in the risk assessment for many chemicals; e.g. IARC integrated mechanistic evidence with evidence from other data streams to support conclusions regarding carcinogenicity (Smith et al. 2016), and recently the Dutch Health Council Committee recommended a health-based occupational exposure limit for benzene that was (primarily) based on hematological effects in humans (Health Council of the Netherlands 2014).

To date, however, there have not been many attempts to assess in a systematic and quantitative way how exposure affects some of these intermediate endpoints, such as the induction of genetic damage. As an example, Angelini et al. (2016) recently reviewed the published data on benzene exposure and MN, but the authors restricted their findings to a non-quantitative comparison between exposed and unexposed individuals.

We hypothesize that it would be useful to investigate the exposure response curve (ERC) of benzene-induced CA or MN. An advantage of using biomarkers of effect for a (more) quantitative risk assessment is that they may offer more precise information on the exact shape of the exposure-response relation at low exposures, especially when the clinical outcomes of interest are relatively rare at low exposure levels. For benzene this is important because there is evidence that the exposure-response relation with cancers might be non-linear (Kim et al. 2006; Rothman et al. 1998; Vlaanderen et al. 2009). CAs and MN are additionally relevant to study because both have been found to be predictive of an increased risk of cancer in large prospective cohort studies (Bonassi et al. 2008; 2007).

We aimed to summarize all currently available human observational data and use this information to quantitatively describe the relation between occupational exposure to benzene and CA and MN frequencies. This would be a first step in the use of pre-clinical cancer-predictive endpoints in risk assessment. We also assessed the

between-study heterogeneity effect estimates and conducted sensitivity analyses to assess how various study characteristics affect the estimated ERC. This information is important when evidence from this literature (either from a single study or from a meta-analysis) is interpreted in a risk assessment context.

Methods

Study identification and selection

A literature search was conducted until August 2018 in PubMed and Scopus, using the following search terms for “benzene” AND chromosomal aberrations: (“chromosomal aberrations” OR “chromosome*”), and the following search terms for “benzene” AND micronuclei: (“micronuclei” OR “micronucleus”). References in all identified publications were checked for additional relevant studies. Publications were selected for inclusion in the meta-analysis if they had been peer reviewed and when they fulfilled the following criteria: 1) the study population had been occupationally exposed to benzene 2) exposure assessment was based on quantitative benzene exposure estimates in air or relevant benzene biomarker measurements; and 3) results from cytogenetic tests were available and shown per exposure category. Studies that were excluded were those that measured general air pollution (e.g. occupational exposure to traffic air), studies with considerable co-exposure to other carcinogenic substances (IARC classification 1 or 2a), studies that reported benzene exposure at levels that reasonably could not be distinguished from general background levels (~5 ppb; IARC 2018). Case-control studies were also excluded from further analysis. If more than one publication was published on the same cohort, the most recent update was included.

We decided to restrict our analyses to classical CA assessments and excluded fluorescent *in situ* hybridization (FISH) CA studies. FISH technology is based on fluorescent DNA probes to allow for the detection of aberrations in specific chromosomes (e.g. an alteration in the number of specific chromosomes and/or loss of particular chromosomal regions). FISH has been reported to be more sensitive to detect the genotoxic effects of benzene, compared to classical CA analysis (Smith *et al*. 1998). Eleven publications (of which about half were from a single study) have employed FISH to report changes in specific chromosomes in relation to quantitatively assessed benzene exposure. However, the large heterogeneity in the various FISH protocols used (e.g. in terms of detecting specific chromosomes and/or measuring specific changes ranging from alterations in the number of specific chromosomes, acquisition of specific translocations or loss of particular regions of certain chromosomes, and assessments in different stages of the cell cycle (interphase vs. metaphase) prohibited the inclusion of these studies into this meta-analysis.

Study evaluation

Each study to be included in the meta-analysis was evaluated for overall study quality and for the (expected) quality of the exposure and cytogenetic assessments. For both endpoints we selected a number of criteria.

Exposure assessment

A tiered approach was used to evaluate exposure assessment quality (Table S1). First, we assessed whether benzene exposure was the main exposure or if other co-exposures may have existed. Studies designed primarily to evaluate the effects of benzene exposure, i.e. in which benzene exposure was the main exposure were classified as A. These were mostly studies in the shoe-making or petroleum industries. Studies that were classified as B were studies in jobs where significant co-exposure to other genotoxic (air) pollutants was not reported, but could not be ruled out, e.g. in filling station attendants (FSA). Studies that were classified A were further divided based on the quality of their exposure assessment, with A+ studies assessed exposure with the specific aim of exposure assessment for the respective study with at least some (personal or stationary) benzene exposure measurements and/or data on benzene metabolites that had been measured in blood or urine. All other A studies were classified as A.

Cytogenic assessment

The ranking of CA and MN assessments focused on the quality of the cytogenetic analyses. Studies that were classified A met the following requirements: for CA the number of metaphases counted should be above 100 and the culture time below 50 hours (Carrano and Natarajan 1988). For MN, the number of bi-nucleated cells scored should be above 1000, and cells had to be incubated initially for 44 hours, followed by the addition of Cytochalasin B, and subsequently another 24-48 hours of incubation (Fenech 2007). Classification as an A class study further required that both the frequency and variability (SE or SD) of CA or MN were reported. All studies not meeting these criteria were classified as B (Table S1).

Primary study set

Studies that scored an A+ for exposure assessment and an A for cytogenic assessment were included in the primary analyses. The other studies were included in sensitivity analyses (the “full study set”). Evaluation of the quality of eligible studies was evaluated jointly by four reviewers for exposure assessment (R.V., J.V., L.P., B.S.) and three reviewers for cytogenetic analysis (J.V., R.S., B.S.).

Extraction and preparation of data from selected papers

The following information was extracted from each relevant publication that was used: the country where the study was performed, the method used for exposure

assessment and cytogenetic analyses, including culture time and the types of aberrations included, and the availability of biomonitoring data. For each exposure group within a study we also extracted: the number of subjects, benzene exposure level, the recorded CA and/or MN frequencies with either the standard error (SE), standard deviation (SD), or range of frequencies. We also extracted information on benzene exposure levels and CA or MN frequencies separately by smoking status (yes/no) when this information was available.

Several studies lacked quantitative estimates of benzene exposure, but did provide information on blood or urine levels of benzene or its metabolites. For those studies, we estimated benzene exposure levels through reverse physiologically based (PBK) modelling based on the model by Knutsen *et al*. (2013) . When the publication provided no benzene exposure levels for controls, we assumed a background level of 5 ppb.

Frequency of CA and MN were expressed as the number of cells with aberrations per 100 cells (%).Reported CA or MN in other units were calculated back to 100 cells. For studies that did not report the SE for CA or MN frequencies, we calculated the SE based on the reported SD and sample size. For the studies that did not report either the SE or SD (and, hence, were not included in the primary study set), we estimated the SE under the assumption that the SD equaled the median SD for the primary study set (n= 5 for CA).

Statistical analysis

The main goal of this paper was to summarize the available information on the exposure-response relation between benzene and markers of cell damage. Because most of the individual studies report aberration frequencies for only two groups (one of which is either unexposed or low-exposed), there is very little opportunity to investigate non-linear exposure response relations using this data, unless one is willing to make strong assumptions. We therefore decided to perform a meta-analysis on summary estimates of the available exposure-response information within each study (i.e., the study-specific slope), rather than attempt to build a (hierarchical) meta-regression model (for the group-level data).

The information available consisted, for most studies, of two average exposure estimates (denoted as X0 and X1) and 2 estimates of the average aberration frequencies (denoted as Y0 and Y1) together with an estimate of the precision of these latter estimates (denoted as variance V0 and V1). Although the aberration frequency as a proportion could theoretically be modelled as a binomial outcome for an individual, this was not the case for the reported average aberration frequencies, and would require quasi-likelihood methods to account for overdispersion even if the individual level data had been available. We therefore regard the average aberration frequencies Y0 and Y1 as random variables with known variances. Standard algebra provides the rules to calculate the expected value and variance of the random variable

that results from subtraction of Y0 from Y1 and dividing it by the difference between X0 and X1 (i.e. the slope factor), without reference to any specific (e.g. normal) distribution. The expected value of this random variable equals (Y1-Y0)/(X1-X0), with variance (V0 + V1)/(X1 - X0)2.

For example, for Bogadi (1997) the reported average proportion plus or minus the standard error (SE) of chromosomal aberrations was 1 ± 0.12% in the unexposed and 1.5 ± 0.11% in the exposed (with an average exposure of 5.9 ppm). The estimated slope (β) is then (1.5-1)/5.9 = 0.08 (units %/ppm) with an estimated variance of (0.122 + 0.112)/(5.92) = 0.0007. This slope factor and associated variance was then used in the meta-analysis.

Meta-analyses were performed using the metafor package (Viechtbauer 2019) in R (version 3.6.1; R Development Core Team). Meta-analysis using this approach failed for the full MN study set as the range of sampling variances was extremely large, making the results of the algorithm used numerically unstable. We therefore chose to analyze the MN primary study set only. We calculated the I^2 statistic to assess the heterogeneity in slope estimates between studies. Funnel plots, the trim and fill method and the Egger test were used to investigate potential publication bias.

Several sensitivity analyses were performed to explore the potential bias/variance trade-off of including all studies (i.e. the full study set) in the analysis, and to evaluate the impact of individual studies on the estimated meta-slope, or consensus ERC. In the first sensitivity analysis we allowed the full set of studies to contribute to the estimation of the consensus ERC. The second (set of) sensitivity analyses was performed by leaving out one study at a time from the sample set (i.e. jackknifing) for both the primary- and full study set.

To evaluate potential effect modification by smoking, we restricted the analyses to data from studies that reported exposure and genotoxic effects by smoking status, and evaluated the interaction between smoking and benzene exposure on the frequency of CA and/or MN by including smoking as a moderator variable.

Results

CA

Of the 745 identified studies on CA and benzene exposure, 16 (1,356 individuals) were considered eligible for inclusion in the meta-analysis. An overview of all eligible studies and a flow-chart indicating how studies were selected is provided in the appendix (Figure S1)). Four studies (477 individuals) were included in the primary study set. Of the 16 selected studies, the vast majority (n=15), across a range of occupations, showed higher CA frequencies in jobs involving benzene exposure when compared to unexposed jobs (see Table S2). Only the benzene exposed fuel-tanker-drivers in

Lovreglio *et al*. (2014) had lower CA frequencies in comparison with their control group.

The individual ERCs for all 16 studies (full study set) are presented in Figure 1, which clearly shows the large variation across studies in baseline CA frequencies and the derived benzene ERCs.

A meta-analysis of the slopes on the primary study set confirms that benzene exposure is positively and significantly associated with an increase in CA [β = 0.27%, 95% CI: 0.08%, 0.47%) but with an I^2 of 94% indicating very strong heterogeneity between studies (Figure 2). Expanding the analysis to include the full study set resulted in a similar positive but heterogenous association [β = 0.29% (95% CI: 0.16 – 0.42), I^2: 91.1%].

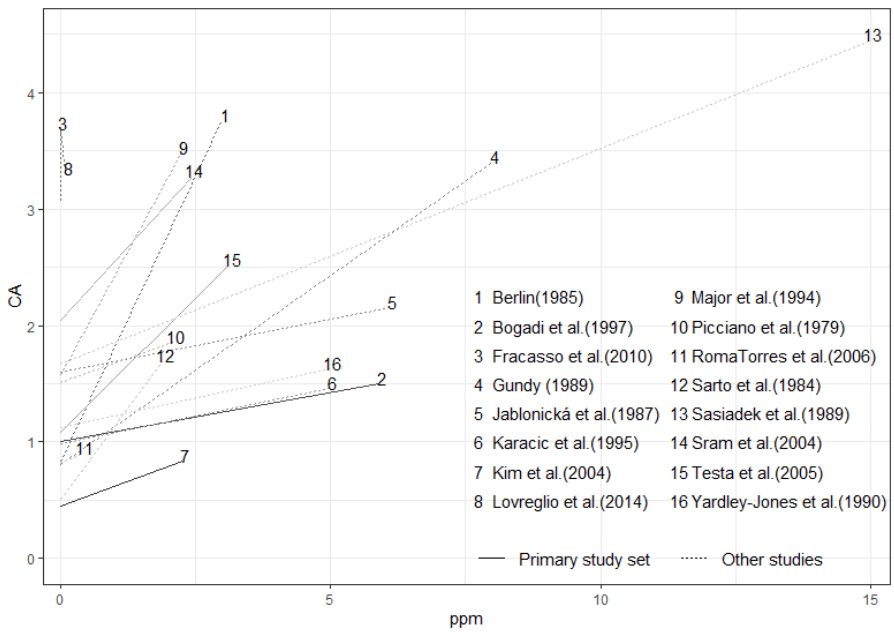


Figure 1. Chromosome aberrations (number of aberrations per 100 cells) and benzene exposure (ppm). The primary study set contained studies where benzene was the primary genotoxic exposure and that had adequate assessment of both exposure and outcomes. The term other studies refers to studies not meeting these criteria. The numbers shown with each line indicate the study name. Note: CA, chromosomal aberration.

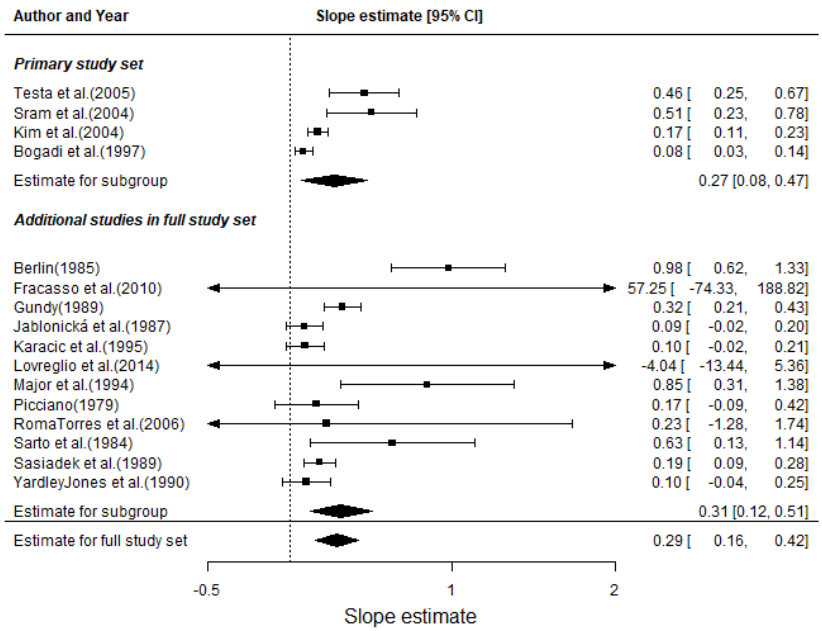


Figure 2. Meta-analysis of slope estimates (in units%/ppm) of benzene induced chromosomal aberrations. The primary study set contained studies where benzene was the primary genotoxic exposure and that had adequate assessment of both exposure and outcomes. The term other studies refers to studies not meeting these criteria. The I2 of the primary study set was: 94%, the I2 of the full study set was 91.1%. Note: CI, confidence interval.

Sensitivity analyses

To assess the sensitivity of the outcome of the meta-analysis, one study at the time was excluded from both the primary and the full study set using a jackknife approach. The results are presented in the appendix (Table S3). These analyses showed that excluding a single study had a considerable impact on the point estimate of the meta-slope (range 0.22 - 0.35) when limiting analysis to the primary study set. When considering all studies, Berlin *et al*. (1995) was found to be influential on the meta-slope estimate (i.e. average slope was 0.29, whereas excluding Berlin *et al*. (1995) would result in a slope estimate of 0.22, I^2: 78%).

Of the selected CA studies from the full study set, four (Jablonicka, Vargova, and Karelova 1987; Major *et al*. 1994; Fracasso *et al*. 2010; Sram *et al*. 2004), contained specific information on CA aberrations for smokers and non-smokers in the control and benzene-exposed group (Figure S2). There was no statistical interaction effect of smoking (p-value: 0.6).

There was evidence of publication bias for the CA studies (p = 0.02). We attempted to adjust the results for this bias using the trim-and-fill method (see Figure S3).

MN

Of the 315 MN studies that were found, 13 (2097 study subjects in total) were considered eligible for inclusion in the meta-analysis (Table S4). Nine of these studies (1672 study subjects) were included in the primary study set (Figure S4). The individual studies (Figure 3) indicate that most studies show higher MN frequencies in the exposed group (n=9), but the slopes vary considerably.

The meta-slope of the primary study set of MN studies is 0.27% (95% CI: -0.23%, 0.76%) with an I^2 of: 99.5% (Figure 4). We were unable to derive a consensus slope for the full dataset owing to the large range in sampling variances for some of the slope estimates.

Sensitivity analysis

When applying jackknifing on the primary study set, the range of slopes varied between 0.17 (when Basso *et al*. (2011) was excluded) to 1.03 (when Ren *et al*. (2018) was excluded) (see Table S5).

Three MN studies from the full study set (Bukvic *et al*. 1998; Basso *et al*. 2011; Sha *et al*. 2014) contained information on smoking status and corresponding MN frequencies in both the benzene exposed and unexposed group. There was no statistical interaction effect of smoking (p-value: 0.5). (Figure S5)

There was evidence of publication bias for the MN studies (p = 0.001). The trim-and-fill method pooled estimate was 0.24 with a corresponding SE of 0.98, whereas the meta slope of the primary study set was 0.27 ± 0.19. (see Figure S6)

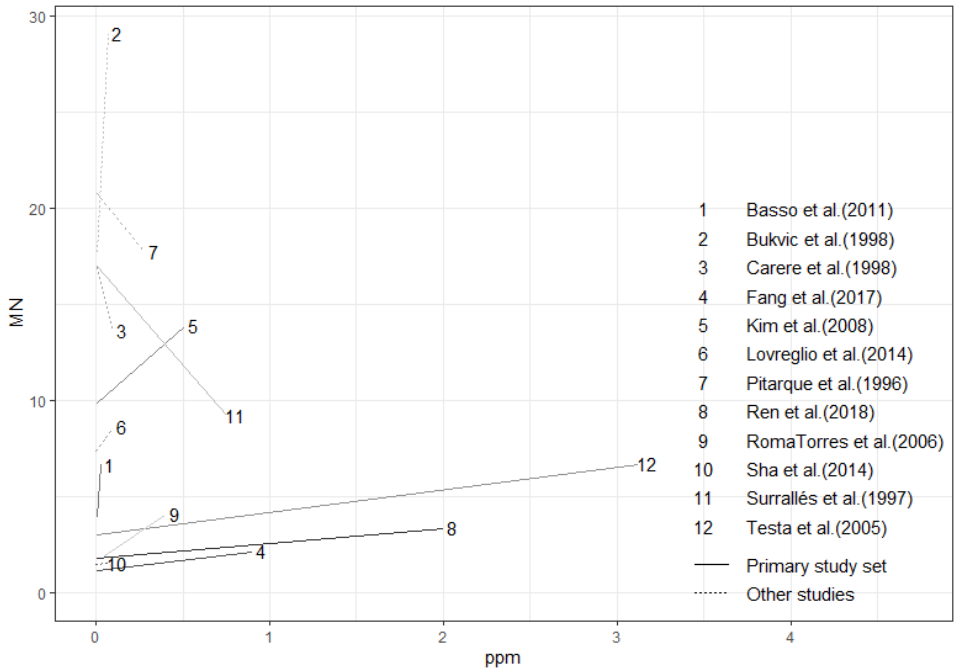


Figure 3. Micronuclei and benzene exposure. The primary study set contains studies where benzene was the primary genotoxic exposure and that had adequate assessment of both exposure and outcomes. The "other studies" refer to studies not meeting these criteria. The numbers shown with each line indicate the study name. The study by Liu et al. (1996) is not displayed in this figure because it was the only study that measured very high exposure levels (up to 60 ppm).

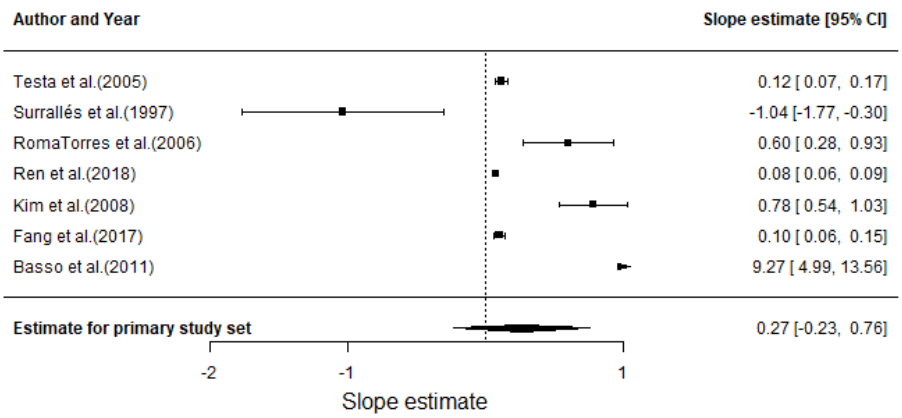


Figure 4. Meta-analysis of slope estimates (in units %/ppm) of benzene induced micronuclei based on primary study set.
The primary study set contains studies where benzene was the primary genotoxic exposure and that had adequate assessment of both exposure and outcomes. The I2 of the study set was: 99.5%.

Discussion

A substantial number of occupational studies has been undertaken in the past decades to study the effect of benzene on genotoxicity, covering a range of occupations and entailing various co-exposures and different levels of benzene exposure. Although in the vast majority CA and MN frequencies were associated with jobs involving benzene exposure (IARC 2018), a consensus quantitative relationship between exposure levels and CA or MN frequency is not yet defined.

To quantitatively summarize data on occupational benzene exposure and markers of genetic damage, we conducted a systematic review, followed by a meta-analysis and a set of sensitivity analyses. We found a positive slope for both CA [β = 0.27% (95% CI: 0.08%, 0.47%)] and MN [β = 0.27% (95% CI: -0.23%, 0.76%)] in relation to benzene exposure for the high quality studies (primary study set), but the between-study heterogeneity in slope estimates was large with an estimated I^2 of over 90% for both endpoints.

Including also studies with less stringent quality criteria (full study set) the slope was comparable for CA [β =0.29% for all studies (95% CI: 0.16, 0.42, I^2: 91.1%] in comparison to the primary study set. We were unable to derive a meta-slope for the full dataset for MN due to the large range in variances for the study-specific slopes.

There are several possible explanations for the observed differences in individual slopes, including the small number of individuals per study (on average about 40), the limited dose groups (in most cases there was only one exposure estimate for the exposed group), differences in exposure assessment methods, and uncertainty in both exposure levels and CA or MN counts. This latter issue is reflected by the results of the HUman MicroNucleus project, which was established early 2000 to gather MN data from 25 labs representative of many countries and populations (Bonassi *et al*. 2001). An overall median MN frequency in non-exposed subjects of 6.5‰ was found with an interquartile range between 3 and 12‰. The authors conclude that most of the observed total variance could be explained by laboratory methods. The issue of uncertainty in measuring DNA damage rates might be even more critical for classical CA since no harmonized protocol has been developed for this biomarker, and CA are known to be more difficult to score in comparison to MN (Fenech 2002). Indeed we found that analyzing and comparing all CA studies included in this meta-analysis was difficult because the interpretation of “(total) chromosomal aberrations” varied amongst studies. A range of structural aberrations were included by some studies, whereas others reported only a specific type of aberration, thereby making comparisons more difficult. Still, all authors aimed to collect the same information (i.e., total classical CAs), whereas this was not the case for the FISH studies, where different endpoints in different chromosomes were selected in different division stages of the cells (e.g., interphase or metaphase). Hence, we decided to exclude studies

based on FISH technology because they could not be sensibly combined. However, potentially, the observed analytical variance in CA measurements could be less for FISH-based studies because scoring is easier and can be automated. In addition, a chromosome-wide aneuploidy study described heterogeneity in monosomy and trisomy rates for specific chromosomes. (Zhang et al. 2011). Focusing on the more sensitive chromosomes may then provide a more accurate estimate of benzene's genotoxic effect which maybe be partly masked in CA and MN assays where all chromosomes are considered.

Exposure assessments applied in the different studies were not equivalent. For example, some studies described the measurement of personal exposure in the breathing zone of workers, whilst others estimated exposure based on stationary samples. Although we did take uncertainty in measured CA or MN frequencies into account in this meta-analysis, this was not possible for the exposure estimates because information on the exposure assessment error was often missing (as we see commonly in meta-analyses). This could potentially add to the observed heterogeneity between studies. Note that it has been reported that I^2 has a substantial bias when the number of studies is small; as an alternative confidence intervals could replace the I^2 estimates (Von Hippel 2015). We provide both estimates in this paper.

The studies in the meta-analysis were weighed on the corresponding inverse variance. We did so because we assumed the SD described the variability of counted CA or MN between individuals; this was confirmed by a number of authors of the studies used for meta-analysis in this paper (R. Sram and P. Lovreglio, personal communication). Due to the limited number of studies, the limited number of estimates per study, and the high heterogeneity between studies, we refrained from explorations on nonlinearity of the ERC. However, some evidence is available that the deleterious effects of benzene might, in fact, be nonlinear (Kim et al. 2006; Rothman et al. 1998).

We also investigated the potential interaction effect between smoking and benzene exposure. The results were not significant for either CA or MN; however, the number of studies available with information on CA or MN frequencies specifically for smokers and non-smokers was limited. There was some evidence of publication bias, most notably for the MN studies where the trim-and-fill method indicated a (relatively small) change in pooled estimate but a very large increase in the corresponding SE (i.e., the slope estimate decreased from 0.27 to 0.24, and the corresponding standard error became much larger (from 0.19 to 0.98). This effect is mainly driven by the study of Basso et al. (2011), where the observed outcome is much larger for the reported standard error in comparison to other studies (Figure S6). Publication bias thus influenced the MN slope estimate, adding to the uncertainty of the meta slope. Evidence for publication bias for MN was also found by Angelini et al. (2016). Finally, in the past years, several studies explored possibilities beyond measuring classical CA or MN frequencies, that is, studying other biomarkers [e.g., telomere length assays

(Bassig *et al*. 2014)], or other cell types [e.g., cells closer to acute myeloid leukemia/ myelodysplastic syndrome-related stem cells (Zhang *et al*. 2012)]. Potentially, these studies can also be used to derive ERCs but we chose to focus on CA and MN assays because the evidence base for these studies is larger and their association to benzene exposure is widely accepted.

We quantitively estimated the relationship between benzene exposure and induced genetic damage to inform future risk assessments; for example, because both MN and CA have been found to be predictive of an increase in future cancer risk (although not for hematopoietic malignancies specific, given the lack of power to investigate these associations), our derived ERC might be used to inform the shape of the benzene-leukemia curve in exposure ranges where epidemiological data is limited. One strategy to include our ERC in a benzene-leukemia risk assessment would involve a Bayesian approach that uses information on the meta-slope of CA and MN as prior in estimating the benzene leukemia ERC. We believe the inclusion of biomarkers could potentially enhance risk assessment because these studies can be informative for early biological effects (which might not be included in cohort studies on, e.g., cancer occurrence) or relevant biological pathways, and in addition they can inform the ERC at lower exposure levels because biomonitoring studies have generally more power than outcome studies at lower exposure levels.

In conclusion, we summarized all available data on CA and MN frequencies in benzene exposed workers, and provided a quantitative estimate of the exposure response relationship. As far as the authors are aware, no other study applied the quantitative approach used in this paper for other chemicals. We believe that further insight into the quantitative relationship between real-life benzene exposure and genetic damage could provide a starting point for further inclusion of this data in benzene risk assessments; however, the considerable observed heterogeneity between studies should be carefully considered.

References Chapter 5

Angelini, Sabrina, Justo Lorenzo Bermejo, Gloria Ravegnini, Giulia Sammarini, and Patrizia Hrelia. 2016. “Application of the Lymphocyte Cytokinesis-Block Micronucleus Assay to Populations Exposed to Petroleum and Its Derivatives: Results from a Systematic Review and Meta-Analysis.” *Mutation Research - Reviews in Mutation Research*. https://doi.org/10.1016/j.mrrev.2016.03.001.

Bassig, Bryan A., Luoping Zhang, Richard M. Cawthon, Martyn T. Smith, Songnian Yin, Guilan Li, Wei Hu, *et al*. 2014. “Alterations in Leukocyte Telomere Length in Workers Occupationally Exposed to Benzene.” *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.21880.

