Advancing the Contribution of Occupational Epidemiology to Risk Assessment

Jelle Vlaanderen

2011

J.J. Vlaanderen, 2011
Advancing the Contribution of Occupational Epidemiology to Risk Assessment
Thesis Utrecht University
ISBN: 978-90-393-5532-9

Cover by Joost Kroese
Lay-out by Harry Otter, Division Multimedia, Veterinary Medicine, Utrecht University
Printed by Drukkerij Ridderprint, Ridderkerk

Advancing the Contribution of Occupational Epidemiology to Risk Assessment

Bevordering van de bijdrage van de arbeidsepidemiologie aan de risicobeoordeling van chemische stoffen

(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. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op dinsdag 19 april 2011
des middags te 4.15 uur

door

Jelle Joost Vlaanderen

geboren op 11 januari 1981 te Soest

Promotor: Prof. dr. ir. H. Kromhout
Co-promotor: Dr. ir. R.C.H. Vermeulen

The studies described in this thesis were financially supported by ECNIS (European Cancer Risk, Nutrition and Individual Susceptibility), a Network of Excellence operating within the European Union 6th Framework Program, Priority 5: 'Food Quality and Safety' (Contract No 513943).

Table of Contents

Chapter 1

General Introduction

The risk assessment paradigm

In 1983 the United States National Research Council defined risk assessment as "…the characterization of the potential adverse health effects of human exposures to environmental hazards." (1). Four components, known as the 'risk assessment paradigm', were suggested (1). *Hazard identification* is the evaluation of all available evidence to determine whether an exposure can cause adverse health effects in humans, *dose-response assessment* is the characterization of the relationship between the level of exposure and the occurrence of an adverse health effect, *exposure assessment* determines the extent to which humans are exposed to an agent, and *risk characterization* integrates the information from *hazard identification, dose-response assessment,* and *exposure assessment* to describe the nature and magnitude of the risk in humans (1, 2). Risk assessment relies heavily on data from scientific studies such as long-term bioassays in experimental animals, mechanistic studies focused on toxicokinetics and disease mechanisms, and epidemiological studies.

The role of occupational epidemiology in risk assessment

This thesis is focused on the role of occupational epidemiological studies of chemical exposures in risk assessment. The main advantage of occupational epidemiological studies over other potential sources of information for risk assessment (primarily animal bioassays) is that humans are studied. As a result no species to species extrapolation is necessary (as is the case for animal bioassays); the context, levels and scenarios of exposure are realistic; and the variability in susceptibility to adverse health effects is represented (3). Furthermore, occupational populations frequently experience high exposures which facilitate the detection of a potential increase in disease risk above background. In addition, historical industrial hygiene measurement data are sometimes available that can be used for the retrospective quantification of exposure levels. This is important because epidemiological studies of chronic health effects generally require retrospective exposure assessment. Indeed, occupational epidemiological data has played an important role in risk assessment over the years especially in the identification of new risk factors (hazard identification). For example, in the hazard identification program of the International Agency for Research on Cancer (IARC Monograph series (4)) sufficient epidemiological evidence of carcinogenicity in humans is generally required before an agent is classified as 'carcinogenic to humans'.
The major advantage of occupational epidemiological studies for risk assessment (humans are studied) is also a major disadvantage. Occupational epidemiological studies are mostly observational by nature which makes them prone to bias. The limitations of using occupational epidemiological data in risk assessment are most pronounced in dose-response assessment. Ideally, results from a quantitative exposure-response analysis conducted in occupational epidemiological studies could be applied almost directly in this step of the risk

assessment paradigm. However, as a result of their quantitative nature, exposure-response analyses are particularly sensitive to a lack of available exposure data, exposure misclassification, potential biases and limited statistical power; issues that frequently occur in occupational epidemiological studies (5). Therefore, increased attention to potential sources of uncertainty in occupational epidemiological data might considerably increase their value for dose-response assessment (3, 5-7).

The current challenge for risk assessment

In the past decades occupational epidemiological studies successfully contributed to the hazard identification of major environmental risk factors (8, 9). The risk factors that have been identified to date are generally characterized by strongly increased risks (e.g., smoking and lung cancer), high exposure levels (e.g., occupational benzene exposure and leukemia) or a specific relation between exposure and disease (e.g., asbestos and mesothelioma). With seemingly all conspicuous risk factors identified the current challenge for risk assessment is the identification and quantification of (environmental) risk factors that are characterized by lower exposure levels, moderately increased risks, and less specific exposure-disease relations (10). While these risk factors will be more difficult to study, their impact on public health can be considerable depending on the prevalence of exposure in the general population (11).
Occupational epidemiological studies can play a role in identification and quantification of more subtle links between environment and disease. However, this will put high demands on the quality of the evidence generated by occupational epidemiological studies. For example, in many existing occupational epidemiological studies random error in exposure estimates and limited statistical power prohibit the observation of moderately increased risks in lower exposure ranges. Furthermore, the potential impact of systematic error in exposure estimates, essential for the interpretation of a study in dose-response assessment, is often not clear (12).

Advancing the use of occupational epidemiology in risk assessment

Although some limitations of the use of occupational epidemiological studies in risk assessment are inherent to the discipline, improvements in the design, conduct and interpretation of studies will likely enhance their use in risk assessment. Furthermore, the developments in the field of molecular biology and the related increase in the understanding of carcinogenesis and other adverse health effects have opened opportunities to further advance the contribution of occupational epidemiological studies to risk assessment by applying molecular information in the epidemiological analyses.

Incorporating the weight of evidence principle

Weight of evidence is defined here as the principle of using as much evidence as possible to support decision making, while explicitly focusing on and accounting for the quality and relevance of the evidence (13). The weight of evidence principle should be applied to the evaluation and synthesis of occupational epidemiological data for risk assessment. However, approaches that systematically incorporate the weight of evidence principle into the evaluation and synthesis of occupational epidemiological data for risk assessment are currently lacking. As a result, to date the weight of evidence principle has found limited application in risk assessment based on occupational epidemiological data.

Evaluation of occupational epidemiological studies

Not all occupational epidemiological studies are equally informative for risk assessment. Which studies are informative and which studies are not greatly depends on the purpose for which studies are collected. While an occupational epidemiological study that reports an increased disease risk for an 'exposed' group relative to an 'unexposed' group (without any quantification of the exposure levels) can still be informative for hazard identification, the same study would be uninformative for dose-response assessment. Furthermore, even within a set of occupational epidemiological studies that is informative for dose-response assessment differences in the quality of the evidence will be present and should be acknowledged in a weight of evidence approach. The weight of evidence of an occupational epidemiological study for risk assessment is determined by the quality of the design, conduct and reporting of the study and by the relevance of the study hypothesis for the risk assessment in question. Differences in the quality of design and conduct are often not reflected in parameters of statistical uncertainty (e.g., confidence intervals). It is therefore important that a rigorous evaluation of study quality is performed before occupational epidemiological study data are used in a risk assessment.

Synthesis of occupational epidemiological data

Systematic synthesis of evidence from occupational epidemiological studies to summarize results and to address potential heterogeneity in the results is a crucial aspect of the application of the weight of evidence principle in risk assessment based on occupational epidemiological studies. Synthesis might include a qualitative description of the evidence (narrative review), a quantitative summary of the evidence (meta-analysis), or a re-analysis based on the original data of the individual studies (pooled analysis) (14). For hazard identification meta-analysis and pooled analysis would generally be preferred over the narrative review because these approaches will result in summary risk estimates that are in theory directly applicable in risk assessment and provide quantitative measures of the precision of summary risk estimates. However, although meta-analysis and pooled analysis contribute to an increase in the precision of a summary risk estimate by addressing the variability in study results that is due to random variation, it is important to realize that

systematic error (of more or less similar direction in all studies) could result in a more precise, but equally biased summary risk estimate (15, 16). Careful evaluation of study quality and the sources of heterogeneity between study results is therefore crucial for the use of meta-analysis and pooled analysis in risk assessment based on occupational epidemiological studies (15).

Meta-regression methods can be used to quantitatively synthesize exposure-response relations derived in occupational epidemiological studies (17). In addition to a role in hazard identification (to explore the exposure-response relation; one of Bradford Hill's guidelines to assess causality (18)), meta-regression can also directly inform dose-response assessment. Meta-regression methods can contribute to the precision of the estimated summary exposure-response relation and can be used to quantify heterogeneity between studies with regard to the reported exposure-response relations. However, the limitations of the application of meta-analysis to observational epidemiological studies apply equally well to meta-regression and a close examination of the sources of heterogeneity is warranted (15, 17). The application of meta-regression methods requires extra careful evaluation of the quality of the exposure assessment in the contributing studies as systematic differences in exposure measurement error are likely to occur and might easily distort the outcomes of a meta-regression. Furthermore, while in previous applications of meta-regression to occupational epidemiological data predominantly linear exposure-response relations were assumed, increased flexibility in the type of regression models that are fitted to occupational epidemiological data will increase the insight into exposure-response relations and thereby advance the use of quantitative occupational exposure-response data in dose-response assessment.

The use of biomarkers in occupational epidemiological studies

Biomarkers have the potential to play an important role in advancing the contribution of occupational epidemiological studies to risk assessment. Revolutions in the field of molecular biology have provided epidemiologists many tools to move beyond the classic study design of correlating external exposure to clinically manifested disease (19). In occupational epidemiological studies the measurement of biomarkers might contribute to improved quality of the assessment of exposure and outcome, help to better understand the distribution of susceptibility to health outcomes in study populations, and might increase the ability to observe adverse health effects in the early stages of a disease (20). In addition to the direct implementation in occupational epidemiological studies, biomarkers might also provide a means to formally integrate evidence from toxicological studies and animal bioassays into occupational epidemiological studies to further increase the use of these types of evidence in risk assessment (21). Examples of such integration are linkage of exposure levels experienced in human populations to outcomes from animal bioassays through the use of physiologically based pharmacokinetic models (21), or the use of biomarkers (22).

However, similar to any other measure of exposure or effect the actual value of biomarkers for occupational epidemiological studies (and ultimately risk assessment) depends largely on the accuracy and reliability of the assay of biomarkers, (insight into) variability of the biomarkers within and between individuals, and the study design in which they are applied (20, 23).

An example: the risk assessment of benzene

Benzene is one of the exposures for which risk assessment is primarily based on occupational epidemiological studies (24-28). Benzene is an aromatic hydrocarbon used in industrial processes (e.g., in the production of rubber products, or as a solvent in paints and glues) and to which the general public is exposed via (environmental) tobacco smoke, car exhaust, use of solvents and paints, among other sources (29). There is general consensus that exposure to benzene causes acute non-lymphocytic leukemia, which consists primarily of acute myeloid leukemia (AML) and might be related to other types of lymphohematopoietic diseases such as acute lymphocytic leukemia, chronic lymphocytic leukemia, multiple myeloma, and non-Hodgkin lymphoma (30, 31).

Although the mechanisms by which benzene exerts its toxic effects are not yet fully elucidated (32, 33), metabolism of benzene into toxic metabolites is thought to play an important role (34-42). The major metabolic pathways are shown in Figure 1. Benzene is metabolized by CYP enzymes (primarily CYP2E1) to benzene oxide (BO, which is in equilibrium with its tautomer, oxepin), and is the source of all other metabolites. Spontaneous rearrangement of BO produces phenol (PH), which can undergo another CYP oxidation to give hydroquinone (HQ). Hydrolysis of BO via epoxide hydrolase produces benzene dihydrodiol which can be converted to catechol (CA), via dihydrodiol dehydrogenases, or to benzene diolepoxides via CYP oxidation. HQ and CA can be oxidized to 1,4-benzoquinone (1,4-BQ) and 1,2-benzoquinone (1,2-BQ), respectively. A second CYP oxidation of oxepin, followed by ring opening eventually results in the production of E,E-muconic acid (MA). S-phenylmercapturic acid (SPMA) is a minor benzene metabolite that is produced following a reaction of BO with gluthathione. For airborne exposure levels between 0.1 and 10 ppm PH represents 70-85% of the urinary metabolites, HQ, MA, and CA each represent 5-10% and SPMA represents < 1% (34). The mechanisms by which benzene metabolites affect lymphohematopoiesis are thought to involve one or more of the electrophilic metabolites (e.g., BO, 1,2-BQ, or 1,4-BQ) that are capable of binding to DNA and other macromolecules and/or reactive oxygen species, produced by redox cycling of CA, HQ and the respective benzoquinones (42). However, the involvement of other metabolites and mechanisms of action has been proposed as well and it is likely that benzene has a multimodal mechanism of action (33).

The evaluation of benzene's carcinogenicity by the IARC Monograph program (30, 43, 44) (hazard identification) and the quantitative risk assessment of benzene by the Health Council

of the Netherlands (27, 28) are discussed below to provide examples of the current use of occupational epidemiological evidence in the risk assessment process.

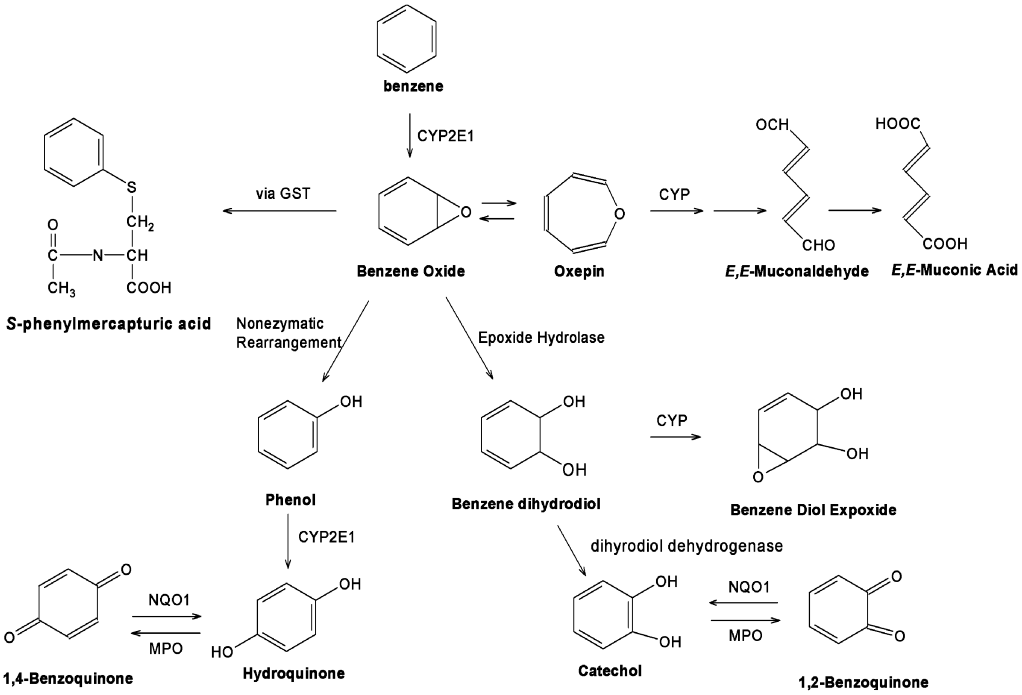


Figure 1: Simplified metabolic scheme for benzene showing major pathways and metabolizing genes.

Risk assessment of benzene

In 1982 the IARC Monograph on benzene stated: "The relationship between benzene exposure and the development of acute myelogenous leukemia has been established in epidemiological studies." (43). As a result IARC concluded that: "There is sufficient evidence that benzene is carcinogenic to man." (43). This conclusion, which was confirmed in subsequent evaluations in 1987 (44) and 2009 (30), was based on epidemiological studies conducted in occupational populations exposed to high levels of benzene and is an example of hazard identification: the 1982 IARC Monograph stated that benzene is a carcinogen, but did not quantify the exposure level at which there is a relevant increase in risk. The Monograph also noted that "Reports linking exposure to benzene with other malignancies were considered to be inadequate for evaluation" (43). This statement reflects the difficulty that the working group had to observe moderately increased risks for other malignancies based on the then available evidence. In the 2009 evaluation, however, the working group

noted that there was limited human evidence that benzene was linked to acute lymphocytic leukemia, chronic lymphocytic leukemia, multiple myeloma, and non-Hodgkin lymphoma, which illustrates the increased body of evidence that was available by 2009 (30).
As a result of the identification of benzene as a confirmed carcinogen regulatory agencies had to quantify what the 'acceptable' exposure level for benzene was. The Health Council of the Netherlands derived an acceptable environmental exposure limit for benzene in 1987 (28). Their approach involved selecting the 'best' study that quantified the relation between exposure to benzene and AML. The selected study reported on the incidence of leukemia (primarily AML) in a cohort of rubber hydrochloride workers from Ohio, USA (the "Pliofilm" cohort) that experienced high exposures to benzene (45). The findings of that study were linearly extrapolated down to exposure levels relevant for the general population. Interestingly, the Health Council noted in their report that the approach followed most likely overestimated the risk of benzene at low exposure levels and decided to multiply the derived acceptable exposure level (i.e., the exposure level that corresponded to one additional case of leukemia per one million individuals with a lifelong exposure to benzene) by a factor of one hundred. This was a reflection of the Council's limited confidence in the exposure estimates that were used in the Pliofilm study and the uncertainties regarding the validity of the approach that was used to extrapolate the study findings to the general population. In a 1997 update of the original risk assessment the Dutch Health Council did note that new epidemiological evidence was available but decided to use the original 'acceptable exposure level' that was derived in 1987 (27).

Benzene was used as a case substance in the approaches to advance the contribution of occupational epidemiology to risk assessment presented in this thesis. The primary reason for selecting benzene as case substance was the large role that occupational epidemiology played in the risk assessment of benzene (24-28).

This thesis

In this thesis a framework to advance the contribution of occupational epidemiological studies to risk assessment is proposed. The presented approaches address aspects that are central to the interpretation of occupational epidemiological data in risk assessment and have the potential to improve the value of future occupational epidemiological studies for risk assessment. In Chapter 2 a set of guidelines to evaluate the quality of (occupational) epidemiological studies for risk assessment is presented. The evaluation guidelines are subsequently used to evaluate the quality of studies in two evidence synthesis approaches discussed in Chapter 4 (meta-analysis) and Chapter 5 (meta-regression). The guidelines specifically address the importance of the quality of exposure assessment because of its large impact on the overall relevance of occupational epidemiological studies for risk assessment.

In Chapter 3 a specific aspect of exposure assessment in occupational epidemiological studies is addressed: the temporal coverage of occupational histories by exposure measurement data. A graphical tool that visualises differences between studies in temporal coverage of exposure history by exposure measurements is presented. Implementation of the graphical tool in standard reporting of occupational epidemiological studies will facilitate the evaluation of study quality as it relates to exposure assessment and the relevance of such studies for risk assessment.
Chapter 4 is an example of a transparent approach to integrate evaluation of study quality into evidence synthesis. Three different proxies were developed to assess study quality and stratification on these proxies was used to assess the relation between study quality and the outcome of a meta-analysis.
In Chapter 5 an approach for a flexible meta-regression of aggregated risk estimates from occupational epidemiological studies is introduced. This approach facilitates evidence synthesis of quantitative exposure-response relations reported in occupational epidemiological studies.
Chapters 6 and 7 address the potential role that biomarkers can play in improving the quality of occupational epidemiological studies for risk assessment. Chapter 6 provides an overview of a new generation of biomarker research (collectively described with the term OMICS); provides examples of their application in epidemiological studies of occupational and environmental health; and describes current difficulties that exist with the implementation of biomarkers (especially OMICS) in (occupational) epidemiological studies. Chapter 7 is a practical example of how information from biomarker studies can be incorporated into occupational epidemiological studies to improve exposure assessment and to address etiological questions.
In Chapter 8 the degree to which the approaches introduced in this thesis advance the use of occupational epidemiological studies in risk assessment and possibilities for further improvements are discussed.

References

1. National Research Council (U.S.). Committee on the Institutional Means for Assessment of Risks to Public Health. Risk assessment in the federal government: managing the process. Washington, D.C.: National Academy Press, 1983.
2. Cogliano VJ. Principles of Cancer Risk Assessment: The Risk Assessment Paradigm. In: Edler L, Kitsos CP, eds. Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment. Chichester: John Wiley, 2005.
3. Hertz-Picciotto I. Epidemiology and quantitative risk assessment: a bridge from science to policy. Am J Public Health. 1995;85(4):484-91.
4. International Agency for Research on Cancer. Preamble to the IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon: International Agency for Research on Cancer, 2006.
5. Stayner L, Bailer AJ, Smith R, et al. Sources of uncertainty in dose-response modeling of epidemiological data for cancer risk assessment. Ann N Y Acad Sci. 1999;895:212-22.