Basso, E., C Cevoli, M Papacchini, G Tranfo, A Mansi, and A Testa. 2011. “Cytogenetic Biomonitoring on a Group of Petroleum Refinery Workers.” *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.20641.

Bonassi, Stefano, Michael Fenech, Cecilia Lando, Yi ping Lin, Marcello Ceppi, Wushou Peter Chang, Nina Holland, *et al*. 2001. “Human Micronucleus Project: International Database Comparison for Results with the Cytokinesis-Block Micronucleus Assay in Human Lymphocytes: I. Effect of Laboratory Protocol, Scoring Criteria, and Host Factors on the Frequency of Micronuclei.” *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/1098-2280(2001)37:1<31::AID-EM1004>3.0.CO;2-P.

Bonassi, Stefano, Hannu Norppa, Marcello Ceppi, Ulf Strömberg, Roel Vermeulen, Ariana Znaor, Antonina Cebulska-Wasilewska, *et al*. 2008. “Chromosomal Aberration Frequency in Lymphocytes Predicts the Risk of Cancer: Results from a Pooled Cohort Study of 22 358 Subjects in 11 Countries.” *Carcinogenesis*. https://doi.org/10.1093/carcin/bgn075.

Bonassi, Stefano, Ariana Znaor, Marcello Ceppi, Cecilia Lando, Wushou Peter Chang, Nina Holland, Micheline Kirsch-Volders, *et al*. 2007. “An Increased Micronucleus Frequency in Peripheral Blood Lymphocytes Predicts the Risk of Cancer in Humans.” *Carcinogenesis* 28 (3): 625–31. https://doi.org/10.1093/carcin/bgl177.

Bukvic, Nenad, Pietro Bavaro, Giuseppe Elia, Filippo Cassano, Margherita Fanelli, and Ginevra Guanti. 1998. “Sister Chromatid Exchange (SCE) and Micronucleus (MN) Frequencies in Lymphocytes of Gasoline Station Attendants.” *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. https://doi.org/10.1016/S1383-5718(98)00045-X.

Carrano, A. V., and A. T. Natarajan. 1988. “Considerations for Population Monitoring Using Cytogenetic Techniques.” *Mutation Research/Genetic Toxicology*. https://doi.org/10.1016/0165-1218(88)90036-5.

Fenech, Michael. 2002. “Biomarkers of Genetic Damage for Cancer Epidemiology.” *Toxicology*. https://doi.org/10.1016/S0300-483X(02)00480-8.

———. 2007. “Cytokinesis-Block Micronucleus Cytome Assay.” *Nature Protocols*. https://doi.org/10.1038/nprot.2007.77.

Fracasso, Maria Enrica, Denise Doria, Giovanni Battista Bartolucci, Mariella Carrieri, Piero Lovreglio, Andrea Ballini, Leonardo Soleo, Giovanna Tranfo, and Maurizio Manno. 2010. “Low Air Levels of Benzene: Correlation between Biomarkers of Exposure and Genotoxic Effects.” *Toxicology Letters*. https://doi.org/10.1016/j.toxlet.2009.04.028.

Health Council of the Netherlands. 2014. “Benzene - Health-Based Recommended Occupational Exposure Limit.”

Hippel, Paul T. Von. 2015. “The Heterogeneity Statistic I2 Can Be Biased in Small Meta-Analyses.” *BMC Medical Research Methodology*. https://doi.org/10.1186/s12874-015-0024-z.

IARC. 1979. “Monographs on the Evaluation of Carcinogenic Risks of Chemicals to Humans.” IARC Monog.

———. 2009. “Monographs on the Evaluation of Carcinogenic Risks to Humans.” Vol. 100F.

———. 2018. “Benzene (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans).” Vol. 120.

Jablonicka, A, M Vargova, and J Karelova. 1987. “Cytogenetic Analysis of Peripheral Blood Lymphocytes in Workers Exposed to Benzene.” *Journal of Hygiene, Epidemiology, Microbiology, and Immunology*.

Kim, Sungkyoon, Roel Vermeulen, Suramya Waidyanatha, Brent A. Johnson, Qing Lan, Nathaniel Rothman, Martyn T. Smith, *et al*. 2006. “Using Urinary Biomarkers to Elucidate Dose-Related Patterns of Human Benzene Metabolism.” *Carcinogenesis*. https://doi.org/10.1093/carcin/bgi297.

Knutsen, Jeffrey S., Brent D. Kerger, Brent Finley, and Dennis J. Paustenbach. 2013. “A Calibrated Human PBPK Model for Benzene Inhalation with Urinary Bladder and Bone Marrow Compartments.” *Risk Analysis* 33 (7): 1237–51. https://doi.org/10.1111/j.1539-6924.2012.01927.x.

Lovreglio, Piero, Francesca Maffei, Mariella Carrieri, Maria N. D’Errico, Ignazio Drago, Patrizia Hrelia, Giovanni B. Bartolucci, and Leonardo Soleo. 2014. “Evaluation of Chromosome Aberration and Micronucleus Frequencies in Blood Lymphocytes of Workers Exposed to Low Concentrations of Benzene.” *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. https://doi.org/10.1016/j.mrgentox.2014.04.022.

Major, J., M. Jakab, Gabriella Kiss, and Anna Tompa. 1994. “Chromosome Aberration, Sister-chromatid Exchange, Proliferative Rate Index, and Serum Thiocyanate Concentration in Smokers Exposed to Low-dose Benzene.” *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.2850230211.

Mchale, C.M., Luoping Zhang, and M.T. Smith. 2012. “Current Understanding of the Mechanism of Benzene-Induced Leukemia in Humans: Implications for Risk Assessment.” *Carcinogenesis*. https://doi.org/10.1093/carcin/bgr297.

NHGRI. 2016. “Chromosome Abnormalities.” 2016.

Ren, Jing Chao, Yu Xia Wu, Zhenzhen Wu, Guang Hui Zhang, Hongjian Wang, Hongzhou Liu, Jun Peng Cui, *et al*. 2018. “MTHFR Gene Polymorphism Is Associated with DNA Hypomethylation and Genetic Damage among Benzene-Exposed Workers in Southeast China.” *Journal of Occupational and Environmental Medicine*. https://doi.org/10.1097/JOM.0000000000001288.

Rothman, N., W. E. Bechtold, S. N. Yin, M. Dosemeci, G. L. Li, Y. Z. Wang, W. C. Griffith, M. T. Smith, and R. B. Hayes. 1998. “Urinary Excretion of Phenol, Catechol, Hydroquinone, and Muconic Acid by Workers Occupationally Exposed to Benzene.” *Occupational and Environmental Medicine*. https://doi.org/10.1136/oem.55.10.705.

Sha, Yan, Wei Zhou, Zhenyu Yang, Xiaoling Zhu, Yingping Xiang, Tiandi Li, Dexiang Zhu, and Xinyue Yang. 2014. “Changes in Poly (ADP-Ribosyl)Ation Patterns in Workers Exposed to BTX.” *PLoS ONE*. https://doi.org/10.1371/journal.pone.0106146.

Smith, Martyn T., Kathryn Z. Guyton, Catherine F. Gibbons, Jason M. Fritz, Christopher J.

Portier, Ivan Rusyn, David M. DeMarini, *et al*. 2016. “Key Characteristics of Carcinogens as a Basis for Organizing Data on Mechanisms of Carcinogenesis.” *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.1509912.

Smith, Martyn T., Luoping Zhang, Yunxia Wang, Richard B. Hayes, Guilan Li, Joseph Wiemels, Mustafa Dosemeci, *et al*. 1998. “Increased Translocations and Aneusomy in Chromosomes 8 and 21 among Workers Exposed to Benzene.” *Cancer Research*.

Sram, Radim J., Olena Beskid, Blanka Binkova, Pavel Rossner, and Zdenek Smerhovsky. 2004. “Cytogenetic Analysis Using Fluorescence in Situ Hybridization (FISH) to Evaluate Occupational Exposure to Carcinogens.” In *Toxicology Letters*. https://doi.org/10.1016/j.toxlet.2003.12.043.

Viechtbauer, W. 2019. “Meta-Analysis Package for R.” https://cran.r-project.org/web/packages/metafor/metafor.pdf.

Vlaanderen, Jelle, Nathaniel Rothman, Lützen Portengen, Qing Lan, Hans Kromhout, and Roel Vermeulen. 2009. “Flexible Meta-Regression to Assess the Shape of the Benzene–Leukemia Exposure–Response Curve.” *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.0901127.

Zhang, L., Q. Lan, Z. Ji, G. Li, M. Shen, R. Vermeulen, W. Guo, *et al*. 2012. “Leukemia-Related Chromosomal Loss Detected in Hematopoietic Progenitor Cells of Benzene-Exposed Workers.” *Leukemia*. https://doi.org/10.1038/leu.2012.143.

Zhang, Luoping, David A. Eastmond, and Martyn T. Smith. 2002. “The Nature of Chromosomal Aberrations Detected in Humans Exposed to Benzene.” *Critical Reviews in Toxicology*. https://doi.org/10.1080/20024091064165.

Zhang, Luoping, Qing Lan, Weihong Guo, Alan E. Hubbard, Guilan Li, Stephen M. Rappaport, Cliona M. McHale, *et al*. 2011. “Chromosome-Wide Aneuploidy Study (CWAS) in Workers Exposed to an Established Leukemogen, Benzene.” *Carcinogenesis*. https://doi.org/10.1093/carcin/bgq286.

Supplemental Material - Scoring criteria

Table S1. Scoring criteria exposure and cytogenic assessment for CA and MN studies Note that only for class A "Exposure" a distinction between A+ and A+ is made. Studies that scored an A+ for exposure assessment and an A for cytogenic assessment were included in the primary analyses. C studies were excluded from further analyses. All other studies were included in the full study set.

Class	Sub class	Exposure	Cytogenic assessment
A	A+	Benzene main source of exposure: Personal exposure measurements or multiple air samples, biomonitoring studies	>100 metaphases (CA), or > 1000 binucleated cells (MN) scored Culture time < 50 hours (CA), cells incubated for 44 hrs and subsequently another 24-28 hrs after addition of Cyt B (MN). Reporting SD or range
	A-	Not fulfilling criteria of A+	
B		Co-exposure to other genotoxic substances	Not fulfilling criteria of A
C		Studies that scored a B on both exposure and cytogenic assessment were ranked C and excluded from further analyses.	

Supplemental Material - CA studies

Table S2. Evaluation CA studies for inclusion in meta-analysis.
1: mean or otherwise median exposure. If instead of an average a range is given the midpoint is calculated, and the range indicated in brackets, 2: Estimated using PBK modelling, 3: Information could not be ascertained. FSA: filling station attendant. n.a.: (data) not available

Job category	Average CA (exposed/ control)%	Exposure controls ppm	Exposure exposed ppm [1]	Exposure method	Sample size exposed	Exposure group	Exposure assessment scoring	CA assessment scoring	Overall scoring group	Reference
Shoe-making industry	1.5/1	n.a.	5.9	Air sampling (Air sampling at 10 stationary sampling locations, measured in one work shift. Pumps situated on the work tables at nose height. For each employee, benzene exposure was ascertained according to the position of each workplace within a radius of 3 m around the sampling pump).	49	A	+	A	A+A	(Bogadi-Šare *et al.* 1997)
FSA	3.69/3.07	0.002	0.01	Personal sampling (measured during work shift by personal diffusive samples, worn at the breathing zone level)	19	B (FSA)	+	A	B+A	(Fracasso *et al*. 2010)
Petroleum-refining industry	0.84/ 0.448	n.a.	2.3 (0.004 - 4.52)	Job exposure matrix	178	A	+	A	A+A	(Kim *et al.* 2004)
Fuel-tanker drivers	3.3/3.7	0.001	0.1	Personal sampling (measured during work shift by personal diffusive samples, worn at the breathing zone level)	18	B (potential co-exposure to diesel exhaust)		A	BA	(Lovreglio *et al*. 2014)
FSA	4/3.7	0.001	0.007	Personal sampling (measured during work shift by personal diffusive samples, worn at the breathing zone level)	24	C (exposure below/around background-concentrations)	+	A	C	(Lovreglio *et al*. 2014)
Oil-refining industry	3.48/1.57	n.a.	2.3	Air sampling (Benzene concentrations of ambient air samples at working places were regularly measured by the safety department	42	A	- (because it appears that measurements were not	A	A-A	(Major *et al.* 1994)

Job category	Average CA (exposed/ control)%	Exposure controls ppm	Exposure exposed ppm [1]	Exposure method	Sample size exposed	Exposure group	Exposure assessment scoring	CA assessment scoring	Overall scoring group	Reference
				according to the NIOH (Hungary) analytic protocol for area monitoring, and according to MSZ for personal monitoring)			collected with specific aim of exposure assessment for this study)			
Benzene production	1.7/0.5	n.a.	3.6 (1.9-5.3)	Personal sampling	22	A	+	B (80 – 100 cells scored)	A+B	(Sarto *et al.* 1984)
Chemical production	3.28/2.05	n.a.	2.4 (0.1-4.7)	Stationary monitoring	39	A	+	A	A+A	(Sram *et al.* 2004)
Auto-mobile painters	2.52/1.08	n.a.	3	Air sampling (repeated sampling in the workplaces)	25	A (authors sufficiently excluded potential exposure to other carcino-genic compounds)	+	A	A+A	(Testa *et al.* 2005)
Shipping, oil, aromatic production	1.63/ 1.12	n.a.	5 (<10)	Personal and workplace air sampling ("Exposures in this study were assumed to be similar to those of previous personal and workplace air samples collected over a period of years from individuals in similar jobs")	48	A	-(exposure assessment not conducted in same study population)	A	A-A	(Yardley-Jones *et al.* 1990)
Road tanker drivers	3.76 /0.83	n.a.	3 (1-5)	Personal sampling	10	B (co-exposure to diesel exhaust cannot be ruled out)	[3]	B - No SD or range	C	(Berlin 1985)
FSA	0.66/0.83	n.a.	1	Personal sampling	10	B (FSA)	[3]	B - No SD or range	C	(Berlin 1985)
FSA	1.43/0.66	0.4[2]	5[2]	Phenol in urine	30	B (FSA)	+	B - Culture time: 72 hrs	C	(Çelık and Akbaş 2005)
Chemical industry	2.75/ 0.81	n.a.	8	Personal sampling (personal exposure 8hr TWA)	10	A	+	B - No SD or range	A+B	(Gundy 1989)

Job category	Average CA (exposed/ control)%	Exposure controls ppm	Exposure exposed ppm [1]	Exposure method	Sample size exposed	Exposure group	Exposure assessment scoring	CA assessment scoring	Overall scoring group	Reference
Ethyl-benzene production	2.2/1.6	n.a.	6.1 (0.5-11.7)	Stationary sampling	66	A	+	B - No SD or range, Culture time too long	A-B	(Jablonicka, Vargova, and Karelova 1987)
Shoe-workers	1.46/0.98	n.a.	5	Stationary sampling (Samples of air were collected at 11 stationary sampling locations. Pumps were fixed on the working tables (one to five workers were located around one table) at nose height, and the air was collected throughout the working day.)	45	A	+	B - No SD or range	A+B	(Karačić *et al.* 1995)
Chemical industry	3.46/3	n.a.	2.1	Multiple measurements (from both fixed stations and personnel air monitors) over a 4-yr period	52	A	+	B - No SD or range, aberrations not clearly described	A+B	(Picciano 1979)
FSA	3.48, 6.86, 8.03/ 2.39	0.04	0.38, 0.42, 0.45	Authors themselves conducted exposure reconstruction based on biomonitoring data (benzene, phenol, t,t-MA and SPMA in urine)	200	B (FSA)	+	B – no SD for CA (control)	C	(Rekhadevi *et al*. 2011)
Petroleum refinery aromatics plant	0.9/0.82	0.05[2]	0.4[2]	t,t-MA in urine	48	A	+	B - CA data only available with gaps	A+B	(Roma-Torres *et al.* 2006)
Several occupations	4.45/1.67	n.a.	15	Stationary sampling	33	A	-	B - Culture time 72 hrs	A-B	(Sąsiadek, Jagielski, and Smolik 1989)

Supplemental Material -Jackknife CA

Table S3. Jackknifing CA studies

Excluded study	Slope estimate (CI)	I^2
Primary study set		
Bogadi	0.35 (0.13-0.57)	78.4
Kim	0.33 (0.05-0.61)	86.5
Sram	0.22 (0.02-0.41)	94.5
Testa	0.22 (0.008-0.43)	94.8
Full study set		
Bogadi	0.32 (0.17-0.46)	88.7
Kim	0.31 (0.16-0.46)	90.1
Lovreglio	0.29 (0.16-0.43)	91.7
Sram	0.28 (0.14-0.41)	91.4
Testa	0.28 (0.14-0.42)	91.7
Berlin	0.22 (0.13-0.31)	78.0
Fracasso	0.29 (0.16-0.42)	91.7
Gundy	0.30 (0.15-0.45)	92.1
Jablonicka	0.32 (0.17-0.46)	91.4
Karacic	0.32 (0.17-0.46)	91.6
Major	0.27 (0.14-0.39)	90.0
Picciano	0.31 (0.16-0.45)	92.5
RomaTorres	0.29 (0.16-0.43)	91.8
Sarto	0.28 (0.15-0.41)	91.3
Sasiadek	0.31 (0.16-0.46)	91.8
Yardley Jones	0.31 (0.17-0.46)	91.9

Supplemental Material - MN studies

Table S4. Evaluation MN studies for inclusion in meta-analysis. 1: mean or otherwise median exposure. 2: If a range is given the midpoint is calculated, and the range indicated in brackets. N.a. (information) not available

Job category	Average MN (exposed/ control) ‰	Exposure controls ppm	Exposure exposed ppm[1]	Exposure method	Sample size exposed	Exposure group	Exposure assessment scoring	MN assessment scoring	Overall scoring group	Reference
Shoe manufacturing	3.38/1.82	n.a.	2	Air monitoring (Collected from different worksites at the plant, 3 times during study)	410	A	+	A	A+A	(Ren *et al.* 2018)
Car manufacturing company	2.12/1.19	n.a.	0.9[2] (< 6 mg/m3)	Air monitoring (multiple measurements by local CDC)	461	A	+	A	A+A	(Fang *et al.* 2017)
Decorators (container-manufacturing plant)	1.56/1.51	0.002	0.009	Air sampling (An individual air samples was placed near the subject's breathing zone to collect air samples during a work-shift)	132	C (exposure below/ around background concentrations)	+	A	C	(Sha *et al.* 2014)
Painters (container manufacturing plant)	1.53/1.51	0.002	0.07	Air sampling (An individual air samples was placed near the subject's breathing zone to collect air samples during a work-shift)	129	B (co-exposure to other carcinogens could not be ruled out)		A	BA	(Sha *et al.* 2014)
Petroleum refinery	6.66/3.96	n.a.	0.03	Personal sampling	79	A	+	A	A+A	(Basso *et al.* 2011)
Petroleum refinery	13.9/9.9	n.a.	0.51	Estimated using job-time exposure matrix	108	A	+	A	A+A	(Kim *et al.* 2008)
Automobile painters	6.68/3	n.a.	3.1	Air monitoring (repeated sampling in the workplaces)	25	A (authors sufficiently excluded potential exposure to other carcinogenic compounds)	+	A	A+A	(Testa *et al.* 2005)
Fuel tanker drivers	8.6/7.3	0.001	0.096	Personal sampling (measured during work shift by personal diffusive	19	B (co-exposure to diesel exhaust cannot be ruled out)		A	BA	(Lovreglio *et al*. 2014)

Job category	Average MN (exposed/ control) ‰	Exposure controls ppm	Exposure exposed ppm[1]	Exposure method	Sample size exposed	Exposure group	Exposure assessment scoring	MN assessment scoring	Overall scoring group	Reference
				samples, worn at the breathing zone level)						
FSA	8/7.3	0.001	0.007	Personal sampling (measured during work shift by personal diffusive samples, worn at the breathing zone level)	24	C (exposure below/around background concentrations)	+	A	C	(Lovreglio *et al*. 2014)
FSA	17.72/20.81	n.a.	0.28	Calculating atmospheric benzene levels with collected personal exposure measurements (personal sampling badges worn by FSA)	50	B (FSA)	+	A	B+A	(Pitarque *et al.* 1996)
FSA	13.62/16.96	n.a.	0.1	Personal sampling	12	B (FSA)	+	A	B+A	(Carere *et al.* 1998)
FSA	29.05/17.7	n.a.	0.07	Personal sampling	21	B (FSA)	+	A	B+A	(Bukvic *et al.* 1998)
Petro-chemical industry	9.23/17	0.009	0.75 (0.5-1) **	Air monitoring (three individual exposure assessments were conducted during a period of 1 year)	35	A	+	A	A+A	(Surrallés *et al.* 1997)
FSA	2.18/0.45	0.05	3.69	Air monitoring (samples were collected 1.5 meters above ground, about 2-3 meters from the fuel pump by active sampling. At each station , 2-3 samples were collected)	62	B (FSA)	+	B (only 500 cells scored for MN)	C	(Salem *et al*. 2017)
Petroleum refinery	4.08/1.97	0.05	0.4	t,t-MA in urine	48	A	+	A	A+A	(Roma-Torres *et al.* 2006)
Shoe factory	3.98/ 2.64	n.a.	0.77	Personal sampling	35	A	+	B – Information on MN assay very limited (e.g. no	A+B	(Liu *et al*. 1996)

Job category	Average MN (exposed/ control) ‰	Exposure controls ppm	Exposure exposed ppm[1]	Exposure method	Sample size exposed	Exposure group	Exposure assessment scoring	MN assessment scoring	Overall scoring group	Reference
								information on culture time or number of cells counted)		
Paint workers of a car factory	7.89/2.64	n.a.	32.2	Personal sampling	24	B (significant co-exposure cannot be ruled out)		B – Information on MN assay very limited (e.g. no information on culture time or number of cells counted)	BB	(Liu *et al*. 1996)
Shoe factory	8.15/2.64	n.a.	66.4	Personal sampling	28	A	+	B – Information on MN assay very limited (e.g. no information on culture time or number of cells counted)	A+B	(Liu *et al*, 1996)
Gasoline pump workers	6.7/2.2	unclear	unclear	Measuring phenol in urine	50	C	-(not possible to estimate benzene exposure using reverse modelling, because urinary phenol levels were so close to background values included in model that a reliable estimate could not be made.	A	C	(Priya *et al.* 2015)

Supplemental Material - Jackknife MN

Table S5. Jackknifing MN studies

Excluded study	Slope estimate (CI)	I^2
Primary study set		
Basso	0.17 (-0.24 – 0.58)	99.68
Fang	1.02 (-0.98 – 3.03)	99.97
Kim	0.91 (-1.12 – 2.94)	99.99
Ren	1.03 (-0.98 – 3.03)	99.96
Roma-Torres	0.94 (-1.09 – 2.98)	99.99
Surrallés	0.36 (0.06 – 0.66)	99.30
Testa	1.02 (-0.99 – 3.03)	99.98

Supplemental Material - Flowchart CA

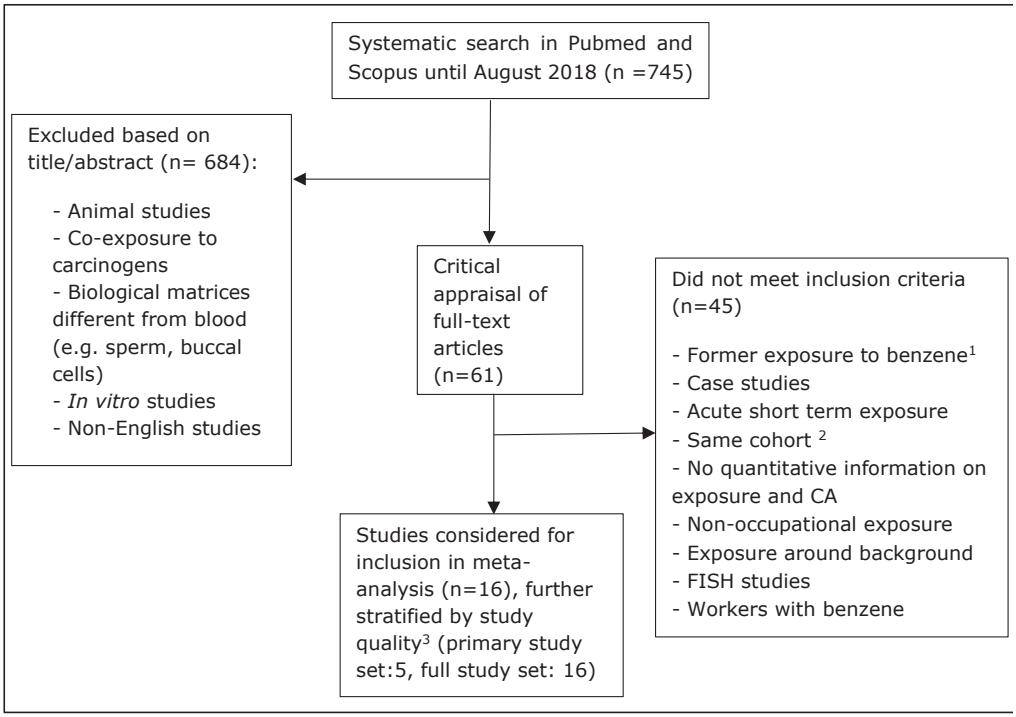


Figure S18. Flowchart of selection of studies on benzene exposure and chromosomal aberrations, for inclusion in meta-analysis. 1: Former exposure describes studies in which workers are tested for CA months or years after benzene exposure. 2: Same cohort refers to studies that have been conducted in the same cohort but are published over several publications. 3: It occurred that a study contained more exposed groups with one exposed group being included in the meta-analysis and another exposed group from the same study not. Therefore the number of studies stratified by study quality is not identical to the number of studies considered for inclusion in the meta-analysis.

Supplemental Material - Smoking interaction CA

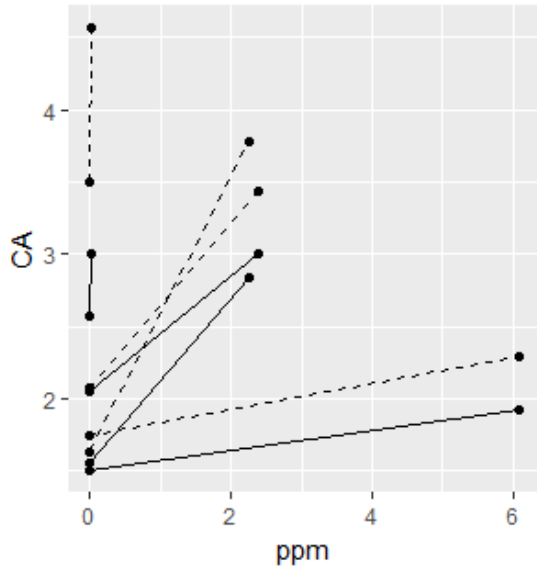


Figure S2. Individual studies (n=4) on CA for smokers (dashed curves) and non-smokers (solid line). The studies included: Fracasso et al. (2010), Jablonicka et al. (1987), Major et al. (1994), Sram et al. (2004).

Supplemental Material - Trim and fill CA

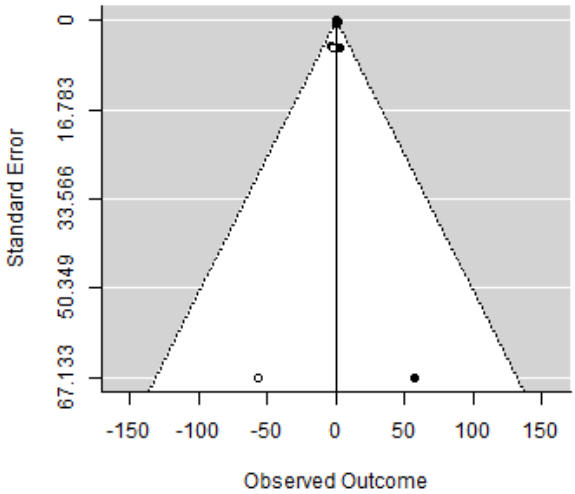


Figure S3. Funnelplot Trim and Fill for CA. The black circles represent the different study outcomes and their standard error. De open circles represent the imputed studies.

Supplemental Material - Flowchart MN studies

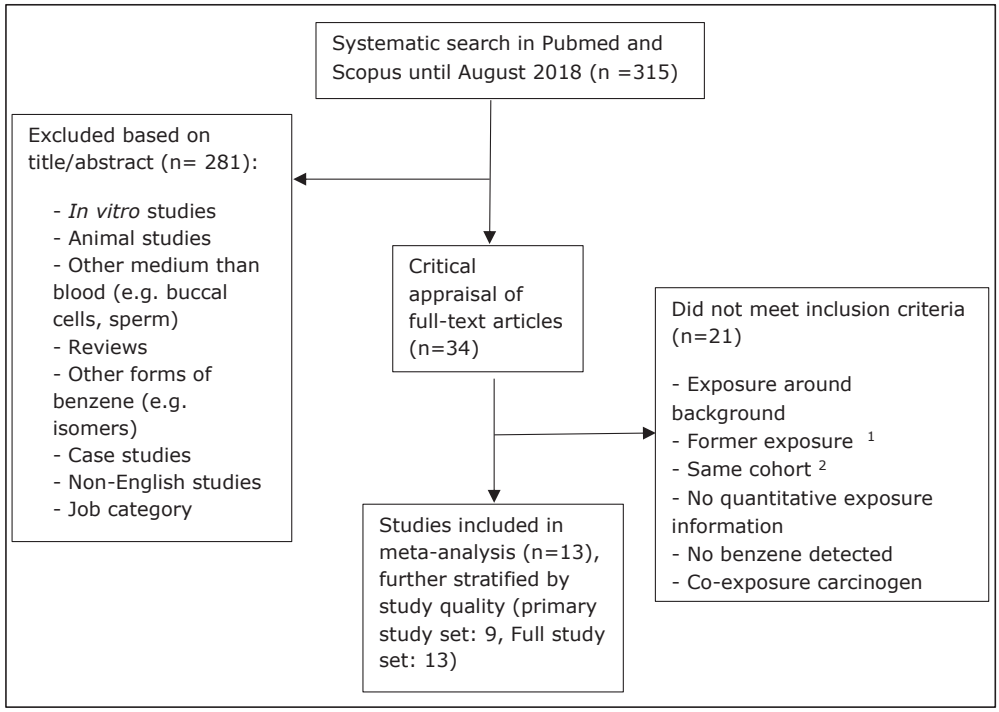


Figure S4. Flowchart of selection of studies, on benzene exposure and micro nuclei aberrations, for inclusion in meta-analysis. 1: Former exposure describes studies in which workers are measured months or years after benzene exposure. 2: Same cohort refers to studies that have been conducted in the same cohort but are published over several publications.

Supplemental Material - Smoking interaction MN

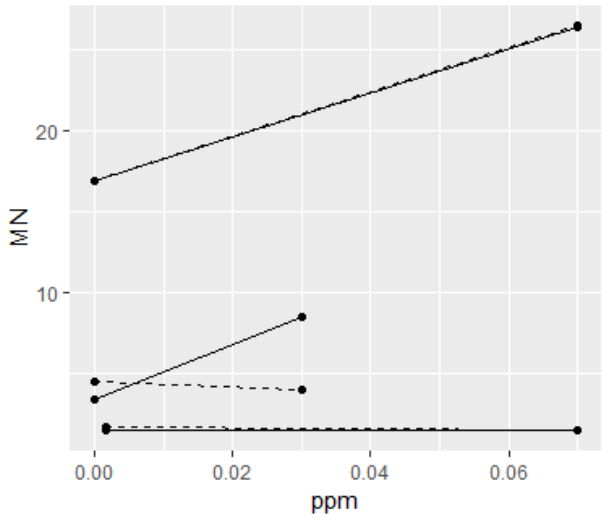


Figure S5. Individual studies (n=3) on MN for smokers (dashed curves) and non-smokers (solid line). The studies included: Basso et al. (2011), Bukvic et al. (1998), Sha et al. (2014).

Supplemental Material - Trim and fill MN

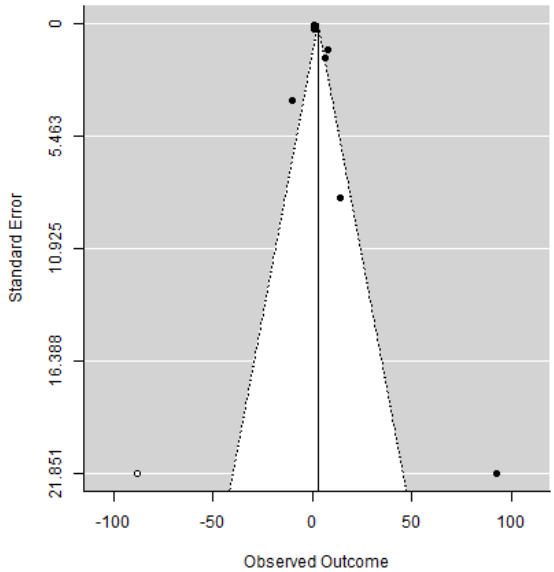


Figure S6. Funnelplot Trim and Fill for MN. The black circles represent the different study outcomes and their standard error. De open circles represent the imputed studies. The outlier at approximately 80 (for the observed outcome) represents the Basso et al. (2011) study, where the observed outcome is much larger for the reported standard error in comparison to the other studies.

References Supplemental Material

Angelini, Sabrina, Rajiv Kumar, Justo Lorenzo Bermejo, Francesca Maffei, Anna Barbieri, Francesca Graziosi, Fabio Carbone, *et al.* 2011. "Exposure to Low Environmental Levels of Benzene: Evaluation of Micronucleus Frequencies and S-Phenylmercapturic Acid Excretion in Relation to Polymorphisms in Genes Encoding Metabolic Enzymes." *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. https://doi.org/10.1016/j.mrgentox.2010.10.002.

Basso, E., C Cevoli, M Papacchini, G Tranfo, A Mansi, and A Testa. 2011. "Cytogenetic Biomonitoring on a Group of Petroleum Refinery Workers." *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.20641.

Berlin, Maths. 1985. "Low Level Benzene Exposure in Sweden: Effect on Blood Elements and Body Burden of Benzene." *American Journal of Industrial Medicine*. https://doi.org/10.1002/ajim.4700070503.

Bogadi-Šare, A, V Brumen, R Turk, … V KARACIC - Industrial, and Undefined 1997. 1997. "Genotoxic Effects in Workers Exposed to Benzene: With Special Reference to Exposure Biomarkers and Confounding Factors." *Jstage.Jst.Go.Jp*. https://www.jstage.jst.go.jp/article/indhealth1963/35/3/35_3_367/_article/-char/ja/.

Bukvic, Nenad, Pietro Bavaro, Giuseppe Elia, Filippo Cassano, Margherita Fanelli, and Ginevra Guanti. 1998. "Sister Chromatid Exchange (SCE) and Micronucleus (MN) Frequencies in Lymphocytes of Gasoline Station Attendants." *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. https://doi.org/10.1016/S1383-5718(98)00045-X.

Carere, A., A. Antoccia, D. Cimini, R. Crebelli, F. Degrossi, P. Leopardi, F. Marcon, A. Sgura, C. Tanzarella, and A. Zijno. 1998. "Genetic Effects of Petroleum Fuels: II. Analysis of Chromosome Loss and Hyperploidy in Peripheral Lymphocytes of Gasoline Station Attendants." *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/(SICI)1098-2280(1998)32:2<130::AID-EM8>3.0.CO;2-#.

Carere, A, A Antoccia, R Crebelli, … DC Di - Epidemiologia e, and Undefined 1995. 1995. "Exposure to Benzene and Genotoxic Effects among Filling Station Attendants." *Europepmc.Org*. https://europepmc.org/abstract/med/7601233.

Carrano, A. V., and A. T. Natarajan. 1988. "Considerations for Population Monitoring Using Cytogenetic Techniques." *Mutation Research/Genetic Toxicology*. https://doi.org/10.1016/0165-1218(88)90036-5.

Çelık, A, and E Akbaş. 2005. "Evaluation of Sister Chromatid Exchange and Chromosomal Aberration Frequencies in Peripheral Blood Lymphocytes of Gasoline Station Attendants." *Ecotoxicology and Environmental Safety*. https://www.sciencedirect.com/science/article/pii/S0147651303002045.

Fang, Yan, Han Tian Wu, Yun Jie Ye, Li Fang Zhou, Wei Hu, Guang Hui Zhang, Pin Sun, William Au, and Zhao Lin Xia. 2017. "Association between Polymorphisms of Metabolic Enzyme Genes and Chromosomal Damage in Benzene-Exposed Workers in China." *Journal of Occupational and Environmental Medicine*. https://doi.org/10.1097/JOM.0000000000001148.

Fenech, Michael. 2007. "Cytokinesis-Block Micronucleus Cytome Assay." *Nature Protocols*. https://doi.org/10.1038/nprot.2007.77.

Fracasso, Maria Enrica, Denise Doria, Giovanni Battista Bartolucci, Mariella Carrieri, Piero Lovreglio, Andrea Ballini, Leonardo Soleo, Giovanna Tranfo, and Maurizio Manno.

2010. “Low Air Levels of Benzene: Correlation between Biomarkers of Exposure and Genotoxic Effects.” *Toxicology Letters*. https://doi.org/10.1016/j.toxlet.2009.04.028.

Gundy, S. 1989. “Cytogenetical Studies on a Large Control Poplution and on Persons Occupatioannly Exposed to Radiation and/or to Chemicals.” *Ann.Lst.Super. Sanita* 25: 549–56. https://www.researchgate.net/profile/Sarolta_Gundy/publication/20523837_Cytogenetic_studies_on_a_large_control_population_and_on_persons_occupationally_exposed_to_radiation_andor_chemicals/links/00463524be489967fe000000.pdf.

Holz, O., G. Scherer, S. Brodtmeier, F. Koops, K. Warncke, T. Krause, A. Austen, *et al*. 1995. “Determination of Low Level Exposure to Volatile Aromatic Hydrocarbons and Genotoxic Effects in Workers at a Styrene Plant.” *Occupational and Environmental Medicine*. https://doi.org/10.1136/oem.52.6.420.

IARC. 2018. “Benzene (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans).” Vol. 120.

Jablonicka, A, M Vargova, and J Karelova. 1987. “Cytogenetic Analysis of Peripheral Blood Lymphocytes in Workers Exposed to Benzene.” *Journal of Hygiene, Epidemiology, Microbiology, and Immunology*.

Jamebozorgi, I., F. Mahjoubi, G. Pouryaghoub, R. Mehrdad, T. Majidzadeh, Z. Saltanatpour, and F. Nasiri. 2016. “Micronucleus, Nucleoplasmic Bridge, and Nuclear Budding in Peripheral Blood Cells of Workers Exposed to Low Level Benzene.” *International Journal of Occupational and Environmental Medicine*. https://doi.org/10.15171/ijoem.2016.785.

Karačić, Višnja, Ljiljana Skender, Branka Bosner-Cucančić, and Ana Bogadi-Sare. 1995. “Possible Genotoxicity in Low Level Benzene Exposure.” *American Journal of Industrial Medicine* 27 (3): 379–88. https://doi.org/10.1002/ajim.4700270307.

Kim, Yang Jee, Yoon Hee Cho, Domyung Paek, and Hai Won Chung. 2004. “DETERMINaTION OF CHROMOSOME ABERRATIONS IN WORKERS IN A PETROLEUM REFINING FACTORY.” *Journal of Toxicology and Environmental Health, Part A*. https://doi.org/10.1080/15287390490513319.

Kim, Yang Jee, Jun Yeol Choi, Domyung Paek, and Hai Won Chung. 2008. “Association of the NQO1, MPO, and XRCC1 Polymorphisms and Chromosome Damage among Workers at a Petroleum Refinery.” *Journal of Toxicology and Environmental Health - Part A: Current Issues*. https://doi.org/10.1080/15287390701738558.

Leopardi, P., A. Zijno, F. Marcon, L. Conti, A. Carere, A. Verdina, R. Galati, F. Tomei, T. P. Baccolo, and R. Crebelli. 2003. “Analysis of Micronuclei in Peripheral Blood Lymphocytes of Traffic Wardens: Effects of Exposure, Metabolic Genotypes, and Inhibition of Excision Repair in Vitro by ARA-C.” *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.10138.

Liu, L, Q Zhang, J Feng, L Deng, N Zeng, … A Yang - Mutation Research, and Undefined 1996. 1996. “The Study of DNA Oxidative Damage in Benzene-Exposed Workers.” *Elsevier*. https://www.sciencedirect.com/science/article/pii/S0165121896000481.

Lovreglio, Piero, Francesca Maffei, Mariella Carrieri, Maria N. D’Errico, Ignazio Drago, Patrizia Hrelia, Giovanni B. Bartolucci, and Leonardo Soleo. 2014. “Evaluation of Chromosome Aberration and Micronucleus Frequencies in Blood Lymphocytes of Workers Exposed to Low Concentrations of Benzene.” *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. https://doi.org/10.1016/j.mrgentox.2014.04.022.

Maffei, Francesca, Patrizia Hrelia, Sabrina Angelini, Fabio Carbone, Giorgio Cantelli Forti, Anna Barbieri, Giovanni Sanguinetti, Stefano Mattioli, and Francesco Saverio Violante. 2005. “Effects of Environmental Benzene: Micronucleus Frequencies and Haematological Values in Traffic Police Working in an Urban Area.” *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. https://doi.org/10.1016/j.mrgentox.2005.01.011.

Major, J., M. Jakab, Gabriella Kiss, and Anna Tompa. 1994. “Chromosome Aberration, Sister-chromatid Exchange, Proliferative Rate Index, and Serum Thiocyanate Concentration in Smokers Exposed to Low-dose Benzene.” *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.2850230211.

Pandey, Alok K., Mahima Bajpayee, Devendra Parmar, Rakesh Kumar, Subodh K. Rastogi, Neeraj Mathur, Paul Thorning, *et al*. 2008. “Multipronged Evaluation of Genotoxicity in Indian Petrol-Pump Workers.” *Environmental and Molecular Mutagenesis*. https://doi.org/10.1002/em.20419.

Picciano, Dante. 1979. “Cytogenetic Study of Workers Exposed to Benzene.” *Environmental Research*. https://doi.org/10.1016/0013-9351(79)90031-8.

Pitarque, M., E. Carbonell, N. Lapeña, M. Marsá, M. Torres, A. Creus, N. Xamena, and R. Marcos. 1996. “No Increase in Micronuclei Frequency in Cultured Blood Lymphocytes from a Group of Filling Station Attendants.” *Mutation Research - Genetic Toxicology*. https://doi.org/10.1016/0165-1218(95)00091-7.

Pitarque, M., A. Creus, R. Marcos, J. A. Hughes, and D. Anderson. 1999. “Examination of Various Biomarkers Measuring Genotoxic Endpoints from Barcelona Airport Personnel.” *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. https://doi.org/10.1016/S1383-5718(99)00026-1.

Priya, Kanu, Anita Yadav, Neeraj Kumar, Sachin Gulati, Neeraj Aggarwal, and Ranjan Gupta. 2015. “Glutathione S-Transferase Gene Polymorphisms: Modulator of Genetic Damage in Gasoline Pump Workers.” *International Journal of Toxicology*. https://doi.org/10.1177/1091581815603935.

Rekhadevi, P.V., M. Mahboob, M.F. Rahman, and Paramjit Grover. 2011. “Determination of Genetic Damage and Urinary Metabolites in Fuel Filling Station Attendants.” *Environmental and Molecular Mutagenesis* 52 (4): 310–18. https://doi.org/10.1002/em.20622.

Ren, Jing Chao, Yu Xia Wu, Zhenzhen Wu, Guang Hui Zhang, Hongjian Wang, Hongzhou Liu, Jun Peng Cui, *et al*. 2018. “MTHFR Gene Polymorphism Is Associated with DNA Hypomethylation and Genetic Damage among Benzene-Exposed Workers in Southeast China.” *Journal of Occupational and Environmental Medicine*. https://doi.org/10.1097/JOM.0000000000001288.

Roma-Torres, Joana, João P. Teixeira, Susana Silva, Blanca Laffon, Luís M. Cunha, Josefina Méndez, and Olga Mayan. 2006. “Evaluation of Genotoxicity in a Group of Workers from a Petroleum Refinery Aromatics Plant.” *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*. https://doi.org/10.1016/j.mrgentox.2005.12.005.

Salem, Eman, Islam El-Garawani, Heba Allam, Bahiga Abd El-Aal, and Mofrih Hegazy. 2017. “Genotoxic Effects of Occupational Exposure to Benzene in Gasoline Station Workers.” *Industrial Health*. https://doi.org/10.2486/indhealth.2017-0126.

Sarto, F, I Cominato, AM Pinton, … PG Brovedani -, and Undefined 1984. 1984. “A Cytogenetic Study on Workers Exposed to Low Concentrations of Benzene.” *Academic.Oup.Com*. https://academic.oup.com/carcin/article-abstract/5/6/827/2477417.

Sąsiadek, M, J Jagielski, and R Smolik. 1989. “Localization of Breakpoints in the Karyotype of Workers Professionally Exposed to Benzene.” *Mutation Research/Genetic Toxicology* 224(2): 235–40. https://www.sciencedirect.com/science/article/pii/0165121889901614.

Sha, Yan, Wei Zhou, Zhenyu Yang, Xiaoling Zhu, Yingping Xiang, Tiandi Li, Dexiang Zhu, and Xinyue Yang. 2014. “Changes in Poly (ADP-Ribosyl)Ation Patterns in Workers Exposed to BTX.” *PLoS ONE*. https://doi.org/10.1371/journal.pone.0106146.

Sram, Radim J., Olena Beskid, Blanka Binkova, Pavel Rossner, and Zdenek Smerhovsky. 2004. “Cytogenetic Analysis Using Fluorescence in Situ Hybridization (FISH) to Evaluate Occupational Exposure to Carcinogens.” In *Toxicology Letters*. https://doi.org/10.1016/j.toxlet.2003.12.043.

Šrám, RJ, L Dobiáš, A Pastorkova, … P Rössner - Mutation Research, and Undefined 1983. 1983. “Effect of Ascorbic Acid Prophylaxis on the Frequency of Chromosome Aberrations in the Peripheral Lymphocytes of Coal-Tar Workers.” *Mutation Research Letters* 120 (2–3): 181–86. https://www.sciencedirect.com/science/article/pii/0165799283901616.

Surrallés, J., K. Autio, L. Nylund, H. Järventaus, H. Norppa, T. Veidebaum, M. Sorsa, and K. Peltonen. 1997. “Molecular Cytogenetic Analysis of Buccal Cells and Lymphocytes from Benzene-Exposed Workers.” *Carcinogenesis*. https://doi.org/10.1093/carcin/18.4.817.

Testa, A., F. Festa, R. Ranaldi, M. Giachelia, D. Tirindelli, A. De Marco, M. Owczarek, M. Guidotti, and R. Cozzi. 2005. “A Multi-Biomarker Analysis of DNA Damage in Automobile Painters.” *Environmental and Molecular Mutagenesis* 46 (3): 182–88. https://doi.org/10.1002/em.20147.

Yardley-Jones, A, D Anderson, … DP Lovell - Occupational and, and Undefined 1990. 1990. “Analysis of Chromosomal Aberrations in Workers Exposed to Low Level Benzene.” *Oem.Bmj.Com*. https://oem.bmj.com/content/47/1/48.short.

Chapter 6

Estimation of the exposure response relation between benzene and acute myeloid leukemia by combining epidemiological, human biomarker, and animal data

Bernice Scholten*, Lützen Portengen*, Anjoeka Pronk, Rob Stierum, George S. Downward, Jelle Vlaanderen*, Roel Vermeulen*.

*authors contributed equally

Published in: Cancer Epidemiology, Biomarkers & Prevention, 31(4), 751-757.

Abstract

Background

Chemical risk assessment can benefit from integrating data across multiple evidence bases, especially in exposure-response curve (ERC) modeling when data across the exposure range is sparse.

Methods

We estimated the ERC for benzene and acute myeloid leukemia (AML), by fitting linear and spline-based Bayesian meta-regression models that included summary risk estimates from non-AML and non-human studies as prior information. Our complete dataset included six human AML studies, three human leukemia studies, ten human biomarker studies, and four experimental animal studies.

Results

A linear meta-regression model with intercept best predicted AML risks after cross-validation, both for the full dataset and AML studies only. Risk estimates in the low exposure range [<40 parts per million (ppm)-years] from this model were comparable, but more precise, when the ERC was derived using all available data than when using AML data only. Allowing for between-study heterogeneity, RRs and 95% prediction intervals [95%PI] at 5 ppm-years were 1.58 [1.01, 3.22]) and 1.44 [0.85, 3.42], respectively.

Conclusions

Integrating the available epidemiologic, biomarker, and animal data resulted in more precise risk estimates for benzene exposure and AML, although the large between-study heterogeneity hampers interpretation of these results. The harmonization steps required to fit the Bayesian meta-regression model involve a range of assumptions that need to be critically evaluated, as they seem crucial for successful implementation.

Impact

By describing a framework for data-integration and explicitly describing the necessary data harmonization steps, we hope to enable risk assessors to better understand the advantages and assumptions underlying a data integration approach.

Introduction

There is international consensus that benzene exposure is causally related to acute myeloid leukemia (AML) (IARC 2018), but accurate description of the AML-benzene exposure response curve (ERC) is still needed for impact and risk assessment. Vlaanderen and colleagues (2009) used meta-regression to derive an ERC for total leukemia, which includes AML, but also subtypes for which a causal relation to benzene exposure has not been shown. Derivation of a precise ERC for AML was not considered possible because of the limited number of studies and low case numbers at low levels of exposure. Combining human AML data with data from closely related study domains, for example, human epidemiologic data on leukemia or cancer biomarkers, and animal experimental data could be used to increase precision of the estimated ERC in this, and other cases (National Academies of Sciences Engineering and Medicine 2017).

The aim of this article is to estimate the benzene-AML ERC using data from both human and animal studies and from both experimental and observational study designs. We also aim to identify and highlight some of the assumptions underlying the harmonization steps required for integrating data across these different study domains. We hypothesize that the shape of the benzene-AML ERC may be estimated more precisely when using a larger evidence base, but also that it will improve our ability to address concerns regarding the generalizability of results obtained from a small set of studies. We focus on the exposure-response relation at relatively low exposure levels, because the risks at low occupational and environmental benzene exposure levels are still being debated (Mchale, Zhang, and Smith 2012).

For this purpose, we add to the available epidemiological studies that directly investigated benzene and AML, i) human studies on benzene induced leukemia, ii) human biomarker studies in benzene exposed workers on the induction of chromosomal aberrations (CA) and micronuclei (MN), and (iii) experimental animal studies on benzene-induced hematopoietic and lymphoid cancers. Biomarker studies for CA and MN were included because these were found to be associated with increased risk of cancer in large prospective cohort studies (Bonassi *et al*. 2008; 2007).

We choose a Bayesian meta-regression approach because it can be readily adapted to include prior information on likely effect sizes and between-study heterogeneity (Turner *et al*. 2015), while also accounting more directly for the imprecisely estimated between-study heterogeneity than frequentist models (Jackson, Bowden, and Baker 2010).

Methods

We refer to the Supplementary for a description of how individual studies were selected for the epidemiologic, human biomarker and animal experimental study domains (Supplementary Study Selection).

A “reference” ERC for benzene and AML was estimated using a Bayesian version of the meta-regression model used by Vlaanderen *et al.*(2011) for estimating the ERC for benzene and leukemia, and using only data from epidemiologic studies that directly investigated risk of AML in benzene exposed workers. As usual for this type of (meta-regression) model, the input data consisted of reported summary risk estimates (i.e. log-Hazard Ratios) and standard errors for each (AML) study and for each (average) cumulative benzene exposure level for which it was reported. Additional details on the structure of the meta-regression model are provided further below.

To estimate an “augmented” ERC, i.e. an ERC that includes information from studies in some or all of the related study domains, we converted the available exposure-response information to fit the same format as that used for the AML studies, i.e. into log relative effect estimates with standard errors at an estimated cumulative benzene exposure level. Detailed examples of how the available information was processed in order to fit this format (i.e. the data harmonization steps) are provided in the Supplementary Data Harmonization (including Supplementary Table S1), but the approach is also outlined below, separately for each study domain.

Epidemiological studies

For each study, risk estimates were selected from models where cumulative exposure was entered as a categorical variable, with standard errors estimated from reported confidence intervals. No distinction was made between rate or hazard ratios (HRs), standardized mortality ratios (SMRs), or odds ratios (ORs), and we will refer to each of these using the term “relative risk” (RR) in the remainder of the article. When no cases were observed in one of the categories [as was the case for the study by Collins *et al*. (2015)], we imputed half a case as a continuity correction to allow calculation of the logRR for the meta-regression. An overview of all included AML or leukemia studies can be found in the Supplementary AML studies (Supplementary Table S2) and Supplementary Leukemia studies (Supplementary Table S3), respectively.

Exposure estimates were based either on reported average cumulative exposures or, when these were not available, by assigning the midpoint of the reported range or, for open-ended upper categories, the lower category boundary multiplied by 5/3.

Human biomarker studies

The meta-analytical approach of Scholten *et al*. (2020) (where benzene exposure effects are estimated on an additive scale) was modified in two ways to allow inclusion of the study data in our meta-regression framework. First, to quantify effects on a

multiplicative scale, we log-transformed the reported proportions and used the delta method to estimate their variance on the log-scale (Oehlert 1992). We then subtracted the log-transformed proportion of aberrant cells in the unexposed (or low-exposed) from that in the exposed and estimated its standard error using standard variance rules. The main assumption underlying this approach is that the ratio of CA (or MN) in exposed to unexposed categories can be used to inform our prior estimate for the RR of benzene-induced AML. Second, to harmonize exposure levels in the biomarker studies to those in the epidemiologic studies, (average) exposure levels were multiplied by the reported average working histories, as a way to estimate cumulative exposures. If for a specific study no working history was available (which was the case for three CA and five MN studies) we assigned the average working history across all other studies. The main assumption underlying this step is that group-differences in average exposures calculated from cross-sectional studies are equivalent to the differences observed in the epidemiologic studies.

Animal experimental studies

Effect estimates for most animal studies were either available as, or could be easily expressed as, risk ratios. Concordance between experimental animal and human epidemiologic data for AML is supported by results presented in the recent International Agency for Research on Cancer (IARC) Monograph on tumor site concordance and mechanisms of carcinogenesis (IARC 2019). We estimated cumulative benzene exposure levels by multiplying reported exposure levels (ppm) by reported exposure durations without applying any further conversion factors. Results from multiple experiments reported in a single paper (n=2) were considered separate studies. An overview of all animal studies can be found in the Supplementary file (Supplementary Table S4).

Bayesian meta-regression model

Our meta-regression model is formulated as a hierarchical two-level random intercept and slope model. The (level 1) model for the observed (log) relative risks (Y) is:

$$Y \sim MVN(E(Y), \Sigma_y)$$
$$E(Y_{ij}) = \delta_{0,i} + \sum_{k=1}^{K} \delta_{k,i} * X_{k,ij}$$

Where MVN is the multivariate normal distribution and E(Y) the expected value of Y, *i* and *j* index different studies and exposure levels, $\delta_{0,i}$ and $\delta_{k,i}$ are the study-specific intercept and slope coefficients, and where X_k is either the benzene exposure level itself (for linear models; K=1) or the k^{th} basis of a K-dimensional regression spline. The covariance matrix of residual errors (Σ_y) is assumed to be known and was estimated using the method proposed by Greenland and Longnecker (1992).

The (level 2) model for the random intercept ($\delta_{0,i}$) and slope coefficient(s) ($\delta_{k,i}$) is:

$$\begin{bmatrix} \delta_{0,i} \\ \vdots \\ \delta_{k,i} \end{bmatrix} \sim MVN(\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}, \Sigma_\delta)$$

We aimed to specify priors that are strong enough to rule out unreasonable values, while still allowing the data to dominate the prior when it provides enough information. We assigned a weakly informative normal prior centered at 0 and with a scale of 1 to the overall intercept (β_0), and normal priors centered at 0 with a scale of 2 to the slope parameter(s) ($\beta_1 \cdots \beta_K$). This reflects our prior belief that the RR at zero exposure is likely to be exp(0)=1 and rather unlikely (i.e. with prior probability <20%) to be outside the range exp(±1.28) = [1/3.6, 3.6] and that the RR per 100 ppm-years is unlikely to be outside the range exp(±2*1.28) = [1/13, 13] (for the linear model).