6. Nurminen M, Nurminen T, Corvalan CF. Methodologic issues in epidemiologic risk assessment. Epidemiology. 1999;10(5):585-93.
7. van den Brandt P, Voorrips L, Hertz-Picciotto I, et al. The contribution of epidemiology. Food Chem Toxicol. 2002;40(2-3):387-424.
8. Rushton L, Hutchings S, Brown T. The burden of cancer at work: estimation as the first step to prevention. Occup Environ Med. 2008;65(12):789-800.
9. Lopez AD, Mathers CD, Ezzati M, et al. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet. 2006;367(9524):1747-57.
10. Taubes G. Epidemiology faces its limits. Science. 1995;269(5221):164-9.
11. Benichou J. Biostatistics and epidemiology: measuring the risk attributable to an environmental or genetic factor. C R Biol. 2007;330(4):281-98.
12. Blair A, Stewart P, Lubin JH, et al. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med. 2007;50(3):199-207.
13. Weed DL. Weight of evidence: a review of concept and methods. Risk Anal. 2005;25(6):1545-57.
14. Blettner M, Sauerbrei W, Schlehofer B, et al. Traditional reviews, meta-analyses and pooled analyses in epidemiology. Int J Epidemiol. 1999;28(1):1-9.
15. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. BMJ. 1998;316(7125):140-4.
16. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283(15):2008-12.
17. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. Epidemiology. 1993;4(3):218-28.
18. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295-300.
19. Boffetta P. Biomarkers in cancer epidemiology: an integrative approach. Carcinogenesis. 2010;31(1):121-6.
20. Garcia-Closas M, Vermeulen R, Sherman M, et al. Application of biomarkers in cancer epidemiology. In: Schottenfeld D, Fraumeni JF, Jr. eds. Cancer Epidemiology and Prevention. New York, NY: Oxford University Press, 2006:70-88.
21. Conolly RB, Kimbell JS, Janszen D, et al. Human respiratory tract cancer risks of inhaled formaldehyde: dose-response predictions derived from biologically-motivated computational modeling of a combined rodent and human dataset. Toxicol Sci. 2004;82(1):279-96.
22. Hack CE, Haber LT, Maier A, et al. A Bayesian network model for biomarker-based dose response. Risk Anal. 2010;30(7):1037-51.
23. Vineis P, Gallo V. The epidemiological theory: principles of biomarker validation. In: Vineis P, Gallo V, eds. Epidemiological Concepts of Validation of Biomarkers for the Identification/Quantification of Environmental Carcinogenic Exposures. Lodz: Nofer Institute of Occupational Medicine, 2007.
24. California EPA. Public Health Goal for Benzene In Drinking Water. Scramento, California: Office of Environmental Health Hazard Assessment, 2001.
25. U.S. EPA. Carcinogenic Effects of Benzene: An Update. EPA/600/P-97/001F. Washington, DC: United States Environmental Protection Agency, 1998.
26. ATSDR. Toxicological Profile for Benzene. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, 2007.
27. Gezondheidsraad. Comissie Beoordeling carcinogeniteit van stoffen. Benzeen. Rijswijk, 1997.
28. Gezondheidsraad. Benzeen. Toetsing van een basisdocument en voorstel risico-evaluatie. Den Haag, 1987.
29. Wallace L. Environmental exposure to benzene: an update. Environ Health Perspect. 1996;104 Suppl 6:1129-36.
30. Baan R, Grosse Y, Straif K, et al. A review of human carcinogens--Part F: chemical agents and related occupations. Lancet Oncol. 2009;10(12):1143-4.
31. U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program. Report on Carcinogens, Eleventh Edition. http://ntp.niehs.nih.gov/ntp/roc/toc11.html. Accessed November 2010.
32. Ahmad Khan H. Benzene's toxicity: a consolidated short review of human and animal studies. Hum Exp Toxicol. 2007;26(9):677-85.
33. Smith MT. Advances in understanding benzene health effects and susceptibility. Annu Rev Public Health. 2010;31:133,48 2 p following 148.

34. Kim S, Vermeulen R, Waidyanatha S, et al. Modeling human metabolism of benzene following occupational and environmental exposures. Cancer Epidemiol Biomarkers Prev. 2006;15(11):2246-52.
35. Henderson RF, Sabourin PJ, Bechtold WE, et al. The effect of dose, dose rate, route of administration, and species on tissue and blood levels of benzene metabolites. Environ Health Perspect. 1989;82:9-17.
36. Lin YS, Vermeulen R, Tsai CH, et al. Albumin adducts of electrophilic benzene metabolites in benzene-exposed and control workers. Environ Health Perspect. 2007;115(1):28-34.
37. Rappaport SM, Kim S, Lan Q, et al. Evidence that humans metabolize benzene via two pathways. Environ Health Perspect. 2009;117(6):946-52.
38. Rothman N, Li GL, Dosemeci M, et al. Hematotoxicity among Chinese workers heavily exposed to benzene. Am J Ind Med. 1996;29(3):236-46.
39. Snyder R. Benzene's toxicity: a consolidated short review of human and animal studies by HA Khan. Hum Exp Toxicol. 2007;26(9):687-96.
40. Yokley K, Tran HT, Pekari K, et al. Physiologically-based pharmacokinetic modeling of benzene in humans: a Bayesian approach. Risk Anal. 2006;26(4):925-43.
41. Rappaport SM, Waidyanatha S, Qu Q, et al. Albumin adducts of benzene oxide and 1,4-benzoquinone as measures of human benzene metabolism. Cancer Res. 2002a;62(5):1330-7.
42. Rappaport SM, Waidyanatha S, Yeowell-O'Connell K, et al. Protein adducts as biomarkers of human benzene metabolism. Chem Biol Interact. 2005;153-154:103-9.
43. IARC (International Agency for Research on Cancer). Some industrial chemicals and dyestuffs. IARC Monogr Eval Carcinog Risk Chem Hum. 1982;29:93-148
44. IARC (International Agency for Research on Cancer). Overall evaluations of carcinogenicity: an updating of IARC Monographs volumes 1 to 42. IARC Monogr Eval Carcinog Risk Chem Hum Suppl. 1987;7:1-440.
45. Rinsky RA, Smith AB, Hornung R, et al. Benzene and leukemia. An epidemiologic risk assessment. N Engl J Med. 1987;316(17):1044-50.

Chapter 2

Guidelines to Evaluate Human Observational Studies for Quantitative Risk Assessment

Jelle Vlaanderen
Roel Vermeulen
Dick Heederik
Hans Kromhout

on behalf of the EU Network of Excellence ECNIS Integrated Risk Assessment Group

Environmental Health Perspectives. 2008;116(12):1700-5.

Abstract

Background Careful evaluation of the quality of human observational studies (HOS) is required to assess the suitability of HOS for quantitative risk assessment (QRA). In particular, the quality of quantitative exposure assessment is a crucial aspect of HOS that are to be considered for QRA.

Objective We aimed to develop guidelines for the evaluation of HOS for QRA and to apply these guidelines to case-control and cohort studies on the relation between exposure to benzene and acute myeloid leukemia (AML).

Methods We developed a three-tiered framework specific for the evaluation of HOS for QRA and used it to evaluate HOS on the relation between exposure to benzene and AML.

Results The developed framework consists of 20 evaluation criteria. A specific focus of the framework was on the quality of exposure assessment applied in HOS. Seven HOS on the relation of benzene and AML were eligible for evaluation. Of these studies, five were found suitable for QRA and were ranked based on the quality of the study design, conduct, and reporting on the study.

Conclusion The developed guidelines facilitate a structured evaluation that is transparent in its application and harmonizes the evaluation of HOS for QRA. With the application of the guidelines it was possible to identify studies suitable for QRA of benzene and AML and rank these studies based on their quality. Application of the guidelines in QRA will be a valuable addition to the assessment of the weight of evidence of HOS for QRA.

Introduction

Epidemiologic evidence is the most relevant type of evidence for risk assessment, because limited extrapolation is needed to apply study results to a real-life situation. However, because of ethical considerations epidemiologic assessment of risk of potential hazardous exposures is most often limited to observational studies. This deviation from experimental study conditions (e.g., randomized clinical trials) requires careful evaluation of the quality of the observational evidence. A major issue in human observational studies (HOS) is the more limited control of the circumstances under which studies are performed leading to a potential bias in the estimated association between exposure and health outcome. The quality of design and conduct of a study affects the potential for bias in the study results and thus the value for risk assessment. In quantitative risk assessment (QRA) exposure-response relations are defined in quantitative terms (i.e., risk per unit of exposure). HOS that conducted quantitative exposure-response analysis (i.e., a quantitative description of the relation between exposure to a hazardous agent and a specific health effect) can contribute directly to QRA. Therefore, the quality of quantitative exposure assessment is crucial to HOS that are used in QRA. In recent years, several frameworks have been developed to assess the quality of HOS for risk assessment (1-7). These frameworks have provided broad overviews of different aspects that contribute to HOS quality. However, the existing frameworks lack a specific focus on the evaluation of exposure assessment in HOS for QRA. We developed a structured framework with guidelines for the evaluation of HOS in QRA that have a specific focus on the evaluation of the exposure assessment component of HOS. The approach incorporates exclusion of HOS that do not meet the minimal quality required for QRA and ranking based on the quality of the design, conduct and reporting of the HOS that do meet the minimal quality required for QRA. Subsequently, to demonstrate its usefulness we applied the framework to all case-control and cohort studies on the relation between exposure to benzene and acute myeloid leukemia (AML).

Definition of terms related to quantification of exposure in QRA

The exposure evaluation guidelines are related largely to the assessment and assignment of exposure. Exposure assessment is defined as estimation of the concentration of an agent in a specific medium (e.g., air or soil), during a specific time period (e.g., a working day), and under specific conditions (e.g., type of weather) (8). Examples are the concentration of respirable crystalline silica to which a worker was exposed in his breathing zone on a specific day performing a specific task, or the level of caffeine in a single cup of coffee. The most direct strategy for exposure assessment is to perform quantitative measurements. However, in many HOS, exposure measurements are scant, and other sources of information (e.g., expert judgment, questionnaire data, or predictive models) are used to assess exposure (9). Exposure assignment is defined as the step where exposure estimates are assigned to the

individuals in the study population based on information on, for instance, jobs held or food frequency questionnaires (10).

Description of the framework and evaluation guidelines

The criteria that together form the guidelines for evaluation of HOS for QRA are described in detail in the Supplemental Material (See Supplemental Material I, Evaluation Guidelines). Here we provide an overview of the structure of the framework and discuss the evaluation criteria that are crucial for the quality of the assessment and assignment of exposure. The framework is based on three tiers (Figure 1). The criteria in the first tier are used to exclude studies that are not suitable for QRA and should be applied to all HOS considered for QRA (Table 1). The questions in tier I are all related to crucial aspects of the quality of the design, the quality of conduct, and the quality of the reporting of HOS. Therefore, HOS are suitable for QRA only if all questions are answered affirmatively. A negative answer to one of the questions should result in exclusion of the HOS for QRA.

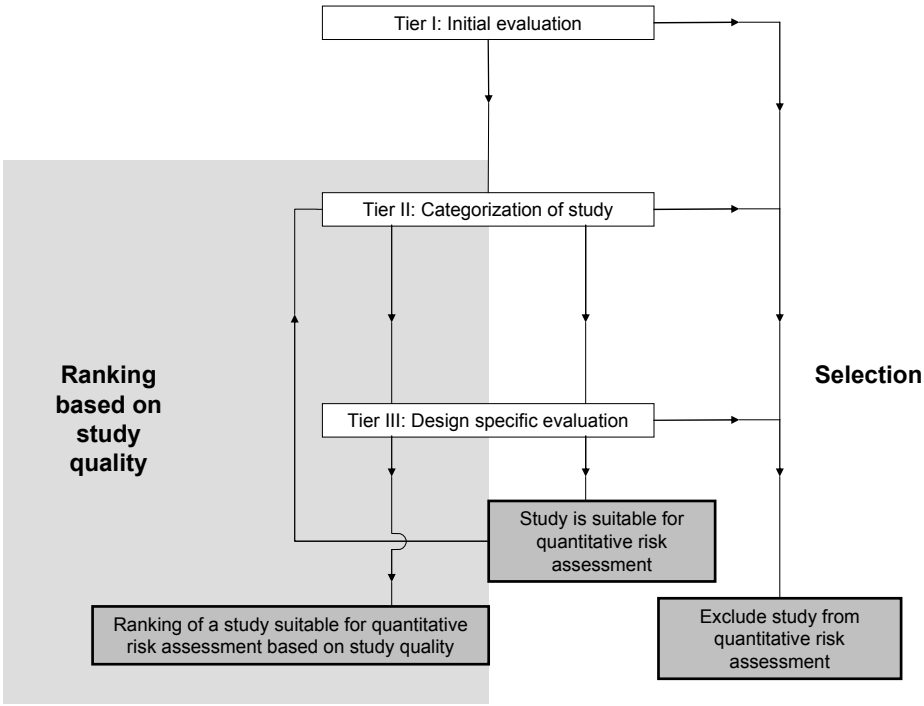


Figure 1 Decision pathway of the framework for evaluation of HOS for QRA. Outcomes of the pathway: exclude study from QRA; study is suitable for QRA; and ranking of a study suitable for QRA based on study quality.

In the second tier, the HOS are categorized based on the type of study design (Table 1). The reason for categorization in tier II is two-fold: exclusion of HOS that have an inappropriate study design for QRA, and selection of appropriate criteria for further evaluation in tier III. In the third tier, a decision is made whether to include HOS in QRA based on a set of design specific criteria. A distinction is made between the criteria intended to assess whether HOS are suitable for QRA and the criteria intended to be used in ranking of the HOS suitable for QRA based on the quality aspects of these HOS. Some criteria in Table 1 are used in both the selection and ranking of HOS. Although this framework has been developed primarily to facilitate objective evaluation of HOS for QRA, the criteria in the framework can also be used as guidelines for the conduct of high-quality HOS suitable for future QRA. To facilitate transparent and objective evaluation of evidence from HOS, risk assessors should *a priori* define minimum requirements for including a study in QRA, such as *a priori* definitions of acceptable levels of the response rate and loss to follow-up. In addition, the minimal follow-up time required to detect the health effect of interest should be defined. Finally, all relevant potential strong confounding factors should be identified. The actual operational definition of these requirements will need to be based on a case-by-case basis depending on the specific exposure-response relation studied.

Criteria related to the quality of assessment and assignment of exposure of HOS

Is exposure expressed on a ratio scale and specific for the agent of interest?
If exposure is expressed on a ratio scale, the units of the scale represent the same magnitude of exposure across the whole range of the scale, and a rational zero is included (11). Quantitative exposure measurements, therefore, should be at the basis of exposure assessment. HOS that present quantitative exposure estimates based solely on expert judgment should not be used in QRA because of difficulties with regard to calibration of these estimates. For QRA, the exposure measures reported in HOS need to be specific for the agent of interest. Only a highly specific measure of exposure can be used to demonstrate a potential causal relation between exposure and health effect.

Quality of the exposure measurement methods
Quantitative measurements used in the exposure assessment in HOS can potentially differ with regard to the quality of the measurement methods and the analytical methods used. A guideline to evaluate HOS based on the quality of exposure measurements is to compare the method(s) used in the study to the method(s) that are currently considered as best practice. Some studies provide information on side-by-side comparisons of the exposure measurement method used with the best practice at the time of the study. Additional information from studies that solely focus on side-by-side comparisons of exposure measurement methods can be used as well (12, 13).

Table 1 Overview of the criteria that are used in the three-tiered evaluation of human observational studies for quantitative risk assessment.[a]

Tier	Evaluation criteria	Outcome	Impact on evaluation	CC[b]	COH[c]	CR[d]
I[e]	1.1 Is the study design case-control, cohort or cross-sectional?	Yes / no	Selection for QRA[f]	X	X	X
I[e]	1.2 Is exposure expressed on a ratio scale and specific for the agent of interest?	Yes / no	Selection for QRA[f]	X	X	X
I[e]	1.3 Is a detailed description of the statistical analysis provided?	Yes / no	Selection for QRA[f]	X	X	X
I[e]	1.4 Are criteria for inclusion of subjects into the study described with sufficient detail?	Yes / no	Selection for QRA[f]	X	X	X
I[e]	1.5 Is the assessment of the health effect performed according to recognized norms?	Yes / no	Selection for QRA[f]	X	X	X
I[e]	1.6 Are all relevant potential strong confounding factors considered in the study design?	Yes / no	Selection for QRA[f]	X	X	X
II[g]	2.1 Type of study design	Case-control / cohort / cross-sectional	Selection for QRA[f] / study quality ranking[h]	X	X	X
III[i]	3.1 Response rate	Numerical	Selection for QRA[f] / study quality ranking[h]	X	X	X
III[i]	3.2 Loss to follow-up	Numerical	Selection for QRA[f] / study quality ranking[h]		X	
III[i]	3.3 Minimum follow-up time	Description	Selection for QRA[f]		X	
III[i]	3.4 Quality of the exposure measurement methods	Description	Selection for QRA[f] / study quality ranking[h]	X	X	X
III[i]	3.5 Insight into the variability of exposure	Description	study quality ranking[h]	X	X	X
III[i]	3.6 Application of exposure measurements in exposure assessment	Description	Selection for QRA[f] / study quality ranking[h]	X	X	X
III[i]	3.7 Type of exposure metric	Description	Study quality ranking[h]	X	X	X
III[i]	3.8 Specificity of the exposure indicator	Category[j]	Study quality ranking[h]	X	X	X
III[i]	3.9 Blinded exposure assessment	Description	Selection for QRA[f]	X	X	X
III[i]	3.10 Quality of the exposure assignment strategy	Description	Study quality ranking[h]	X	X	
III[i]	3.11 Potential for information bias	Description	Study quality ranking[h]	X	X	X
III[i]	3.12 Blinded health outcome assessment?	Description	Selection for QRA[f]		X	X
III[i]	3.13 Insight into the potential for systematic error in study results	Description	Study quality ranking[h]	X	X	X

[a] Evaluation criteria are discussed in detail In Supplemental Material I. [b] Criteria relevant for case-control (CC) study design. [c] Criteria relevant for cohort (COH) study design. [d] Criteria relevant for cross-sectional (CR) study design. [e] Tier I: Initial evaluation. [f] Criteria relevant for selection of HOS for QRA. [g] Tier II: categorization of HOS into the three types of study designs that can potentially be used in QRA. [h] Criteria relevant for ranking of studies based on quality of design, conduct and reporting. [i] Tier III: Specific evaluation of the quality of the design, conduct and reporting of HOS. [j] Categories are constructed based on a combination of: proxy vs. causal exposure and external vs. internal exposure.

Insight into the variability of exposure

For the evaluation of HOS, it is important to realize that exposure measurements used in exposure assessment can be highly variable in level. This variability can be attributed to a combination of variation in exposure levels over time and space. Advanced methodologies to acquire insight into the level of measurement variability on HOS outcomes have been proposed (14-17). Before the evaluation, risk assessors must define a minimum acceptable level of information required to assess whether enough insight into variability of exposure measurements is provided in HOS. Tielemans et al. have developed guidelines to evaluate exposure data from HOS performed in the occupational exposure context (18). Similar approaches should be applied to exposure data from other exposure contexts (e.g., dietary exposure, consumer exposure). Differences between HOS in the ability to assess the relative contribution of the different sources of variability in exposure measurements can be used to rank the HOS.

Application of exposure measurements in exposure assessment

In most HOS, researchers are confronted with a scarcity of exposure measurements. As a result, exposure measurements might not be available for each assignment unit (i.e., a single individual or a group of individuals with assumed similar exposure patterns) for the complete time period of interest. In this situation, exposure measurements performed for assignment unit-time period combinations and information regarding the circumstances of these measurements (e.g., year of measurement, type of weather during measurement, or the task the measured individual performed during the measurement) are used to estimate exposure levels for assignment unit-time period combinations for which exposure measurements are not available. The strategy used to extrapolate measurements over assignment unit-time period combinations determines the validity of the exposure estimates and therefore has a large impact on the overall quality of the quantification of exposure. In most HOS, exposure measurements are extrapolated following a set of decision rules based on expert judgment and/or via a modeling framework. A complete and detailed insight into the applied decision rules in these approaches is essential for evaluation of HOS.

Type of exposure metric

In an ideal situation, an exposure metric captures three aspects that determine exposure: intensity, duration and timing (19). The quality of an exposure metric is based on biological considerations such as the time window of exposure that is relevant to the health effect of interest (16, 19, 20). A guideline to evaluate HOS based on the used exposure metric is to compare the metric used with the current state of knowledge on the nature of the relation between the exposure and health outcome of interest.

Specificity of the exposure indicator

In situations where it is difficult to assess the actual exposure that is assumed to be causally related to the health effect of interest, a causal indicator of exposure, researchers might assess a proxy for the causal exposure. However, it is crucial that the proxy exposure is highly correlated to the exposure of interest. Once absorbed in the human body, distribution, metabolism, and excretion have a large impact on the dose of a specific agent (or metabolite) at the site of action. Application of exposure indicators capable of incorporating these biological influences in exposure estimates will result in increased correlation between the exposure indicator and the dose at the site of action. The application of biomarkers of exposure in HOS potentially provides the possibility to obtain exposure indicators with higher specificity compared to indicators of external exposure. Similarly, as with external exposure, insight into variability of biomarker-based exposure measurements is of utmost importance for QRA.