For the random effect variances we followed the recommendations by Röver *et al*. (2021) in using a half-Cauchy prior for the random effect variances and the prior suggested by Lewandowski, Kurowicka, and Joe (LKJ) (2009) for the correlations between random effects. Our main analyses are based on using a half-Cauchy with a scale of 1. Our choice for this particular prior scale reflects our belief that the between-study variation in logRRs is likely (i.e. with prior probability >80%) within the range [0.16, 6.3], but with a mode at 0, and it allows for considerable heterogeneity in RRs between studies. Based on this prior, and assuming an overall RR of 1, approximately 20% of study-specific RRs (per 100 ppm-years) are expected to be between 0.9 and 1.1, 40% are expected to be between 0.7 and 1.5, 60% are expected to be between 0.4 and 2.4, and 80% are expected to be between and 0.1 and 9.9.

The parameter η of the LKJ prior was set at 2, which mainly serves to exclude very strong correlations between estimated random effect parameters (e.g. 80% of the intercept-slope correlations is expected to be in the range [-0.6, +0.6]).

Exposure response models

To assess the shape of the ERC, we fitted a regression spline model using a natural regression spline basis with interior knots at 10 and 65 ppm-years (the approximate 33% and 67% percentiles of the exposure distribution) and outer knots at 0 and 130 ppm-years (the approximate 0% and 85% percentiles of the exposure distribution). Alternative exposure-response models included a regression spline model without an intercept, forcing the ERC through the origin (RR=1) at zero exposure, and linear models with and without an intercept.

Between-study heterogeneity was accommodated by allowing for study-specific intercepts and slopes (regression coefficients) as (correlated) random effects. We additionally present 95% prediction intervals (95%PI) that take into account the between-study heterogeneity as recommended by Higgins, Thompson, and

Spiegelhalter (2009). Prediction intervals were calculated using estimates for the between-study heterogeneity for AML studies.

We evaluated model fit for each individual study by jackknifing (i.e. leaving out one study at a time), refitting the meta-regression model, and calculating the ratio of the sum of the differences between observed and predicted values over its estimated standard error for each held out study as an externally studentized residual. An absolute value of the ratio exceeding 3 was considered evidence of severe lack of fit. To compare the quality of posterior predictions from different model structures we estimated the sum of the expected log pointwise predictive density (ELPD) (Vehtari, Gelman, and Gabry 2017) using the models fitted during jackknifing. Higher ELPDs indicate better predictions for the left-out studies and can be calculated for the full set of studies, but also for a subset (e.g. only studies in the AML set).

Finally, to combine predictions from models with different model structures (i.e. linear/spline, with/without intercept) we used Bayesian stacking (Yao *et al*. 2018), with weights calculated based on the ELPD estimates. We call the resulting model the "consensus" model and the estimated ERC the "consensus" ERC.

Sensitivity analyses

First, we evaluated the sensitivity of our results to our choice of prior for the between-study covariance by changing the scale of the half-Cauchy prior to 0.5 and 5.

Second, we investigated the effect of excluding risk estimates for relatively high exposures (i.e. over 40 ppm-years) from the model on estimated risks at low exposures.

Third, we used an approach similar to that described in Bartell, Hamra, and Steenland (2017), reducing the precision of point estimates for a subset of studies, to illustrate how our approach could be used to selectively downweigh the impact of some studies, in this case adding an (arbitrary) 10-fold uncertainty factor to results from the experimental animal studies and a 3-fold uncertainty factor to results from both sets of biomarker studies.

Impact assessment

To illustrate how our framework could be used for risk assessment, we calculated the excess risk of AML due to benzene exposure for our "consensus" ERC using a life table analysis. Background incidence rates for AML in the Netherlands were obtained from the Dutch Integraal Kanker Centrum (IKC) and combined with Dutch mortality rates as obtained from Statistics Netherlands. Excess risk of AML was estimated for workers that were exposed to 0.1 ppm benzene for 40 years (between age 20-60), assuming a five year lag period, and evaluated at age 80. We also estimated benzene exposure levels corresponding to the definition of acceptable risk (AR) and maximum tolerable risk (MTR) levels for occupational settings, i.e. the benzene exposure level

at which the number of excess cases is either 40 per 1,000,000 or 40 per 10,000 exposed workers.

Software

The Bayesian meta-regression model was implemented in STAN using the brms package (version 2.10.0). We collected 10,000 samples for each parameter from 4 chains after a burn-in of 5,000 iterations using Markov Chain Monte Carlo (MCMC) techniques to sample from the posterior distribution. The algorithm was tuned by increasing adapt.delta to 0.999 to avoid divergent transitions (Stan Development Team n.d.). The posterior distribution was summarized by calculating the mean, standard deviation and 2.5, 50, and 97.5 percentiles.

Jackknifing (leave-one-group-out cross-validation) was performed using the R function *kfold* from the *brms* package, which was also used to estimate ELPDs for the held-out studies. Weights for Bayesian model averaging (model stacking) used to estimate the "consensus" ERC were estimated using the *stacking_weights* function from the *loo* package in R.

Results

We removed the MN biomarker study by Surrallés *et al*. (1997) from all further analyses because the results of the (first) jackknifing analysis indicated a severe lack of fit for this study even in our most flexible exposure-response model (i.e. a regression spline model with intercept; see Supplementary Table S5). Our final study base therefore consisted of 26 studies: six human epidemiological studies on AML, three human leukemia studies (that had not reported on AML), 10 biomarker studies (4 CA and 6 MN), and seven experimental animal studies [from four publications (Cronkite *et al*. 1989; Kawasaki *et al*. 2009; Farris *et al*. 1996; Li *et al*. 2006)]. Most of the AML and leukemia studies, but only a minority of other studies, had estimates for more than one exposure level.

Table 1 shows estimated RRs at cumulative benzene exposure levels of 0, 5, 10, 20, and 40 ppm-years based on our spline model with intercept for models fitted to data from studies in the reference set (i.e. human AML studies), after adding studies from single additional study domains, and after using all studies from all domains. Using data from the full set of studies resulted in slightly higher relative risk estimates than using the reference set only, and these were estimated more precisely with narrower credible intervals (CIs) and prediction intervals (PIs). There was considerable between-study heterogeneity in risk estimates however, as evident also from the PIs that were much wider than the CIs. Using the full set of studies resulted in narrower PIs than using the reference set only, suggesting that the gain in precision for population-level risk estimates was not at the expense of increased between-study heterogeneity.

Using cross-validation to estimate and compare the quality of model predictions, we found that the linear model with intercept provided the best predictions (i.e. had lowest kICs; see Table 2); both for the full set of studies and when considering predictions only for the AML studies. This was true also when models were fitted using data only from AML studies. Predictions for AML studies were also better from models fitted using all available data than from models fitted to the AML data only. As a result, estimated model weights were highly skewed towards the linear model with intercept, which was assigned a weight of nearly 1 (versus <1e-5 for any of the other model structures), making the "consensus" model virtually identical to the linear model with intercept.

Estimated ERCs using the linear "consensus" model for the AML set only and after adding studies from different domains are presented graphically in Figure 1. Although most datapoints from the biomarker and experimental animal studies were in the lower exposure region, inclusion of these studies had only limited effect on the estimated ERC.

Table 1. Relative risk estimates and 95% credible and prediction intervals for benzene-induced acute myeloid leukemia (AML), for benzene exposure at selected exposure levels. Risk estimates are population-level estimates from a meta-regression model that includes an intercept and allows for non-linear effects of benzene exposure. Estimates are given for only AML studies, and with inclusion of several data domains, including the full dataset. Second row in each cell: credible interval, third row: prediction interval.

Benzene (ppm-yrs)	AML	AML + leukemia	AML + CA biomarker	AML + MN biomarker	AML + animal data	All
0*	1.39	1.40	1.35	1.59	1.44	1.58
	(0.77 – 2.47)	(0.81 – 2.41)	(0.76 – 2.34)	(1.15 – 2.12)	(0.82 – 2.44)	(1.17 – 2.09)
	(0.73 – 3.67)	(0.76 – 3.54)	(0.73 – 3.45)	(0.94 – 3.53)	(0.78 – 3.55)	(0.96 – 3.47)
5	1.40	1.36	1.41	1.60	1.35	1.57
	(0.84 – 2.30)	(0.87 – 2.09)	(0.91 – 2.18)	(1.26 – 1.99)	(0.86 – 2.07)	(1.25 – 1.92)
	(0.72 – 3.89)	(0.73 – 3.49)	(0.78 – 3.68)	(0.92 – 3.69)	(0.73 – 3.39)	(0.93 – 3.51)
10	1.40	1.32	1.48	1.62	1.28	1.57
	(0.79 – 2.50)	(0.84 – 2.08)	(0.98 – 2.19)	(1.24 – 2.09)	(0.81 – 2.03)	(1.25 – 1.90)
	(0.60 – 4.83)	(0.61 – 4.05)	(0.71 – 4.48)	(0.78 – 4.40)	(0.58 – 3.80)	(0.81 – 3.98)
20	1.44	1.32	1.61	1.69	1.23	1.61
	(0.66 – 3.32)	(0.72 – 2.45)	(0.95 – 2.52)	(1.13 – 2.45)	(0.68 – 2.28)	(1.18 – 2.05)
	(0.43 – 7.60)	(0.45 – 5.63)	(0.57 – 6.73)	(0.60 – 6.18)	(0.41 – 5.19)	(0.65 – 5.28)
40	1.64	1.56	1.79	1.94	1.37	1.78
	(0.70 – 4.08)	(0.81 – 3.05)	(1.05 – 2.86)	(1.20 –2.97)	(0.75 – 2.63)	(1.28 – 2.34)
	(0.43 – 10.21)	(0.47 – 7.57)	(0.57 – 8.55)	(0.60 – 8.12)	(0.40 – 6.73)	(0.63 – 6.52)

** intercept*

Table 2. Comparison of predictive model quality for models with different model structures. Results are presented as the sum of expected log pointwise predictive densities ($\sum ELPD$), with higher values indicating better model predictions. ELPDs were estimated using jackknifing (leave-one-study-out cross-validation). Results are shown for models fitted to the full set of studies from all study domains for all studies and for AML studies only and for models fitted to data only from AML studies.

Model	Full dataset	Full dataset : AML only	AML
Linear model without intercept	121.6	54.52	56.82
Linear model with intercept	85	49.12	52.76
Spline model without intercept	113.2	54.02	58.88
Spline model with intercept	96	52.16	56.56

Similar to the more flexible regression spline model, estimated RRs at low cumulative benzene exposure levels were generally lower and less precise when using data only from the AML studies, when compared to using all available studies (Table 3). As an example, the risk estimate (95%PI) at 5 ppm-years was 1.44 (95% PI: 0.85,3.42) when using only AML studies, while it was 1.58 (95% PI: 1.01,3.22) when using all available studies.

Parameter estimates for the consensus parameters and between-study (co)variance for the linear model with an intercept fitted to the full dataset are provided in the Supplemental Materials (Table s6). Between-study variance in slopes was large relative to the consensus slope estimate. There were no large differences in estimated between-study variance for slopes for different study domains, but these were estimated rather imprecisely.

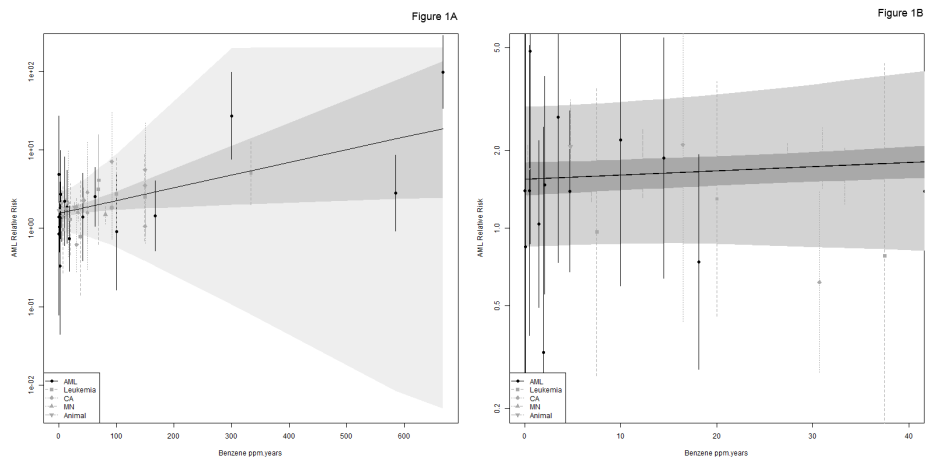


Figure 1. Benzene exposure and RR of developing acute myeloid leukemia (AML) based on various data domains. Linear model with intercept. Dark ribbon: Confidence interval for all studies, light ribbon: prediction interval for all studies. Dashed line represents knots. Left panel: all study points, right panel: zoomed 0 - 40 ppm-yrs.

Table 3. Risk estimates and 95% prediction intervals for benzene induced acute myeloid leukemia (AML) at selected exposure levels. Risk estimates are population-level estimates from a linear meta-regression model that includes an intercept. Estimates are given for only AML studies, and with inclusion of several data domains, including the full dataset. Second row in each cell: credible interval, third row: prediction interval.

Benzene exposure (ppm-yrs)	AML	AML + leukemia	AML + CA biomarker	AML + MN biomarker	AML + animal data	All
0	1.42	1.42	1.54	1.56	1.33	1.55
(intercept)	(0.91 – 2.21)	(0.99 – 2.06)	(1.11 – 2.12)	(1.31 – 1.90)	(0.91 – 1.93)	(1.34 – 1.80)
	(0.84 – 3.40)	(0.86 – 3.22)	(0.94 – 3.28)	(1.01 – 3.18)	(0.80 – 3.03)	(1.00 – 3.16)
5	1.44	1.45	1.56	1.60	1.36	1.58
	(0.94 – 2.22)	(1.01 – 2.09)	(1.14 – 2.11)	(1.36 – 1.92)	(0.93 – 1.96)	(1.37 – 1.82)
	(0.85 – 3.42)	(0.88 – 3.29)	(0.97 – 3.32)	(1.03 – 3.27)	(0.82 – 3.09)	(1.01 – 3.22)
10	1.46	1.48	1.58	1.63	1.38	1.61
	(0.95 – 2.24)	(1.04 – 2.12)	(1.17 – 2.10)	(1.39 – 1.94)	(0.96 – 1.98)	(1.40 – 1.84)
	(0.85 – 3.50)	(0.89 – 3.35)	(0.97 – 3.38)	(1.04 – 3.35)	(0.83 – 3.17)	(1.02 – 3.30)
20	1.50	1.54	1.63	1.71	1.44	1.67
	(0.97 – 2.31)	(1.08 – 2.19)	(1.23 – 2.10)	(1.43 – 2.03)	(1.01 – 2.04)	(1.46 – 1.90)
	(0.81 – 3.77)	(0.88 – 3.58)	(0.95 – 3.58)	(1.02 – 3.68)	(0.83 – 3.40)	(1.02 – 3.51)
40	1.59	1.67	1.72	1.86	1.55	1.80
	(0.96 – 2.55)	(1.16 – 2.38)	(1.29 – 2.24)	(1.44 – 2.37)	(1.10 – 2.21)	(1.56 – 2.07)
	(0.66 – 4.72)	(0.81 – 4.39)	(0.82 – 4.37)	(0.90 – 4.86)	(0.76 – 4.21)	(0.94 – 4.21)

Sensitivity analyses

Prior choice for the between-study (co)variance

Detailed results from sensitivity analyses regarding our prior choice for the between-study (co)variance are presented in the Supplementary file (tables S7-S9).Using a more diffuse prior had relatively little effect on the width of credible intervals (CIs) for predicted exposure effects, but resulted in significantly wider posterior prediction intervals (PIs). As an example, for the linear model with intercept the CIs for the predicted exposure effect at 40 ppm-yrs were (1,58 – 2.04) and (1.51 – 2.11) for prior scales of 0.5 and 5 respectively, while the corresponding PIs were (1.03 – 3.72) and (0.80 – 5.54).

Prior-posterior plots (Figure S1) suggest that there is no large mismatch between any of these priors and information provided by the data itself.

Excluding high exposure datapoints

Risk estimates were comparable but less precise after excluding all datapoints with benzene exposures over 40 ppm-years, with a noticeable drop in risk estimates for exposures of 40 ppm-years for the regression-spline based models (Supplementary Table S10).

Uncertainty factors

Estimated risks were marginally lower and considerably less precise when 3-fold uncertainty factors were used for the biomarker studies and a 10-fold uncertainty factor

for the experimental animal studies (Supplementary Table S11). The effect was more pronounced for CIs: risk estimates at 5 ppm-years were reduced from 1.58 (95% CI: 1.37,1.82) to 1.52 (95% CI: 1.16,2.00) when uncertainty factors were applied.

Impact assessment

Results from the life table analyses that were used to estimate the excess risk of AML due to benzene exposure using the linear model exposure-response model with intercept are presented in Table 4. These results indicate that, when workers are exposed to 0.1 ppm for 40 years, the estimated number of excess cases is higher than that corresponding to AR levels, except when the most conservative approach (i.e. subtracting the intercept) is used on the linear model fitted to data from only the reference set, where it is just slightly lower.

Exposure levels corresponding to AR and the MTR levels are shown in the Supplementary (Table S12, S13), and range from 0.0003 ppm (for the interpolated linear model) to 0.098 ppm (for the model with intercept subtracted) when models are fitted to the full set of studies.

Table 4. Results of a life table analysis for calculating excess cases per 1,000,000 at age 80 when exposed to benzene at 0.1 ppm for 40 years.

	AML	**AML + leukemia**	**AML+ All data**	**<40 ppm.yrs**
Linear with intercept	1289	1322	1698	1711
Linear with intercept – intercept subtracted	30	42	41	3
Linear with intercept – interpolation	1222	1253	1606	1619
Spline no intercept	295	204	379	n.a.
Linear no intercept	64	65	75	184

Discussion

To estimate the benzene-AML ERC, we collected and summarized human and animal studies. The ERC derived using studies from this broader evidence base was very similar to that based on data from the human AML studies only, but with more precise risk estimates. Based on results from cross-validation, prediction of risks observed in single (held-out) AML studies was improved by using data from other study domains.

We included a detailed description of data harmonization steps and our prior motivation to improve understanding of the assumptions underlying the use of these models by risk assessors. The need to harmonize exposure and outcome variables required us to make strong and mostly untestable assumptions. For human studies, exposure metrics used in the, mostly cross-sectional, biomarker studies (average exposure levels) were fundamentally different from those in the long-term prospective epidemiological studies (cumulative exposure levels). In addition, effects recorded in

the biomarker studies had to be converted to relative effects to allow combination with epidemiological data. For the animal data, cumulative exposure could be readily estimated, but it is unclear whether any inter-species extrapolation factors should have been applied. Although others [e.g., Bartell, Hamra, and Steenland (2017)] used conversion factors, there seems to be no general agreement on either use or value. We chose not to apply any conversion factor, based on the argument that metabolic rate and cell division are roughly inversely correlated to lifespan(Dutta and Sengupta 2016).

Uncertainties stemming from the extrapolation of risks from animal to human studies in current risk assessment procedures are typically addressed by using a fixed set of inter-species extrapolation factors (Jones *et al*. 2009). In an analogous approach, we (arbitrarily) used a 10-fold uncertainty factor for the animal studies and a 3-fold uncertainty factor for the biomarker studies to downweigh evidence from these study domains, resulting in slightly less precise risk estimates. While safeguarding against inappropriate over-reliance on animal or biomarker data, this may also result in sub-optimal use of the available data.

Prior choice for the between study variation is an important ingredient of Bayesian meta-analyses, and may be crucial when there is little information. There were only few studies per study domain in the analysis, and we therefore used a half-Cauchy prior with a scale parameter of 1, to exclude implausible high values for the random effects variance. We evaluated the sensitivity of our findings to this choice in sensitivity analyses (Table S7, Figure S1,S2), with the results confirming our prior is broadly compatible with the data, but also that estimated heterogeneity, which affects precision of e.g. common slope factors, is quite sensitive to the prior scale. We discuss the importance of heterogeneity for model inference further below, but note that with more or more precise data, the meta-regression model could be formulated with hyperpriors for the half-Cauchy scale. Our approach also allows evaluation and comparison of study heterogeneity across different study domains. We found no large differences in heterogeneity between study domains in our study, this could well be due to the fact that these were estimated rather poorly, as there were only few studies per study domain.

We used life table calculation to evaluate the impact of our estimated “consensus” ERC for further risk assessment. It should be noted that, in the presence of significant between-study heterogeneity, use of an ERC based on consensus (population-level) parameters may be difficult to justify (Dahabreh et al. 2020). Without knowing, or at least suspecting, what the reasons for the apparent heterogeneity may be, the mean parameters may be difficult to interpret. In case of strong heterogeneity, the mean parameters are also estimated with (near) equal weights for smaller and larger studies, which could be problematic when smaller studies are more likely to suffer from small-

sample or publication bias (Richardson et al. 2020). Alternatively, the full (random effects) distribution may be used e.g. to average lifetable results across a random sample of study-specific ERCs or for choosing an upper quantile of the between-study distribution under the assumption that the higher risks are observed in better studies (similar to what is done in benchmark dose modelling; BMDL).

Sobel, Madigan, and Wang (2017) and Dahabreh et al. (2020) recently discussed the problem of casual interpretation of results from meta-analyses. Both papers stress the importance of investigating sources of heterogeneity and rely on using additional individual-level covariate data to account for differential selection and exposure effects. Our approach is flexible enough to include further covariates as moderators.

We found that a linear model with intercept provided the best predictions. Possible reasons for this intercept include exposure measurement error, uncontrolled confounding, and healthy worker selection. Models with an intercept present considerable interpretational difficulties from a risk assessment perspective, and risk assessors therefore often prefer (meta-regression) models that do not allow intercepts. We therefore evaluated risks using a number of different approach to account for the intercept and also assessed risks based on the regression-spline model without an intercept. Estimated AR levels from these models estimated using all avalilable data were in the range 0.0003 – 0.098 ppm, which includes the limit for benzene exposure that was recently proposed by the Risk Assessment Comittee of 0.05ppm (RAC 2018).

Our approach differs from earlier proposals for using Bayesian methods to integrate data for risk assessment [e.g.,Dumouchel and Harris (1983;) Bartell, Hamra, and Steenland (2017)]. Most notably these authors used the additional data to derive a single prior for a slope coefficient from a linear meta-analytical model. In contrast, we aimed to include data from several different study domains simultaneously and wanted to allow for potential non-linear effects using a meta-regression model. Our approach is more easy to use with more complex model structures and allows more explicit evaluation of between-study heterogeneity for studies from different study sources.

Our approach can be seen as a first step towards quantitative integration of human and animal data in risk assessment. It involves a trade-off between a potential gain in statistical precision that comes with a more complete evaluation of the evidence base, but it also runs the risk of introducing bias or increasing heterogeneity due to the harmonization steps that may be required. Risk assessors may therefore choose to use this approach primarily when evaluating compounds with weak epidemiologic evidence (i.e. insufficient data) but a large animal or molecular evidence base. Implementation of our approach in regulatory chemical risk assessment exercises would require additional knowledge (such as better understanding of toxicokinetics, better evidence for pre-diagnostic biomarker-cancer associations, better

understanding of reasonable approaches to extrapolate animal evidence to the human setting) to reduce the assumptions that had to be made in the current evaluation.

To conclude, we provide a first step towards the quantitative integration of data from different study domains, into human risk assessment and identified a number of gaps that need to be addressed in further research.

References Chapter 6

Bartell, Scott M., Ghassan Badri Hamra, and Kyle Steenland. 2017. “Bayesian Analysis of Silica Exposure and Lung Cancer Using Human and Animal Studies.” *Epidemiology*. https://doi.org/10.1097/EDE.0000000000000602.

Bonassi, Stefano, Hannu Norppa, Marcello Ceppi, Ulf Strömberg, Roel Vermeulen, Ariana Znaor, Antonina Cebulska-Wasilewska, *et al*. 2008. “Chromosomal Aberration Frequency in Lymphocytes Predicts the Risk of Cancer: Results from a Pooled Cohort Study of 22 358 Subjects in 11 Countries.” *Carcinogenesis*. https://doi.org/10.1093/carcin/bgn075.

Bonassi, Stefano, Ariana Znaor, Marcello Ceppi, Cecilia Lando, Wushou Peter Chang, Nina Holland, Micheline Kirsch-Volders, *et al*. 2007. “An Increased Micronucleus Frequency in Peripheral Blood Lymphocytes Predicts the Risk of Cancer in Humans.” *Carcinogenesis* 28 (3): 625–31. https://doi.org/10.1093/carcin/bgl177.

Collins, James J., Steven E. Anteau, Gerard M.H. Swaen, Kenneth M. Bodner, and Catherine M. Bodnar. 2015. “Lymphatic and Hematopoietic Cancers Among Benzene-Exposed Workers.” *Journal of Occupational and Environmental Medicine*. https://doi.org/10.1097/JOM.0000000000000324.

Cronkite, E. P., R. T. Drew, T. Inoue, Y. Hirabayashi, and J. E. Bullis. 1989. “Hematotoxicity and Carcinogenicity of Inhaled Benzene.” *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.898297.

Dahabreh, I.J., L.C. Petito, S.E. Robertson, M.A. Hernán, and J.A. Steingrimsson. 2020. “Toward Causally Interpretable Meta-Analysis: Transporting Inferences from Multiple Randomized Trials to a New Target Population.” *Epidemiology* 31 (3): 334–44.

Dumouchel, William H., and Jeffrey E. Harris. 1983. “Bayes Methods for Combining the Results of Cancer Studies in Humans and Other Species.” *Journal of the American Statistical Association* 78 (382): 293–308. https://doi.org/10.1080/01621459.1983.10477968.

Dutta, Sulagna, and Pallav Sengupta. 2016. “Men and Mice: Relating Their Ages.” *Life Sciences*. https://doi.org/10.1016/j.lfs.2015.10.025.

Farris, G M, S N Robinson, K W Gaido, B A Wong, V A Wong, L Leonard, and R Shah. 1996. “Effects of Low Concentrations of Benzene on Mouse Hematopoietic Cells in Vivo: A Preliminary Report.” *Environmental Health Perspectives* 104 Suppl (December): 1275–76. https://doi.org/10.1289/ehp.961041275.

Greenland, Sander, and Matthew P. Longnecker. 1992. “Methods for Trend Estimation from Summarized Dose-Response Data, with Applications to Meta-Analysis.” *American Journal of Epidemiology*. https://doi.org/10.1093/oxfordjournals.aje.a116237.

Higgins, Julian P.T., Simon G. Thompson, and David J. Spiegelhalter. 2009. “A Re-Evaluation of Random-Effects Meta-Analysis.” *Journal of the Royal Statistical Society. Series A: Statistics in Society*. https://doi.org/10.1111/j.1467-985X.2008.00552.x.

IARC. 2018. “Benzene (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans).” Vol. 120.

———. 2019. “Tumour Site Concordance and Mechanisms of Carcinogenesis. IARC Scientific Publication No. 165.” http://publications.iarc.fr/578.

Jackson, D, J Bowden, and R Baker. 2010. “How Does the DerSimonian and Laird Procedure for Random Effects Meta-Analysis Compare with Its More Efficient but

Harder to Compute Counterparts?" *Journal of Statistical Planning and Inference* 140 (4): 961–70. https://www.sciencedirect.com/science/article/pii/S0378375809003073.

Jones, David R., Jaime L. Peters, Lesley Rushton, Alex J. Sutton, and Keith R. Abrams. 2009. "Interspecies Extrapolation in Environmental Exposure Standard Setting: A Bayesian Synthesis Approach." *Regulatory Toxicology and Pharmacology*. https://doi.org/10.1016/j.yrtph.2009.01.011.

Kawasaki, Yasushi, Yoko Hirabayashi, Toyozo Kaneko, Jun Kanno, Yukio Kodama, Yuuko Matsushima, Yukio Ogawa, *et al*. 2009. "Benzene-Induced Hematopoietic Neoplasms Including Myeloid Leukemia in Trp 53-Deficient C57BL/6 and C3H/He Mice." *Toxicological Sciences*. https://doi.org/10.1093/toxsci/kfp107.

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe. 2009. "Generating Random Correlation Matrices Based on Vines and Extended Onion Method." *Journal of Multivariate Analysis*. https://doi.org/10.1016/j.jmva.2009.04.008.

Li, Guang Xun, Yoko Hirabayashi, Byung Il Yoon, Yasushi Kawasaki, Isao Tsuboi, Yukio Kodama, Yuji Kurokawa, Junji Yodoi, Jun Kanno, and Tohru Inoue. 2006. "Thioredoxin Overexpression in Mice, Model of Attenuation of Oxidative Stress, Prevents Benzene-Induced Hemato-Lymphoid Toxicity and Thymic Lymphoma." *Experimental Hematology*. https://doi.org/10.1016/j.exphem.2006.08.005.