Blinded exposure assessment

Exposure assessment should always be performed blinded for the health outcome of interest to avoid observer bias. If exposure assessment was performed on the individual level, omission of a statement regarding blinded exposure assessment is a reason to exclude HOS from QRA. If exposure assessment was performed to assess exposure for previously defined homogeneous exposure categories, there is no direct connection between the individuals in the study population and the exposure assessment, and therefore this criterion needs less stringent application.

Quality of the exposure assignment strategy

In the exposure assignment step, exposure levels assessed for specific assignment-unit time-period combinations are translated into exposure estimates for each individual in the study population. Assignment is based on information related to the individuals in the study population and related to the assignment-unit time-period combinations for which exposure levels have been assessed. Examples of this information are the jobs an individual performed during his or her working career, a description of daily diet, or information on other factors potentially affecting exposure levels. The exposure context in which HOS are performed determines which type of information is available for exposure assignment. A proper evaluation of the quality of exposure assignment requires insight into the proportion of the assignment-unit time-period combinations used for assignment for which no or little exposure measurements were available and exposure levels had to be inferred. In addition, the overlap between the assignment-unit time-period combinations for which exposure measurements were available and the exposure time periods that are assumed to be relevant to the assessed health risk needs to be evaluated.

Application of the guidelines on benzene case-control and cohort studies

Selection of studies eligible for evaluation

To test the usefulness and practical implications of our guidelines, we applied the developed framework to all case-control and cohort studies that have reported on a dose-response relation between exposure to benzene and acute nonlymphocytic leukemia (ANLL) or AML. In this example we will ignore the small differences in disease classification between ANLL and AML and consider both as the same health outcome (referred to as AML). A detailed report of the selection of publications that were eligible for evaluation is presented in the Supplemental Material (See Supplemental Material II, Search Strategy). All identified publications were reviewed for eligibility of application of the evaluation guidelines (Figure 2).

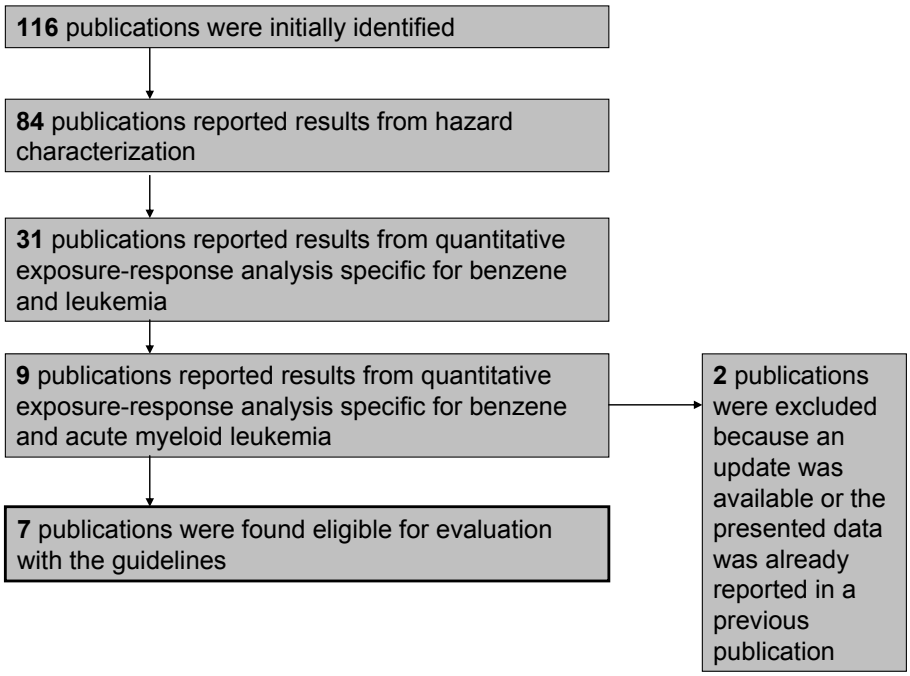


Figure 2 Overview of the strategy that was applied to select publications that report on the relation between exposure to benzene and acute myeloid leukemia and are eligible for evaluation with the guidelines.

Thirty-two publications were found not eligible because results from hazard characterization were not reported. From the 84 publication that did report results from hazard characterization, 53 publications were excluded because no quantitative exposure-response analysis specific for benzene and leukemia was reported. Finally, 22 publications did not report results from quantitative exposure-response analysis specific for benzene and AML. Therefore, the selection strategy resulted in only seven studies eligible for evaluation. Details of these studies are presented in Table 2.

Evaluation

A detailed report of the evaluation is presented in the Supplemental Material (See Supplemental Material III, Outcome of the Evaluation). Here we discuss the aspects that contributed to the ranking of the seven remaining HOS on benzene and AML that were evaluated with the use of our guidelines.

Definition of minimal requirements for QRA and identification of potential strong confounding factors

Before the evaluation we defined minimal requirements for inclusion into QRA: response rate > 60%; loss to follow-up < 10%; and follow-up time > 10 years. We considered exposure to ionizing radiation as the only factor for which there is evidence of potential confounding on the relation between exposure to benzene and AML (21)

Initial evaluation

Two studies, Guénel (22) and Monsanto (23, 24), did not pass the initial evaluation. The Guénel study was excluded because exposure was not presented on a ratio scale, but in unit-years (criterion 1.2). This limitation prohibits the use of this study in QRA and therefore further evaluation was not done. The Monsanto study was excluded because of the very limited information that was provided on the performed statistical analysis performed (criterion 1.3). All other studies passed initial evaluation. It was assumed that exposure to ionizing radiation was not above background level in all the populations studied. Therefore, no potential strong confounding factors needed to be considered in the evaluation (criterion 1.6)

Categorization

From the studies that passed initial evaluation, two were case-control studies AHW (25-27), and U.K. Petrol (28, 29), and three were cohort studies CAPM-NCI (30-33), Dow (34, 35) and Pliofilm (36-40). The case-control studies were all nested in large occupational cohorts.

Table 2 Summary details of the quantitative benzene-AML case-control and cohort studies ranked based on the outcome of the evaluation.

Ranking based on evaluation of study quality	Name of the study	Type of study design	Publications used for evaluation	Date of publication of hazard characterization	Evaluation outcomes that contributed to the differentiation of the evaluated HOS
1	U.K. Petrol[a]	Nested case-control	(28, 29)	1997	+ Detailed insight into methodology for assessment and assignment of exposures + Limitations of exposure measurements were assessed and discussed + Potential for systematic error was assessed
2	AHW[b]	Nested case-control	(25-27)	2003	+ Detailed insight into methodology for assessment and assignment of exposures + Limitations of exposure measurements were assessed and discussed - Potential for systematic error was not assessed
3	CAPM-NCI[c]	Cohort	(30-33)	1997	+ Insight into methodology for assessment and assignment of exposure - Limited insight into quality and use of exposure measurements
4	Pliofilm[d]	Cohort	(36-40)	1995	+ Insight into methodology for assessment and assignment of exposure - Limited insight into quality and use of exposure measurements
5	Dow[e]	Cohort	(34, 35)	2004	- Limited insight into methodology for assessment and assignment of exposure - Actual use of exposure measurements in exposure assessment is unclear
–	Guénel[f]	Nested case-control	(22)	2002	Study not suitable for QRA
–	Monsanto[g]	Cohort	(23, 24)	2003	Study not suitable for QRA

[a] Study performed on petroleum distribution workers in U.K. [b] Australian Health Watch study. [c] Study performed by Chinese Academy of Preventive Medicine (CAPM) and the U.S National Cancer Institute (NCI). [d] Study performed on workers employed at two Ohio factories producing hydrochloride. [e] Study performed on Dow Chemical Michigan Operations employees. [f] Study performed by Guénel et al. on men employed at EDF-GDF. [g] Study performed on Monsanto plant employees.

Design specific evaluation

Design specific criteria that contributed to the ranking based on quality were related to exposure assessment, exposure assignment, and insight into systematic error in exposure assessment/assignment. All studies (n = 5) reported the use of exposure measurements in the exposure assessment. However, there was a wide range in the amount of information that was provided regarding the quality of the measurements, insight into the variability of the measurements, and the use of measurements in exposure assessment. The AHW study and the U.K. Petrol study provided the most detailed information and apparently applied the most stringent quality criteria for inclusion of measurements in exposure assessment. The CAPM-NCI study reported the use of short-term area measurements but provided very little information regarding the quality and variability of these measurements. The Dow study reported that an industrial hygienist categorized all job-titles into exposure categories that were defined in an earlier study on the same cohort with the use of industrial hygiene measurements. However, the actual relation between exposure measurements and exposure assessment is unclear. The Pliofilm study provided limited information on the measurements used in exposure assessment. However, it was reported that the measurements used for the Pliofilm cohort reflected benzene concentrations in workplace area and no personal sampling was performed. Exposure assignment strategy was most detailed in the U.K. Petrol, AHW and CAPM-NCI studies. These studies reported the use of job- or task-specific and time-specific information for assignment. The Pliofilm study applied a less detailed assignment strategy which was based on a job title-exposure class matrix and provided limited insight into the exposure assignment strategy. Dow reported very limited information regarding assignment of exposure, which made proper evaluation impossible. Only one study performed a sensitivity analysis to acquire insight into the potential of systematic error due to potential biases such as misclassification of exposure and quality of work histories used in assignment (U.K. Petrol).

Ranking of the evaluated studies

Based on our evaluation, the two case-control studies, the U.K. Petrol and AHW studies, have received the highest relative ranking for QRA (Table 2). Although the study designs of the U.K. Petrol and AHW studies were comparable, the U.K. Petrol study was ranked higher because this study reported results from a sensitivity analysis used to evaluate the impact of several crucial decisions made in the assessment of exposure. The rationale to assign a lower ranking to the CAPM-NCI and the Pliofilm studies is that in both studies considerable uncertainty existed regarding the quality of the exposure measurements used and the methods used to incorporate exposure measurements in the assessment and assignment of exposures. The CAPM-NCI study provided more detailed information on the methods used for exposure assessment and was therefore ranked higher than the Pliofilm study. Although the Dow study was considered suitable for QRA, large uncertainty remained regarding the potential contribution of this study to QRA. This uncertainty was largely determined by the lack of

information on the actual use of exposure measurements in assessment and assignment of exposure. Therefore the Dow study received the lowest ranking.

Discussion of the application of the guidelines in the benzene-AML example

In our example, differentiation of the five studies suitable for QRA was largely based on the quality of assessment and assignment of exposure. In general, evaluation was difficult because of the limited information provided in the evaluated publications. Therefore, it is possible that the evaluation outcome of this example is partly based on the absence of information. Recently, the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) initiative provided general requirements for reporting of HOS (41). Application of such requirements in the publication of studies will facilitate the evaluation of HOS. Unfortunately STROBE proposes only limited guidelines for the reporting of exposure assessment in HOS and is therefore of limited use for the evaluation of HOS for QRA. In our example we evaluated only publications published in the peer-reviewed scientific literature. An alternative approach is to contact the researchers responsible for the studies selected for evaluation in order to acquire as much detailed information as possible. In our evaluation, each study included had specific limitations with regard to the quality of the estimation of quantitative exposure levels. As a result of this situation, several studies have been the subject of discussion regarding the quality and validity of exposure estimates (42, 43). We think that a thorough sensitivity analysis that provides insight into the level of uncertainty of the estimated exposure levels and a detailed description of the approach used for assessment and assignment of exposure could have left less room for discussion and thereby would have increased the quality of all evaluated HOS for QRA. For the design of future quantitative HOS in this field, researchers should be aware of the specific requirements of QRA to HOS with regard to study design and reporting of results.

Impact for human regulatory risk assessment of benzene

We compared the outcome of our evaluation with the selection of studies used in the regulatory QRA performed by the U.S. Environmental Protection Agency (EPA) in 1985 and updated in 1998 (44). The U.S. EPA QRA is based on the study by Rinksy (38) (Pliofilm), Wong et al. (45), and Ott et al. (35, 46) (Dow). A difference between the U.S. EPA QRA and our evaluation is the health endpoint that was considered. Whereas we evaluated only studies that reported specific risk estimates for AML, the U.S. EPA QRA focused on all leukemias together as a single health outcome. Therefore, the study by Wong (45) was not considered in our evaluation because this study did not report specific risk estimates for AML. Based on our evaluation three additional studies should be considered for a regulatory QRA of benzene U.K. Petrol, AHW and CAPM-NCI. Interestingly, these three studies were all regarded as providing higher quality evidence than the Pliofilm and the Dow study using our proposed

framework. To assess the contribution of evidence from a single HOS to regulatory QRA, the assessment of the quality of the evidence needs to be combined with an assessment of the relevance of the evidence for QRA. The combination of quality and relevance of evidence is defined as the weight of evidence for QRA (47). Aspects that contribute to the relevance of evidence for QRA are the exposure context in which the study was performed (e.g., occupational exposure vs. dietary exposure), the range of exposure levels included in the study, and the potential impact of random error on the study findings, usually quantified with confidence intervals (CIs). In Table 3 an overview of these aspects that contribute to the relevance of a study to QRA are presented for the five studies that we evaluated. In our example, all included studies were performed in the occupational exposure context. However, the U.K. Petrol, AHW and CAPM-NCI and Dow studies included ranges of benzene exposure levels that are thought to be more relevant for the current work population and the general population than the range of exposures that was included in the Pliofilm study (Table 3). Therefore, these studies require less extrapolation to calculate relevant risk estimates. To assess the potential impact of random error on the study findings the fold range of the 95% CIs surrounding the relevant risk estimates is reported for each relevant risk estimate that was reported in the evaluated studies (Table 3). Relatively large differences in fold ranges were observed. We expect that a renewed QRA that included all quantitative epidemiological evidence available at this time and incorporated a weight of evidence approach would significantly increase the confidence in unit risk estimates for exposure to benzene. Our approach contributes to a transparent qualitative insight into the differences in the weight of evidence of HOS for QRA. Quantification of the weight of evidence based on a review of the quality and the relevance of the available studies will be highly subjective and, if performed at all, should be as transparent as possible. Although existing approaches acknowledge the importance of exposure assessment in HOS for QRA (1, 2), we attempted to improve these methods by providing a detailed discussion of the aspects that collectively determine the quality of assessment and assignment of exposure in HOS. The outcome of the benzene-AML example indicated that, in this case, there were large differences between HOS with regard to the quality of the exposure assessment that would not have been detected with the application of the existing evaluation approaches.

Table 3 Aspects that contribute to the relevance of HOS to regulatory quantitative risk assessment.

Name of the study	Exposure context in which the study was performed	Size of the study population	Exposure categories included in the study (ppm-years)[a]	Fold range of the 95% confidence intervals that were reported for relevant risk estimates[b]
U.K. Petrol	Occupational exposure	31 cases / 121 controls	0.26–0.59	14.6
			0.60–1.64	13.3
			1.65–4.78	13.2
			≥ 4.79	13.4
AHW	Occupational exposure	11 cases / 44 Controls	4– 8	100.0
			> 8	31.8
CAPM-NCI	Occupational exposure	110,633 individuals (21 cases)	< 40	14.0
			40–99	14.5
			≥ 100	10.5
Pliofilm	Occupational exposure	1,868 individuals (6 cases)	< 40	221
			40–200	–[c]
			200–400	29.9
			> 400	14.2
Dow	Occupational exposure	2,266 individuals (4 cases)	< 28.3	28.5
			28.3–79.1	204.3
			> 79.1	223.8

[a] Exposure categories for which a risk estimate was reported for AML in the evaluated publications. [b] Fold range was calculated as (upper bound of the 95% CI) / (lower bound of the 95% CI) for each exposure group for which a risk estimate was reported for AML in the evaluated publications. [c] No cases were observed in this study for this exposure category; therefore, the lower bound of the 95% CI was 0 and a fold range could not be calculated.

Acknowledgements

This work was performed as part of the work package “Integrated Risk Assessment” of the ECNIS Network of Excellence (Environmental Cancer Risk, Nutrition and Individual Susceptibility), operating within the European Union 6th Framework Program, Priority 5: “Food Quality and Safety” (FOOD-CT-2005-513943).
Partners within this ECNIS work package “Integrated Risk Assessment” are: W. Sobala and W. Hanke, Nofer Institute of Occupational Medicine; P. Boffetta, International Agency for Research on Cancer; U. Strömberg, B. Jönsson and H. Tinnerberg, University of Lund; S.A. Kyrtopolus, National Hellenic Research Foundation; A. Hirvonen, Finnish Institute of Occupational Health; P.A. van den Brandt, University of Maastricht; P. Vineis, Imperial College London; D. Segerback, Karolinska Institutet; S. Loft, University of Copenhagen.

References

1. Goldbohm RA, Tielemans EL, Heederik D, et al. Risk estimation for carcinogens based on epidemiological data: a structured approach, illustrated by an example on chromium. Regul Toxicol Pharmacol. 2006;44(3):294-310.
2. Hertz-Picciotto I. Epidemiology and quantitative risk assessment: a bridge from science to policy. Am J Public Health. 1995;85(4):484-91.
3. Money CD, Margary SA. Improved use of workplace exposure data in the regulatory risk assessment of chemicals within Europe. Ann Occup Hyg. 2002;46(3):279-85.
4. Shore RE, Iyer V, Altshuler B, et al. Use of human data in quantitative risk assessment of carcinogens: impact on epidemiologic practice and the regulatory process. Regul Toxicol Pharmacol. 1992;15(2 Pt 1):180-221.
5. Swaen GM. A framework for using epidemiological data for risk assessment. Hum Exp Toxicol. 2006;25(3):147-55.
6. van den Brandt P, Voorrips L, Hertz-Picciotto I, et al. The contribution of epidemiology. Food Chem Toxicol. 2002;40(2-3):387-424.
7. WHO-Working-Group. Evaluation and use of epidemiological evidence for environmental health risk assessment: WHO guideline document. Environ Health Perspect. 2000;108(10):997-1002.
8. Zartarian VG, Ott WR, Duan N. A quantitative definition of exposure and related concepts. J Expo Anal Environ Epidemiol. 1997;7(4):411-37.
9. Ott MG. Exposure assessment as a component of observational health studies and environmental risk assessment. Scand J Work Environ Health. 2005;31 Suppl 1:110,4; discussion 63-5.
10. Loomis D, Kromhout H. Exposure variability: concepts and applications in occupational epidemiology. Am J Ind Med. 2004;45(1):113-22.
11. Stevens SS. On the theory of scales of measurement. Science. 1946;103(2684):677-80.
12. de Vocht F, Huizer D, Prause M, et al. Field comparison of inhalable aerosol samplers applied in the European rubber manufacturing industry. Int Arch Occup Environ Health. 2006;79(8):621-9.
13. Stephenson DJ, Lillquist DR, DeRosso FD, et al. Side-by-side comparison of three sampling methods for aerosolized endotoxin in a wastewater treatment facility. J Environ Health. 2004;67(4):16-9.
14. Heederik D, Miller BG. Weak associations in occupational epidemiology: adjustment for exposure estimation error. Int J Epidemiol. 1988;17(4):970-4.
15. Kromhout H, Loomis DP, Kleckner RC. Uncertainty in the relation between exposure to magnetic fields and brain cancer due to assessment and assignment of exposure and analytical methods in dose-response modeling. Ann N Y Acad Sci. 1999;895:141-55.
16. Loomis A, Kromhout H, Kleckner RC, et al. Effects of the analytical treatment of exposure data on associations of cancer and occupational magnetic field exposure. Am J Ind Med. 1998;34(1):49-56.