Mchale, C.M., Luoping Zhang, and M.T. Smith. 2012. "Current Understanding of the Mechanism of Benzene-Induced Leukemia in Humans: Implications for Risk Assessment." *Carcinogenesis*. https://doi.org/10.1093/carcin/bgr297.

National Academies of Sciences Engineering, and Medicine. 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. National Academies Press. https://doi.org/10.17226/24635.

Oehlert, Gary W. 1992. "A Note on the Delta Method." *The American Statistician*. https://doi.org/10.2307/2684406.

RAC. 2018. "Opinion on Scientific Evaluation of Occupational Exposure Limits for Benzene." https://echa.europa.eu/documents/10162/13641/benzene_opinion_en.pdf/4fec9aac-9ed5-2aae-7b70-5226705358c7.

Richardson, D, S.R. Cole, R.K Ross, C Poole, H Chu, and A Keil. 2020. "Meta-Analysis and Sparse-Data Bias." *American Journal of Epidemiology* 190 (2): 336–40.

Röver, C, R Bender, S Dias, C.H. Schmid, H Schmidli, S Sturtz, S Weber, and T Friede. 2021. "On Weakly Informative Prior Distributions for the Heterogeneity Parameter in Bayesian Random-Effects Meta-Analysis." *Res Synth Methods* 12 (4): 448–74. https://doi.org/10.1002/jrsm.1475.

Scholten, B, J Vlaanderen, R Stierum, L Portengen, N Rothman, Q Lan, A Pronk, and R Vermeulen. 2020. "A Quantitative Meta-Analysis of the Relation between Occupational Benzene Exposure and Biomarkers of Cytogenetic Damage." *Environmental Health Perspectives* 128 (https://doi.org/10.1289/EHP6404).

Sobel, M, D Madigan, and W Wang. 2017. "Causal Inference for Meta-Analysis and Multi-Level Data Structures, with Application to Randomized Studies of Vioxx." *Psychometrika* 82 (2): 459–74.

Stan Development Team. n.d. "Brief Guide to Stan's Warnings." 2019. Accessed June 22, 2019. https://mc-stan.org/misc/warnings.html.

Surrallés, J., K. Autio, L. Nylund, H. Järventaus, H. Norppa, T. Veidebaum, M. Sorsa, and K. Peltonen. 1997. "Molecular Cytogenetic Analysis of Buccal Cells and Lymphocytes from

Benzene-Exposed Workers." *Carcinogenesis*. https://doi.org/10.1093/carcin/18.4.817.

Turner, Rebecca M., Dan Jackson, Yinghui Wei, Simon G. Thompson, and Julian P. T. Higgins. 2015. "Predictive Distributions for Between-Study Heterogeneity and Simple Methods for Their Application in Bayesian Meta-Analysis." *Statistics in Medicine* 34 (6): 984–98. https://doi.org/10.1002/sim.6381.

Vehtari, A, A Gelman, and J Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC." *Statistics and Computing* 27: 1413–32.

Vlaanderen, Jelle, Nathaniel Rothman, Lützen Portengen, Qing Lan, Hans Kromhout, and Roel Vermeulen. 2009. "Flexible Meta-Regression to Assess the Shape of the Benzene–Leukemia Exposure–Response Curve." *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.0901127.

Yao, Y, A Vehtari, D Simpson, and A Gelman. 2018. "Using Stacking to Average Bayesian Predictive Distributions (with Discussion)." *Bayesian Anal.* 13 (3): 917–1007.

Supplemental Material - Data Harmonization

We provide detailed examples of data harmonization for data from an epidemiological cancer study (Stenehjem et al. 2015)), a biomarker study (Testa et al. 2005), and an experimental animal study (Li et al. 2006).

Stenehjem et al. (2015) is part of our primary set of studies that directly investigated the exposure-response relation between benzene and AML incidence. The authors examined the incidence of lympho-haematopoietic cancers due to benzene exposure among 24,917 male offshore oil workers followed during 1999-2011. The authors provide hazard ratios (HRs) and 95% confidence intervals (95%CIs) for three categories of cumulative benzene exposure in ppm-years when compared to unexposed workers. These can be directly used to calculate estimates for the logRR and its standard error by log-transforming reported HRs and 95%CIs as is customarily done for meta-analyses. The (3x3) covariance matrix for the combined set of logRRs was estimated using the approach outlined by Greenland and Longnecker (1992) and implemented in the *covar.logrr* function from the *dosresmeta* R package. To use this function, the user is required to provide additional details on the number of cases and the total number of subjects or person-years of follow-up, which were available from the paper by Stenehjem et al. (2015). Data harmonization steps for studies included in the leukemia set were the same as those used for the AML set.

Testa et al. (2005) studied chromosomal aberration frequencies (a cancer biomarker) in 62 benzene exposed workers. Benzene exposure was monitored using stationary sampling and expressed as a time-weighted average (TWA) exposure level. The prevalence of chromosomal aberrations was assessed in (at least) 200 metaphases and reported as the proportion of aberrations. For the controls this was 1.08% (sd: 0.81%) and for exposed workers 2.52% (sd: 1.58%). Cumulative benzene exposure was estimated by multiplying the reported average exposure for exposed workers with the duration of exposure. A logRR was estimated by subtracting the log-transformed average proportion of aberrations in unexposed workers (controls) from that in exposed workers. The SE was estimated by applying the delta rule to standard errors of the reported average proportions to account for the log-transformation and standard variance rules for the difference between two independent random variables. Data harmonization steps for other studies in the CA and MN sets were similar. For the (few) studies that included multiple exposure levels, the full covariance matrix of logRRs was estimated using the same approach as for AML and leukemia.

Li et al. (2006) exposed 10 mice to 300 ppm benzene for 6 hours/day, 5 days/week, for 26 weeks. Six out of 8 animals in the unexposed and 8 out of 10 animals in the exposed group developed lymphoma's in haematopoietic and lymphoid tissues. Cumulative exposure was estimated by multiplying the benzene exposure level by the exposure duration (in years). The RR [95%CI] was estimated to be 107 [0.64, 1.77] using standard relative risk calculations and were converted to a logRR and SE using the same approach as used for the AML and leukemia studies.

Table S1. Results from the data harmonization step for three selected studies.

Authors (year)	**Cumulative benzene exposure (ppm-years)**	**logRR**	**Covariance matrix**
Stenehjem *et al.* (2015)	0.02 0.08 0.54	0.34 -0.16 1.58	$\begin{bmatrix} 1{,}11 & 0.54 & 0.56 \\ 0.54 & 1.49 & 0.57 \\ 0.56 & 0.57 & 0.77 \end{bmatrix}$
Testa *et al.* (2005)	43	0.85	[[0.03]
Li *et al.* (2006)	150	0.06	[0.07]

Supplemental Material - Study Selection

Epidemiological studies

We used a study selection strategy similar to that used in Vlaanderen et al. (2009). Briefly, a PubMed and Scopus search was conducted for published peer-reviewed papers from 2008 onwards, using the key terms “benzene,” “humans,” “leukemia” or “AML” in combination with either “cohort studies” or “case–control studies”. Publications were selected if they had been peer reviewed and fulfilled the following criteria: 1) the study population had been occupationally exposed to benzene and 2) exposure assessment was based on quantitative benzene exposure measurements in air.

AML

All publications that had not been considered earlier by Vlaanderen et al. (2009) were reviewed for information on AML. Studies that included information only on combined AML and myelodysplastic syndromes (MDS) were included in the AML set, but note that MDS was classed separately from AML in the recent International Classification of Diseases 10. We did not consider publications where the authors applied a *post hoc* updated or alternative exposure assessment without extended follow-up.

We reviewed 187 studies of which 6 investigated the exposure-response relation between benzene exposure and AML, and were therefore included in the meta-regression (Table S2). For the study by Collins et al. (2003), we selected the results for acute nonlymphocytic leukemia (ANL) because it is considered to be very similar to AML (National Cancer Institute 2019).

Leukemia

We compiled a dataset of studies that provided information on the ERC between benzene and leukemia, but that were independent from (i.e. not already included) our AML study set. We selected 3 leukemia studies for inclusion in the meta-regression (Table S3).

Human biomarker studies

A review and meta-analysis of studies that investigated chromosomal aberrations (CA) and micronuclei (MN) in workers occupationally exposed to benzene is provided by Scholten et al. (2020). In brief, four studies were considered eligible for inclusion in this meta-analysis for CA, and seven for MN. The authors looked at the exposure-response relationship between current benzene exposure levels and the number of CA or the number of bi-nucleated cells with MN on an additive scale. Estimated slope estimates were an increase of 0.27% CA (95%CI: 0.08% - 0.47%; based on the results from 4 studies) and 0.27% MN (95%CI: -0.23% - 0.76%; based on the results from 7 studies) per ppm of benzene exposure.

Animal studies

We identified studies investigating benzene exposure effects in rodents by reviewing IARC's Monograph 120 (2018). Selected studies were required to have rodents exposed through inhalation for a minimum duration of 16 weeks, with tumors studied in hematopoietic and lymphoid tissues, and with a long observation period (over 22 months). For each study that matched our search criteria we extracted information on the species, strain, sex, the observation- and exposure period, dose, and endpoint. In total four studies were considered eligible for further analysis, accounting for 7 entries (taking into account different strains, sex and species) (Table S4).

Supplemental Material - AML Studies

Table S2. Overview of human epidemiological studies on benzene exposure and AML risk.

Reference (study)	Study design	Risk estimates	Country	Industry	Reference category	exposure category (ppm-years		Study outcome	ICD code	Study population size
						Lowest	upper			
(Stenehjem et al. 2015)	Stratified case-cohort design	HR	Norway	Offshore oil industry workers	Unexposed workers	0.001	0.948	Incidence	C92.0 (10)	24917 (subcohort of 1661 workers) 10 cases
(Schnatter et al. 2012)	Nested case-control (pooled from three individual studies)	OR	Australia, Canada, UK	Petroleum industry	Unexposed workers	<0.348	>2.93	Mortality and incidence	n.a.	46882 (60 case subjects)
(Collins et al. 2015)	Retrospective cohort	SMR	USA	Chemical industry	National and regional specific death rates	<3.9	>25	mortality	AML (10)	2266 (5 cases)
(Linet et al. 2019)	Retrospective cohort	HR	China	A variety of industries	Workers in same work units which were not exposed	<5	>300	mortality and incidence	AML and MDS (9)	1500 (from 110631 cohort members) (44 cases)
(Otto Wong and Raabe 1995)	Cohort	SMR	USA	Chemical industry	National population death rates	<40	>400	Mortality	n.a.	1165*
(Collins et al. 2003)	Cohort	SMR	USA	Chemical Industry	National and regional specific death rates	<1	>6	Mortality	ANL (8)	4417 (7 cases)

*ICD: International Classification of Disease, * based on Rinsky et al. (1987).*

Supplemental Material - Leukemia Studies

Table S3. Overview of human epidemiological studies on benzene exposure and AML or leukemia risk.

Reference (study)	in AML dataset	in leukemia dataset
(Collins et al. 2003)	Yes	No
(Collins et al. 2015)	Yes	No
(Linet et al. 2019)	Yes	No
(Schnatter et al. 2012)	Yes	No
(Stenehjem et al. 2015)	Yes	No
(Wong and Raabe 1995)	Yes	No
(Constantini et al. 2003)	No	Yes
(Glass et al. 2003)	No	No: this study is included Schnatter (2012)
(Hayes et al. 1997)	No	No: this study is updated by Linet et al. (2019)
(Rinsky et al. 2002)	No	No: the pliofilm study is already "represented" in Wong (1994)
(Rushton 1997)	No	No: this study is included Schnatter (2012)
(Schnatter et al. 1996)	No	No: this study is included Schnatter (2012)
(Swaen et al. 2005)	No	Yes
(Wong 1987)	No	Yes
(Bloemen et al. 2004)	No	No: updated by Collins et al. (2015)

** Based on Vlaanderen et al. (2009)*

Supplemental Material - Animal Studies

Table S4. Overview of animal (rodent) studies on inhalatory benzene exposure and occurrence of cancer in haematopoietic and lymphoid tissues.

Reference	Observation period Exposure period	Dose (ppm)	Species Strain (sex)	Tumour incidence	Notes
Cronkite, (1989) (Cronkite et al. 1989)	Life time 16 weeks (6 h/d, 5 d/w)	300	Mice CBA/Ca BNL (male)	7/60, 12/57	Cases of lymphomatous and myelogenous neoplasms added together
		300	Mice CBA/Ca BNL (female)	6/60, 10/54	
		100	Mice CBA/Ca BNL (male)	12/70, 9/85	
Farris (1993) (Farris et al. 1993)	22 months 16 weeks (6h/d, 5 d/w)	300	Mice CBA/Ca (m)	2/119, 14/118	Malignant lymphoma
Kawasaki (2009) (Kawasaki et al. 2009)	Life time 26 weeks (6h/d, 5 d/w)	33	Mice C57BL/6 (wildtype, male)	2/20, 4/19	Combination of neoplasm of haematopoietic and lymphoid tissues
		100		2/20, 3/19	
		300		2/20, 10/18	
		100	C3H/He (wildtype, male)	2/23, 6/24	
		300		2/23, 7/23	
Li (2006) (Li et al. 2006)	Life time 26 weeks (6hr/day , 5 d/w)	300	Mice C57BL/6 (wildtype)	6/8, 8/10	Thymic and non-thymic lymphoma added together

Supplemental Material - Model Fit

Table S5. Model fit for a regression-spline based model with intercept, for the combined dataset either including or excluding the study by Surralles et al (Surralles et al. 1997)

			dataset including Surralles et al.			dataset excluding Surralles et al.		
Outcome	**Study**	**Benzene exposure (ppm-yrs)**	**log(RR)**	**SE log(RR)**	**scaled residual***	**log(RR)**	**SE log(RR)**	**scaled residual***
AML	Stenehjem_2015	0.02	0.34	1.05	0.04	0.34	1.05	0.11
AML	Stenehjem_2015	0.08	-0.16	1.22	0.44	-0.16	1.22	0.50
AML	Stenehjem_2015	0.54	1.58	0.88	-1.34	1.58	0.88	-1.27
AML	Collins_2015	1.95	-1.11	1.02	1.47	-1.11	1.02	1.53
AML	Collins_2015	14.45	0.63	0.55	-0.46	0.63	0.55	-0.29
AML	Collins_2015	41.67	0.33	0.66	0.26	0.33	0.66	0.38
AML	Schnatter_2012	1.47	0.04	0.38	0.94	0.04	0.38	1.11
AML	Schnatter_2012	4.71	0.33	0.37	0.21	0.33	0.37	0.39
AML	Rinsky_1987	100.00	-0.09	0.88	1.07	-0.09	0.88	1.08
AML	Rinsky_1987	300.00	3.30	0.66	-1.62	3.30	0.66	-1.57
AML	Rinsky_1987	666.67	4.59	0.55	-0.91	4.59	0.55	-0.87
AML	Linet_2019	2.10	0.39	0.50	0.01	0.39	0.50	0.15
AML	Linet_2019	18.10	-0.30	0.49	1.38	-0.30	0.49	1.57
AML	Linet_2019	63.10	0.92	0.44	-0.49	0.92	0.44	-0.44
AML	Linet_2019	167.60	0.36	0.53	1.05	0.36	0.53	1.04
AML	Linet_2019	584.60	1.04	0.57	0.46	1.04	0.57	0.46
AML	Collins_2003	0.50	0.34	0.66	0.05	0.34	0.66	0.17
AML	Collins_2003	3.50	0.99	0.66	-0.93	0.99	0.66	-0.81
AML	Collins_2003	10.00	0.79	0.67	-0.66	0.79	0.67	-0.51
Leukemia	Costantini_2003	20.00	0.26	0.53	0.23	0.26	0.53	0.41
Leukemia	Costantini_2003	69.50	1.41	0.65	-1.02	1.41	0.65	-0.99
Leukemia	Costantini_2003	149.50	0.92	0.66	0.08	0.92	0.66	0.08
Leukemia	Costantini 2003	333.33	1.63	0.48	-0.12	1.63	0.48	-0.12

			dataset including Surralles et al.			dataset excluding Surralles et al.		
Outcome	**Study**	**Benzene exposure (ppm-yrs)**	**log(RR)**	**SE log(RR)**	**scaled residual***	**log(RR)**	**SE log(RR)**	**scaled residual***
Leukemia	Wong_1987	37.46	-0.25	0.87	0.84	-0.25	0.87	0.93
Leukemia	Wong_1987	100.00	1.01	0.55	-0.25	1.01	0.55	-0.25
Leukemia	Swaen_2005	68.80	1.14	0.83	-0.49	1.14	0.83	-0.46
Animal	Cronkite_1989_m_300	92.05	0.59	0.44	0.61	0.59	0.44	0.63
Animal	Cronkite_1989_f_300	92.05	0.62	0.48	0.51	0.62	0.48	0.52
Animal	Cronkite_1989_m_100	30.68	-0.48	0.41	2.17	-0.48	0.41	2.38
Animal	Farris_1993	92.05	1.95	0.74	-1.48	1.95	0.74	-1.47
Animal	Kawasaki_2009_C57BL	16.45	0.74	0.80	-0.44	0.74	0.80	-0.33
Animal	Kawasaki_2009_C57BL	49.86	0.46	0.85	0.17	0.46	0.85	0.23
Animal	Kawasaki_2009_C57BL	149.59	1.71	0.70	-0.98	1.71	0.70	-0.98
Animal	Kawasaki_2009_C3H/He	49.86	1.06	0.76	-0.61	1.06	0.76	-0.53
Animal	Kawasaki_2009_C3H/He	149.59	1.25	0.75	-0.27	1.25	0.75	-0.27
Animal	Li_2006	149.59	0.06	0.26	2.65	0.06	0.26	2.65
CA	Bogadi1997	80.83	0.41	0.14	1.58	0.41	0.14	1.97
CA	Kim2004	30.99	0.63	0.14	-1.12	0.63	0.14	-0.56
CA	Sram2004	33.29	0.47	0.13	-0.35	0.47	0.13	0.32
CA	Testa2005	42.88	0.85	0.18	-1.61	0.85	0.18	-1.14
MN	Ren2018	27.38	0.62	0.07	-1.15	0.62	0.07	-0.82
MN	Fang2017	12.26	0.58	0.16	-1.08	0.58	0.16	-0.71
MN	Basso2011	0.33	0.52	0.12	-0.79	0.52	0.12	-0.51
MN	Kim2008	6.92	0.34	0.05	0.22	0.34	0.05	1.08
MN	Testa2005	42.77	0.80	0.17	-1.33	0.80	0.17	-1.01
MN	RomaTorres2006	4.75	0.73	0.21	-1.46	0.73	0.21	-1.25
MN	Surrales1997	8.90	-0.61	0.18	5.23	n.a.	n.a.	n.a.

** scaled residuals were estimated using cross-validation as the ratio of the difference between predicted and observed logRRs and its standard error*

Supplemental Material - Example Parameter Estimates

Table S6. Parameter estimates for the shared parameters and between-study covariance for a linear model with intercept fitted to the full dataset.

		Estimate [95%CI]
Level 1	Shared parameters	
	Intercept (β_0)	0.44 [0.29, 0.59]
	Slope (β_1)	0.37 [0.06, 0.68]
Level 2	Between study (co)variance	
AML	$\sigma^2(\beta_0)^*$	0.05 [0.00, 0.50]
	$\sigma^2(\beta_1)^‡$	0.27 [0.02, 2.34]
	$\sigma(\beta_0, \beta_1)^\#$	0.00 [-0.82, 0.81]
Leukemia	$\sigma^2(\beta_0)^*$	0.13 [0.00, 1.73]
	$\sigma^2(\beta_1)^‡$	0.13 [0.00, 1.90]
	$\sigma(\beta_0, \beta_1)^\#$	-0.07 [-0.86, 0.79]
CA	$\sigma^2(\beta_0)^*$	0.31 [0.00, 1.85]
	$\sigma^2(\beta_1)^‡$	0.19 [0.00, 1.56]
	$\sigma(\beta_0, \beta_1)^\#$	-0.04 [-0.82, 0.80]
MN	$\sigma^2(\beta_0)^*$	0.05 [0.00, 0.56]
	$\sigma^2(\beta_1)^‡$	0.22 [0.00, 2.17]
	$\sigma(\beta_0, \beta_1)^\#$	-0.10 [-0.86, 0.77]
Animal	$\sigma^2(\beta_0)^*$	0.02 [0.00, 0.16]
	$\sigma^2(\beta_1)^‡$	0.31 [0.00, 3.69]
	$\sigma(\beta_0, \beta_1)^\#$	-0.08 [-0.85, 0.79]

* *= between study variation in intercepts*
‡ *= between study variation in slopes*
$^\#$ *= correlation between random intercepts and slopes*

Supplemental Material - Different Scales For The Half-Cauchy Prior

Table S7. Summary of the prior-predictive distribution of Relative Risks (per 100 ppm-yrs of benzene exposure) from a linear model without intercept and assuming an overall (consensus) RR of 1. Random effects were generated using half-Cauchy priors with different scales and are summarized across different ranges of the RR.

	Cumulative distribution			
Scale	[1/2 - 2]	[1/5 - 5]	[1/10 - 10]	[<1/10 or >10]
0.5	70%	85%	89%	11%
1	54%	73%	80%	20%
5	21%	36%	44%	56%

Table S8. Predicted Relative Risks for benzene-induced AML, for benzene exposure at selected exposure levels based on models fitted to the full dataset and using a half-Cauchy prior for the between-study standard deviation with a scale of 0.5. First row for each benzene exposure: point estimate, second row: 95% credible interval, third row: 95% prediction interval.

Benzene exposure (ppm-yrs)	Linear (no intercept)	Linear (with intercept)	Spline (no intercept)	Spline (with intercept)
0	n.a.	1.55	n.a.	1.57
		(1.35 – 1.78)		(1.20 – 2.02)
		(1.05 – 2.79)		(1.01 – 3.02)
5	1.03	1.58	1.19	1.57
	(1.02 – 1.05)	(1.39 – 1.80)	(1.09 – 1.27)	(1.30 – 1.89)
	(0.97 – 1.11)	(1.07 – 2.86)	(0.94 – 1.53)	(1.01 – 3.05)
10	1.07	1.61	1.39	1.58
	(1.04 – 1.11)	(1.42 – 1.82)	(1.19 – 1.57)	(1.32 – 1.87)
	(0.95 – 1.22)	(1.08 – 2.92)	(0.89 – 2.23)	(0.92 – 3.34)
20	1.14	1.67	1.73	1.63
	(1.08 – 1.23)	(1.48 – 1.88)	(1.38 – 2.08)	(1.26 – 2.00)
	(0.90 – 1.49)	(1.08 – 3.11)	(0.86 – 3.66)	(0.79 – 4.13)
40	1.30	1.80	2.01	1.80
	(1.16 – 1.52)	(1.58 – 2.04)	(1.63 – 2.45)	(1.36 – 2.25)
	(0.81 – 2.23)	(1.03 – 3.72)	(0.91 – 4.77)	(0.81 – 4.90)

Table S9. Predicted Relative Risks for benzene-induced AML, for benzene exposure at selected exposure levels based on models fitted to the full dataset and using a half-Cauchy prior for the between-study standard deviation with a scale of 5. First row for each benzene exposure: point estimate, second row: 95% credible interval, third row: 95% prediction interval.

Benzene exposure (ppm-yrs)	Linear (no intercept)	Linear (with intercept)	Spline (no intercept)	Spline (with intercept)
0	n.a.	1.55	n.a.	1.60
		(1.32 – 1.82)		(1.10 – 2.19)
		(0.99 – 3.35)		(0.91 – 3.98)
5	1.03	1.58	1.15	1.57
	(1.02 – 1.06)	(1.35 – 1.84)	(1.00 – 1.31)	(1.16 – 1.98)
	(0.93 – 1.18)	(1.00 – 3.43)	(0.71 – 2.13)	(0.81 – 4.31)
10	1.07	1.61	1.32	1.54
	(1.04 – 1.13)	(1.38 – 1.86)	(1.00 – 1.66)	(1.12 – 1.98)
	(0.86 – 1.40)	(1.00 – 3.59)	(0.53 – 4.14)	(0.59 – 5.77)
20	1.15	1.67	1.61	1.55
	(1.07 – 1.28)	(1.44 – 1.92)	(1.06 – 2.28)	(0.99 – 2.20)
	(0.74 – 1.95)	(0.95 – 4.01)	(0.38 – 9.76)	(0.38 – 9.89)
40	1.31	1.80	1.93	1.75
	(1.15 – 1.63)	(1.54 – 2.11)	(1.28 – 2.78)	(1.09 – 2.57)
	(0.55 – 3.80)	(0.80 – 5.54)	(0.37 – 15.47)	(0.34 – 13.62)

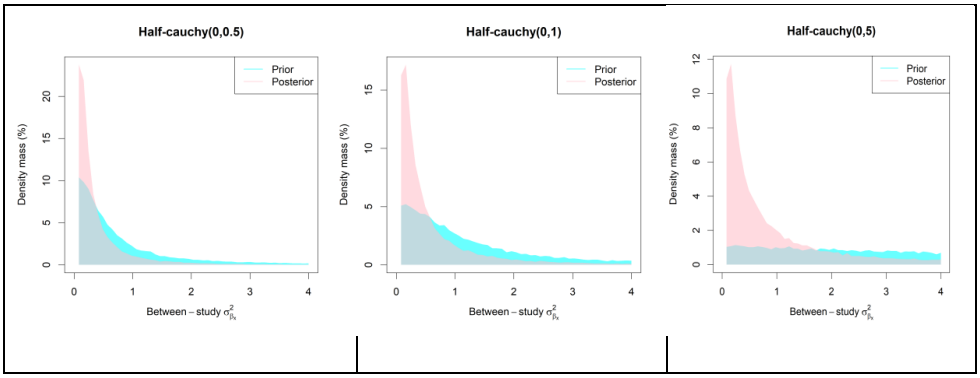


Figure S1. Prior-posterior plots for the estimated between-study variation, for a linear model with intercept, fitted using only the AML data. We see that the use of the default prior used in our analyses (i.e. a Half-Cauchy of 1) is sufficient to cover the posterior estimates. However, a value of 0.5 would have also been an option. A Half-Cauchy of 5 is virtually uninformative.

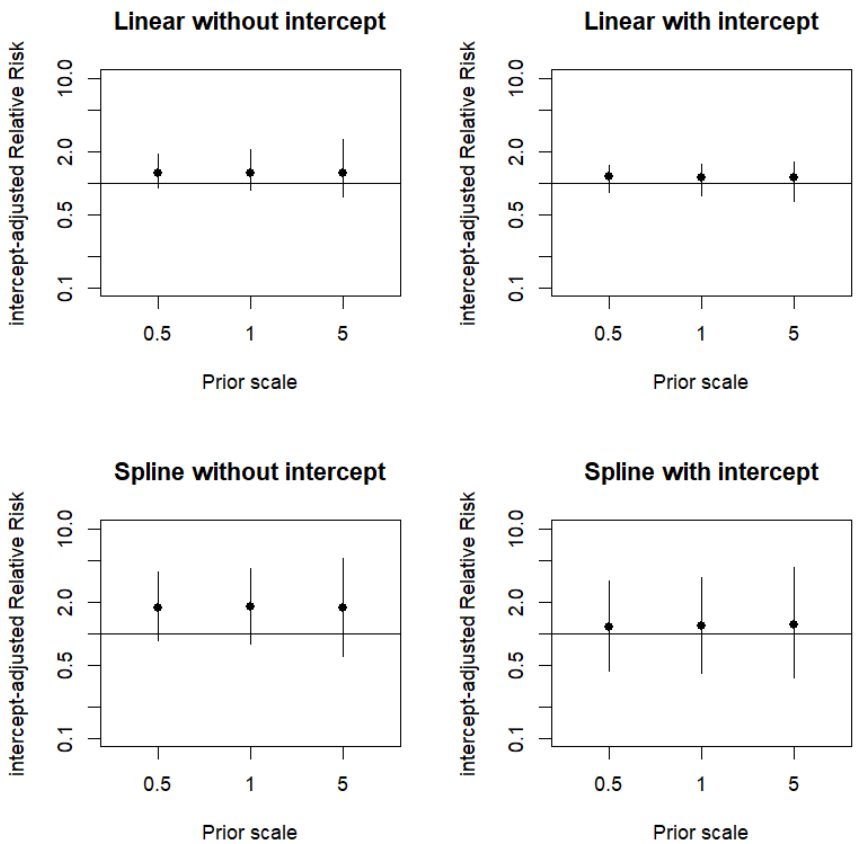


Figure S2. Estimated intercept-adjusted relative risks and 95%CI from models fitted to the AML dataset only, and using different scales for the half-Cauchy prior.

Supplemental Material - Excluding Exposures > 40 PPM-YRS

Table S10. Predicted Relative Risks for benzene-induced AML, for benzene exposure at selected exposure levels based on models fitted to a dataset that included only exposures below 40 ppm-yrs and using a half-Cauchy prior for the between-study standard deviation with a scale of 1. First row for each benzene exposure: point estimate, second row: 95% credible interval, third row: 95% prediction interval

Benzene exposure (ppm-yrs)	Linear (no intercept)	Linear (with intercept)	Spline (no intercept)	Spline (with intercept)
0	n.a.	1.57		1.60
		(1.23 - 2.02)	n.a.	(1.11 – 2.19)
		(0.96 – 3.57)		(0.90 – 3.82)
	1.09	1.58	1.29	1.59
5	(1.02 – 1.15)	(1.28 – 1.94)	(1.08 – 1.46)	(1.22 – 2.01)
	(0.89 – 1.37)	(0.94 – 3.63)	(0.77 – 2.46)	(0.80 – 4.41)
	1.18	1.58	1.58	1.59
10	(1.04 – 1.32)	(1.30 – 1.89)	(1.16 – 1.97)	(1.16 – 2.11)
	(0.79 – 1.89)	(0.88 – 3.93)	(0.61 – 5.27)	(0.59 – 6.22)
	1.39	1.60	1.88	1.58
20	(1.09 – 1.75)	(1.20 – 1.95)	(1.27 – 2.48)	(1.02 – 2.29)
	(0.63 – 3.57)	(0.68 – 5.13)	(0.44 – 12.06)	(0.38 – 10.17)
	1.94	1.64	1.33	1.53
40	(1.19 – 3.08)	(0.89 – 2.43)	(0.70 – 2.33)	(0.67 – 3.04)
	(0.39 – 12.76)	(0.34 – 10.75)	(0.21 – 11.65)	(0.25 – 12.94)

Supplemental Material - Uncertainty Factors

Table S11. Predicted Relative Risks for benzene-induced AML, for benzene exposure at selected exposure levels based on a linear model with intercept fitted to the full dataset and using a half-Cauchy prior for the between-study standard deviation with a scale of 1. Uncertainty factors included a 10-fold factor for the animal-based studies and a 3-fold factor for the CA and MN studies. First row for each benzene exposure: point estimate, second row: 95% credible interval, third row: 95% prediction interval

Benzene exposure (ppm-yrs)	**without uncertainty factors***	**with uncertainty factors**
0	1.55	1.49
(intercept)	(1.34 – 1.80)	(1.12 – 1.97)
	(1.00 – 3.16)	(0.94 – 3.17)
5	1.58	1.52
	(1.37 – 1.82)	(1.16 – 2.00)
	(1.01 – 3.22)	(0.96 – 3.22)
10	1.61	1.55
	(1.40 – 1.84)	(1.19 – 2.03)
	(1.02 – 3.30)	(0.97 – 3.31)
20	1.67	1.61
	(1.46 – 1.90)	(1.25 – 2.10)
	(1.02 – 3.51)	(0.96 – 3.55)
40	1.80	1.75
	(1.56 – 2.07)	(1.35 – 2.28)
	(0.94 – 4.21)	(0.89 – 4.45)

* results identical to those in table 3 in the main paper

Supplemental Material - Impact Assessment

Table S12. Maximum Tolerable Levels of benzene exposure (in ppm) based on different exposure-response models fitted to different datasets.*

	AML	AML + leukemia	All data	<40 ppm-yrs
Linear with intercept	4.9	3.4	3.0	35.9
Linear with intercept – intercept subtracted	8.1	5.8	6.0	74.9
Linear with intercept – interpolation	5.0	3.5	3.1	36.9
Spline no intercept	4.0	2.6	3.4	n.a.
Linear no intercept	3.9	3.8	3.0	1.4

** 40 additional cases of AML per 10,000 exposed workers*

Table S13. Acceptable Risk Levels of benzene exposure (in ppm) based on different exposure-response models fitted to different datasets.*

	AML	AML + Leukemia	AML+ All data	<40 ppm.yrs
Linear with intercept	n.a.	n.a.	n.a.	n.a.
Linear with intercept – intercept subtracted	0.13	0.095	0.098	1.23
Linear with intercept – interpolation	0.0002	0.0002	0.0003	0.0002
Spline no intercept	0.014	0.02	0.011	n.a.
Linear no intercept	0.063	0.062	0.054	0.022

** 40 additional cases of AML per 1,000,000 exposed workers*

References Supplemental Material

Bloemen, L. J., A. Youk, T. D. Bradley, K. M. Bodner, and G. Marsh. 2004. "Lymphohaematopoietic Cancer Risk among Chemical Workers Exposed to Benzene." *Occupational and Environmental Medicine*. https://doi.org/10.1136/oem.2003.007013.

Collins, J. J., B. Ireland, C. F. Buckley, and D. Shepperly. 2003. "Lymphohaematopoeitic Cancer Mortality among Workers with Benzene Exposure." *Occupational and Environmental Medicine*. https://doi.org/10.1136/oem.60.9.676.

Collins, James J., Steven E. Anteau, Gerard M.H. Swaen, Kenneth M. Bodner, and Catherine M. Bodnar. 2015. "Lymphatic and Hematopoietic Cancers Among Benzene-Exposed Workers." *Journal of Occupational and Environmental Medicine*. https://doi.org/10.1097/JOM.0000000000000324.

Constantini, Adele Seniori, Margaret Quinn, Dario Consonni, and Marco Zappa. 2003. "Exposure to Benzene and Risk of Leukemia among Shoe Factory Workers." *Scandinavian Journal of Work, Environment and Health*. https://doi.org/10.5271/sjweh.704.

Cronkite, E. P., R. T. Drew, T. Inoue, Y. Hirabayashi, and J. E. Bullis. 1989. "Hematotoxicity and Carcinogenicity of Inhaled Benzene." *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.898297.

Farris, G. M., J. I. Everitt, R. D. Irons, and J. A. Popp. 1993. "Carcinogenicity of Inhaled Benzene in CBA Mice." *Fundamental and Applied Toxicology*. https://doi.org/10.1006/faat.1993.1061.

Glass, Deborah C., Christopher N. Gray, Damien J. Jolley, Carl Gibbons, Malcolm R. Sim, Lin Fritschi, Geoffrey G. Adams, John A. Bisby, and Richard Manuell. 2003. "Leukemia Risk Associated with Low-Level Benzene Exposure." *Epidemiology*. https://doi.org/10.1097/01.ede.0000082001.05563.e0.

Greenland, Sander, and Matthew P. Longnecker. 1992. "Methods for Trend Estimation from Summarized Dose-Response Data, with Applications to Meta-Analysis." *American Journal of Epidemiology*. https://doi.org/10.1093/oxfordjournals.aje.a116237.

Hayes, Richard B., Song Nian Yin, Mustafa Dosemeci, Gui Lan Li, Sholom Wacholder, Lois B. Travis, Chin Yang Li, Nathaniel Rothman, Robert N. Hoover, and Martha S. Linet. 1997. "Benzene and the Dose-Related Incidence of Hematologic Neoplasms in China." *Journal of the National Cancer Institute*. https://doi.org/10.1093/jnci/89.14.1065.

Kawasaki, Yasushi, Yoko Hirabayashi, Toyozo Kaneko, Jun Kanno, Yukio Kodama, Yuuko Matsushima, Yukio Ogawa, et al. 2009. "Benzene-Induced Hematopoietic Neoplasms Including Myeloid Leukemia in Trp 53-Deficient C57BL/6 and C3H/He Mice." *Toxicological Sciences*. https://doi.org/10.1093/toxsci/kfp107.

Li, Guang Xun, Yoko Hirabayashi, Byung Il Yoon, Yasushi Kawasaki, Isao Tsuboi, Yukio Kodama, Yuji Kurokawa, Junji Yodoi, Jun Kanno, and Tohru Inoue. 2006. "Thioredoxin Overexpression in Mice, Model of Attenuation of Oxidative Stress, Prevents Benzene-Induced Hemato-Lymphoid Toxicity and Thymic Lymphoma." *Experimental Hematology*. https://doi.org/10.1016/j.exphem.2006.08.005.

Linet, Martha S., Ethel S. Gilbert, Roel Vermeulen, Graça M. Dores, Song Nian Yin, Lutzen Portengen, Richard B. Hayes, et al. 2019. "Benzene Exposure Response and Risk of Myeloid Neoplasms in Chinese Workers: A Multicenter Case-Cohort Study." *Journal of the National Cancer Institute*. https://doi.org/10.1093/jnci/djy143.

National Cancer Institute. 2019. "NCI Dictionary of Cancer Terms." 2019.

https://www.cancer.gov/publications/dictionaries/cancer-terms/def/acute-nonlymphocytic-leukemia.

Rinsky, Robert A., R. W. Hornung, S. R. Silver, and C. Y. Tseng. 2002. “Benzene Exposure and Hematopoietic Mortality: A Long-Term Epidemiologic Risk Assessment.” *American Journal of Industrial Medicine*. https://doi.org/10.1002/ajim.10138.

Rinsky, Robert A., Alexander B. Smith, Richard Hornung, Thomas G. Filloon, Ronald J. Young, Andrea H. Okun, and Philip J. Landrigan. 1987. “Benzene and Leukemia.” *New England Journal of Medicine*. https://doi.org/10.1056/nejm198704233161702.

Rushton, Lesley. 1997. “A Case-Control Study to Investigate the Risk of Leukaemia Associated with Exposure to Benzene in Petroleum Marketing and Distribution Workers in the United Kingdom.” *Occupational and Environmental Medicine*. https://doi.org/10.1136/oem.54.3.152.

Schnatter, A. Robert, Thomas W. Armstrong, Mark J. Nicolich, Florence S. Thompson, Arnold M. Katz, Wendy W. Huebner, and Eileen D. Pearlman. 1996. “Lymphohaematopoietic Malignancies and Quantitative Estimates of Exposure to Benzene in Canadian Petroleum Distribution Workers.” *Occupational and Environmental Medicine*. https://doi.org/10.1136/oem.53.11.773.

Schnatter, A. Robert, Deborah C. Glass, Gong Tang, Richard D. Irons, and Lesley Rushton. 2012. “Myelodysplastic Syndrome and Benzene Exposure among Petroleum Workers: An International Pooled Analysis.” *Journal of the National Cancer Institute*. https://doi.org/10.1093/jnci/djs411.

Scholten, B, J Vlaanderen, R Stierum, L Portengen, N Rothman, Q Lan, A Pronk, and R Vermeulen. 2020. “A Quantitative Meta-Analysis of the Relation between Occupational Benzene Exposure and Biomarkers of Cytogenetic Damage.” *Environmental Health Perspectives* 128 (https://doi.org/10.1289/EHP6404).

Stenehjem, J. S., K. Kjærheim, M. Bråtveit, S. O. Samuelsen, F. Barone-Adesi, N. Rothman, Q. Lan, and T. K. Grimsrud. 2015. “Benzene Exposure and Risk of Lymphohaematopoietic Cancers in 25 000 Offshore Oil Industry Workers.” *British Journal of Cancer*. https://doi.org/10.1038/bjc.2015.108.

Surralles, J, K Autio, L Nylund, H Jarventaus, H Norppa, T Veidebaum, M Sorsa, and K Peltonen. 1997. “Molecular Cytogenetic Analysis of Buccal Cells and Lymphocytes from Benzene-Exposed Workers.” *Carcinogenesis* 18 (4): 817–23. https://doi.org/10.1093/carcin/18.4.817.

Swaen, Gerard M.H., Theo Scheffers, Johan De Cock, Jos Slangen, and Hinkelien Drooge. 2005. “Leukemia Risk in Caprolactam Workers Exposed to Benzene.” *Annals of Epidemiology*. https://doi.org/10.1016/j.annepidem.2004.03.007.

Testa, A., F. Festa, R. Ranaldi, M. Giachelia, D. Tirindelli, A. De Marco, M. Owczarek, M. Guidotti, and R. Cozzi. 2005. “A Multi-Biomarker Analysis of DNA Damage in Automobile Painters.” *Environmental and Molecular Mutagenesis* 46 (3): 182–88. https://doi.org/10.1002/em.20147.

Vlaanderen, Jelle, Nathaniel Rothman, Lützen Portengen, Qing Lan, Hans Kromhout, and Roel Vermeulen. 2009. “Flexible Meta-Regression to Assess the Shape of the Benzene–Leukemia Exposure–Response Curve.” *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.0901127.

Wong, O. 1987. “An Industry Wide Mortality Study of Chemical Workers Occupationally Exposed to Benzene. II. Dose Response Analyses.” *British Journal of Industrial Medicine*. https://doi.org/10.1136/oem.44.6.382.

Wong, Otto, and Gerhard K. Raabe. 1995. “Cell-Type-Specific Leukemia Analyses in a Combined Cohort of More than 208,000 Petroleum Workers in the United States and the United Kingdom, 1937-1989.” *Regulatory Toxicology and Pharmacology*. https://doi.org/10.1006/rtph.1995.1044.

Chapter 7

Discussion

Evidence integration for risk assessment is commonly done in a qualitative manner. For more quantitative approaches, the focus on evidence synthesis is typically within the same stream of scientific evidence. The goal of this thesis was to develop and test approaches to integrate different types of evidence for chemical risk assessment in a quantitative and systematic manner, both within and across evidence streams. For this purpose, advanced statistical and biological informed tools and approaches were used including meta-analysis, meta-regression, text mining, systematic reviews, systems biology, physiological based kinetic (PBK) models, and Bayesian inference. The evidence considered was published data of the following evidence streams: epidemiologic studies, molecular epidemiology studies, animal studies and mechanistic assays. Below I first consider how advanced approaches applied in my thesis improve and facilitate the risk assessment of two proof-of-principle chemicals: benzene and diisocyanates. Subsequently I will reflect on overarching methodological issues that emerged over the course of my work.

Implications for the regulatory risk assessment of benzene and diisocyanates

Diisocyanates

Diisocyanates are a group of chemicals containing two isocyanate functional groups. Most used diisocyanates are 4,4'-methylene diphenyl diisocyanate (MDI), toluene diisocyanate (TDI) and hexamethylene diisocyanate (HDI). Recently the Risk Assessment Committee (RAC 2020) identified respiratory health effects as the critical endpoints related to diisocyanate exposure. Measuring inhalation exposure to diisocyanates is technically challenging due to their reactivity (The Dutch Health Council 2018). In addition, air sampling does not provide information on dermal exposure and cannot account for the efficacy of respiratory protective equipment (RPE) that is often used. Therefore a substantial number of cross-sectional biomarker studies have been performed to study diisocyanate exposure, most notably by measuring diamines in urine. Chapter 3 (Biomonitoring for occupational exposure to diisocyanates: a systematic review) aimed to systematically summarize all exposure biomonitoring data in occupational settings to get a better understanding of diisocyanate exposure levels across various occupations, industries, and countries. I also identified research gaps which included understudied sectors as well as large differences in sampling strategies and analytical techniques. This information was used for the design of a new diisocyanate field study undertaken in 2021 in five European countries, with particular emphasis on the understudied sectors and using harmonized study protocols.

Chapter 4 focused on improving the interpretation of diisocyanate exposure biomonitoring data by developing a PBK model to better characterize the relationship between external exposure and urinary excretion of diamines. Using PBK sensitivity analysis, several factors and scenarios were identified and quantified that had an

impact on the relation between airborne diisocyanate levels and urinary amines. PBK models can also be used for derivation of biological limit values (BLVs), so I compared the current available BLVs for diisocyanates to our model predictions. For TDI the Deutsche Forschungsgemeinschaft (DFG) proposed a BLV based on correlation formula's between air TDI and urinary amines (Hartwig and MAK Commission 2021). The DFG derived a value of 6.13 µg TDA/gr creatinine based on the MAK value of 7 µg/m^3 and proposed a BLV of 5 µg TDA/gr creatinine. With our model I predicted urinary values of 8.7 µg TDA/gr creatinine, when modelling the last post-shift urine sample, so comparable to the value based on the correlation formulae. The DFG also used correlation formula's for MDI in air and MDA in urine, and proposed a value of 10 µg MDA/gr creatinine based on exposure to a MAK value of 0.05 mg/m^3 (DFG 2007). With our model we predicted MDA urinary values of 22 µg/gr creatinine, post-shift. Note that recently the Risk Assessment Committee (RAC) proposed a different assessment (RAC 2020). The Committee conclude that exposures > 0.67 µg/m^3 NCO[2] in air would result in 5% excess risk for hyperresponsiveness over a working life period. Given that our model predicts urinary levels of 1 and 1.4 µg/gr creatinine for TDI and MDI respectively, at this critical exposure level, a purely health based BLV should also be lower than the current values of 5 µg TDA/gr creatinine and 10 µg MDA/gr creatinine.

Benzene

Benzene is a human carcinogen with wide industrial applications and a large global production volume. In 2018 the International Agency for Research on Cancer (IARC) reaffirmed the classification of benzene as a group 1 human carcinogen and its role in the etiology of acute myeloid leukemia (AML) (IARC 2018). Occupational exposure to benzene has been associated with toxicity to the bone marrow and blood. There are several candidates of human biomarkers for assessing benzene exposure and its effect. For effect biomarkers, dozens of human biomonitoring studies are available that studied blood cell counts as well as markers of genetic damage, including micronuclei (MN) and chromosomal aberrations (CA) (IARC 2018).

In 2014 the Dutch Health Council Committee recommended a health-based occupational exposure limit of 0.2 ppm for benzene. This exposure limit was primarily based on biomarker studies focusing on hematological effects in humans (Health Council of the Netherlands 2014). The Health Council considered benzene as having an indirect, non-stochastic genotoxic mode of action and applied a threshold approach based on three studies that found hematological effects below 1 ppm of exposure. The Health Council focused on hematological effects because they considered studies on

[2] Isocyanates are characterized by the presence of [NCO] groups. RAC expressed urinary amine concentrations as NCO group equivalents because diisocyanates share a common mechanism of inducing hypersensitivity reactions and RAC concludes there is not enough data to assess differences in potency for different diisocyanates. The conversion from NCO groups to diisocyanates (i.e. µg TDA) and vice versa, is based on the total molecular weight of the diisocyanates: 174.2, 168.2, 250.3 for TDI, HDI and MDI respectively

the induction of CA to be limited in design and therefore, could not serve as a basis to derive a reliable point of departure.

In Chapter 5, I similarly recognize and describe large differences in study quality for CA, an issue on which I reflect in more detail in the third paragraph below. With regards to benzene's mechanisms of action, there is evidence that benzene has direct genotoxic properties (i.e. DNA double-stranded breaks) (IARC 2018). This is also illustrated in the benzene network for key characteristics of cancer (KCC) on genotoxicity and DNA repair in Chapter 2, which was developed using a text mining approach combined with systems biology. For benzene the preferred choice would be to derive an ERC for a recognized endpoint, as was done in Chapter 6.

The benzene biomarker studies of the meta-analysis of Chapter 5 were also included in the meta-regression on benzene induced AML (Chapter 6). AML has a strong causal link to benzene exposure (IARC 2018). Vlaanderen et al. (2009) used meta-regression to derive an exposure response curve (ERC) for total leukemia, which includes AML and other subtypes for which a causal relation to benzene exposure has not been shown. Since an accurate description of the AML-benzene ERC was still needed for impact and risk assessment, Chapter 6 aimed to combine human AML data with data from closely related study domains, e.g. human epidemiological data on leukemia or cancer biomarkers and animal experimental data.

Using the same model and exposure categorization, our predictions were slightly higher and more precise when including all available data as opposed to when using only AML studies. However, as described in detail in Chapter 6 and in the third paragraph below, it is not easy to understand to which extent our approach resulted in more precise estimates given the uncertainty introduced by combining different types of evidence streams.

Implications for risk assessment and evidence integration in general

It is not straightforward to evaluate to which extent the quantitative evidence synthesis approaches described in this thesis increased precision, but they do improve transparency

Evidence synthesis is widely used for summarizing findings of multiple studies. To determine which studies should be given more or less weight, studies are typically evaluated using ranking systems. There is criticism that qualitative weight of evidence approaches contain too much subjectivity (Linkov et al. 2015). Others claim that ranking methods can be too rigorous by *a priori* excluding or down grading observational studies, because these are often judged to have a high potential for bias (Steenland et al. 2020).

Throughout the chapters of this thesis, I applied quantitative and systematic methods for finding and summarizing evidence, using well defined search terms and systematic review approaches. In addition I used quantitative approaches for quantitatively

combining data, such as meta-analysis. Meta-analysis are well-known for summarizing data and providing more precise effects estimates with enhanced statistical power compared to individual studies (Hernández and Tsatsakis 2017). This is particularly true when evidence on a topic is extensive. In addition, meta-analysis can help obtain scientific consensus and reduce the risk of a single "best" study influencing risk assessment and resulting policy decisions (Wittwehr et al. 2020). However, although the use of meta-analysis can increase transparency, a generally acknowledged problem is the variation in results across studies (Song et al. 2001). For the meta-analysis performed to derive exposure-response curves for benzene exposure in relation to CA and MN induction (Chapter 5), I also found considerable heterogeneity between studies (even though the majority of studies indicated a positive association between benzene exposure and cytogenetic endpoints), which hampered the quantitative interpretation of the retrieved slope estimate. Possible contributors for the observed heterogeneity in the CA and MN studies included the limited number of exposure groups, differences in exposure assessment methods, and uncertainty in both exposure levels and protocols to derive CA or MN counts. Using a quantitative approach, I provide risk assessors insight into the heterogeneity that exists in the benzene-cytogenetic damage literature, information that is crucial when evidence from this literature (either from a single study or from a meta-analysis) is interpreted in a risk assessment. Furthermore, by carefully investigating heterogeneity I could identify research gaps and suggest directions for future research. For example, issues with regards to study design could be (partly) resolved by using more harmonized protocols. This highlights the value of projects such as HBM4EU, in which a significant effort was placed on developing and using harmonized standard operating procedures (SOPs) for sample collection, analysis, and reporting (Galea et al. 2021).

I also attempted to characterize study heterogeneity in Chapter 6, where I used a meta-regression approach to integrate epidemiology, molecular epidemiology and animal data. There were no large differences in heterogeneity between the various evidence streams considered, but this could be because it was difficult to precisely estimate heterogeneity as there were only a few studies per evidence stream. A Bayesian approach was used which required the use of a specified prior for between study heterogeneity. I showed that the choice of this prior is influential while there is no general guidance on the best prior value. In a commentary on our evidence integration paper, Keil (2022) argued that heterogeneity should have been further examined, for example by stratifying studies based on age or gender, rather than assuming heterogeneity is random. However this is less straightforward for an evidence integration approach based on different evidence streams. For instance, how should we stratify information on age from animal and human studies?

I further found that by including all available data for benzene the retrieved exposure-response curve (ERC) for AML was analytically more precise, but the uncertainty due to all required harmonization steps could not be quantified. It therefore remains difficult

to evaluate to which extent the derived ERC is an improvement for the risk assessment of benzene because it also relies on the assumptions risk assessors are willing to make, particularly with regards to the use of animal studies to inform on human health outcomes. For making such decisions it will be helpful if the principles of the methods chosen can be readily understood by assessors and risk managers (EFSA 2017). This is what my original aim was, by clearly identifying and highlighting the methodologies used together with the assumptions underlying the harmonization steps required for integrating data within and across different evidence streams.

Our evidence synthesis approaches were mostly resource intensive and complex; the use of artificial intelligence and generic open source tooling should further be explored

For the work described in this thesis on evidence synthesis, several statistical (systematic reviews, meta-analysis, meta-regression) and in silico chemical-specific (PBK modelling) approaches were applied to be able to quantitatively use evidence within and across evidence streams. These approaches are typically resource intensive. For example, a chemical-centered systematic review takes between 6 months to 2 years to complete, so the work burden limits the scope of what can be covered (Tsertsvadze et al. 2015; Miake-Lye et al. 2016). At the same time, many international chemical management policies have increasing demands for data on the safety of chemicals, and a high number of chemicals remain to be evaluated (Kavlock et al. 2018).

I explored whether an automated approach using text mining in conjunction with systems biology could increase efficiency of collecting and organizing mechanistic evidence (Chapter 2). Although developing the approach was resource intensive, the tooling can now be readily applied for a more efficient process of evidence identification for hazard identification, specifically for the identification of key characteristics of carcinogens, for which the Python programming pipeline was published.

Text mining is an example of artificial intelligence (AI), which is a set of theories and techniques used to create machines capable of simulating intelligence, performing functions that require intelligence when executed by humans (Lavecchia 2019). AI is considered promising for use in chemical risk assessment (Luechtefeld, Rowlands, and Hartung 2018; Wittwehr et al. 2020). An example is the establishment of the International Collaboration for the Automation of Systematic Reviews (ICASR) to coordinate efforts in natural language processing, text mining and machine learning for automatic review purposes (Beller et al. 2018). Beller et al. (2018) argued that many task of a systematic review are amenable for automation - from screening titles and abstracts to collation of meta-analysis results – but also recognized the need for manually reviewing and correcting automatically generated results in between systematic review steps. In addition, there should be attention for integrating evidence quality in automated approaches (Beller et al. 2018). I did not evaluate the studies

selected by our automated approach for informativeness or study quality after filtering steps. Although this could be adopted in the process by modifying search queries, it requires experts to stratify or limit the evidence base.

The most complex and resource intensive endeavor described in this thesis is the Bayesian evidence integration approach (Chapter 6). The work involved compiling evidence from all evidence streams, harmonizing data, and developing a Bayesian meta-regression model including prior elicitation. Perhaps AI tooling may help, although I foresee considerable more efforts are needed to use automated approaches for the evidence integration approach described in Chapter 6, given all required harmonization steps.

With regards to resource intensive endeavors for evidence synthesis, I have put considerable time and resource efforts in developing a diisocyanate specific PBK model. An alternative option could have been the use of open access generic PBK models (Paini et al. 2017). Recently Pletz et al. (2020) investigated the suitability and limitations of generic PBK models in deriving biological equivalent values for several compounds. The researchers evaluated two generic models and found high uncertainties around simulated metabolite concentrations in urine, and conclude that the approach might be more straightforward for more persistent chemicals that are analysed as parent compounds in blood. As diisocyanates readily degrade and have complicated kinetics, and I had specific research aims including exposure reconstruction for chronic exposures, there seemed to be no generic PBK tooling alternative available. But in general, the use of generic PBK open source tooling could be a solution for some applications, such as the derivation of a biological limit value for a persistent chemical.

Reporting of molecular epidemiological data should be enhanced to increase re-usability of this type of data stream within chemical risk assessment

In Chapters 4 and 5 where I aimed to quantitatively re-use published data, I regularly had to omit a molecular biomarker study because information was missing. In some cases this was because study reporting was not optimal. For example when there was no data on variation of estimates (i.e. a standard deviation or standard error), or a lack of adequate information on the used laboratory methods to analyse urine or blood samples. Those issues are generally covered by reporting guidelines for observational research such as the STrengthening Reporting of OBservational studies in Epidemiology (STROBE) initiative (Gallo et al. 2012).

Here I would like to reflect on studies where information was missing because it was simply outside the scope of that specific study to report specific information. For example, to be able to use diisocyanate biomarker studies for PBK model validation it is imperative to have access to contextual exposure information in addition to air and urine data. I demonstrated this in Chapter 4 where I compared scenarios and found

that if a worker is exposed for 4 hours in the morning without exposure in the afternoon, predicted post-shift urine levels after 8 hours were considerably higher compared to a worker with no morning exposure but 4 hours exposure in the afternoon. Diisocyanate biomarker studies that did not report when exposure started and/or how long exposure lasted were thus difficult to use for comparing with model predictions. The same counts for reporting on the use of personal protective equipment, or working history (to understand potential accumulation due to binding to protein adducts). Also for Chapter 6 I relied on information reported in molecular epidemiological studies, for example on working history to be able to convert daily exposure to ppm-years. These specific information requests are covered by the STROBE guidelines, but implicitly (i.e. “give characteristics of study participants and information on exposure and potential confounders”). Perhaps a standard framework with specific guidance on which data to report might help to re-use data for those studies where information is collected but not reported. I believe that, if this information is readily available, results from molecular epidemiology studies could be better re-used for evidence synthesis in various applications such as PBK model validation, meta-analysis or derivation of biological limit values from biomarkers of exposure or effect.