17. Xue X, Kim MY, Shore RE. Estimation of health risks associated with occupational radiation exposure: addressing measurement error and minimum detectable exposure level. Health Phys. 2006;91(6):582-91.
18. Tielemans E, Marquart H, De Cock J, et al. A proposal for evaluation of exposure data. Ann Occup Hyg. 2002;46(3):287-97.
19. Vacek PM. Assessing the effect of intensity when exposure varies over time. Stat Med. 1997;16(5):505-13.
20. Seixas NS, Robins TG, Becker M. A novel approach to the characterization of cumulative exposure for the study of chronic occupational disease. Am J Epidemiol. 1993;137(4):463-71.
21. Pagano L, Caira M, Fianchi L, et al. Environmental risk factors for MDS/AML. Haematologica reports. 2006;2(15)42-45.
22. Guenel P, Imbernon E, Chevalier A, et al. Leukemia in relation to occupational exposures to benzene and other agents: a case-control study nested in a cohort of gas and electric utility workers. Am J Ind Med. 2002;42(2):87-97.
23. Collins JJ, Ireland B, Buckley CF, et al. Lymphohaematopoeitic cancer mortality among workers with benzene exposure. Occup Environ Med. 2003;60(9):676-9.
24. Ireland B, Collins JJ, Buckley CF, et al. Cancer mortality among workers with benzene exposure. Epidemiology. 1997;8(3):318-20.
25. Glass DC, Adams GG, Manuell RW, et al. Retrospective exposure assessment for benzene in the Australian petroleum industry. Ann Occup Hyg. 2000;44(4):301-20.
26. Glass DC, Gray CN, Jolley DJ, et al. Health Watch exposure estimates: do they underestimate benzene exposure? Chem Biol Interact. 2005;153-154:23-32.
27. Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology. 2003;14(5):569-77.
28. Rushton L, Romaniuk H. 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. Occup Environ Med. 1997;54(3):152-66.
29. Lewis SJ, Bell GM, Cordingley N, et al. Retrospective estimation of exposure to benzene in a leukaemia case-control study of petroleum marketing and distribution workers in the United Kingdom. Occup Environ Med. 1997;54(3):167-75.
30. Dosemeci M, Li GL, Hayes RB, et al. Cohort study among workers exposed to benzene in China: II. Exposure assessment. Am J Ind Med. 1994;26(3):401-11.
31. Hayes RB, Yin SN, Dosemeci M, et al. Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine—National Cancer Institute Benzene Study Group. J Natl Cancer Inst. 1997;89(14):1065-71.
32. Travis LB, Li CY, Zhang ZN, et al. Hematopoietic malignancies and related disorders among benzene-exposed workers in China. Leuk Lymphoma. 1994;14(1-2):91-102.
33. Yin SN, Linet MS, Hayes RB, et al. Cohort study among workers exposed to benzene in China: I. General methods and resources. Am J Ind Med. 1994;26(3):383-400.
34. Bloemen LJ, Youk A, Bradley TD, et al. Lymphohaematopoietic cancer risk among chemical workers exposed to benzene. Occup Environ Med. 2004;61(3):270-4.
35. Ott MG, Townsend JC, Fishbeck WA, et al. Mortality among individuals occupationally exposed to benzene. Arch Environ Health. 1978;33(1):3-10.
36. Paxton MB, Chinchilli VM, Brett SM, et al. Leukemia risk associated with benzene exposure in the pliofilm cohort: I. Mortality update and exposure distribution. Risk Anal. 1994;14(2):147-54.
37. Paxton MB, Chinchilli VM, Brett SM, et al. Leukemia risk associated with benzene exposure in the pliofilm cohort. II. Risk estimates. Risk Anal. 1994;14(2):155-61.
38. Rinsky RA, Smith AB, Hornung R, et al. Benzene and leukemia. An epidemiologic risk assessment. N Engl J Med. 1987;316(17):1044-50.
39. Rinsky RA. Benzene and leukemia: an epidemiologic risk assessment. Environ Health Perspect. 1989;82:189-91.
40. Wong O. Risk of acute myeloid leukaemia and multiple myeloma in workers exposed to benzene. Occup Environ Med. 1995;52(6):380-4.
41. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology. 2007;18(6):805-35.
42. Paustenbach DJ, Price PS, Ollison W, et al. Reevaluation of benzene exposure for the Pliofilm (rubberworker) cohort (1936-1976). J Toxicol Environ Health. 1992;36(3):177-231.

43. Wong O. A critique of the exposure assessment in the epidemiologic study of benzene-exposed workers in China conducted by the Chinese Academy of Preventive Medicine and the US National Cancer Institute. Regul Toxicol Pharmacol. 1999;30(3):259-67.
44. U.S. EPA. Carcinogenic Effects of Benzene: An Update. EPA/600/P-97/001F. Washington, DC: United States Environmental Protection Agency, 1998.
45. Wong O. An industry wide mortality study of chemical workers occupationally exposed to benzene. II. Dose response analyses. Br J Ind Med. 1987;44(6):382-95.
46. Bond GG, McLaren EA, Baldwin CL, et al. An update of mortality among chemical workers exposed to benzene. Br J Ind Med. 1986;43(10):685-91.
47. Weed DL. Weight of evidence: a review of concept and methods. Risk Anal. 2005;25(6):1545-57.

Supplemental material I, Initial evaluation - Tier I

Tier I consists of six dichotomous questions (table 1). HOS are only suitable for QRA if all questions are answered affirmatively. A negative answer to one of the questions should result in exclusion of the HOS for QRA.

T1.1 Is the study design case-control, cohort or cross-sectional?
With the exception of rare examples such as asbestos-mesothelioma few exposures are exclusively related to a specific health effect (1). This situation underlines the desirability to approach experimental conditions as close as possible in HOS designs. An important aspect of experimental conditions is the ability to compare an observed effect to a reference situation. For HOS considered for QRA this has two implications. In order to discuss acceptable exposure levels or exposure thresholds studies need to include exposure contexts in which health effects are representative of background levels. In addition, the exposure levels of a group of individuals with similar observed health effects (cases) can only be interpreted if information on the exposure levels of individuals without the observed health effects (controls) is available. Therefore, HOS that do not include a relevant reference situation should not be used for QRA. In addition, HOS based on an ecological study design should not be used in QRA to avoid the ecological fallacy (2). Consequently, the guidelines will focus on case-control, cohort or cross-sectional study designs.

T1.2 Is exposure expressed on a ratio scale and specific for the agent of interest?
To allow comparison of exposure-response relationships derived from multiple data sets in meta-analyses or systematic reviews exposure needs to be expressed on a ratio scale (3). Therefore, quantitative exposure measurements should be at the basis of exposure assessment. HOS that present quantitative exposure estimates based solely on expert judgment should not be used in QRA. The range of exposures that is reported in HOS has no direct bearing on the inclusion of HOS into QRA. A small range of exposures observed in a single study does indeed preclude the derivation of an exposure-response relation within that study. However, in combination with other studies such a study might provide valuable information regarding observed health effects at a specific point on the ratio scale. For QRA the exposure measures reported in HOS need to be specific for the agent of interest. Only a highly specific measure of exposure can be used to demonstrate a potential causal relation between exposure and health effect.

T1.3. Is a detailed description of the statistical analysis provided?
An accurate description of the applied statistical methods is essential for the interpretation of the results of HOS. Many different aspects contribute to the quality of the statistical analysis. In general, a good initial approach to the assessment of statistical analysis in HOS is to question whether enough details have been provided to allow replication of the analysis

(provided that the crude data would be available). Without a sufficient detailed description of the statistical analysis it is impossible for risk assessors to assess the quality of statistical analysis and HOS should be excluded from QRA. A sufficiently described, but flawed, statistical analysis should also result in exclusion from QRA.

T1.4. Are criteria for inclusion of subjects into the study described with sufficient detail?
Without accurately defined inclusion criteria it is impossible to assess the impact of selection bias on the quality of HOS. Selection bias potentially could affect both internal and external validity of HOS (4). Therefore HOS that lack a sufficiently detailed description of the subject inclusion criteria should be excluded from QRA.

T1.5. Is the assessment of the health effect performed with a validated method?
QRA requires objective assessment of health effects in HOS. Therefore health effects should be assessed according to recognized norms and the methods that are used for assessment should have been validated with a current 'best practice'. Studies that incorporate a clinically manifested disease as health endpoint should provide a detailed description of the definition that is used to classify the disease (e.g., ICD classification, pathological confirmation). Studies that incorporate an intermediate endpoint (e.g., chromosomal aberrations, blood pressure) should specify the laboratory methods used to quantify the endpoint as well as (if applicable) information on sample collection, timing of sample collection and storage of biological materials.

T1.6. Are all relevant potential strong confounding factors considered in the study design?
The possibility of confounding is a commonly used argument to point out the limitations of HOS. Confounding occurs when a factor associated to the health outcome and the exposure of interest modifies the observed study outcome (5). Potential confounding factors can be identified based on previous epidemiological evidence and/or knowledge on the underlying pathophysiological process of the health effect of interest. When potential confounding factors are included in the study design, it is possible to test for the possible effect of confounders on the study outcome and if necessary to adjust study outcomes for this effect. It is important to realize that the impact of confounding factors on the study outcome is only substantial when the association with the health outcome and exposure of interest is strong (6). As such HOS should only be excluded when there is convincing evidence that a factor is a potential strong confounder of the exposure-response relation of interest and the factor is not considered in the HOS. In addition, if one assumes that there is not a strong probability of imbalance between study groups for a potentially strong confounder, (e.g., RR > 5), non-inclusion of the confounder in the study design study should not be a basis for exclusion for QRA (6) .

Supplemental material I, Categorization of the study - Tier II

The categorization of HOS based on the type of study design (table 1) is used to exclude HOS and to select the appropriate criteria for evaluation in tier III.

T2.1 Type of study design:
To determine whether the type of study design is appropriate for QRA risk assessors need to make an assumption on the nature of the exposure-response relation of interest. The assumed nature of exposure-response relation determines whether a longitudinal component is necessary in the study design in order to be able to detect a potential response resulting from the exposure of interest. If it is decided that a longitudinal component is required cross-sectional study designs should be excluded from QRA. Study designs with a longitudinal component are categorized into case-control (including case-cohort, and nested case-control studies) or cohort designs.

Supplemental material I, Design specific evaluation - Tier III

In tier III (table 1) guidelines for the design specific evaluation of HOS are listed. A distinction is made between the criteria that are intended to assess whether HOS are suitable for QRA and the criteria that are intended to be used in ranking of the HOS suitable for QRA based on the quality of these HOS. Criteria 3.2-3.5 (table 1) are intended to be used in both the selection and ranking of HOS. For an objective and transparent evaluation decisions concerning acceptable levels for these criteria should be made *a priori* to the evaluation. To provide an example of this approach we consider the response rate of a study. If a risk assessor decides that the acceptable level of the response rate is 80% all studies with a response rate < 80% should be excluded from QRA. For studies with a response rate 80% - 100% the response rate should be used in the ranking based on study quality.

T3.1 Response rate
A low response rate has a large effect on the potential for bias in HOS and therefore affects the quality of HOS (4, 7-9). The potential effect of a low response rate on bias in HOS has been demonstrated by Callahan et a (10)l. In their example a disease prevalence of 20% is assumed. With a response rate of 90% the potential bias ranged from -9% to +2%. However the potential bias increased to -20% to +13% when the response rate was reduced to 60% (10). Aspects that contribute to the impact of response rate on the estimated risks are the 'true' prevalence of the studied health effect and the underlying causes for the low response rate. For an objective evaluation it is important that risk assessors define a minimum acceptable response rate *a priori* to evaluation of HOS. However, an exception should be made for studies that report a response rate below the minimum acceptable response rate but are able to demonstrate that the studied population is representative for the population of interest, e.g., with a non-response analysis.

T3.2 Loss to follow-up
A high loss to follow-up is in many ways comparable to low response rate. The mechanism causing loss to follow-up and the 'true' prevalence of a studied health effect determine the potential impact of loss to follow-up on estimated risk levels (11-13). Therefore the approach to the evaluation of loss to follow-up is the same as the approach to the response rate. Again, for objective evaluation it is important that risk assessors define maximum acceptable loss to follow-up *a priori* to evaluation of HOS.

T3.3 Minimum follow-up time
The follow-up time in a study should be based on the estimated latency between exposure and the development of a health effect. For most chronic health effects, especially certain cancer types, considerable latency is expected between exposure and disease. For example, a study by Agalliu et al. explored the latency between exposure to metalworking fluids and

prostate cancer incidence and suggested that a latency period of 25 years was plausible (14). While the exact latency between exposure and occurrence of a health effect is usually not known it is clear that insufficient follow-up time will certainly lead to a considerable bias in the exposure-response relation. Therefore, studies that have not incorporated sufficient follow-up time should be excluded from QRA.

T3.4 Quality of the exposure measurement methods

Quantitative measurements used in exposure assessment in HOS can potentially differ with regard to the quality of the measurement methods and the analytical methods that have been used. A guideline to evaluate HOS based on the quality of exposure measurements is to compare the method(s) used in the study under evaluation to the method(s) that are currently considered as best practice. Some studies provide information on side by side comparisons of the used exposure measurement method with the best practice at the time of the study. Additional information from studies that solely focus on side by side comparisons of exposure measurement methods can be used as well (15, 16). If there is solid evidence that a specific method is unable to provide high quality exposure measurements HOS that used this method in their exposure assessment strategy should not be used for QRA.

T3.5 Insight into the variability of the exposure

For the evaluation of HOS it is important to realize that the exposure measurements used in exposure assessment can be highly variable. This variability can be attributed to a combination of measurement error and variation in exposure levels over time and space. Classical measurement error, by which analytical and sampling error is covered, usually only plays a marginal role because its magnitude is often orders of magnitude smaller than the variability in exposure over time and space. It is therefore preferred to use the terminology variability in exposure instead of measurement error although the effect of the variability on measures of associations between exposure and disease is a measurement error issue. The influence of exposure variability is dependent on the exposure assessment strategy. Simply speaking two quantitative exposure assessment strategies exist, measurements for each individual in the population or measurements for so called homogeneous exposure categories (17). Often, exposure assessment on the individual level is considered the gold standard. However, this strategy is most sensitive to intra-individual variability of the exposure. If intra-individual variability is not correctly addressed in an individual based exposure assessment strategy strong underestimation of the exposure-response relationship might occur when this variability is large relative to the variability between individuals in the population. However, this underestimation becomes smaller when more repeated measurements per individual have been taken (18). Categorization of the population in *a priori* assigned exposure groups, and use of the measured average exposure per exposure group in an exposure-response relationship is less sensitive for intra-individual variability. In most cases this strategy is known to lead to unbiased relations between exposure and response, however, unexplained

differences in health risks within exposure groups and unaddressed differences in exposure levels within exposure groups can lead to a reduction of power of this strategy in comparison with the individual exposure assessment strategy (19). Advanced methodologies to acquire insight into the level of measurement variability on HOS outcomes have been proposed (20-23). *A priori* to the evaluation risk assessors need to define a minimum acceptable level of information that is required to be able to assess whether enough insight into variability of exposure measurements is provided in HOS. Tielemans et al have developed guidelines to evaluate exposure data from HOS performed in the occupational exposure context (24). Similar approaches should be applied to exposure data from other exposure contexts. Differences between HOS in the ability to assess the relative contribution of the different sources of variability in exposure measurements can be used to rank the HOS. One should also realize that the two examples given above of individual exposure assessment and categorization approaches are only two contrasting examples out of a wide range of exposure assessment strategies. Specific theory exists for other exposure metrics and specific situations such as conversion of exposure measured on a continuous scale into exposure categories, unmeasured exposure correlated to the measured exposure, et cetera (25, 26). For specific situations the theoretical background needs to be considered in detail to evaluate a strategy.

T3.6 Application of exposure measurements in the exposure assessment

In most HOS researchers are confronted with a scarcity of exposure measurements. As a result exposure measurements might not be available for each 'assignment unit' (e.g., a single individual or a group of individuals with assumed similar exposure patterns) for the complete time period of interest. In this situation exposure measurements performed for 'assignment-unit-time-period' combinations and information regarding the circumstances of these measurements (e.g., year of measurement, type of weather during measurement or the task the measured individual performed during the measurement) is used to estimate exposure levels for 'assignment-unit-time-period' combinations for which exposure measurements are not available. The strategy that is used to extrapolate measurements over assignment-unit-time-period combinations determines the validity of the exposure estimates and therefore has a large impact on the overall quality of the quantification of exposure. In most HOS exposure measurements are extrapolated following a set of decision rules based on expert judgment. The use of expert judgment requires that a complete and detailed insight into the applied decision rules is essential for evaluation of HOS.

T3.7 Type of exposure metric

In an ideal situation an exposure metric captures three aspects that determine exposure: intensity, duration and timing (27). The quality of an exposure metric is based on biological considerations such as the time window of exposure that is relevant to the health effect of interest (22, 27, 28). A guideline to evaluate HOS based on the used exposure metric is to

compare the used metric with the current state of knowledge on the nature of the relation between the exposure and health outcome of interest.

T3.8 Specificity of the exposure indicator

In situations where it is difficult to assess the actual exposure that is assumed to be causally related to the health effect of interest, a '*causal*' indicator of exposure, researchers might assess a '*proxy*' for the causal exposure. However, it is crucial that the proxy exposure is highly correlated to the exposure of interest. An example of the use of an exposure proxy is the use of elemental carbon as proxy for exposure to diesel engine exhausts (29). Once absorbed in the human body distribution, metabolism and excretion have a large impact on the dose of a specific agent (or metabolite) at the site of action. Given they can be measured accurately, application of exposure indicators capable of incorporating these biological influences in exposure estimates will result in increased correlation between the exposure indicator and the dose at the site of action. The application of biomarkers of exposure in HOS potentially provides the possibility to obtain exposure indicators with higher specificity compared to indicators of external exposure. Similar, as with external exposure, insight into variability of biomarker based exposure measurements is of utmost importance for QRA. We suggest a categorization of exposure indicators based on two decisions '*proxy*' vs. '*causal*' indicator of exposure and '*external*' vs. '*internal*' indicator of exposure. Although the exposure indicator combination '*causal-internal*' in theory provides the highest quality of evidence for QRA the actual quality will still depend on the assumed accuracy of the exposure measurement/classification based on the level of insight into the variability of exposure.

T3.9 Blinded exposure assessment

To avoid that observer bias occurs exposure assessment should always be performed blinded for the health outcome of interest. If exposure assessment was performed on the individual level, omission of a statement regarding blinded exposure assessment is a reason to exclude HOS from QRA. If exposure assessment was performed to assess exposure for *a priori* defined 'homogeneous exposure categories', there is no direct connection between the individuals in the study population and the exposure assessment and therefore this criterion needs less stringent application.

T3.10 Quality of the exposure assignment strategy

In the exposure assignment step exposure levels assessed for specific 'assignment-unit-time-period' combinations are translated into exposure estimates for each individual in the study population. Assignment is based on information that is related to the individuals in the study population and related to the 'assignment-unit-time-period' combinations for which exposure levels have been assessed. Examples of this information are the jobs an individual performed during his working career, a description of daily diet or information on other factors potentially affecting exposure levels. The exposure context in which HOS are performed

determines which type of information is available for exposure assignment. A proper evaluation of the quality of exposure assignment requires insight into the proportion of the 'assignment-unit-time-period' combinations used for assignment for which no or little exposure measurements were available and exposure levels had to be inferred. In addition, the overlap between the 'assignment-unit-time-period' combinations for which exposure measurements were available and the exposure time periods that are assumed to be relevant to the assessed health risk needs to be evaluated. Miller et al. have demonstrated that, if enough information is available, differences with regard to the use of measurement data in exposure assignment can be made evident with the use of a simple tool (30). In this example the availability of exposure measurements was directly compared to the distribution of person years in the study population. However, a more detailed analysis on the level of 'assignment-unit-time-period' combinations is needed to provide a more accurate insight into the quality of exposure assignment. While a high quality exposure assignment strategy contributes considerably to the quality of the evidence from HOS, at this moment, most HOS do not provide enough information to enable such a detailed evaluation.

T3.11 Potential for information bias

In studies in which potentially more detailed information is available for cases than for controls or in cohort studies in which an index population is compared to a reference population there is a potential for information bias. Information bias has a large impact on HOS study outcomes and therefore on the quality of the evidence from HOS for QRA (4). A large potential for information bias in HOS will result in a decrease in the potential weight of evidence for QRA and should therefore be used in the ranking of HOS based on quality.

T3.12 Blinded health outcome assessment

Blinded determination of health outcomes in HOS regardless of the exposure status of the observed individual reduces the probability of observer bias in the study results. Non-blinded health outcome assessment should result in exclusion from QRA.

T3.13 Insight into the potential for systematic error in the study results

Most HOS provide statistics such as confidence intervals surrounding an estimate or *p* values for the interpretation of sampling error in the study results. However, these statistics do not provide insight into the potential for systematic error in study results. A more sophisticated approach to acquire insight into the potential for systematic error in a HOS is to perform sensitivity analyses. The basic idea of a sensitivity analysis is to quantify the uncertainty in the estimated effect based on insight into the possible variation of all sources that contribute to systematic error in a study. To quantify this uncertainty one needs to define an overall structure that relates all the individual sources of systematic error to the study outcome (31). Insight into the potential for systematic error in study results contributes to the quality of HOS.

Supplemental Material II, Selection of Studies that are Eligible for Application of the Evaluation Guidelines

Publications eligible for evaluation were identified as follows: 81 publications were identified with a Pubmed search which included the following MESH keywords *benzene, humans, leukemia* in combination with either *cohort studies* or case-*control studies*, 12 publications were added by following references included in a literature review that was identified in the original Pubmed search (32), finally 23 publications were added by following references included in regulatory risk assessments by the Canadian Centre for Occupational Health and Safety (33), the U.S. National institute for Occupation and Health (34), the U.S. Agency for Toxic Substances and Disease Registry (35) and the U.S. Environmental Protection Agency (36). All the identified publications were reviewed for eligibility of application of the evaluation guidelines (*table 1)*. For one of the study populations (the Pliofilm cohort) several re-analyses were performed (37-39). These re-analyses were based on sets of exposure estimates that were different from the publications on this cohort by the principal investigators. Because the discussion on which exposure estimates were 'best' remains unresolved and because only publications based on one of the sets of exposure estimates should be included in QRA, we chose to include only publications that were based on the original exposure estimates (40, 41). Additionally, two other re-analyses of the pliofilm data were excluded from evaluation as well (42, 43). Preference was given to the analyses performed in the original publications on this cohort, which were more compatible to the analyses performed in the other included publications. 32 publications were found not eligible because results from hazard characterization were not reported. From the 84 publication that did report results from hazard characterization, 53 publications were excluded because no quantitative exposure-response analysis specific for benzene and leukemia was reported. Finally 22 publications did not report results from quantitative exposure-response analysis specific for benzene and AML.