Besides data reporting issues for molecular epidemiology studies, I also experienced text mining approaches rely on study reporting. For example, it is undesirable for a mining system to report a hypothetical question (e.g. “Here we investigate whether benzene causes ROS damage”) as a true finding (e.g. “Here we demonstrated that benzene causes ROS damage”). Our text mining systems tried to remove hypothetical statements but was not always effective in doing so. Some text mining systems (e.g. the one by Carvaillo et al. (2019)) give more weight to results which are mentioned more at the end of a text, so they more likely represent a true finding.

In addition I encountered some issues with regards to annotation of chemicals. Generally text mining systems ground biochemical entities to IDs of actual entities through the use of knowledge bases. The text mining system I have used processes chemicals using the ChEBI database (www.ebi.ac.uk/chebi/) but for one chemical (lindane) I could not proceed with further network analyses because this chemical was not well annotated yet and thus not recognized by the text mining system. As entities can be written in many different forms (i.e. synonyms, acronyms), entity extraction is a common problem when using text mining approaches for building biomedical networks (Conceição et al. 2021).

Both issues (reusing and grounding of data) are of central importance to the FAIR (Findability, Accessibility, Interoperability, and Reuse) initiative, which was introduced in 2016 (Wilkinson et al. 2016). The FAIR guidelines on how to find, access, interoperate, and reuse data are now more generally being acknowledged, for example by the PARC (European Partnership for the Assessment of Risks from Chemicals) initiative which was recently launched. Within this program data

FAIRification is a key activity. The PARC Partnership will promote transparency and harmonization of data, and facilitate the exchange of data between different disciplines and the wider reuse of obtained data.

Our quantitative evidence synthesis approaches are data hungry: there is a need for additional steps for data-poor substances

The methods and tooling introduced in this thesis were mostly tested on benzene and diisocyanates, which are chemicals with a large evidence base and extensive hazard and risk assessments, so they could act as proof of principle chemicals to evaluate the usefulness of the applied tools and approaches. Not all of these approaches are useful for chemicals with a limited evidence base. For example, although it is theoretically possible to conduct a meta-analysis with only two studies, conclusion based on only a few studies might not be received as strong evidence by the research community (Cheung and Vijayakumar 2016). Since for most chemicals there is little to no information available (Persson et al. 2022; Wang et al. 2020), there is a need for complementary approaches for data-poor substances.

One opportunity which could be further explored is the use of quantitative structure-activity relationships (QSARS). QSARs aim to relate features of physico-chemical structure to a property or biological activity, so that the biological activity of new or untested chemical can be inferred from the molecular structure of similar compounds whose activities have already been assessed (e.g. (Schultz et al. 2003). The concept of QSARs could be interesting for our text mining approach. Currently only known chemicals with a chemical identifier are included, but perhaps the INDRA platform could be expanded with a module on finding chemical similarities. Guha et al. (2016) for example, combined information on chemical structure with database integration and automated text mining. Expanding the networks for chemical similarities is especially relevant for chemicals with a small evidence base.
QSARs could also play a role in PBK modelling, most notably for deriving chemical-specific data. Zang et al. (2017), for example, generated an open-source quantitative structure-property relationship workflow to predict a variety of physico-chemical properties using machine learning.
Furthermore, the amount of high-throughput *in vitro* data available is rapidly increasing due to programs such as US EPA ToxCast and US Tox21. Generating toxicity data for more chemicals at a faster pace helps to address the issue that many chemicals are not yet well studied. Simultaneously, there is a need for analyzing this type of data, for example by using in vitro-in vivo extrapolation (IVIVE) approaches based on PBK modelling (e.g. Cohen Hubal et al. 2018).
Lastly the evidence integration approach (Chapter 6), might be of interest to use when evaluating compounds with weak epidemiologic evidence (i.e. insufficient data) but a large animal or molecular evidence base.

Adverse Outcome Pathways provides a useful framework for evidence synthesis, but more research is needed for quantitative evidence integration approaches

The focus of chemical risk assessment will increasingly be on perturbations of biological pathways (National Academies of Sciences Engineering and Medicine 2017). One concept that supports biological network understanding and that has attracted a large amount of attention in recent years, is adverse outcome pathways (AOPs) (Ankley et al. 2010). The AOP concept provides a useful framework for integrating diverse lines of evidence by visualizing the relationships and connections between different evidence streams (Wittwehr et al. 2020; OECD 2017), and is therefore interesting to consider in light of the work described in this thesis.

The evidence integration approach in Chapter 6 could also be viewed within the framework of AOPs since it involved data on intermediate events for predicting an adverse outcome (AO). I assumed proportionality (i.e. an X increase in CA equals an X increase in developing AML) for converting information on intermediate events (biomarker studies on genotoxicity, but also animal studies) to AML outcomes as typically reported in epidemiological studies. By doing so I harmonized datasets and could apply meta-regression techniques for sensitivity analysis and disease burden calculation using life table analysis. Progressing towards the inclusion of more mechanistic data requires a different set of assumptions that in most cases is even larger than for animal studies. For instance, if omics data were to be incorporated, proportionality assumptions such as an *X* % increase in gene induction equals *X*% increased risk of developing AO would be needed.

The AOP framework might further be of interest for our work on text mining and systems biology in Chapter 2. The focus of this work was on (in) activation of KCCs based on specific filtering steps. But if I looked further from the compound of interest, the networks were able to identify relations between additional events. For example, while benzene activates DNA damage, DNA damage in turn may activate cell death, and necrotic cell death may be associated with activation of an inflammatory response (Figure 5, Chapter 2). Thus, the adverse outcome pathway concept may benefit from AI/textmining approaches.

Concluding remarks

There is a general recognition of the need for innovations in chemical risk assessment. To progress from the traditional animal-based assays, it is important to consider additional evidence streams such as molecular epidemiology, toxicogenomics, in silico data, and mechanistic assays. However, contrary to animal studies, for these evidence streams there is generally no detailed guidance for conducting and reporting, and these studies are not yet widely accepted by risk assessment agencies. The use of evidence streams other than animal studies requires i) a range of computational and biological informed tools for processing the data (i.e. AI for summarizing the

abundance of mechanistic studies, or PBK models for forward and reverse calculations of biomarker data), ii) more advanced statistical methods to account for uncertainty (i.e. Bayesian methods for defining priors on between study heterogeneity, global sensitivity analysis based on parameter distributions), iii) knowledge on relationships between evidence streams (i.e. on the linkage between intermediate effects and ultimate disease), but also iv) understanding and trust of risk assessors in the new methods and evidence (i.e. understanding which steps and assumptions are required when synthesizing evidence from different evidence streams).

In this thesis these issues have been addressed in the various Chapters, which all aimed to increase the use of various types of evidence streams within the risk assessment paradigm.
I demonstrated how the process of identifying and summarizing evidence for hazard identification could be more efficient when using a text mining approach combined with systems biology, as opposed to manually reading and selecting abstracts and articles (Chapter 2). Exposure assessment is generally based on air or dermal measures outside the human body, but can be complemented with biomonitoring studies that reflect the absorbed dose (Chapter 3). PBK models are tools helpful for the interpretation of results from biomonitoring studies and understanding variation for various exposure scenarios (Chapter 4). Biomonitoring studies on intermediate health effects can be quantitatively meta-analysed to provide summary estimates of exposure-response curves (Chapter 5) in contrast to qualitatively using these studies for weight of evidence approaches, or even combined with other evidence streams including human epidemiology and animal studies (Chapter 6).

Throughout the discussion sections of the various chapters I reflect on the possibilities and limitations of considering various evidence streams to advance risk assessment. By using systematic and quantitative methods I was able to quantify uncertainty on the available evidence, and was forced to be transparent and explicit on the assumptions that I had to make to combine data. The use of the evidence synthesis methods for chemical risk assessment is promising but will benefit from the use of more harmonized methods for either conducting a field study or reporting on results. In addition for chemicals with a more limited evidence base, I identified various research areas that could be further explored in conjunction with the approaches and methods used in this thesis, including the use of QSARS, read across methods and generic PBK models and the consideration of omics data for evidence synthesis approaches.

References Discussion

Ankley, Gerald T., Richard S. Bennett, Russell J. Erickson, Dale J. Hoff, Michael W. Hornung, Rodney D. Johnson, David R. Mount, et al. 2010. “Adverse Outcome Pathways: A Conceptual Framework to Support Ecotoxicology Research and Risk Assessment.” *Environmental Toxicology and Chemistry* 29 (3): 730–41. https://doi.org/10.1002/ETC.34.

Beller, Elaine, Justin Clark, Guy Tsafnat, Clive Adams, Heinz Diehl, Hans Lund, Mourad Ouzzani, et al. 2018. “Making Progress with the Automation of Systematic Reviews: Principles of the International Collaboration for the Automation of Systematic Reviews (ICASR).” *Systematic Reviews* 7 (1): 1–7. https://doi.org/10.1186/S13643-018-0740-7/TABLES/1.

Carvaillo, Jean Charles, Robert Barouki, Xavier Coumoul, and Karine Audouze. 2019. “Linking Bisphenol S to Adverse Outcome Pathways Using a Combined Text Mining and Systems Biology Approach.” *Environmental Health Perspectives*. https://doi.org/10.1289/EHP4200.

Cheung, Mike W.L., and Ranjith Vijayakumar. 2016. “A Guide to Conducting a Meta-Analysis.” *Neuropsychology Review* 26 (2): 121–28. https://doi.org/10.1007/S11065-016-9319-Z/TABLES/1.

Cohen Hubal, Elaine A., Barbara A. Wetmore, John F. Wambaugh, Hisham El-Masri, Jon R. Sobus, and Tina Bahadori. 2018. “Advancing Internal Exposure and Physiologically-Based Toxicokinetic Modeling for 21st-Century Risk Assessments.” *Journal of Exposure Science & Environmental Epidemiology 2018 29:1* 29 (1): 11–20. https://doi.org/10.1038/s41370-018-0046-9.

Conceição, Sofia I R, Francisco M Couto, Francisco Rodrigues Pinto, Javier De, and Las Rivas. 2021. “Text Mining for Building Biomedical Networks Using Cancer as a Case Study.” *Biomolecules 2021, Vol. 11, Page 1430* 11 (10): 1430. https://doi.org/10.3390/BIOM11101430.

DFG. 2007. “Addendum to 4,4’-Methylene Diphenyl Diisocyanate (Inhalable Fraction)[BAT Value Documentation, 2007].”

EFSA. 2017. “Guidance on the Use of the Weight of Evidence Approach in Scientific Assessments.” *EFSA Journal* 15 (8). https://doi.org/10.2903/J.EFSA.2017.4971.

Galea, Karen S., Simo P. Porras, Susana Viegas, Beatrice Bocca, Radia Bousoumah, Radu Corneliu Duca, Lode Godderis, et al. 2021. “HBM4EU Chromates Study - Reflection and Lessons Learnt from Designing and Undertaking a Collaborative European Biomonitoring Study on Occupational Exposure to Hexavalent Chromium.” *International Journal of Hygiene and Environmental Health* 234 (May): 113725. https://doi.org/10.1016/J.IJHEH.2021.113725.

Gallo, Valentina, Matthias Egger, Valerie McCormack, Peter B. Farmer, John P.A. Ioannidis, Micheline Kirsch-Volders, Giuseppe Matullo, et al. 2012. “STrengthening the Reporting of OBservational Studies in Epidemiology – Molecular Epidemiology (STROBE-ME): An Extension of the STROBE Statement.” *Mutagenesis* 27 (1): 17–29. https://doi.org/10.1093/MUTAGE/GER039.

Guha, Neela, Kathryn Z. Guyton, Dana Loomis, and Dinesh Kumar Barupal. 2016. “Prioritizing Chemicals for Risk Assessment Using Chemoinformatics: Examples from the IARC Monographs on Pesticides.” *Environmental Health Perspectives*. https://doi.org/10.1289/EHP186.

Hartwig, A, and MAK Commission. 2021. “The MAK Collection for Occupational Toluylendiisocyanate MAK-Begründung, Nachtrag.” *Health and Safety* 6 (2). https://doi.org/10.34865/mb58484ismd6_2ad.

Health Council of the Netherlands. 2014. “Benzene - Health-Based Recommended Occupational Exposure Limit.”

Hernández, Antonio F., and Aristidis M. Tsatsakis. 2017. “Human Exposure to Chemical Mixtures: Challenges for the Integration of Toxicology with Epidemiology Data in Risk Assessment.” *Food and Chemical Toxicology* 103 (May): 188–93. https://doi.org/10.1016/J.FCT.2017.03.012.

IARC. 2018. “Benzene (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans).” Vol. 120.

Kavlock, Robert J., Tina Bahadori, Tara S. Barton-Maclaren, Maureen R. Gwinn, Mike Rasenberg, and Russell S. Thomas. 2018. “Accelerating the Pace of Chemical Risk Assessment.” *Chemical Research in Toxicology* 31 (5): 287–90. https://doi.org/10.1021/ACS.CHEMRESTOX.7B00339/ASSET/IMAGES/MEDIUM/TX-2017-003397_0001.GIF.

Keil, Alexander P. 2022. “Real and Illusory Precision Gains in Meta-Analysis to Speed Action on Carcinogens.” *Cancer Epidemiology, Biomarkers & Prevention* 31 (4): 695–97. https://doi.org/10.1158/1055-9965.EPI-21-1286.

Lavecchia, Antonio. 2019. “Deep Learning in Drug Discovery: Opportunities, Challenges and Future Prospects.” *Drug Discovery Today* 24 (10): 2017–32. https://doi.org/10.1016/J.DRUDIS.2019.07.006.

Linkov, Igor, Olivia Massey, Jeff Keisler, Ivan Rusyn, and Thomas Hartung. 2015. “From ‘Weight of Evidence’ to Quantitative Data Integration Using Multicriteria Decision Analysis and Bayesian Methods.” *ALTEX* 32 (1): 3. https://doi.org/10.14573/ALTEX.1412231.

Luechtefeld, Thomas, Craig Rowlands, and Thomas Hartung. 2018. “Big-Data and Machine Learning to Revamp Computational Toxicology and Its Use in Risk Assessment.” *Toxicology Research* 7 (5): 732–44. https://doi.org/10.1039/C8TX00051D.

Miake-Lye, Isomi M, Susanne Hempel, Roberta Shanman, and Paul G Shekelle. 2016. “What Is an Evidence Map? A Systematic Review of Published Evidence Maps and Their Definitions, Methods, and Products.” https://doi.org/10.1186/s13643-016-0204-x.

National Academies of Sciences Engineering, and Medicine. 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. National Academies Press. https://doi.org/10.17226/24635.

OECD. 2017. “Integrated Approaches to Testing and Assessment (IATA).” 2017. https://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm.

Paini, Alicia, Jeremy A. Leonard, Tomas Kliment, Yu Mei Tan, and Andrew Worth. 2017. “Investigating the State of Physiologically Based Kinetic Modelling Practices and Challenges Associated with Gaining Regulatory Acceptance of Model Applications.” *Regulatory Toxicology and Pharmacology* 90 (November): 104–15. https://doi.org/10.1016/J.YRTPH.2017.08.019.

Persson, Linn, Bethanie M. Carney Almroth, Christopher D. Collins, Sarah Cornell, Cynthia A. de Wit, Miriam L. Diamond, Peter Fantke, et al. 2022. “Outside the Safe Operating Space of the Planetary Boundary for Novel Entities.” *Environmental Science and*

Technology 56 (3): 1510–21. https://doi.org/10.1021/ACS.EST.1C04158/ASSET/IMAGES/LARGE/ES1C04158_0002.JPEG.

Pletz, Julia, Samantha Blakeman, Alicia Paini, Nikolaos Parissis, Andrew Worth, Anna Maria Andersson, Hanne Frederiksen, Amrit K. Sakhi, Cathrine Thomsen, and Stephanie K. Bopp. 2020. “Physiologically Based Kinetic (PBK) Modelling and Human Biomonitoring Data for Mixture Risk Assessment.” *Environment International* 143 (October): 105978. https://doi.org/10.1016/J.ENVINT.2020.105978.

RAC. 2020. “Opinion on Scientific Evaluation of Occupational Exposure Limits for Diisocyanates.”

Schultz, T. Wayne, Mark T.D. Cronin, John D. Walker, and Aynur O. Aptula. 2003. “Quantitative Structure–Activity Relationships (QSARs) in Toxicology: A Historical Perspective.” *Journal of Molecular Structure: THEOCHEM* 622 (1–2): 1–22. https://doi.org/10.1016/S0166-1280(02)00614-0.

Song, Fujian, Trevor A. Sheldon, Alex J. Sutton, Keith R. Abrams, and David R. Jones. 2001. “Methods for Exploring Heterogeneity in Meta-Analysis:” *Evaluation & the Health Professions* 24 (2): 126–51. https://doi.org/10.1177/016327870102400203.

Steenland, Kyle, M. K. Schubauer-Berigan, R. Vermeulen, R. M. Lunn, K. Straif, S. Zahm, P. Stewart, W. D. Arroyave, S. S. Mehta, and N. Pearce. 2020. “Risk of Bias Assessments and Evidence Syntheses for Observational Epidemiologic Studies of Environmental and Occupational Exposures: Strengths and Limitations.” *Environmental Health Perspectives* 128 (9): 1–10. https://doi.org/10.1289/EHP6980.

The Dutch Health Council. 2018. “Di- and Triisocyanates. Health-Based Recommendation on Occupational Exposure Limits (No. 2018/20).” The Hague.

Tsertsvadze, Alexander, Yen Fu Chen, David Moher, Paul Sutcliffe, and Noel McCarthy. 2015. “How to Conduct Systematic Reviews More Expeditiously?” *Systematic Reviews* 4 (1): 1–6. https://doi.org/10.1186/S13643-015-0147-7/TABLES/1.

Vlaanderen, Jelle, Nathaniel Rothman, Lützen Portengen, Qing Lan, Hans Kromhout, and Roel Vermeulen. 2009. “Flexible Meta-Regression to Assess the Shape of the Benzene–Leukemia Exposure–Response Curve.” *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.0901127.

Wang, Zhanyun, Glen W. Walker, Derek C.G. Muir, and Kakuko Nagatani-Yoshida. 2020. “Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories.” *Environmental Science and Technology* 54 (5): 2575–84. https://doi.org/10.1021/ACS.EST.9B06379/ASSET/IMAGES/LARGE/ES9B06379_0002.JPEG.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data 2016 3:1* 3 (1): 1–9. https://doi.org/10.1038/sdata.2016.18.

Wittwehr, Clemens, Paul Blomstedt, John Paul Gosling, Tomi Peltola, Barbara Raffael, Andrea Nicole Richarz, Marta Sienkiewicz, Paul Whaley, Andrew Worth, and Maurice Whelan. 2020. “Artificial Intelligence for Chemical Risk Assessment.” *Computational Toxicology* 13 (February): 100114. https://doi.org/10.1016/J.COMTOX.2019.100114.

Zang, Qingda, Kamel Mansouri, Antony J. Williams, Richard S. Judson, David G. Allen, Warren M. Casey, and Nicole C. Kleinstreuer. 2017. “In Silico Prediction of

Physicochemical Properties of Environmental Chemicals Using Molecular Fingerprints and Machine Learning." *Journal of Chemical Information and Modeling* 57 (1): 36–49. https://doi.org/10.1021/ACS.JCIM.6B00625/SUPPL_FILE/CI6B00625_SI_002.XLSX.

Appendices

Summary

Background

Every day, we are surrounded by hundreds to thousands of synthetic chemicals. Despite their useful properties, we know that many of these substances can have negative impacts on our health and the environment. Chemical risk assessment is therefore important for deriving safe levels of exposure. The process of chemical risk assessment is typically divided into four steps. The first step is **hazard identification**, that is an evaluation of all available evidence to determine whether a certain chemical exposure can cause adverse human health effects; it reflects the intrinsic capacity of the chemical to cause harm. The second step is **exposure assessment**, that is measuring or modelling the degree to which humans are exposed to a chemical. Subsequently, the relationship between exposure levels and the occurrence of an adverse health effect should be quantitatively characterized in the third step: **hazard characterization**. This provides information at which exposure levels an adverse health effect can be expected, for example: at *X* exposure there is an *X* increased risk of developing *X* (a negative health effect). In the fourth step all previous information is then considered in a **risk characterization**. Here a comparison is made between human exposures to chemicals, and the exposure levels at which these chemicals can cause harm.

Traditionally, mostly animal studies were used for the hazard identification and hazard characterization step. Since the 1950s-1960s, these studies have been executed and created general consensus and detailed guidance on how these studies should be carried out and interpreted. However, animal studies are economically costly and it is often unclear to which extent observed effects in animals are predictive for humans. As more-and-more human (observational) studies are becoming available their use is preferred as it directly provides information on the target population: humans. These epidemiologic studies investigate for a large group of people if they have a higher risk of developing a negative health effect, often based on measured or estimated exposures outside the human body (for example, in the breathing zone of workers). However, for human studies it is often difficult to adequately characterize exposure. Another approach to study exposure to chemicals and the induction of negative health effects, is taking biological samples from human tissues, such as blood or urine or saliva. This is commonly referred to as molecular epidemiology, or human biomonitoring. Instead of estimating or measuring external exposure, exposure biomarkers reflect the concentration of a chemical or its metabolites that is taken up by the body. And instead of observing the negative health effect, for example cancer, effect biomarkers reflect the effects that have occurred after a chemical is circulating inside the human body. In this Thesis I have further considered mechanistic studies,

which are performed in a laboratory to examine chemical induced effects at the tissue, cellular and molecular level.

The starting point of this thesis is that all types of evidence (i.e. all data, such as scientific publications, from the various categories of studies), can be valuable for the various steps of a chemical risk assessment. However, the consideration of different types of evidence does require additional methods, tooling and approaches. For example, the guidance for conducting and evaluating animal studies for chemical risk assessment, is not applicable for conducting and evaluating human biomonitoring studies. The goal of this thesis was therefore to further develop and test approaches to use different heterogeneous types of evidence for chemical risk assessment. Below, these are explained for different steps in chemical risk assessment.

Hazard identification

To fully evaluate the available evidence on chemical induced health effects it is important to systematically collect, screen and interpret all available data. This can be challenging, especially for mechanistic data, because for some chemicals there are over 1000 studies that report on mechanistic findings for that particular chemical. Since it takes considerable efforts to collect and screen all these studies manually, the aim of the research in Chapter 2 was to use computational systems that read text from scientific publications automatically and very fast (text mining techniques) and combine these with network visualization tools to show the hazard information from the publications in a comprehensive 'network' manner. This computational approach was tested for 13 chemicals. The results of the computational approach (in the format of networks for several chemicals) were compared to reviews on the hazard that were already performed for those chemicals by human experts from a well-recognized institute: the International Agency for Research on Cancer (IARC). The results were comparable, i.e. for those chemicals for which IARC found strong evidence for a hazard, the computational approach also indicated that the chemical was likely to have this hazard. Chapter 2 thus provides an example of how text mining tooling can assist in the process of hazard identification, by quickly condensing available mechanistic literature into searchable and interpretable networks.

Exposure assessment

For human epidemiology studies it is important to adequately characterize exposure. However this is not always straightforward, for example when it is technically challenging to measure external exposure (i.e. exposure outside the human body). Also, sometimes various exposure routes such as exposure via the skin and the lungs may exist and these may not always be known upfront. In those cases exposure biomarkers have been suggested as a useful source of information. For the work of this thesis on exposure assessment, diisocyanates were used as a case study.

Diisocyanates are widely used in occupational setting in different job sectors such as the construction and car repair industry, but may cause allergic reactions. It is therefore important to characterize the exposure. However for diisocyanates it is challenging to determine external exposure and therefore a large number of diisocyanate biomonitoring studies have been performed. To be able to consider all those studies, rather than cherry picking “one best study”, a systematic review of the publications from these studies was conducted (Chapter 3). By doing so, the levels of exposure biomarkers from these biomonitoring studies were summarized across several job sectors. In addition, the review identified job sectors for which the number of studies was low (and by doing so identified sectors were additional research is needed).

A challenge of biomonitoring data using exposure biomarkers lays in the interpretation of the data, especially for chemicals like diisocyanates which directly bind to various cells in the human body and are then quickly excreted. In that case, biomonitoring depends on factors such as the exposure scenario on that day (for example the moment of exposure, exposure duration and exposure route) as well as the behavior of the worker being measured (for example the use of personal protective equipment, or the moment of urination via which the biomarker may be excreted). To be able to understand how these factors affect the relationship between chemical exposure outside the human body, and subsequently biomarkers of exposure inside, a physiologically-based kinetic (PBK) model was developed for diisocyanates in Chapter 4. PBK models are mathematical models that reflect a human body upon exposure to a chemical. These models describe in which parts of the body (organs, tissues) the chemical will distribute to, this in relation to the external exposure via inhalation and/or uptake via the skin. By using various scenarios in the model and performing a so called “sensitivity analysis”, it was possible to identify which parameters or factors in exposure scenarios contributed most to variation in predicted urinary values for diisocyanates. In other words, the modelling helped to identify how predicted urinary values for biomarkers of exposure for diisocyanate are related to different external exposure situations.

Hazard characterization

In chapter 5, a method is described for combining information from human effect biomarker studies, to better characterize the exposure level at which a health effect is to be expected. Benzene was used as a case-study. Benzene is a well-known chemical that can ultimately cause cancer. The development of cancer occurs via intermediate steps; it is possible to study these intermediate steps by measuring effect biomarkers of genotoxicity. Genotoxicity is the capacity of a chemical to damage the genetic material within a cell causing mutations, which can lead to cancer. An advantage of these effect biomarker studies is that they may offer more precise information on early health effects at lower external exposure levels. However, to date

there have not been many attempts to combine evidence on biomarkers of effect in a systematic and quantitative way. Chapter 5 provides a case study of such an approach, by summarizing all currently available human biomonitoring data on occupational exposure to benzene and biomarkers of effect. From the integration of the data, an estimation was made for the quantitative relationship between benzene exposure and the genetic biomarkers of effect. The results showed that although the majority of the retrieved studies indicated a positive association between exposure and genetic damage, still the studies differed considerably from each other - this is referred to as heterogeneity - and the combined summary effect was thus difficult to interpret.

The aim of Chapter 6 was to provide an example of combining different types of evidence, including epidemiology, human biomonitoring and animal data. Most often, in risk assessment, these data are considered separately by experts (toxicologists, epidemiologists) in scientific discussion. It is not common to combine such different types of data directly into a quantitative approach in order to estimate an exposure level above which it is likely that a harmful health effect will occur. Therefore, the aim of Chapter 6 was to explore an example of how this approach could work, and what the consequences and limitations are when doing so. The approach chosen was to convert (i.e. harmonize) all evidence to risk estimates reported in epidemiology studies. This introduced difficulties because the different types of data are not easy to combine. For example, benzene exposure in animal studies was well controlled and generally for a couple of hours on consecutive days, for several months and to relatively high concentrations. The endpoints in these studies recorded were the occurrence of tumors. In contrast, human biomonitoring studies measured benzene exposure at one occasion only and reported intermediate health effects as biomarker of effect. Epidemiology studies, on the other hand, followed large groups of humans for many years and often reported cumulative benzene exposures. Although I found that the exposure response relationship was statistically more precise when I combined all data, it was difficult to quantify the uncertainty which was introduced due to all the required harmonization steps.

Concluding remarks

Throughout the chapters various examples have been given for (quantitatively) considering various types of evidence. The quantitative approaches used gave insight into research gaps. It also forced me to be explicit, hence making the approaches transparent and reproducible for other researchers. For all work performed in this Thesis I relied on published data. Especially for the human biomonitoring studies of both exposure and effect this introduced difficulties because the studies were different in design, and thus not always easy to combine. In addition, in some cases information was not available on parameters that were required for re-using the data for other purposes. Throughout the chapters, research gaps were identified and

recommendations given for data collection and reporting. Although study differences made interpretation of results often challenging, still, by combining data and providing summary estimates, risk assessors will have the option to select an overall, best “average” estimate, even if that estimate is imprecise.

The case-studies of this thesis illustrate that human biomonitoring studies can be a valuable source of information for both enhancing exposure assessment and hazard characterization. Tooling such as text mining and network visualization tools can help to make evidence identification for hazard identification more efficient. And evidence from human (epidemiology and biomonitoring) and animal studies can also be combined together, but this does require a large number of assumptions. The chapters in this thesis demonstrate that a wide range of interdisciplinary approaches are required to be able to integrate data, for it to become optimally useful for chemical risk assessment. This can be challenging for risk assessors, so there should be a clear guidance on how to combine different types of evidence and how the results can be interpreted. The methods and tooling in this thesis are mostly applied to chemicals for which a large amount of data is known. In the Discussion chapter of this thesis, suggestions are given for approaches that can be used for chemicals for which only little information is known, possibly in conjunction with the new scientific approaches described in the other Chapters of this Thesis.