Supplemental Material II, Table 1 Overview identified studies.

	Citation	Hazard characterization	Quantitative exposure-response analysis specific for benzene and leukemia	Quantitative exposure-response analysis specific for benzene and AML	Included in the evaluation	Remarks	Source
1	Girard, R. & Revol, L. La frequence d'une exposition benzenique au cours des hempathies graves. Nouv Rev Fr Hematol 10, 477-83 (1970).	YES	NO		NO		a
2	Ishimaru, T. et al. Occupational factors in the epidemiology of leukemia in Hiroshima and Nagasaki. Am J Epidemiol 93, 157-65 (1971).	YES	NO		NO		a
3	Aksoy, M., Erdem, S. & DinCol, G. Leukemia in shoe-workers exposed chronically to benzene. Blood 44, 837-41 (1974).	YES	NO		NO		b
4	Thorpe, J. J. Epidemiologic survey of leukemia in persons potentially exposed to benzene. J Occup Med 16, 375-82 (1974).	YES	NO		NO		c
5	McMichael, A. J., Spirtas, R., Kupper, L. L. & Gamble, J. F. Solvent exposure and leukemia among rubber workers: an epidemiologic study. J Occup Med 17, 234-9 (1975).	YES	NO		NO		c
6	Brown, S. M. Letters to the editor: Leukemia and potential benzene exposure. J Occup Med 17, 5-6 (1975).	NO			NO		d
7	McMichael, A. J., Spirtas, R., Gamble, J. & Tousey, P. Mortality among rubber workers- Relationship to specific jobs. J Occup Med 18, 178-185 (1976).	YES	NO		NO		c
8	Infante, P. F., Rinsky, R. A., Wagoner, J. K. & Young, R. J. Leukaemia in benzene workers. Lancet 2, 76-8 (1977).	YES	NO		NO		d
9	Aksoy, M. & Erdem, S. Followup study on the mortality and the development of leukemia in 44 pancytopenic patients with chronic exposure to benzene. Blood 52, 285-92 (1978).	YES	NO		NO		d
10	Brandt, L., Nilsson, P. G. & Mitelman, F. Occupational exposure to petroleum products in men with acute non-lymphocytic leukaemia. Br Med J 1, 553 (1978).	YES	NO		NO		a
11	Infante, P. F. Leukemia among workers exposed to benzene. Tex Rep Biol Med 37, 153-61 (1978).	YES	NO		NO		e
12	Nicholson, W. J., Selikoff, I. J. & Seidman, H. Mortality experience of styrene-polystyrene polymerization workers. Initial findings. Scand J Work Environ Health 4 Suppl 2, 247-52 (1978).	YES	NO		NO		d

13	Ott, M. G., Townsend, J. C., Fishbeck, W. A. & Langner, R. A. Mortality among individuals occupationally exposed to benzene. Arch Environ Health 33, 3-10 (1978).	YES	YES	NO	NO	d
14	Linos, A., Kyle, R. A., O'Fallon, W. M. & Kurland, L. T. A case-control study of occupational exposures and leukaemia. Int J Epidemiol 9, 131-5 (1980).	YES	NO		NO	b
15	Ott, M. G., Kolesar, R. C., Scharnweber, H. C., Schneider, E. J. & Venable, J. R. A mortality survey of employees engaged in the development or manufacture of styrene-based products. J Occup Med 22, 445-60 (1980).	YES	NO		NO	d
16	Rinsky, R. A., Young, R. J. & Smith, A. B. Leukemia in benzene workers. Am J Ind Med 2, 217-45 (1981).	YES	NO		NO	d
17	Rushton, L. & Alderson, M. R. A case-control study to investigate the association between exposure to benzene and deaths from leukaemia in oil refinery workers. Br J Cancer 43, 77-84 (1981).	YES	NO		NO	a
18	Schottenfeld, D., Warshauer, M. & Zauber, A. in Quantification of Occupational Cancer (eds. R, P. & M, S.) (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1981).	YES	NO		NO	a
19	Thomas, T. L., Waxweiler, R. J., Moure-Eraso, R., Itaya, S. & Fraumeni, J. F., Jr. Mortality patterns among workers in three Texas oil refineries. J Occup Med 24, 135-41 (1982).	YES	NO		NO	a
20	Arp, E. W., Jr., Wolf, P. H. & Checkoway, H. Lymphocytic leukemia and exposures to benzene and other solvents in the rubber industry. J Occup Med 25, 598-602 (1983).	YES	NO		NO	a
21	Decoufle, P., Blattner, W. A. & Blair, A. Mortality among chemical workers exposed to benzene and other agents. Environ Res 30, 16-25 (1983).	YES	NO		NO	b
22	Tsai, S. P. et al. Retrospective mortality and medical surveillance studies of workers in benzene areas of refineries. J Occup Med 25, 685-92 (1983).	YES	NO		NO	b
23	Checkoway, H., Wilcosky, T., Wolf, P. & Tyroler, H. An evaluation of the associations of leukemia and rubber industry solvent exposures. Am J Ind Med 5, 239-49 (1984).	YES	NO		NO	a
24	Shaw, G., Lavey, R., Jackson, R. & Austin, D. Association of childhood leukemia with maternal age, birth order, and paternal occupation. A case-control study. Am J Epidemiol 119, 788-95 (1984).	YES	NO		NO	d
25	Austin, H. & Cole, P. Cigarette smoking and leukemia. J Chronic Dis 39, 417-21 (1986).	NO			NO	d
26	Austin, H., Cole, P. & McCraw, D. S. A case-control study of leukemia at an oil refinery. J Occup Med 28, 1169-73 (1986).	NO			NO	d

27	Bond, G. G., McLaren, E. A., Baldwin, C. L. & Cook, R. R. An update of mortality among chemical workers exposed to benzene. Br J Ind Med 43, 685-91 (1986).	YES	YES	NO	NO	2004 update is included	a
28	Flodin, U., Fredriksson, M., Persson, B., Hardell, L. & Axelson, O. Background radiation, electrical work, and some other exposures associated with acute myeloid leukemia in a case-referent study. Arch Environ Health 41, 77-84 (1986).	YES	NO		NO		b
29	Linet, M. S., Stewart, W. F., Van Natta, M. L., McCaffrey, L. D. & Szklo, M. Comparison of methods for determining occupational exposure in a case-control interview study of chronic lymphocytic leukemia. J Occup Med 29, 136-41 (1987).	YES	NO		NO		b
30	Rinsky, R. A. et al. Benzene and leukemia. An epidemiologic risk assessment. N Engl J Med 316, 1044-50 (1987).	YES	YES	NO	NO		a
31	Wong, O. An industry wide mortality study of chemical workers occupationally exposed to benzene. II. Dose response analyses. Br J Ind Med 44, 382-95 (1987).	YES	YES	NO	NO		a
32	Wong, O. An industry wide mortality study of chemical workers occupationally exposed to benzene. I. General results. Br J Ind Med 44, 365-81 (1987).	YES	NO	NO	NO		a
33	Yin, S. N. et al. Leukaemia in benzene workers: a retrospective cohort study. Br J Ind Med 44, 124-8 (1987).	YES	NO		NO		d
34	Malone, K. E. et al. Chronic lymphocytic leukemia in relation to chemical exposures. Am J Epidemiol 130, 1152-8 (1989).	YES	NO		NO		b
35	Paci, E. et al. Aplastic anemia, leukemia and other cancer mortality in a cohort of shoe workers exposed to benzene. Scand J Work Environ Health 15, 313-8 (1989).	YES	NO		NO		a
36	Rinsky, R. A. Benzene and leukemia: an epidemiologic risk assessment. Environ Health Perspect 82, 189-91 (1989).	YES	YES	NO	NO		d
37	Rinsky, R. A., Hornung, R. W. & Landrigan, P. J. Re: "Benzene and Leukemia: a Review of the Literature and a Risk Assessment". Am J Epidemiol 129, 1084-6 (1989).	NO			NO		d
38	Vai, T. et al. [A follow-up study of 304 cases of suspected pathology caused by benzene seen in 1950-71]. Med Lav 80, 397-404 (1989).	YES	NO		NO		d
39	Wongsrichanalai, C., Delzell, E. & Cole, P. Mortality from leukemia and other diseases among workers at a petroleum refinery. J Occup Med 31, 106-11 (1989).	YES	NO		NO		a

40	Yin, S. N. et al. A retrospective cohort study of leukemia and other cancers in benzene workers. Environ Health Perspect 82, 207-13 (1989).	YES	NO		NO		d
41	Young, N. Benzene and lymphoma. Am J Ind Med 15, 495-8 (1989).	NO			NO	Editorial	d
42	Collins, J. J. et al. A study of the hematologic effects of chronic low-level exposure to benzene. J Occup Med 33, 619-26 (1991).	YES	NO		NO		e
43	Hurley, J. F., Cherrie, J. W. & Maclaren, W. Exposure to benzene and mortality from leukaemia: results from coke oven and other coal product workers. Br J Ind Med 48, 502-3 (1991).	YES	NO		NO		d
44	McKinney, P. A., Alexander, F. E., Cartwright, R. A. & Parker, L. Parental occupations of children with leukaemia in west Cumbria, north Humberside, and Gateshead. Bmj 302, 681-7 (1991).	YES	NO		NO		d
45	Crane, M. M., Godwin, J. E., Annegers, J. F. & Keating, M. J. Is histological subtype a marker for environmental exposures in acute myelogenous leukemia? Cancer Epidemiol Biomarkers Prev 1, 183-8 (1992).	YES	NO		NO		d
46	Hayes, R. B. Biomarkers in occupational cancer epidemiology: considerations in study design. Environ Health Perspect 98, 149-54 (1992).	NO			NO		d
47	Heineman, E. F. et al. Occupational risk factors for multiple myeloma among Danish men. Cancer Causes Control 3, 555-68 (1992).	YES	YES	NO	NO		
48	Jakobsson, R., Ahlbom, A., Bellander, T. & Lundberg, I. [Follow-up of leukemia in drivers is of interest]. Lakartidningen 89, 1557 (1992).	NO			NO		d
49	Paustenbach, D. J. et al. Reevaluation of benzene exposure for the Pliofilm (rubberworker) cohort (1936-1976). J Toxicol Environ Health 36, 177-231 (1992).	YES	YES	NO	NO	No Rinksy exposure estimates; not considered in the evaluation	f
50	Richardson, S. et al. Occupational risk factors for acute leukaemia: a case-control study. Int J Epidemiol 21, 1063-73 (1992).	YES	NO		NO		b
51	Ciccone, G. et al. Myeloid leukemias and myelodysplastic syndromes: chemical exposure, histologic subtype and cytogenetics in a case-control study. Cancer Genet Cytogenet 68, 135-9 (1993).	YES	NO		NO		d

52	Schnatter, A. R., Katz, A. M., Nicolich, M. J. & Theriault, G. A retrospective mortality study among Canadian petroleum marketing and distribution workers. Environ Health Perspect 101 Suppl 6, 85-99 (1993).	YES	NO		NO		a
53	Crump, K. S. Risk of benzene-induced leukemia: a sensitivity analysis of the pliofilm cohort with additional follow-up and new exposure estimates. J Toxicol Environ Health 42, 219-42 (1994).	YES	YES	NO	NO	No Rinksy exposure estimates; not considered in the evaluation	d
54	Li, G. L. et al. Gender differences in hematopoietic and lymphoproliferative disorders and other cancer risks by major occupational group among workers exposed to benzene in China. J Occup Med 36, 875-81 (1994).	YES	NO		NO		d
55	Paxton, M. B., Chinchilli, V. M., Brett, S. M. & Rodricks, J. V. Leukemia risk associated with benzene exposure in the pliofilm cohort. II. Risk estimates. Risk Anal 14, 155-61 (1994).	YES	YES	NO	NO		d
56	Paxton, M. B., Chinchilli, V. M., Brett, S. M. & Rodricks, J. V. Leukemia risk associated with benzene exposure in the pliofilm cohort: I. Mortality update and exposure distribution. Risk Anal 14, 147-54 (1994).	NO			NO		d
57	Travis, L. B. et al. Hematopoietic malignancies and related disorders among benzene-exposed workers in China. Leuk Lymphoma 14, 91-102 (1994).	NO			NO		d
58	Bithell, J. F. & Draper, G. J. Apparent association between benzene and childhood leukaemia: methodological doubts concerning a report by Knox. J Epidemiol Community Health 49, 437-9 (1995).	NO			NO		d
59	Mele, A. et al. Epidemiology of acute promyelocytic leukemia. Haematologica 80, 405-8 (1995).	YES	NO		NO		d
60	Utterback, D. F. & Rinsky, R. A. Benzene exposure assessment in rubber hydrochloride workers: a critical evaluation of previous estimates. Am J Ind Med 27, 661-76 (1995).	NO			NO		d
61	**Wong, O. Risk of acute myeloid leukaemia and multiple myeloma in workers exposed to benzene. Occup Environ Med 52, 380-4 (1995).**	**YES**	**YES**	**YES**	**YES**		**d**

62	Wong, O. & Raabe, G. K. 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. Regul Toxicol Pharmacol 21, 307-21 (1995).	NO			NO	Pooled analysis	d
63	Armstrong, T. W. et al. Retrospective benzene and total hydrocarbon exposure assessment for a petroleum marketing and distribution worker epidemiology study. Am Ind Hyg Assoc J 57, 333-43 (1996).	NO			NO		d
64	Clavel, J. et al. Hairy cell leukaemia and occupational exposure to benzene. Occup Environ Med 53, 533-9 (1996).	YES	YES	NO	NO		d
65	Crump, K. S. Risk of benzene-induced leukemia predicted from the Pliofilm cohort. Environ Health Perspect 104 Suppl 6, 1437-41 (1996).	YES			NO	No Rinksy exposure estimates; not considered in the evaluation	d
66	Hayes, R. B. et al. Mortality among benzene-exposed workers in China. Environ Health Perspect 104 Suppl 6, 1349-52 (1996).	YES	YES	NO	NO		d
67	Linet, M. S. et al. Clinical features of hematopoietic malignancies and related disorders among benzene-exposed workers in China. Benzene Study Group. Environ Health Perspect 104 Suppl 6, 1353-64 (1996).	NO			NO		d
68	Mahendra, P., Richards, E. M., Sinclair, P., Nacheva, E. & Marcus, R. E. t(9;13)(q34;q12) chromosomal translocation persisting 4 years post autologous bone marrow transplantation for secondary AML despite morphological remission. Clin Lab Haematol 18, 121-2 (1996).	NO			NO		d
69	Paxton, M. B. Leukemia risk associated with benzene exposure in the Pliofilm cohort. Environ Health Perspect 104 Suppl 6, 1431-6 (1996).	YES	YES	NO	NO		d
70	Raabe, G. K. & Wong, O. Leukemia mortality by cell type in petroleum workers with potential exposure to benzene. Environ Health Perspect 104 Suppl 6, 1381-92 (1996).	NO			NO	Meta analysis	d

71	Rushton, L. Benzene exposure in the petroleum distribution industry associated with leukemia in the United Kingdom: overview of the methodology of a case-control study. Environ Health Perspect 104 Suppl 6, 1371-4 (1996).	NO			NO		d
72	Schnatter, A. R. et al. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med 53, 773-81 (1996).	YES	YES	NO	NO		d
73	Schnatter, A. R. et al. The relationship between low-level benzene exposure and leukemia in Canadian petroleum distribution workers. Environ Health Perspect 104 Suppl 6, 1375-9 (1996).	YES	YES	NO	NO		d
74	Schnatter, A. R., Nicolich, M. J. & Bird, M. G. Determination of leukemogenic benzene exposure concentrations: refined analyses of the Pliofilm cohort. Risk Anal 16, 833-40 (1996).	YES	YES	YES	NO	Re-analysis of pliofilm data; not considered in the evaluation	d
75	Yin, S. N. et al. A cohort study of cancer among benzene-exposed workers in China: overall results. Am J Ind Med 29, 227-35 (1996).	YES	NO		NO		d
76	Yin, S. N. et al. An expanded cohort study of cancer among benzene-exposed workers in China. Benzene Study Group. Environ Health Perspect 104 Suppl 6, 1339-41 (1996).	NO			NO		d
77	Collins, J. J., Ireland, B. K., Easterday, P. A., Nair, R. S. & Braun, J. Evaluation of lymphopenia among workers with low-level benzene exposure and the utility of routine data collection. J Occup Environ Med 39, 232-7 (1997).	YES	YES	NO	NO		e
78	**Hayes, R. B. et al. Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine--National Cancer Institute Benzene Study Group. J Natl Cancer Inst 89, 1065-71 (1997).**	**YES**	**YES**	**YES**	**YES**		**b**

79	Ireland, B., Collins, J. J., Buckley, C. F. & Riordan, S. G. Cancer mortality among workers with benzene exposure. Epidemiology 8, 318-20 (1997).	YES	YES	YES	NO	2003 follow up was used	d
80	Lewis, S. J., Bell, G. M., Cordingley, N., Pearlman, E. D. & Rushton, L. Retrospective estimation of exposure to benzene in a leukaemia case-control study of petroleum marketing and distribution workers in the United Kingdom. Occup Environ Med 54, 167-75 (1997).	NO			NO		d
81	Lynge, E., Anttila, A. & Hemminki, K. Organic solvents and cancer. Cancer Causes Control 8, 406-19 (1997).	NO			NO		d
82	**Rushton, L. & Romaniuk, H. 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. Occup Environ Med 54, 152-66 (1997).**	**YES**	**YES**	**YES**	**YES**		**d**
83	Nilsson, R. I., Nordlinder, R., Horte, L. G. & Jarvholm, B. Leukaemia, lymphoma, and multiple myeloma in seamen on tankers. Occup Environ Med 55, 517-21 (1998).	YES	NO		NO		d
84	Albin, M. et al. Acute myeloid leukemia and clonal chromosome aberrations in relation to past exposure to organic solvents. Scand J Work Environ Health 26, 482-91 (2000).	YES	NO		NO		b
85	Finkelstein, M. M. Leukemia after exposure to benzene: temporal trends and implications for standards. Am J Ind Med 38, 1-7 (2000).	YES	YES	NO	NO	Re-analysis of pliofilm data; not considered in the evaluation	d
86	Glass, D. C., Adams, G. G., Manuell, R. W. & Bisby, J. A. Retrospective exposure assessment for benzene in the Australian petroleum industry. Ann Occup Hyg 44, 301-20 (2000).	NO			NO		d
87	Hayes, R. B. et al. Benzene and lymphohematopoietic malignancies in China. J Toxicol Environ Health A 61, 419-32 (2000).	NO			NO		d

88	Korte, J. E., Hertz-Picciotto, I., Schulz, M. R., Ball, L. M. & Duell, E. J. The contribution of benzene to smoking-induced leukemia. Environ Health Perspect 108, 333-9 (2000).	NO			NO	d
89	Schnatter, R. Petroleum worker studies and benzene risk assessment. J Toxicol Environ Health A 61, 433-7 (2000).	NO			NO	d
90	Glass, D. C. & Gray, C. N. Estimating mean exposures from censored data: exposure to benzene in the Australian petroleum industry. Ann Occup Hyg 45, 275-82 (2001).	NO			NO	d
91	Raaschou-Nielsen, O., Hertel, O., Thomsen, B. L. & Olsen, J. H. Air pollution from traffic at the residence of children with cancer. Am J Epidemiol 153, 433-43 (2001).	YES	YES	NO	NO	d
92	**Guenel, P., Imbernon, E., Chevalier, A., Crinquand-Calastreng, A. & Goldberg, M. Leukemia in relation to occupational exposures to benzene and other agents: a case-control study nested in a cohort of gas and electric utility workers. Am J Ind Med 42, 87-97 (2002).**	**YES**	**YES**	**YES**	**YES**	**d**
93	Li, K. & Yu, S. Leukemia mortality and occupational exposure to rubber: a nested case-control study. Int J Hyg Environ Health 204, 317-21 (2002).	YES	NO		NO	d
94	Rinsky, R. A., Hornung, R. W., Silver, S. R. & Tseng, C. Y. Benzene exposure and hematopoietic mortality: A long-term epidemiologic risk assessment. Am J Ind Med 42, 474-80 (2002).	YES	YES	NO	NO	d
95	Silver, S. R., Rinsky, R. A., Cooper, S. P., Hornung, R. W. & Lai, D. Effect of follow-up time on risk estimates: a longitudinal examination of the relative risks of leukemia and multiple myeloma in a rubber hydrochloride cohort. Am J Ind Med 42, 481-9 (2002).	YES	YES	NO	NO	d
96	Adegoke, O. J. et al. Occupational history and exposure and the risk of adult leukemia in Shanghai. Ann Epidemiol 13, 485-94 (2003).	YES	NO		NO	b
97	**Collins, J. J., Ireland, B., Buckley, C. F. & Shepperly, D. Lymphohaematopoeitic cancer mortality among workers with benzene exposure. Occup Environ Med 60, 676-9 (2003).**	**YES**	**YES**	**YES**	**YES**	**b**
98	**Glass, D. C. et al. Leukemia risk associated with low-level benzene exposure. Epidemiology 14, 569-77 (2003).**	**YES**	**YES**	**YES**	**YES**	**d**
99	Seniori Costantini, A., Quinn, M., Consonni, D. & Zappa, M. Exposure to benzene and risk of leukemia among shoe factory workers. Scand J Work Environ Health 29, 51-9 (2003).	YES	YES	NO	NO	d
100	Adegoke, O. J. et al. Agreement of job-exposure matrix (JEM) assessed exposure and self-reported exposure among adult leukemia patients and controls in Shanghai. Am J Ind Med 45, 281-8 (2004).	NO			NO	d

101	**Bloemen, L. J., Youk, A., Bradley, T. D., Bodner, K. M. & Marsh, G. Lymphohaematopoietic cancer risk among chemical workers exposed to benzene. Occup Environ Med 61, 270-4 (2004).**	**YES**	**YES**	**YES**	**YES**	**d**
102	Crosignani, P. et al. Childhood leukemia and road traffic: A population-based case-control study. Int J Cancer 108, 596-9 (2004).	YES	YES	NO	NO	d
103	Glass, D. C. et al. Leukemia risk and relevant benzene exposure period-Re: follow-up time on risk estimates, Am J Ind Med 42:481-489, 2002. Am J Ind Med 45, 222-3; author reply 224-5 (2004).	NO			NO	d
104	Patel, A. S. et al. Risk of cancer as a result of community exposure to gasoline vapors. Arch Environ Health 59, 497-503 (2004).	YES	NO		NO	d
105	Steffen, C. et al. Acute childhood leukaemia and environmental exposure to potential sources of benzene and other hydrocarbons; a case-control study. Occup Environ Med 61, 773-8 (2004).	YES	NO		NO	d
106	Glass, D. C., Gray, C. N., Jolley, D. J., Gibbons, C. & Sim, M. R. Health Watch exposure estimates: do they underestimate benzene exposure? Chem Biol Interact 153-154, 23-32 (2005).	NO			NO	d
107	Kasim, K., Levallois, P., Abdous, B., Auger, P. & Johnson, K. C. Lifestyle factors and the risk of adult leukemia in Canada. Cancer Causes Control 16, 489-500 (2005).	YES	NO		NO	d
108	Mirer, F. E. Comment on Caprolactam study. Ann Epidemiol 15, 735; author reply 736 (2005).	NO			NO	d
109	Schnatter, A. R., Rosamilia, K. & Wojcik, N. C. Review of the literature on benzene exposure and leukemia subtypes. Chem Biol Interact 153-154, 9-21 (2005).	NO			NO	d
110	Sorahan, T., Kinlen, L. J. & Doll, R. Cancer risks in a historical UK cohort of benzene exposed workers. Occup Environ Med 62, 231-6 (2005).	YES	NO		NO	d
111	Swaen, G. M., Scheffers, T., de Cock, J., Slangen, J. & Drooge, H. Leukemia risk in caprolactam workers exposed to benzene. Ann Epidemiol 15, 21-8 (2005).	YES	YES	NO	NO	d
112	Glass, D. C., Gray, C. N., Jolley, D. J., Gibbons, C. & Sim, M. R. The health watch case-control study of leukemia and benzene: the story so far. Ann N Y Acad Sci 1076, 80-9 (2006).	NO			NO	d
113	Li, G. & Yin, S. Progress of epidemiological and molecular epidemiological studies on benzene in China. Ann N Y Acad Sci 1076, 800-9 (2006).	NO			NO	d

114	Wiwanitkit, V. Classification of risk occupation for benzene exposure by urine trans, trans-munconic acid level. Asian Pac J Cancer Prev 7, 149-50 (2006).	YES	NO	NO	d
115	Schubauer-Berigan, M. K. et al. Risk of chronic myeloid and acute leukemia mortality after exposure to ionizing radiation among workers at four U.S. nuclear weapons facilities and a nuclear naval shipyard. Radiat Res 167, 222-32 (2007).	YES	NO	NO	d
116	Zhang, L. et al. Aberrations in chromosomes associated with lymphoma and therapy-related leukemia in benzene-exposed workers. Environ Mol Mutagen 48, 467-74 (2007).	YES	NO	NO	d

[a] Referenced in 1995 risk assessment by Canadian Centre for Occupational Health and Safety (CCOHS) http://www.canoshweb.org/odp/html/rp7.htm (accessed on 02/12/2008).
[b] Referenced in literature by Schnatter et al. (Schnatter, A. R., Rosamilia, K. & Wojcik, N. C. Review of the literature on benzene exposure and leukemia subtypes. Chem Biol Interact 153-154, 9-21 (2005). [c] Referenced in 1976 risk assessment by NIOSH http://www.cdc.gov/niosh/pdfs/76-benz.pdf (accessed on 02/12/2008). [d] Pubmed search including the following MESH keywords: *benzene, humans, leukaemia* in combination with either *cohort studies* or case- *studies* (performed on 02/12/2008). [e] Referenced in 2007 literature review by U.S. Agency for Toxic Substances and Disease Registry (ATSDR) http://www.atsdr.cdc.gov/toxprofiles/tp3.pdf (accessed on 02/12/2008). [f] Referenced in risk assessment by U.S. EPA 'Carcinogenic Effects of Benzene: an Update' (1998) http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=2806 (accessed on 02/12/2008)

Supplemental Material III, Outcome of the Evaluation of the Studies that were Eligible for Evaluation with the Guidelines

(1) AUSTRALIAN HEALTH WATCH STUDY		
Evaluated papers	1. Glass, D. C., Adams, G. G., Manuell, R. W. & Bisby, J. A. Retrospective exposure assessment for benzene in the Australian petroleum industry. Ann Occup Hyg 44, 301-20 (2000). 2. Glass, D. C. et al. Leukemia risk associated with low-level benzene exposure. Epidemiology 14, 569-77 (2003). 3. Glass, D. C., Gray, C. N., Jolley, D. J., Gibbons, C. & Sim, M. R. Health Watch exposure estimates: do they underestimate benzene exposure? Chem Biol Interact 153-154, 23-32 (2005).	
Criteria	**Description**	**Outcome**
T1.1	A case-control design was used in this study	Yes
T1.2	Exposure was expressed on a ratio scale and specific for benzene	Yes
T1.3	Sufficient details were provided regarding the performed statistical analysis (Conditional logistic regression)	Yes
T1.4	Criteria for inclusion of subjects were described with sufficient detail. Inclusion criteria for cases were: sex = male, member of the Health Watch cohort and reported and confirmed lympho-hematopoietic cancer. Controls were randomly selected from the Health Watch cohort and matched to the cases based on year of birth	Yes
T1.5	Assessment of the health effect was performed following recognized norms. A hierarchical classification strategy was applied in which histological confirmation of AML was considered to be a higher level of evidence than classification based on a doctor's letter, which was considered to be a higher level of evidence than information from the cancer registry, which was considered to be a higher level of evidence than information from a death certificate	Yes
T1.6	Ionizing radiation is not considered as potential strong confounder in this study. It is assumed that the level of ionizing radiation received by the studied workers is on background level. Other factors were tested for a potential association with total leukemia: tobacco smoking, drinking alcohol, duration of employment and employment starting date	Yes
T2.1	This study was conducted following a case-control design nested in the Health Watch cohort	
T3.1	The response rate of this study was 100%	
T3.4	The quality of the methods for exposure measurements and criteria for inclusion and exclusion and limitations of exposure measurements were discussed. Results from exposure measurements were used for exposure assessment if personal exposure measurements were performed, if measurements were not corrected to 8 hour time weighted averages and if information on job site-location, job title and duration of monitoring was available for the measurements. In addition, exposure measurements were not used for exposure assessment if adequate information on the used measuring method was not available, if no units were presented, if no information on the year of measurement was available, if detailed information regarding the measured tasks was not available, if the purpose of the measurement was not specified and if information whether the measured exposure situation was typical for the assessed task was not available. Limitations of the exposure measurements used for exposure assessment were discussed as well. More frequently occurring limitations were: no information on the limit of detection, no information on the technology used for exposure	

	measurements, no information on the type of products that were handled by the measured worker and limited information on the specificity of the measurements for the measurement site	
T3.5	The variability in exposure measurements and potential for measurement error was not discussed	
T3.6	Task based exposure measurements were used to generate a workplace exposure estimate defined as the 'time weighted average of different activity exposures normalized to 35 hour work week'. Exposure modifying factors were applied to the exposure measurements in order to estimate exposures for workplaces and time periods for which no exposure measurements were available	
T3.7	The reported exposure metrics included cumulative lifetime exposure (ppm-years)	
T3.8	Benzene concentration (causal agent) in the breathing zone (external exposure) was used as indicator of exposure	Causal / External
T3.9	It was reported that the occupational hygienists performing the exposure assessment did so without knowledge of the case or control status of the subjects	
T3.10	Exposure was assigned based on the jobs workers performed during their lifetime. For each job title a workplace exposure estimate was multiplied with the years spent in a specific job. The estimates for specific job-titles were aggregated into cumulative lifetime exposure. In addition, the level of extrapolation needed to assign measured exposures to the individuals in the studied population was discussed	
T3.11	Because both cases and controls were selected from the Health Watch cohort it is assumed that there was limited potential for information bias	
T3.13	No sensitivity analysis was performed	

(2) CAPM-NCI STUDY		
Evaluated papers	1. Yin, S. N. et al. Cohort study among workers exposed to benzene in China: I. General methods and resources. Am J Ind Med 26, 383-400 (1994). 2. Dosemeci, M. et al. Cohort study among workers exposed to benzene in China: II. Exposure assessment. Am J Ind Med 26, 401-11 (1994). 3. Travis, L. B. et al. Hematopoietic malignancies and related disorders among benzene-exposed workers in China. Leuk Lymphoma 14, 91-102 (1994). 4. Hayes, R. B. et al. Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine--National Cancer Institute Benzene Study Group. J Natl Cancer Inst 89, 1065-71 (1997).	
Criteria	**Description**	**Outcome**
T1.1	A cohort design was used in this study	Yes
T1.2	Exposure was expressed on a ratio scale and specific for benzene	Yes
T1.3	Sufficient details were provided regarding the performed statistical analysis (Poisson regression analysis)	Yes
T1.4	Criteria for inclusion of subjects were described with sufficient detail. Included workers were employed in 672 factories in 12 cities in China	Yes
T1.5	An extensive method for the classification of the health outcome was used. The method included evaluation of medical records, pathology reports and histopathological material	Yes
T1.6	Ionizing radiation was not considered as potential strong confounder in this study. It is assumed that the level of ionizing radiation received by the individuals in the studied population is on background level	Yes
T2.1	This study was conducted following a cohort design	
T3.1	The response rate of this study was 100%	
T3.2	The loss to follow-up reported in this study was 0.2% for exposed individuals and 0.3% for unexposed individuals	
T3.3	The average follow up time in this study was 10.5 years for individuals exposed to benzene and 11.7 years for individuals not exposed to benzene	
T3.4	Although the number of exposure measurements used for exposure assessment was reported, there was limited insight into the quality of these measurements. It was reported that most of the benzene measurements were based on short-term area sampling and that there was a lack of personal sampling	
T3.5	There was limited discussion on the variability in exposure measurements. The mean concentration of the available exposure measurements was reported for seven calendar periods. This provides some insight into the variability of exposure over time. The potential for measurement error in the exposure measurements was not discussed	
T3.6	For each combination of job-title and time-period, exposure was estimated based on estimates by local industrial hygienist and other occupational health personnel. The experts used the available ambient benzene exposure measurements in combination with detailed production and related process information for seven calendar periods	
T3.7	The reported exposure metrics included cumulative lifetime exposure (ppm-years) and average exposure (ppm)	
T3.8	Benzene concentration (causal agent) in the breathing zone (external exposure) was used as indicator of exposure	Causal / External
T3.9	Exposure assessment was performed on the level of factory/work unit/job title/calendar-year. Only in the assignment stage were results from	

	exposure assessment linked to the individuals in the study population based on work histories. It is therefore assumed that exposure assessment was performed blinded for the disease status of the individuals in the population
T3.10	Exposure was assigned to individuals in the study population based on the jobs that were performed in their work history
T3.11	There was no difference in the level of information available for exposed individuals and unexposed individuals. Therefore it is assumed that the potential for information bias was low
T3.12	It was reported that clinical laboratory and pathology data for all patients were abstracted onto standardized forms by physician investigators who were not aware of the exposure status of the subjects, nor of the numbers of exposed and non-exposed cases
T3.13	No sensitivity analysis was performed

(3) DOW COHORT STUDY		
Evaluated papers	1. Ott, M. G., Townsend, J. C., Fishbeck, W. A. & Langner, R. A. Mortality among individuals occupationally exposed to benzene. Arch Environ Health 33, 3-10 (1978). 2. Bloemen, L. J., Youk, A., Bradley, T. D., Bodner, K. M. & Marsh, G. Lymphohaematopoietic cancer risk among chemical workers exposed to benzene. Occup Environ Med 61, 270-4 (2004).	
Criteria	**Description**	**Outcome**
T1.1	A cohort design was used in this study	Yes
T1.2	Exposure was expressed on a ratio scale and specific for benzene	Yes
T1.3	Sufficient details were provided regarding the performed statistical analysis (OCMAP-PLUS modified life table procedure)	Yes
T1.4	Criteria for inclusion of subjects were described with sufficient detail. Employees with at least one month's work experience in any of three relevant production areas on or after 1 january 1938 were included in the study	Yes
T1.5	Assessment of the health effect was performed following recognized norms. Cause of death was determined based on a death certificate and was coded by a certified nosologist according to the ICD in effect at the time of death. For this study all original codifications were recoded to ICD-9	Yes
T1.6	Ionizing radiation was not considered as potential strong confounder in this study. It is assumed that the level of ionizing radiation received by the individuals in the studied population is on background level	Yes
T2.1	This study was conducted following a cohort design	
T3.1	The response rate of this study was 100%	
T3.2	The loss to follow-up reported in this study was 0.6%	
T3.3	68% of the members of the cohort were followed at least 30 years	
T3.4	The quality criteria for exposure measurements to be included in the exposure assessment were not discussed	
T3.5	The number of exposure measurements, range of the measurements and estimated time weighted averages were presented. But these results were not discussed in the text	
T3.6	The relation between job-title and time period specific exposure estimates and the use of industrial hygiene measurements is unclear. Job-titles were assigned to exposure categories by an industrial hygienist. The exposure categories were based on the industrial hygiene measurements reported in the 1978 study	
T3.7	The reported exposure metrics included cumulative lifetime exposure (ppm-years) and average exposure (ppm)	
T3.8	Benzene concentration (causal agent) in the breathing zone (external exposure) was used as indicator of exposure	Causal / External
T3.9	Exposure assessment was performed on the level of job title/time-period. Only in the assignment stage were results from exposure assessment linked to the individuals in the study population based on work histories. It is therefore assumed that exposure assessment was performed blinded for the disease status of the individuals in the population	
T3.10	Individual employee histories were linked to job and time specific benzene exposure estimates to compute the summary exposure measures	
T3.11	The exposed population is compared to a local population which is assumed to be exposed at background level. Therefore, some potential for information bias regarding health outcome assessment exists	

T3.12	The cause of death was derived from company human resource records or population mortality registries. Therefore the health outcome assessment is assumed to have been performed blinded for the exposure status
T3.13	No sensitivity analysis was performed

(4) GUÉNEL STUDY		
Evaluated papers	1. Guenel, P., Imbernon, E., Chevalier, A., Crinquand-Calastreng, A. & Goldberg, M. Leukemia in relation to occupational exposures to benzene and other agents: a case-control study nested in a cohort of gas and electric utility workers. Am J Ind Med 42, 87-97 (2002).	
Criteria	**Description**	**Outcome**
T1.1	A case-control design was used in this study	Yes
T1.2	Exposure was not expressed on a ratio scale but presented in unit-years.	No, Study should be excluded from QRA
T1.3	Sufficient details were provided regarding the performed statistical analysis (Unadjusted conditional logistic regression)	Yes
T1.4	Criteria for inclusion of subjects were described with sufficient detail. Cases were workers diagnosed with leukemia active at the time of diagnosis. For each case 4 controls were selected. The controls were also active EDF-GDF workers matched to the cases by year of birth	Yes
T1.5	Assessment of the health effect was performed following recognized norms. A pathology report was used to code cases following ICD-O	Yes
T1.6	Ionizing radiation was not considered as potential strong confounder in this study. It is assumed that the level of ionizing radiation received by the studied workers is on background level. In addition, there is no indication that the cases in this study are differently exposed to ionizing radiation than the controls	Yes

(5) MONSANTO COHORT STUDY		
Evaluated papers	1. Ireland, B., Collins, J. J., Buckley, C. F. & Riordan, S. G. Cancer mortality among workers with benzene exposure. Epidemiology 8, 318-20 (1997). 2. Collins, J. J., Ireland, B., Buckley, C. F. & Shepperly, D. Lymphohaematopoeitic cancer mortality among workers with benzene exposure. Occup Environ Med 60, 676-9 (2003).	
Criteria	**Description**	**Outcome**
T1.1	A cohort design was used in this study	Yes
T1.2	Exposure was expressed on a ratio scale and specific for benzene	Yes
T1.3	Insufficient details were provided regarding the performed statistical analysis in this study. Therefore, there is no insight into the decisions that were made in the statistical analysis (e.g., stratification for age and time period)	No, Study should be excluded from QRA
T1.4	The criteria for inclusion of subjects were described with sufficient detail. The study population consisted of all hourly workers that began employment between 1940 and 1977 at the Monsanto (Solutia) plant	Yes
T1.5	Assessment of the health effect was performed following recognized norms. Death certificates were used to assess the health effect	Yes
T1.6	Ionizing radiation was not considered as potential strong confounder in this study. It is assumed that the level of ionizing radiation received by the individuals in the studied population is on background level	Yes

(6) PLIOFILM COHORT STUDY		
Evaluated papers	1. Rinsky, R. A. et al. Benzene and leukemia. An epidemiologic risk assessment. N Engl J Med 316, 1044-50 (1987). 2. Rinsky, R. A. Benzene and leukemia: an epidemiologic risk assessment. Environ Health Perspect 82, 189-91 (1989). 3. Paxton, M. B., Chinchilli, V. M., Brett, S. M. & Rodricks, J. V. Leukemia risk associated with benzene exposure in the pliofilm cohort: I. Mortality update and exposure distribution. Risk Anal 14, 147-54 (1994a). 4. Paxton, M. B., Chinchilli, V. M., Brett, S. M. & Rodricks, J. V. Leukemia risk associated with benzene exposure in the pliofilm cohort. II. Risk estimates. Risk Anal 14, 155-61 (1994b). 5. Wong, O. Risk of acute myeloid leukaemia and multiple myeloma in workers exposed to benzene. Occup Environ Med 52, 380-4 (1995).	
Criteria	**Description**	**Outcome**
T1.1	A cohort design was used in this study	Yes
T1.2	Exposure was expressed on a ratio scale and specific for benzene	Yes
T1.3	Sufficient details were provided regarding the performed statistical analysis (Standardized mortality ratios were calculated with the NIOSH lifetable analysis program)	Yes
T1.4	Criteria for inclusion of subjects were described with sufficient detail. All nonsalaried white men employed in a rubber hydrochloride department for at least one day between jan 1, 1940 and December 31, 1965 were included in the study population	Yes
T1.5	Assessment of the health effect was performed following recognized norms. Health effect classification was based on death certificates and codification by a qualified nosologist	Yes
T1.6	Ionizing radiation was not considered as potential strong confounder in this study. It is assumed that the level of ionizing radiation received by the individuals in the studied population is on background level	Yes
T2.1	This study was conducted following a cohort design.	
T3.1	The response rate of this study was 100%	
T3.2	The loss to follow-up reported in this study was 0.9%	
T3.3	A minimum follow-up time of 22 years was reported in this study	
T3.4	Industrial hygiene measurements were used for exposure assessment. Limited discussion regarding the quality of the measurements and the number of measurements that was available is presented in the reviewed papers. However, it was mentioned that industrial hygiene measurements consisted primarily of area samples and not personal samples	
T3.5	Variability in exposure measurements and the potential for measurement error was not discussed	
T3.6	Job titles were grouped into exposure classes. In general exposure classes represented areas in which industrial-hygiene data had been collected. In some instances, job titles did not readily fit into a single area; in such situations hybrid classes were defined. Cells were defined based on an exposure class-year combination. Cells for which no data was available were completed by interpolation between available previous and	

	subsequent values. When interpolation could not be performed because no measured value existed for an exposure class in the first or last year of the study, the nearest measured value for that exposure class was projected forward or backward	
T3.7	The reported exposure metrics included cumulative lifetime exposure (ppm-years)	
T3.8	Benzene concentration (causal agent) in the breathing zone (external exposure) was used as indicator of exposure	Causal / External
T3.9	Exposure assessment was performed on the level of 'exposure class'. Only in the assignment stage were results from exposure assessment linked to the individuals in the study population based on work histories. It is therefore assumed that exposure assessment was performed blinded for the disease status of the individuals in the population	
T3.10	Person's daily benzene exposure was obtained from the appropriate cell in the exposure class-year matrix. These daily values were then summed for a workers entire career	
T3.11	Mortality in studied population was compared to mortality in the general population for which background exposure levels were assumed. Therefore there is some potential for information bias	
T3.12	There was no specific mention of blinded health outcome assessment	
T3.13	No sensitivity analysis was performed	