Samenvatting

Achtergrond

Elke dag, worden we omringd door honderden tot duizenden synthetische chemicaliën. Ondanks hun nuttige eigenschappen weten we dat veel van deze stoffen een negatieve invloed kunnen hebben op onze gezondheid en het milieu. Chemische risicobeoordeling is daarom belangrijk voor het afleiden van veilige blootstellingsniveaus. Het proces van chemische risicobeoordeling is verdeeld in vier stappen. De eerste stap is **identificatie van mogelijke stof specifieke gevaren voor de gezondheid**; het weerspiegelt het intrinsieke vermogen van de chemische stof om schade aan te richten. De tweede stap is de **blootstellingsbepaling**, dat wil zeggen het meten of modelleren van de mate waarin mensen worden blootgesteld aan een chemische stof. De derde stap is de **karakterisering van blootstellingsniveaus waarbij een nadelig gezondheidseffect te verwachten is**, bijvoorbeeld: bij *X* blootstelling is er een *X* verhoogde kans op het ontwikkelen van *X* (een negatief gezondheidseffect). In de vierde stap wordt vervolgens alle voorgaande informatie meegenomen in een **risicobeoordeling**. Hier wordt een vergelijking gemaakt tussen menselijke blootstelling aan chemicaliën, en de blootstellingsniveaus waarbij deze chemicaliën schade kunnen veroorzaken.

Traditioneel gezien werden voornamelijk dierstudies gebruikt voor het identificeren en karakteriseren van stof specifieke gevaren voor de gezondheid. Sinds de jaren 1950-1960 zijn deze onderzoeken uitgevoerd en hebben ze geleid tot algemene consensus en gedetailleerde richtlijnen over hoe deze onderzoeken moeten worden uitgevoerd en geïnterpreteerd. Dierstudies zijn echter economisch kostbaar en het is vaak onduidelijk in hoeverre waargenomen effecten bij dieren voorspellend zijn voor de mens. Aangezien er steeds meer humane (observationele) studies beschikbaar komen, verdient het gebruik van deze studies de voorkeur omdat het rechtstreeks informatie verschaft over de doelpopulatie: de mens. Deze epidemiologische studies onderzoeken voor een grote groep mensen of zij een hoger risico hebben op het ontwikkelen van een negatief gezondheidseffect, vaak op basis van gemeten of geschatte blootstellingen buiten het menselijk lichaam (bijvoorbeeld in de ademzone van werknemers). Voor humane studies is het wel vaak moeilijk om de blootstelling adequaat te karakteriseren. Een andere benadering om blootstelling aan chemicaliën en de inductie van negatieve gezondheidseffecten te bestuderen, is het nemen van biologische monsters van menselijk weefsel, zoals bloed of urine of speeksel. Dit wordt doorgaans moleculaire epidemiologie of humane biomonitoring genoemd. In plaats van externe blootstelling te schatten of te meten, weerspiegelen blootstellings-biomarkers de concentratie van een chemische stof, of zijn metabolieten, die door het lichaam wordt opgenomen. En in plaats van het negatieve effect op de gezondheid, bijvoorbeeld kanker, te observeren, weerspiegelen effect-biomarkers de effecten die

zijn opgetreden nadat een chemische stof in het menselijk lichaam circuleert. In dit proefschrift heb ik verder mechanistische studies meegenomen, die in een laboratorium worden uitgevoerd om chemische geïnduceerde effecten op weefsel-, cellulair en moleculair niveau te onderzoeken.

Het uitgangspunt van dit proefschrift is dat alle soorten bewijs (d.w.z. alle data, zoals wetenschappelijke publicaties, van de verschillende categorieën van studies) waardevol kunnen zijn voor de verschillende stappen van een chemische risicobeoordeling. De overweging van verschillende soorten bewijs vereist echter aanvullende methoden, tooling en benaderingen. Bijvoorbeeld, de richtlijnen voor het uitvoeren en evalueren van dierstudies voor chemische risicobeoordeling zijn niet van toepassing op het uitvoeren en evalueren van humane biomonitoringstudies. Het doel van dit proefschrift was daarom het verder ontwikkelen en testen van benaderingen om verschillende soorten van bewijs te kunnen gebruiken voor chemische risicobeoordeling. Hieronder worden deze toegelicht voor verschillende stappen in de chemische risicobeoordeling.

Identificatie mogelijk stof specifiek gevaar voor de gezondheid

Om het beschikbare bewijs over door chemicaliën veroorzaakte gezondheidseffecten volledig te kunnen evalueren, is het belangrijk om alle beschikbare data systematisch te verzamelen, screenen en interpreteren. Dit kan een uitdaging zijn, vooral voor mechanistische data, omdat er voor sommige chemicaliën meer dan 1000 onderzoeken zijn die verslag doen van mechanistische bevindingen voor die specifieke chemische stof. Aangezien het aanzienlijke inspanningen kost om al deze onderzoeken handmatig te vinden en screenen, was het doel van het onderzoek in Hoofdstuk 2 om computersystemen te gebruiken die tekst uit wetenschappelijke publicaties automatisch en zeer snel kunnen lezen (text mining-technieken) en deze te combineren met netwerkvisualisatietools om de informatie uit de publicaties op een overzichtelijke 'netwerk'-manier weer te geven. Deze aanpak is getest voor 13 chemicaliën. De resultaten van de computationele benadering (in de vorm van netwerken voor verschillende chemicaliën) werden vergeleken met beoordelingen die al voor die chemicaliën waren uitgevoerd door menselijke experts van een gerenommeerd instituut: het International Agency for Research on Cancer (IARC). De resultaten waren vergelijkbaar, d.w.z. voor die chemicaliën waarvoor IARC sterke aanwijzingen vond voor een gevaar, gaf de computationele benadering ook aan dat de chemische stof waarschijnlijk dit gevaar met zich meebracht. Hoofdstuk 2 geeft dus een voorbeeld van hoe een tool zoals text mining kan helpen bij het identificeren van gevaren van chemicaliën, door beschikbare mechanistische literatuur snel samen te vatten in doorzoekbare en interpreteerbare netwerken.

Blootstellingsbepaling

Voor humane epidemiologische studies is het belangrijk om de blootstelling adequaat te karakteriseren. Dit is echter niet altijd eenvoudig, bijvoorbeeld wanneer het

technisch moeilijk is om externe blootstelling (d.w.z. blootstelling buiten het menselijk lichaam) te meten. Ook kunnen er soms verschillende blootstellingsroutes bestaan, zoals blootstelling via de huid en de longen, die niet altijd vooraf bekend zijn. In die gevallen kunnen blootstellings-biomarkers een nuttige informatiebron zijn. Voor het werk van dit proefschrift rondom blootstellingsbepaling werden diisocyanaten gebruikt als casestudy. Diisocyanaten worden veel gebruikt in de beroepsomgeving in verschillende beroepssectoren, zoals de bouw- en autoreparatie-industrie, maar diisocyanaten kunnen allergische reacties veroorzaken. Het is daarom belangrijk om de blootstelling te karakteriseren. Voor diisocyanaten is het echter een uitdaging om externe blootstelling te bepalen en daarom is er een groot aantal biomonitoringsonderzoeken voor diisocyanaten uitgevoerd. Om al die studies mee te nemen, in plaats van "one best study" uit te kiezen, is een systematische review van de publicaties van deze studies uitgevoerd (Hoofdstuk 3). Op deze manier werden de blootstellingsniveaus uit de biomonitoringstudies samengevat voor de verschillende beroepssectoren. Bovendien identificeerde de review beroepssectoren waarvoor het aantal studies laag was (en daarmee werd ook aangetoond waarvoor aanvullend onderzoek nodig is).

Een uitdaging van biomonitoringsdata ligt in de interpretatie van de gegevens, vooral voor chemicaliën zoals diisocyanaten die zich rechtstreeks binden aan verschillende cellen in het menselijk lichaam en vervolgens snel worden uitgescheiden. In dat geval is blootstellings- biomonitoring data afhankelijk van het blootstellingsscenario op die dag (bijvoorbeeld het moment van blootstelling, blootstellingsduur en blootstellingsroute) en het gedrag van de werknemer die wordt gemeten (bijvoorbeeld het gebruik van persoonlijke beschermingsmiddelen, of het moment van urineren via welke de biomarker uitgescheiden kan worden). Om te begrijpen hoe deze factoren de relatie beïnvloeden tussen chemische blootstelling buiten het menselijk lichaam en vervolgens in het lichaam, is in Hoofdstuk 4 een fysiologisch gebaseerd kinetisch (PBK) model ontwikkeld voor diisocyanaten. PBK-modellen zijn wiskundige modellen die een menselijk lichaam nabootsen bij blootstelling aan een chemische stof. Deze modellen beschrijven in welke delen van het lichaam (organen, weefsels) de chemische stof zich zal verspreiden, dit in relatie tot de externe blootstelling via inademing en/of opname via de huid. Door verschillende scenario's in het model te gebruiken en een zogenaamde "gevoeligheidsanalyse" uit te voeren, was het mogelijk om te identificeren welke parameters of factoren in blootstellingsscenario's het meest bijdroegen aan variatie in voorspelde urinaire waarden voor diisocyanaten. Met andere woorden, de modellering hielp om te identificeren hoe voorspelde urinaire waarden voor biomarkers van blootstelling aan diisocyanaten in relatie staan tot verschillende externe blootstellingssituaties.

Karakteriseren van blootstellingsniveaus waarbij een nadelig gezondheidseffect te verwachten is

In hoofdstuk 5 wordt een methode beschreven voor het combineren van informatie uit humane effect-biomarker studies, om het blootstellingsniveau waarbij een gezondheidseffect te verwachten is, beter te karakteriseren. Benzeen werd gebruikt als casestudy. Benzeen is een bekende chemische stof die uiteindelijk kanker kan veroorzaken. De ontwikkeling van kanker vindt plaats via tussenstappen; het is mogelijk om deze tussenstappen te bestuderen door effect-biomarkers van genotoxiciteit te meten. Genotoxiciteit is het vermogen van een chemische stof om het genetisch materiaal in een cel te beschadigen en mutaties te veroorzaken die tot kanker kunnen leiden. Een voordeel van deze effect-biomarker studies is dat ze wellicht meer nauwkeurige informatie kunnen bieden over vroege gezondheidseffecten bij lagere niveaus van externe blootstelling. Tot op heden zijn er echter niet veel pogingen gedaan om studies van biomarkers van effect op een systematische en kwantitatieve manier te combineren. Hoofdstuk 5 biedt een casestudy voor een mogelijke benadering, door een samenvatting te geven van alle beschikbare humane biomonitoringsstudies over beroepsmatige blootstelling aan benzeen en biomarkers van effect. Uit de integratie van de data werd een schatting gemaakt voor de kwantitatieve relatie tussen benzeen-blootstelling en de genetische biomarkers van effect. De resultaten toonden aan dat hoewel de meerderheid van de studies een positief verband aangaf tussen blootstelling en genetische schade, de onderzoeken toch aanzienlijk van elkaar verschilden - dit wordt heterogeniteit genoemd - en het gecombineerde effect dus moeilijk te interpreteren was.

Het doel van Hoofdstuk 6 was om een voorbeeld te geven van het combineren van verschillende soorten bewijs, waaronder epidemiologie, humane biomonitoring en dier studies. Meestal worden deze gegevens bij risicobeoordeling afzonderlijk door experts (toxicologen, epidemiologen) in wetenschappelijke discussies beschouwd. Het is niet gebruikelijk om dit soort verschillende type data direct en kwantitatief te combineren om zo een blootstellingsniveau te kunnen schatten waarboven het waarschijnlijk is dat een schadelijk gezondheidseffect zal optreden. Daarom was het doel van hoofdstuk 6 om een voorbeeld te verkennen van hoe een gecombineerde en kwantitatieve benadering zou kunnen werken, en wat de consequenties en beperkingen zijn van een dergelijke aanpak. De gekozen benadering was om al het bewijs om te zetten (d.w.z. te harmoniseren) in risicoschattingen zoals die in epidemiologische studies worden gerapporteerd. Dit bracht moeilijkheden met zich mee omdat de verschillende soorten data niet eenvoudig te combineren zijn. Zo was de blootstelling aan benzeen in dierstudies goed gecontroleerd en in het algemeen gedurende een paar uur op opeenvolgende dagen, gedurende enkele maanden en tot relatief hoge concentraties. De eindpunten die in deze onderzoeken werden geregistreerd, waren het optreden van tumoren. In studies met menselijke biomonitoring daarentegen werd de blootstelling aan benzeen slechts één keer gemeten en werden tussentijdse

gezondheidseffecten gerapporteerd als biomarker van effect. Epidemiologische studies daarentegen volgden grote groepen mensen jarenlang en rapporteerden vaak cumulatieve benzeenblootstellingen. Hoewel ik ontdekte dat de blootstellingsresponse-relatie statistisch nauwkeuriger was toen ik alle data combineerde, was het moeilijk om de onzekerheid te kwantificeren die werd geïntroduceerd vanwege alle vereiste harmonisatiestappen.

Afsluitende opmerkingen

Door de hoofdstukken heen zijn verschillende voorbeelden gegeven voor het (kwantitatief) overwegen van verschillende soorten bewijs. De gebruikte kwantitatieve benaderingen gaven inzicht in onderzoeks-hiaten. Het dwong me ook om expliciet te zijn, waardoor de benaderingen transparant en reproduceerbaar werden voor andere onderzoekers. Voor al het werk dat in dit proefschrift is uitgevoerd, heb ik me gebaseerd op gepubliceerde gegevens. Vooral voor de humane biomonitoringstudies van zowel blootstelling als effect leverde dit problemen op omdat de studies verschillend van opzet waren en daardoor niet altijd gemakkelijk te combineren waren. Daarnaast was in sommige gevallen geen informatie beschikbaar over parameters die nodig zijn om de gegevens voor andere doeleinden te kunnen hergebruiken. Door de hoofdstukken heen werden lacunes in het onderzoek geïdentificeerd en werden aanbevelingen gedaan voor het verzamelen en rapporteren van gegevens. Hoewel verschillen tussen studies de interpretatie van resultaten vaak een uitdaging maakten, hebben risicobeoordelaars door het combineren van gegevens en het geven van samenvattende schattingen toch de mogelijkheid om een algemene, beste "gemiddelde" schatting te selecteren, zelfs als die schatting onnauwkeurig is.

De casestudy's van dit proefschrift illustreren dat humane biomonitoringstudies een waardevol bron van informatie kunnen zijn voor zowel het verbeteren van de blootstellingsbepaling, als voor het karakteriseren van blootstellingsniveaus waarbij een nadelig gezondheidseffect te verwachten is. Tooling zoals text mining en netwerkvisualisatiehulpmiddelen kunnen helpen om relevante studies over mogelijke stof specifiek gezondheidseffecten meer efficiënt te identificeren. En bewijs uit humane studies (epidemiologie en biomonitoring) en dierstudies kan ook gecombineerd worden, maar daarvoor zijn wel een groot aantal aannames nodig. De hoofdstukken in dit proefschrift laten zien dat er een breed scala aan interdisciplinaire benaderingen nodig is om data te kunnen integreren, zodat ze optimaal bruikbaar worden voor chemische risico beoordeling. Dit kan een uitdaging zijn voor risicobeoordelaars, dus er moet een duidelijke leidraad zijn voor het combineren van verschillende soorten bewijs en hoe de resultaten kunnen worden geïnterpreteerd. De methoden en tooling in dit proefschrift worden meestal toegepast op chemicaliën waarvoor een grote hoeveelheid data bekend is. In het Discussie hoofdstuk van dit proefschrift worden suggesties gegeven voor benaderingen die kunnen worden gebruikt voor chemicaliën waarover slechts weinig informatie bekend is, mogelijk in

samenhang met de nieuwe wetenschappelijke benaderingen beschreven in de andere hoofdstukken van dit proefschrift.

Scientific Publications

Scholten, Bernice, Laura Kenny, Radu-Corneliu Duca, Anjoeka Pronk, Tiina Santonen, Karen S. Galea, Miranda Loh et al. "Biomonitoring for occupational exposure to diisocyanates: a systematic review." Annals of work exposures and health 64, no. 6 (2020): 569-585.

Scholten, Bernice, Jelle Vlaanderen, Rob Stierum, Lützen Portengen, Nat Rothman, Qing Lan, Anjoeka Pronk, and Roel Vermeulen. "A quantitative meta-analysis of the relation between occupational benzene exposure and biomarkers of cytogenetic damage." Environmental health perspectives 128, no. 8 (2020): 087004.

Scholten, Bernice, Laura Guerrero Simón, Shaji Krishnan, Roel Vermeulen, Anjoeka Pronk, Benjamin M. Gyori, John A. Bachman, Jelle Vlaanderen, and Rob Stierum. "Automated network assembly of mechanistic literature for informed evidence identification to support cancer risk assessment." Environmental Health Perspectives 130, no. 3 (2022): 037002.

Scholten, Bernice, Lützen Portengen, Anjoeka Pronk, Rob Stierum, George S. Downward, Jelle Vlaanderen, and Roel Vermeulen. "Estimation of the Exposure–Response Relation between Benzene and Acute Myeloid Leukemia by Combining Epidemiologic, Human Biomarker, and Animal Data." Cancer Epidemiology, Biomarkers & Prevention 31, no. 4 (2022): 751-757.

Jones, Kate, Karen S. Galea, **Bernice Scholten**, Marika Loikala, Simo P. Porras, Radia Bousoumah, Sophie Ndaw et al. "HBM4EU Diisocyanates Study—Research Protocol for a Collaborative European Human Biological Monitoring Study on Occupational Exposure." International Journal of Environmental Research and Public Health 19, no. 14 (2022): 8811.

Scholten, B., J. Westerhout, A. Pronk, R. Stierum, J. Vlaanderen, R. Vermeulen, K. Jones, T. Santonen, and L. Portengen. "A physiologically-based kinetic (PBK) model for work-related diisocyanate exposure: relevance for the design and reporting of biomonitoring studies." Environment International (2023): 107917.

Huuskonen, Pasi, Simo P. Porras, **Bernice Scholten**, Lützen Portengen, Sanni Uuksulainen, Katriina Ylinen, and Tiina Santonen. "Occupational Exposure and Health Impact Assessment of Diisocyanates in Finland." Toxics 11, no. 3 (2023): 229.

Dankwoord

Ik wil graag starten met het bedanken van de promotor van mijn PhD: Roel Vermeulen. We spraken elkaar niet bijzonder vaak maar je was steeds in staat om in een korte tijd te begrijpen wat de stand van zaken was, wat eventuele problemen waren en ook wist je hier dan met concrete voorstellen oplossingen voor aan te dragen. Je input kwam altijd op het juiste moment, en jouw visie heeft mijn PhD gevormd. Mijn dank gaat ook uit naar mijn begeleiders Jelle, Rob en Anjoeka: Jelle, jouw kritische blik heeft geholpen bij het uitvoeren en uitwerken van het onderzoek in alle hoofdstukken, maar zeker ook bij het schrijven van de Introductie en Discussie. Rob en Anjoeka, ook jullie bedankt voor jullie input in de afgelopen jaren, en ook dank voor de ruimte en vertrouwen om mijn weg te zoeken in de werkzaamheden van het HBM4EU project en het werk voor mijn PhD. Rob, het was een uitdagend en leerzaam proces om samen te werken aan hoofdstuk 2.

Lützen, ik kan niet genoeg benadrukken hoe belangrijk jij bent geweest in elk stadium van mijn PhD. Je hebt onvermoeid(?) antwoord gegeven op mijn vele vragen over statistiek en epidemiologie, en geholpen met het programmeer werk. Ondanks het feit dat je me in gedachte altijd twintig stappen vooruit was, gaf je me toch het gevoel dat we als gelijkwaardige collega's aan het discussiëren waren, en dat maakte dat ik steeds meer vertrouwen en plezier kreeg in het onderzoek. Zonder jou was dit hele PhD avontuur niet gelukt. Mariana en Tanja, bedankt dat jullie mijn paranimfen willen zijn. Tanja, ik vind je niet alleen een hele goede statisticus, maar ook een fijne en inspirerende vriendin. Ik kijk uit naar nog vele gezellige etentjes samen. Mariana, it was a bit of struggle to find my way at IRAS given my joint position at TNO, but you reached out and initiated the best imaginable IRAS breaks: yoga-halloumi session. I also enjoyed our many discussions on combing motherhood with work, and I hope we remain friends for many more years. Mijn dank gaat verder uit naar alle IRAS collega's met wie ik regelmatig pauze wandelingen maakte, al waren deze meestal kort naar de SPAR en terug omdat we allemaal druk waren.

Ik werk al jaren graag bij TNO en dan komt onder meer door alle fijne collega's. Specifiek wil ik nog de volgende mensen noemen. Eelco, onze gesprekken helpen me regelmatig bij het maken van keuzes, of het oplossen van problemen, in mijn werkzaamheden. Bedankt voor je betrokkenheid. Susan, gezellig en leerzaam dat we steeds meer samen werken. Ik kijk uit naar al het werk rondom safe and sustainable by design. Joost, het was erg leerzaam om met jou samen te werken aan PBK modellen. Remy, jouw zwart-galige humor draagt in mijn ogen zeker bij aan een goede werksfeer. Jody en Wouter, dankzij jullie initiatieven kon ik na mijn PhD onder meer aan de slag binnen het pre-pmc hitte en het meeschrijven aan een Novel Point Of View. Harrie, bedankt voor de spoedcursus "in-silico methods for data poor substances". Monique dank voor de mogelijkheid die ik heb gehad om een PhD traject te volgen vanuit TNO, en dank voor de ruimte die je gaf toen bleek dat er thuis veel

extra zorg voor Jens nodig was. Helena, dank voor de inspirerende coachings-sessies, en Calvin, thank you for critically reviewing my Introduction en Discussion chapters, én for introducing good coffee at the workplace.

Mijn dank gaat uit naar mijn ouders: Ik ben dankbaar voor jullie onvoorwaardelijke liefde door alle jaren heen, jullie betrokkenheid, en de goede zorgen voor ons hele gezin als we bij jullie zijn. Mam, bedankt voor het helpen bij het ontwerp van dit boek, ik vind het resultaat heel gaaf geworden. Florina, ik waardeer het dat je ons altijd het allerbeste gunt en dat je regelmatig komt logeren en dan zorgt voor gezelligheid, opgeruimdheid en luchtigheid in ons gezin. Jefta en Sylvana, ik vind onze wandelingen en spelletjesavonden heel gezellig, maar ook bedankt voor de goede zorgen voor Jens toen hij zoveel in het ziekenhuis lag. Ali, je tovert altijd een lach op het gezicht van Jens (en dus van ons) en dat wordt enorm gewaardeerd. En ook de rest van mijn familie (Simon en Conny, Hans en Femke, en Charissa, Joëlle en Lennart en hun gezinnen) wil ik bedanken voor alle gezellige verjaardagen en bezoeken, en de betrokkenheid naar elkaar toe.

Bart: Erwin en ik hebben een meer dan geweldige tijd gehad op jouw Boerderie. En dankzij jouw generositeit konden we in alle rust en toewijding voor Jens zorgen in de eerste jaren na zijn geboorte. Didi en Loes, we zijn inmiddels al een lange tijd vriendinnen en ik waardeer onze vriendschap en de gezelligheid en ontspanning in het samenzijn. Jean-Yves, Annemiek, wat fijn dat we na al die jaren vriendschap nu zo dicht bij elkaar wonen, en in plaats van op avontuur in Oeganda nu van alles meemaken in Wageningen en omgeving met onze kinderen. Hendrik en Lena, bedankt voor de gezellige zeiltochten en etentjes; ik vind het inspirerend om te zien hoe jullie op eigen wijze invulling geven aan zaken als opvoeding, duurzaam leven en zingeving. En Lena, enorm bedankt voor al je hulp bij de lay-out van dit boek. Ook wil ik naast alle leuke mensen die ik heb ontmoet op de Boerderie en tijdens het studeren specifiek Laurens, Herman, Stella, Lotte en Ben en Eva bedanken voor onze vriendschappen door de jaren heen. Liesbeth, je hebt Kaylah een super plekje geboden, dank voor je goede zorgen voor haar.

Tenslotte de drie grote liefdes van mijn leven: Erwin, Jens en Jona. Erwin we zijn al zo lang samen en hebben van alles meegemaakt. Jij moet het vaak als eerste ontgelden als ik druk ben met werk of toen ik me aan het begin van mijn PhD zo vaak gespannen voelde en twijfelde of het wel ging lukken. Bedankt dat je altijd enthousiast bent over mijn keuzes en plannen, en dat je hier ook ruimte voor maakt.
Jens en Jona, het moederschap is het beste wat me is overkomen. Ik slaap vaak slecht, zit regelmatig onder de spuug en heb al duizenden uren naar Nijntje Pluis gekeken, maar niks geeft me een fijner gevoel als bij jullie zijn en voor jullie te zorgen.

Curriculum Vitae

Bernice Scholten werd geboren op 20 juli 1986 in Hellevoetsluis. Na in 2002 haar HAVO diploma behaald te hebben aan het Jacob van Liesveld college te Hellevoetsluis, startte zij met de opleiding leerkracht basisonderwijs aan de Thomas More Hogeschool te Rotterdam. In 2007 behaalde ze hiervoor haar bachelor en in de twee daaropvolgende jaren gaf zij les aan groep 3 en 4 in basisscholen in Brielle en Rotterdam. Tegelijkertijd volgde zij twee thuisstudies voor natuurkunde en wiskunde. In 2008 behaalde zij het VWO staatsexamen voor beiden vakken en in 2009 startte zij met de studie International Land en Waterbeheer aan de Wageningen Universiteit. Voor haar bachelor thesis deed ze onderzoek in Nepal naar de effectiviteit van keramische klei filters in het zuiveren van drinkwater. In 2012 vervolgde zij haar bachelor met een master in Environmental Sciences, specialisatie toxicologie. Haar master stage omhelsde een stage van een half jaar bij de Tanzania Food and Drugs Authority, waar ze meeliep in een project wat onderzoek deed naar mycotoxines in graan- en borstmelk. Tijdens haar master thesis heeft ze gewerkt met in vitro test systemen middels onderzoek naar de toxiciteit van nanodeeltjes op het immuunsysteem. Na het behalen van haar master heeft zij eerst gewerkt als vrijwilliger, en later als consultant, bij de Wereldgezondheidsorganisatie, afdeling Chemical Safety van de afdeling Public Health, Environmental and Social determinants of Health, in Geneve. Nadat ze vervolgens in 2015 voor een half jaar had gewerkt bij RoyalHaskoning DHV als consultant chemicals management, startte ze in 2016 bij TNO, afdeling Risk Analysis for Products in Development. Een jaar later startte zij het promotieonderzoek zoals beschreven in dit proefschrift, vanuit een samenwerking tussen TNO en het Institute for Risk Assessment Sciences (IRAS), onderdeel van de Universiteit Utrecht.

Bernice Scholten was born on July 20, 1986 in Hellevoetsluis, the Netherlands. After completing secondary school at the Jacob van Liesveld College in Hellevoetsluis, she started the bachelor for primary school teacher at the Thomas More Hogeschool in Rotterdam. In 2007 she obtained her bachelor's degree and in the following two years she taught group 3 and 4 in primary schools in Brielle and Rotterdam. At the same time she followed two home studies for physics and mathematics. In 2008 she passed the VWO state exam for both subjects and in 2009 she started studying International Land and Water Management at Wageningen University. For her bachelor thesis she conducted research in Nepal on the effectiveness of ceramic clay filters in purifying drinking water. In 2012 she continued her bachelor's degree with a master's degree in Environmental Sciences, specializing in toxicology. Her master's internship included a six-month internship at the Tanzania Food and Drugs Authority, where she

participated in a project that researched mycotoxins in grain and breast milk. For her master thesis she investigated the immunotoxicity of nanoparticles on in vitro test systems. After obtaining her master's degree, she first worked as a volunteer, and later as consultant, at the World Health Organization, Department of Chemical Safety of the Department of Public Health, Environmental and Social Determinants of Health, in Geneva. Following this experience she held a position as expert chemicals management at RoyalHaskoning DHV. In 2016 she started working at TNO, at the Risk Analysis for Products in Development department. A year later she started the PhD research as described in this thesis, from a collaboration between TNO and the Institute for Risk Assessment Sciences (IRAS), part of Utrecht University.