(7) U.K. PETROLEUM WORKERS STUDY		
Included papers	1. Rushton, L. & Romaniuk, H. 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. Occup Environ Med 54, 152-66 (1997). 2. Lewis, S. J., Bell, G. M., Cordingley, N., Pearlman, E. D. & Rushton, L. Retrospective estimation of exposure to benzene in a leukaemia case-control study of petroleum marketing and distribution workers in the United Kingdom. Occup Environ Med 54, 167-75 (1997).	
Criteria	**Description**	**Outcome**
T1.1	A case-control design was used in this study	Yes
T1.2	Exposure was expressed on a ratio scale and specific for benzene	Yes
T1.3	Sufficient details were provided regarding the performed statistical analysis.(various different statistical approaches were applied)	Yes
T1.4	Criteria for inclusion of subjects were described with sufficient detail. Cases were men from UK oil distribution cohort who died before 1 january 1993 with a mention of leukemia on their death certificate or had an ICD-9 code 204-208 in the cancer registry. Controls were randomly selected from the same cohort and were matched to the cases based on age	Yes
T1.5	Assessment of the health effect was performed following recognized norms. Information from death certificates and information from a cancer registry was used. If the information from the two sources was conflicting information from the death certificate overruled information from the cancer registry	Yes
T1.6	Ionizing radiation was not considered as potential strong confounder in this study. It is assumed that the level of ionizing radiation received by the individuals in the studied population is on background level	Yes
T2.1	In this study a nested case-control design was used	
T3.1	The response rate of this study was 100%	
T3.4	The quality of the exposure measurements used in the exposure assessment was discussed. The discussion was focused on the quality of the exposure measurements, the validity of the used sampling technologies and the used analytical techniques. The authors reported that only high quality personal exposure measurements were used for exposure assessment	
T3.5	The distribution of benzene exposure measurements originating from similar exposure contexts was assessed and tested for log-normality. The potential for measurement error due to inaccuracy of the used measurement techniques was not discussed	
T3.6	Retrospective estimates of workplace exposure for each job reported in the work histories of all the study members were obtained by creating base estimates. Base estimates were estimated based on the exposure measurements and adjusted with the use of modifying factors. Modifying factors represented factors that could have affected the exposure levels (e.g., changes in exposure circumstances over time or between two different work-sites)	
3.7	The reported exposure metrics included cumulative lifetime exposure (ppm-years)	
T3.8	Benzene concentration (causal agent) in the breathing zone (external exposure) was used as indicator of exposure	Causal / External

T3.9	Exposure assessment was performed on the level of workplace. Only in the assignment stage were results from exposure assessment linked to the individuals in the study population based on work histories. It is therefore assumed that exposure assessment was performed blinded for the disease status of the individuals in the population
T3.10	Cumulative lifetime exposure was estimated by summing the cumulative exposure for each different job held by an individual for the individuals entire work history
T3.11	Potential for information bias was tested with a sensitivity analysis and was reported to be low
T3.13	A sensitivity analysis was performed to test the influence of several factors on the study outcomes

References Supplemental Material

1. Swaen GM. A framework for using epidemiological data for risk assessment. Hum Exp Toxicol. 2006;25(3):147-55.
2. Greenland S, Robins J. Invited commentary: ecologic studies—biases, misconceptions, and counterexamples. Am J Epidemiol. 1994;139(8):747-60.
3. Stevens SS. On the Theory of Scales of Measurement. Science. 1946;103(2684):677-80.
4. Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Community Health. 2004;58(8):635-41.
5. Checkoway H. Case-crossover designs in occupational health. Occup Environ Med. 2004;61(12):953-4.
6. Blair A, Stewart P, Lubin JH, et al. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med. 2007;50(3):199-207.
7. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615-25.
8. Mezei G, Kheifets L. Selection bias and its implications for case-control studies: a case study of magnetic field exposure and childhood leukaemia. Int J Epidemiol. 2006;35(2):397-406.
9. Pearce N, Checkoway H, Kriebel D. Bias in occupational epidemiology studies. Occup Environ Med. 2007;64(8):562-8.
10. Callahan MA, Clickner RP, Whitmore RW, et al. Overview of important design issues for a National Human Exposure Assessment Survey. J Expo Anal Environ Epidemiol. 1995;5(3):257-82.
11. Blanker MH, Prins J, Bosch JL, et al. Loss to follow-up in a longitudinal study on urogenital tract symptoms in Dutch older men. Urol Int. 2005;75(1):30-7.
12. Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19(8):751-60.
13. Zunzunegui MV, Beland F, Gutierrez-Cuadra P. Loss to follow-up in a longitudinal study on aging in Spain. J Clin Epidemiol. 2001;54(5):501-10.
14. Agalliu I, Kriebel D, Quinn MM, et al. Prostate cancer incidence in relation to time windows of exposure to metalworking fluids in the auto industry. Epidemiology. 2005;16(5):664-71.
15. de Vocht F, Huizer D, Prause M, et al. Field comparison of inhalable aerosol samplers applied in the European rubber manufacturing industry. Int Arch Occup Environ Health. 2006;79(8):621-9.
16. Stephenson DJ, Lillquist DR, DeRosso FD, et al. Side-by-side comparison of three sampling methods for aerosolized endotoxin in a wastewater treatment facility. J Environ Health. 2004;67(4):16-9.
17. Armstrong BG. The effects of measurement errors on relative risk regressions. Am J Epidemiol. 1990;132(6):1176-84.
18. Heederik D, Attfield M. Characterization of dust exposure for the study of chronic occupational lung disease: a comparison of different exposure assessment strategies. Am J Epidemiol. 2000;151(10):982-90.
19. Tielemans E, Kupper LL, Kromhout H, et al. Individual-based and group-based occupational exposure assessment: some equations to evaluate different strategies. Ann Occup Hyg. 1998;42(2):115-9.
20. Heederik D, Miller BG. Weak associations in occupational epidemiology: adjustment for exposure estimation error. Int J Epidemiol. 1988;17(4):970-4.
21. Kromhout H, Loomis DP, Kleckner RC. Uncertainty in the relation between exposure to magnetic fields and brain cancer due to assessment and assignment of exposure and analytical methods in dose-response modeling. Ann N Y Acad Sci. 1999;895:141-55.
22. Loomis A, Kromhout H, Kleckner RC, et al. Effects of the analytical treatment of exposure data on associations of cancer and occupational magnetic field exposure. Am J Ind Med. 1998;34(1):49-56.
23. Xue X, Kim MY, Shore RE. Estimation of health risks associated with occupational radiation exposure: addressing measurement error and minimum detectable exposure level. Health Phys. 2006;91(6):582-91.
24. Tielemans E, Marquart H, De Cock J, et al. A proposal for evaluation of exposure data. Ann Occup Hyg. 2002;46(3):287-97.
25. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol. 1990;132(4):746-8.
26. Loomis DP, Savitz DA. Effect of incomplete exposure assessment on epidemiologic dose-response analyses. Scand J Work Environ Health. 1994;20(3):200-5.
27. Vacek PM. Assessing the effect of intensity when exposure varies over time. Stat Med. 1997;16(5):505-13.

28. Seixas NS, Robins TG, Becker M. A novel approach to the characterization of cumulative exposure for the study of chronic occupational disease. Am J Epidemiol. 1993;137(4):463-71.
29. Schauer JJ. Evaluation of elemental carbon as a marker for diesel particulate matter. J Expo Anal Environ Epidemiol. 2003;13(6):443-53.
30. Miller BG, Fransman W, Heederik D, et al. A review of the data quality and comparability of case-control studies of low-level exposure to benzene in the petroleum industry (IOM Report TM/05/04). Edinburgh: Institute of Occupational Medicine. http://www.iom-world.org/pubs/IOM_TM0504.pdf, 2005.
31. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. Epidemiology. 2003;14(4):451-8.
32. Schnatter AR, Rosamilia K, Wojcik NC. Review of the literature on benzene exposure and leukemia subtypes. Chem Biol Interact. 2005;153-154:9-21.
33. Penney J. Report to the Occupational Disease Panel (Industrial Disease Standards Panel) on Occupational Exposure to Benzene and Leukaemia. : Canadian Centre for Occupational Health and Safety, 1995.
34. NIOSH. Revised Recommendation for an Occupational Exposure Standard for Benzene. : National Institute for Occupational Safety and Health, 1976.
35. ATSDR. Toxicological Profile for Benzene. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, 2007.
36. U.S. EPA. Carcinogenic Effects of Benzene: An Update. EPA/600/P-97/001F. Washington, DC: United States Environmental Protection Agency, 1998.
37. Crump KS. Risk of benzene-induced leukemia: a sensitivity analysis of the pliofilm cohort with additional follow-up and new exposure estimates. J Toxicol Environ Health. 1994;42(2):219-42.
38. Crump KS. Risk of benzene-induced leukemia predicted from the Pliofilm cohort. Environ Health Perspect. 1996;104 Suppl 6:1437-41.
39. Paustenbach DJ, Price PS, Ollison W, et al. Reevaluation of benzene exposure for the Pliofilm (rubberworker) cohort (1936-1976). J Toxicol Environ Health. 1992;36(3):177-231.
40. Rinsky RA. Benzene and leukemia: an epidemiologic risk assessment. Environ Health Perspect. 1989;82:189-91.
41. Rinsky RA, Young RJ, Smith AB. Leukemia in benzene workers. Am J Ind Med. 1981;2(3):217-45.
42. Finkelstein MM. Leukemia after exposure to benzene: temporal trends and implications for standards. Am J Ind Med. 2000;38(1):1-7.
43. Schnatter AR, Nicolich MJ, Bird MG. Determination of leukemogenic benzene exposure concentrations: refined analyses of the Pliofilm cohort. Risk Anal. 1996;16(6):833-40.

Chapter 3

A Graphical Tool to Evaluate Temporal Coverage of Occupational History by Exposure Measurements

Jelle Vlaanderen
Wouter Fransman
Brian Miller
Igor Burstyn
Dick Heederik
Fintan Hurley
Roel Vermeulen
Hans Kromhout

Occupational and Environmental Medicine. 2010;67(9):636-8.

Abstract

Introduction In occupational epidemiology, differences in the temporal coverage of the exposure history by available exposure measurement data may affect the uncertainty of exposure estimates. In the reporting of results of studies, greater attention should be paid to the extent to which exposure assessments require extrapolation outside the time frame for which exposure measurements are available. We propose a simple graphical method that can be used to visualize the temporal coverage of exposure history with exposure measurements and the extent of temporal extrapolation needed.

Methods We construct a graph that displays the accumulated work history years for which exposure had to be assessed in each calendar year. Years for which exposure measurements were available are shaded. The proportion of work history years covered by exposure measurements and the proportion of work history years accrued before the first measurements are summarized. When available, the actual number of measurements available in each calendar year is shown.

Results We demonstrate the application of the graphical tool in three nested case-control studies that reported on leukemia in relation to low-level benzene exposures in the petroleum industry. Considerable differences in temporal coverage between the studies were illustrated, which may have resulted in differences in the reliability of the retrospective exposure estimates derived for these studies.

Conclusion We introduce a graphical tool for visualising the temporal coverage by available exposure measurement data in epidemiological studies and encourage others to use similar graphs to derive and share better qualitative insights into the uncertainty in exposure assessment.

Introduction

Recently, it has become more common in occupational epidemiology to assign quantitative exposure estimates to the subjects' work histories (1), which allows for the elaboration of quantitative exposure-response associations. However, it is well known that errors in assigning exposures can have a large impact, in terms of bias and precision, on reported exposure-response associations (2-4).

The intensity of exposure that a worker experiences on any working day depends on many factors, and is subject to considerable variability. Ideally, assessment of exposure history would be based on a set of repeated exposure measurements for all study subjects (here: workers) and for all relevant time-periods in each subject's work history (2). However, due to high costs of exposure measurements and the fact that most occupational studies are performed retrospectively, this is rarely the case. In fact, it is common in occupational epidemiology that reliable exposure measurements are available for only a subset of workers or work locations, and for selected periods of time. Therefore, most exposure assessment methods rely on extrapolation, either through statistical models or using subjective judgments.

Much has been written about the effect of between- and within-worker variation in exposure intensity measurements, and on the attenuation bias this can induce in exposure-response relationships (3). Less attention has been paid to temporal extrapolation, such as may be necessary in studies where cumulative exposure is important, and the exposure history of the study population started years before the measurements were made. Temporal trends in exposure levels have been identified where data are available (5), but are obviously hard to infer reliably in the absence of exposure measurements. One could reasonably assume that the overall uncertainty in exposure assessment will be greatest where the degree of temporal extrapolation is the greatest. Currently, publications that report results in occupational epidemiology using quantitative data provide at best limited insight into the temporal coverage of the exposure history with measurements. This hampers the evaluation of the uncertainty introduced in exposure estimation and subsequent exposure-response estimates, and makes comparisons of reliability between studies and appropriate weighting of pooled-/meta-analyses difficult (6).

During a review (7, 8) of the data quality and consistency in three studies of leukemia and benzene exposure in the petroleum industries from Canada (9, 10), United Kingdom (11, 12) and Australia (13, 14), we devised a graphical tool to provide an insight into the temporal coverage of the exposure history by measurements. We have since developed this tool further, and illustrate its use here with data from the three studies.

Methods: a graphical summary

Whether obtained from employment records or by interview, a subject's work history will typically contain start and finish dates of employment (sometimes in individual jobs throughout the subject's career within a company). From these, years (or months) of work history can be logged against calendar years (or months). 'Work history years' can be thus defined as the years of a subject's tenure for which exposure has to be assessed. We cumulate these across study subjects to give a total number of 'work history years' in each calendar year. These can be plotted against calendar year in the form of a bar graph, in which the height of each bar represents the total number of work history years in a given calendar year.

If we know for which years exposure measurements are available, we can distinguish these years on our graph by shading, as in the upper part of Figure 1. We can summarize numerical aspects of the pattern this produces in a number of ways, such as the proportion of work history years covered by exposure measurements and the proportion of work history years accrued before any measurements were taken. The latter is an indicator of how much extrapolation outside the observational data needs to be done to estimate past exposure levels.

As an aid to interpretation, we can add to the graph a representation of the actual number of measurements available in each year. The lower part of Figure 1 shows one way that this can be done, with the measurement numbers plotted, again as year-specific bars, below the horizontal axis. In some cases, the exact numbers of measurements taken in each year may not be available, but we can adopt a convention to show this: here we have drawn a line around the bar where the figure is for a particular year, but omitted the border where the distribution over a block of years is unknown and the height of the block represents the average number per year.

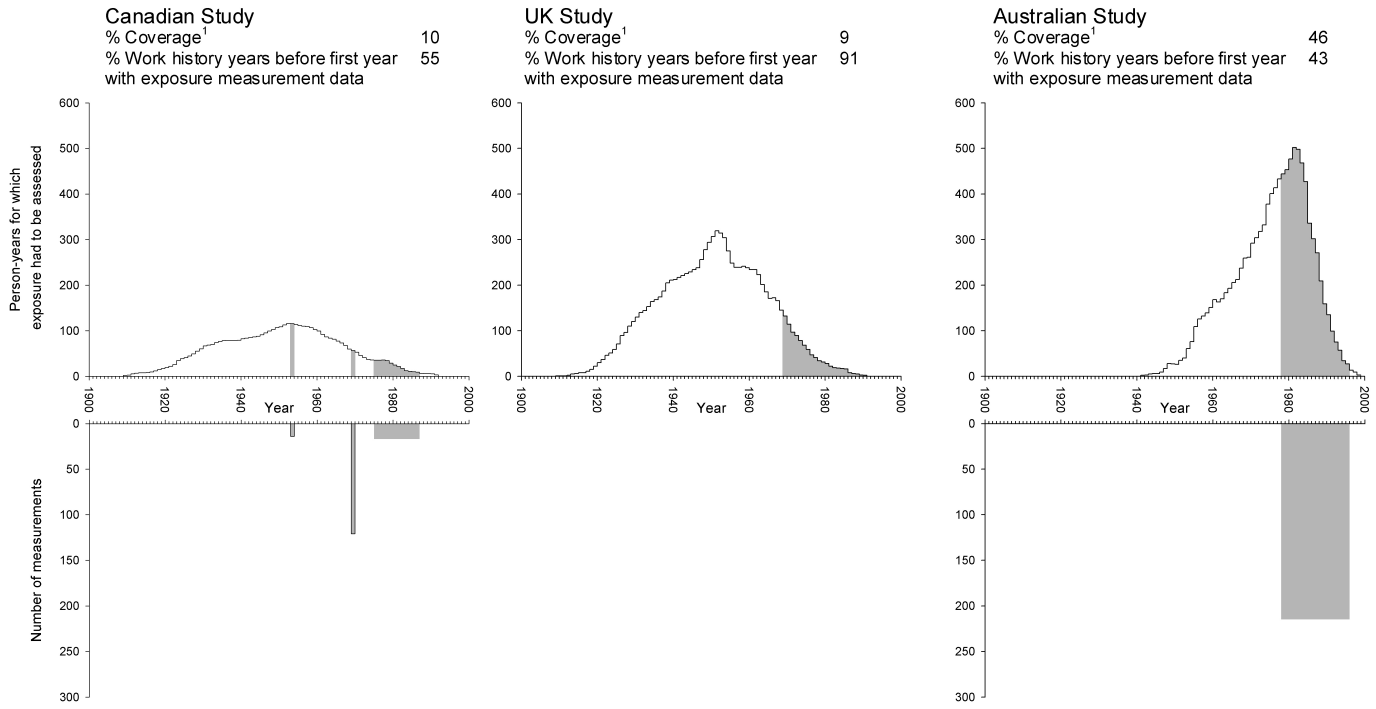


Figure 1 Upper plot is a graphical representation of the temporal distribution of the accumulated work history years in a study. Calendar years for which exposure measurements are available are hatched. The total number of work history years available for each study was 4,557 (Canadian study), 10,258 (UK study), and 10,922 (Australian study). Lower plot is a bar graph of the number of available exposure measurements per calendar year. The Canadian study used 340 exposure measurements, the Australian study used > 3,870 measurements. The number of measurements used in the UK study is unknown.

[1] Coverage of work history years by available exposure measurement data

An example: three studies of benzene and leukemia

As part of a comparative review (7, 8), we devised an early version of the graphs presented here to compare the exposure assessments in three nested case-control studies on leukemia in relation to relation to low-level benzene exposures in the petroleum industry (10, 12, 14). The methods of exposure assessment developed for the Canadian study (9) were adopted by the other two studies working in collaboration (11, 13), so the exposure assessment and assignment methods were, in principle, comparable. The review of the three studies (7, 8) noted that the time periods for which exposure had to be assessed were similar for the studies in Canada (1909-1983) and the UK (1909-1993), but in the Australian study the period to be covered was more recent (1941-1998), and needed less extrapolation backward. The information regarding coverage of the work history years by exposure measurements, as supplied by the principal investigators of the three studies, is summarized in Figure 1. Note that these descriptions do not distinguish between jobs within the industry. As seen from the graph, although the total number of work history years in the UK and Australia were similar, in Australia they were distributed over a more recent period. The Canadian graph is much shallower, reflecting the smaller number of total work history years. For the UK study, only the temporal coverage of work history years with exposure measurements was provided, and no information was available on the number of measurements used for exposure assessment, either by year or in total. The temporal coverage of total work history by exposure measurements was larger for Australian study (46%) than for the Canadian (10%) and UK (9%) studies. The proportion of work history years before the first year with exposure measurements ranged from 91% in the UK study to 55% in the Canadian study and 43% in the Australian study. The figure for Canada would be much higher (87%) without the small proportion of measurements (n = 14, 4%) taken in 1953, many years before the other measurements. Given the similarities between studies in methods for exposure assessment and assignment these results suggest that the uncertainty in exposure assessment in the Canadian and UK studies was larger than in the Australian study.

Discussion

Although differences in the temporal coverage of the exposure history by available exposure measurement data may affect the degree of uncertainty in quantitative exposure estimates that are used in occupational epidemiology, they generally remain unnoticed in the evaluation of epidemiological results. We believe they should be given greater prominence, and have developed a tool that visualizes coverage of exposure history by measurements. Our method can be readily adapted to any situation where temporal coverage of work history by measurements is important, for example at the level of jobs within a particular industry or workplace.

The numerical summary measures we show conceal complexity that may be relevant to uncertainty analysis. For several periods in our example, only the cumulative number of exposure measurements during a certain period was known, and we were not able to assess the exact number of measurements per calendar year. Therefore, we graphed the annual average. The summary value of the proportion of work history years before any measurements is rather sensitive to the appearance of isolated early measurements, as was demonstrated in the Canadian data. Another option would be to count coverage only for years in which a specified minimum number of measurements was available. The graphical tool should be seen as an addition to the evaluation of other factors that may affect the degree of uncertainty in quantitative exposure estimates that are used in occupational epidemiology, such as the precision and accuracy of the assessment of the within- and between-worker variation in exposure intensity (3).

Conclusion

We present a simple tool that can assist in reporting the extent to which exposure assessment relies on extrapolation outside the time frame for which measurements of exposure are available. Various adaptations of the tool we introduce could be developed and we offer it here in the hope of stimulating such reporting.

Acknowledgements

The authors are indebted to the principal investigators who supplied data for use in the comparative review of their studies of low-level benzene exposure and leukemia. The comparative review of the three nested case-control studies was funded by CONCAWE (Boulevard du Souverain 165, B-1160, Brussels, Belgium). Development of the methods described here was partly supported by ECNIS (European Cancer Risk, Nutrition and Individual Susceptibility), a Network of Excellence operating within the European Union 6th Framework Program, Priority 5: “Food Quality and Safety” (Contract No 513943). Igor Burstyn was supported by the Population Health Investigator salary award from the Alberta Heritage Foundation for Medical Research.

References

1. Kriebel D, Checkoway H, Pearce N. Exposure and dose modelling in occupational epidemiology. Occup Environ Med. 2007;64(7):492-8.
2. Heederik D, Attfield M. Characterization of dust exposure for the study of chronic occupational lung disease: a comparison of different exposure assessment strategies. Am J Epidemiol. 2000;151(10):982-90.

3. Loomis D, Kromhout H. Exposure variability: concepts and applications in occupational epidemiology. Am J Ind Med. 2004;45(1):113-22.
4. Tielemans E, Kupper LL, Kromhout H, et al. Individual-based and group-based occupational exposure assessment: some equations to evaluate different strategies. Ann Occup Hyg. 1998;42(2):115-9.
5. Kromhout H, Vermeulen R. Long-term trends in occupational exposure: Are they real? What causes them? What shall we do with them? Ann Occup Hyg. 2000;44(5):325-7.
6. Vlaanderen J, Vermeulen R, Heederik D, et al. Guidelines to evaluate human observational studies for quantitative risk assessment. Environ Health Perspect. 2008;116(12):1700-5.
7. Miller BG, Fransman W, Heederik D, et al. A review of the data quality and comparability of case-control studies of low-level exposure to benzene in the petroleum industry. Int Arch Occup Environ Health. 2010;831:69-76.
8. Miller BG, Fransman W, Heederik D, et al. A review of the data quality and comparability of case-control studies of low-level exposure to benzene in the petroleum industry (IOM Report TM/05/04). Edinburgh: Institute of Occupational Medicine. http://www.iom-world.org/pubs/IOM_TM0504.pdf, 2005.
9. Armstrong TW, Pearlman ED, Schnatter AR, et al. Retrospective benzene and total hydrocarbon exposure assessment for a petroleum marketing and distribution worker epidemiology study. Am Ind Hyg Assoc J. 1996;57(4):333-43.
10. Schnatter AR, Armstrong TW, Nicolich MJ, et al. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med. 1996;53(11):773-81.
11. Lewis SJ, Bell GM, Cordingley N, et al. Retrospective estimation of exposure to benzene in a leukaemia case-control study of petroleum marketing and distribution workers in the United Kingdom. Occup Environ Med. 1997;54(3):167-75.
12. Rushton L, Romaniuk H. 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. Occup Environ Med. 1997;54(3):152-66.
13. Glass DC, Adams GG, Manuell RW, et al. Retrospective exposure assessment for benzene in the Australian petroleum industry. Ann Occup Hyg. 2000;44(4):301-20.
14. Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology. 2003;14(5):569-77.

Chapter 4

Occupational Benzene Exposure and the Risk of Lymphoma Subtypes: A Meta-Analysis of Cohort Studies Incorporating Three Study Quality Dimensions

Jelle Vlaanderen
Qing Lan
Hans Kromhout
Nathaniel Rothman*
Roel Vermeulen*

* Co-senior authors

Environmental Health Perspectives. 2011;119(2):159-167

Abstract

Background The use of occupational cohort studies to assess the association of benzene and lymphoma is complicated by problems with exposure misclassification, outcome classification, and low statistical power.

Objective We performed meta-analyses of occupational cohort studies for five different lymphoma categories: Hodgkin lymphoma (HL), non-Hodgkin lymphoma (NHL), multiple myeloma (MM), acute lymphocytic leukemia (ALL), and chronic lymphocytic leukemia (CLL).

Data extraction We assessed three study quality dimensions to evaluate the impact of study quality variations on meta-relative risks (mRR): stratification by the year-of-start of follow-up, stratification by the strength of the reported acute myelogenous leukemia association, and stratification by the quality of benzene exposure assessment.

Data synthesis mRRs for MM, ALL, and CLL increased with increasing study quality, regardless of the study quality dimension. mRRs for NHL also increased with increasing study quality, although this effect was less pronounced. We observed no association between occupational benzene exposure and HL.

Conclusions Our meta-analysis provides support for an association between occupational benzene exposure and risk of MM, ALL, and CLL. The evidence for an association with NHL is less clear, but this is likely complicated by the etiologic heterogeneity of this group of diseases. Further consideration of the association between benzene and NHL will require delineation of risks by NHL subtype.

Introduction

The International Agency for Research on Cancer (IARC) classified benzene as a group 1 carcinogen (‘*carcinogenic to humans’)* in its 1982 and 1987 evaluations (1, 2) based primarily on reports of an association between occupational exposure to benzene and leukemia, particularly acute non-lymphocytic leukemia (ANLL), which consists primarily of acute myelogenous leukemia (AML). Recently, IARC updated its previous reviews of several chemicals and occupational exposure circumstances, including benzene, to reassess carcinogenicity and to consider potential associations with additional tumor sites (3). In that review, IARC determined for the first time that in addition to the confirmed association with ANLL, there was also limited evidence that benzene causes acute lymphocytic leukemia (ALL), chronic lymphocytic leukemia (CLL), non-Hodgkin lymphoma (NHL), and multiple myeloma (MM) in humans (3). At the same time, in recent years, there has been a plethora of reviews and meta-analyses of benzene and one or more lymphoid neoplasms, at times reaching diametrically opposed conclusions (4-18).

There are two fundamental challenges in using the large number of occupational cohort studies that have been published over the last 30 or so years when considering the relationship between occupational benzene exposure and the risk of lymphoid neoplasms. First, there have been substantial changes in testing procedures, diagnostic criteria, and categorization of lymphoid neoplasms over the last half century (19-21), the time-period in which follow-up of these occupational cohorts took place. Indeed, diagnostic criteria that were used in these cohort studies were based on a range of classification strategies including the International Classification of Diseases (ICD)-7, ICD-8, ICD-9, and ICD-O3. The changing views on the categorization of lymphoid neoplasms is illustrated by the current categorization of ALL and CLL as subtypes of NHL in the most recent World Health Organization (WHO) disease classification (22), although these entities have been reported separately from NHL in essentially all occupational cohort studies of benzene-exposed workers. Second, there is heterogeneity in occupational cohort studies with regard to industry, sample size, documentation and level of benzene exposure, and documentation of the percent of a given cohort that had true, nontrivial exposure to benzene. Inadequate documentation, uncertain quality of follow-up, and, most problematic, potential inclusion of ‘unexposed’ workers in ‘exposed’ categories would have likely resulted in attenuation of the observed associations. Further, for the purpose of reviews or meta-analyses, it can be challenging to separate out informative from potentially non informative cohorts in the face of uncertain documentation of key epidemiological study design and exposure assessment characteristics.

Given the changing nature of the diagnosis of lymphoid neoplasms over time and the heterogeneity of occupational benzene cohort study quality in the literature, it is a challenge to discern the nature of the relationship between benzene and lymphoid neoplasms. To

address this issue we developed three strategies that we employ in a set of meta-analyses of occupational cohort studies for five lymphoma categories defined according to ICD-9: Hodgkin lymphoma (HL) (ICD-9: 201), NHL (ICD-9: 200,202), MM (ICD-9: 203.0), ALL (ICD-9: 204.0), and CLL (ICD-9: 204.1).

We applied the first strategy to assess the potential impact of the gradual increase in the quality of hematological diagnoses over the last decades. This strategy involved stratification of the studies in the meta-analyses based on the reported *start of follow-up*. We used the year 1970 as a cut-off point for stratification (approximate midpoint of follow-up of all studies included in this analysis). We based the second strategy on the established strong association between benzene and AML. We argue that any study that was not able to detect at least a suggestive association between benzene and AML most likely had serious methodological limitations in one or more aspects of study design. Examples of possible limitations are trivial exposure to benzene in the studied cohort, inclusion of 'unexposed' workers in 'exposed categories' or flaws in the assessment (or categorization) of health effects (23). Therefore, we used the direction and significance level of a reported association between benzene and AML as proxies for the overall study quality (*AML significance level).*

We based the third strategy on the evaluation of the quality of the exposure assessment carried out in each cohort. High-quality exposure assessment is essential to discriminate exposed individuals from nonexposed individuals (24). We assigned an *exposure assessment quality* classification to each study based on an *a priori* defined classification scheme and used this classification as an additional proxy of study quality, reasoning that those cohort studies with the highest quality exposure assessment had the greatest ability to identify and include workers who were truly exposed to benzene in their analyses.

We hypothesized that application of the three study quality dimensions—stratification based on the *start of follow-up*, *AML significance level*, and *exposure assessment quality*—would identify a subgroup of occupational cohort studies that is most informative for the evaluation of the possible association between benzene and lymphoid neoplasms.

Methodology

Study identification and data extraction

We conducted a search of Pubmed (http://www.ncbi.nlm.nih.gov/sites/entrez) using the keywords "benzene", and "cohort" or "case-control". We included publications in the meta-analysis if they were published in the peer-reviewed literature, reported results for any of the five lymphoma subtypes (HL, NHL, MM, ALL, and CLL), and were conducted in the occupational setting. We checked references in all identified publications for additional studies. When more than one paper was published on the same cohort, we chose the publication with the highest quality exposure assessment (e.g., in the Australian petroleum

workers cohort for AML we preferred the nested-case control study that included an elaborated exposure assessment approach (25) over a more recent update on the full cohort that included no detailed benzene exposure assessment (26)). When multiple publications with similar *exposure assessment quality* were published on the same cohort, we chose the most recent update (with the longest follow-up time). In this meta-analysis we pooled risk ratios, odds ratios (ORs), and standardized mortality ratios (SMRs). ORs and SMRs can be interpreted as reasonable approximations of the risk ratio when the disease is rare, and these measures have been pooled with risk ratios for meta-analyses before (27). We use the term "relative risk" (RR) to refer to either the risk ratio, the OR, or the SMR. We extracted RRs based both on incidence and mortality. However, if a publication reported both, we chose incidence over mortality in the meta-analysis.

Risk estimates

To allow the inclusion of studies without quantitative exposure assessment in our analysis , we used only RRs for "any occupational benzene exposure" versus "background benzene exposure" in the meta-analyses. If publications only reported RRs stratified for cumulative exposure and not for "any occupational benzene exposure" versus "background benzene exposure", we pooled RRs by summing observed and expected cases for studies that reported SMRs (percentage of RRs: AML, 4.8%; HL, 3.7%; NHL, 3.0%; MM, 3.8%; CLL, 5.6%), or by conducting a within-study random effects meta-analysis of the non-reference exposure groups for studies that reported RRs or ORs (percentage of RRs: AML, 14.3%; NHL, 3.0%; MM, 7.7%; ALL, 5.9%; CLL, 16.7%). If publications reported only observed and expected number of cases and no RRs, we calculated RRs and estimated associated confidence intervals with mid-P exact (28) (percentage of RRs: AML, 4.8%; HL, 7.4%; ALL, 17.6%). For publications that reported no observed cases for any of the lymphoma subtypes, we calculated continuity corrected RRs (observed and expected number of cases plus one) and we estimated associated confidence intervals with mid-P exact (percentage of the RRs: ALL, 11.8%; HL, 11.1%). If studies reported zero for the lower confidence interval, a value of 0.1 was imputed to allow estimation of the variance (percentage of RRs: ALL, 5.9%; MM, 3.8%).

Three strategies for the assessment of study quality dimensions

We stratified by the *start of follow-up* based on the information provided in the included publications (follow-up started before 1970 versus follow-up started in 1970 or later). The median start of follow-up in the stratum with studies that started follow-up before 1970 was 1947, and the median start of follow-up in the stratum with studies that started follow-up in 1970 or later was 1973.

We assigned *AML significance level* to each publication based on a two-sided *p-value* of the *z*-score, which we estimated by dividing the reported log RR for AML by its standard error. Based on the calculated *AML significance level,* we assigned one of the following categories

(A-E) to each publication: A, AML RR > 1, $p < 0.1$; B, AML RR > 1, $0.1 \leq p < 0.2$; C, AML RR > 1, $p \geq 0.2$; D, AML RR ≤ 1; E, no AML RR reported.
We assigned *exposure assessment quality* (A-D) to each publication as follows: A, in the publication explicit quantitative exposure estimates for benzene were reported; B, in the publication semi-quantitative estimates of benzene exposure or quantitative estimates of exposures containing benzene (e.g., gasoline) were reported; C, in the publication some industrial hygiene sampling results to indicate that benzene exposure was present in the cohort that was studied were reported; D, the publication qualitatively indicated that benzene exposure was present in the cohort.

Statistical analyses
We conducted random-effects meta-analyses to pool the RRs reported in the included publications. We used an α of 0.05 to assess whether meta-relative risks (mRRs) were significantly elevated. We conducted the first set of meta-analyses on the full set of studies stratified for the *start of follow-up* (follow-up started before 1970 vs. follow-up started in 1970 or later). We compared mRRs by strata using a test of interaction as suggested by Altman and Bland (29).
We applied the study quality dimensions *AML significance level* and *exposure assessment quality* in two series of meta-analyses. The initial analysis in each series included all studies regardless of quality. In each subsequent analysis, we excluded the group of studies with the lowest *AML significance level* or the lowest *exposure assessment quality*.
We used Cochran's Q test to assess between study heterogeneity in all meta-analyses. A *p*-value < 0.1 was considered to be statistically significant evidence for between study heterogeneity. We used I^2 to describe the percentage of total variation across studies that was due to heterogeneity rather than chance (30). For analyses that displayed significant between study heterogeneity, we assessed the sensitivity of the outcome of the meta-analysis for individual studies by excluding studies one at the time (jackknife analysis). We assessed publication bias with Egger's graphical test (31). We performed all meta-analyses in Stata (version 11; StataCorp LP, College Station, TX, USA).

Results

We identified 44 publications that provided a RR for at least one of the lymphoma subtype-specific meta-analyses. We did not extract data from three publications: one with likely under-ascertainment of cancer deaths as the result of the inability to identify the type of cancer for a number of cancer deaths (32, 33); one study for which we could not estimate the RR variance (a nested case-control study that did not report confidence intervals (34); and one study that reported proportionate mortality ratios, which tend to underestimate the RR (35). Table 1 lists all publications that contributed to the meta-analyses, their (assigned)

cohort name, the (assigned) name of the subcohort (if relevant), the literature reference, the type of industry in which the study was performed, the follow-up period, the lymphoma subtype for which the publication was included (if reported with ICD code and revision), an indicator whether RRs were based on incidence or mortality, the assigned *AML significance level,* and the assigned *exposure assessment quality*. The earliest included publication dates from 1983, and the most recent publication was from 2008. For two cohorts we used non-peer-reviewed publications to extract RRs for MM (and NHL) that were not reported in the peer-reviewed publications (36, 37). Both reports were based on the exact same methodology and follow-up time as reports of these cohorts that appeared in the peer-reviewed literature (25, 38). We included an RR for MM from a study by Decoufle et al. (1983) based on additional information that was reported in the preamble to the final OSHA Benzene Standard of 1987 (39, 40). We extracted NHL RRs for two studies by Wong and colleagues (41, 42) from Wong (1998), a letter that provided results from additional analyses for these studies (43). Finally, there might have been a slight (nonidentifiable) overlap in the cohorts studied by Wong (41, 44) and Collins et al. (45).

Table 2 shows the mRR based on random-effect meta-analyses for all studies and stratified by the *start of follow-up* for AML and the five lymphoma subtypes (i.e., HL, NHL, MM, ALL, and CLL). The overall mRRs (95% CIs) for AML and ALL were significantly increased (mRR=1.68 (1.35-2.10) and mRR=1.44 (1.03-2.02)). The overall mRR for MM and CLL were slightly but not significantly elevated, whereas the overall mRRs for HL and NHL were close to unity. Stratified analyses by start of follow-up showed higher RRs for AML, NHL, and CLL for studies with a follow-up starting in 1970 or later than for studies that started the follow-up before 1970 ($p < 0.10$). We observed no significant difference in mRR between the follow-up strata for HL, MM, and ALL. We observed significant between-study heterogeneity for AML, NHL, and CLL overall and in the studies with *start of follow-up* before 1970 (see Supplemental Material, Figure 1). Exclusion of the most influential studies/RRs (based on the distance of the RR to the mRR and the weight of the study) resulted in mRRs that were essentially similar (data not shown).

Table 1 Overview of publications included in the meta-analyses.

Cohort	Sub-cohort	Reference	Industry	Follow-up period	Included for outcomes (ICD[a])	ICD revision[a,b]	I / M[c]	AML Significance level[d]	Exposure assessment quality[e]
Australian petroleum workers cohort		(37)[f]	Petroleum industry	1980-1998	MM(203)	9	I	C	A
Australian petroleum workers cohort		(25)	Petroleum industry	1980-1998	AML(205.0, 208.0), CLL(204.1)	9	I	C	A
Australian petroleum workers cohort		(26)	Petroleum industry	1981-1999	ALL(204.0), NHL(200-202)	9	I	C	D
Beaumont, Texas petroleum refinery cohort		(70)	Petroleum industry	1945-1996	ALL(204.0), AML(205.0), CLL(204.1), MM(203), NHL(200, 202), HL(*)	8	M	C	D
Canadian petroleum company cohort		(52)	Petroleum industry	1964-1983	MM(203), NHL(200, 202.0, 202.1, 202.2, 202.9)	8	M	E	A
Canadian petroleum company cohort		(71)	Petroleum industry	1964-1994	ALL(204.0), AML(205.0), CLL(204.1), HL(201)	9	I	D	D
Caprolactam workers		(64)	Chemical industry	1951-2001	MM(31), HL(29)	[b]	M	E	A
Conoco chemical plant cohort		(39)	Chemical industry	1947-1977	MM(*)	8	M	E	D
Dow cohort		(51)	Chemical industry	1940-1996	AML(205.0, 206.0, 207.0), CLL(204.1), MM(203), NHL(200.0-200.8, 202.0, 202.8), HL(201)	9	M	C	A
Exxon cohort	Louisiana	(72)	Petroleum industry	1970-1997	ALL(204.0), AML(205.0, 206.0, 207.0, 207.2), CLL(204.1), MM(203.0), NHL(200.0-200.2, 200.8, 202.0-202.2, 202.8-202.9), HL(201)	9	M	B	D
Exxon cohort	Texas	(72)	Petroleum industry	1970-1997	ALL(204.0), AML(205.0, 206.0, 207.0, 207.2), CLL(204.1), MM(203.0), NHL(200.0-200.2, 200.8, 202.0-202.2, 202.8-202.9) HL(201)	9	M	A	D
Finnish oil refinery workers		(73)	Petroleum industry	1967-1994	NHL(*), HL(*)	*	I	E	D
French gas and electric utility workers		(74)	Gas and electric utility industry	1978-1989	ALL(*), AML(*), CLL (*)	O	I	D	B
Italian oil refinery		(75)	Petroleum industry	1949-1991	NHL(200, 202), HL(201)	8	M	E	D

Table 1 continued Overview of publications included in the meta-analyses.

Cohort	Sub-cohort	Reference	Industry	Follow-up period	Included for outcomes (ICD[a])	ICD revision[a,b]	I / M[c]	AML Significance level[d]	Exposure assessment quality[e]
Martinez and Wilmington refineriy and petrochemical plants California		(76)	Petroleum industry	1973-1989	NHL(200), HL(201)	8	M	E	D
Monsanto cohort		(45)	Chemical industry	1940-1999	AML(205.0, 206.0), CLL(204.1), MM(203), NHL(200, 202), HL(201)	8	M	A	A
NCI-CAPM		(65)	Multiple industries	1972-1987	ALL(*), MM(*)	9	I	A	A
NCI-CAPM		(50)	Multiple industries	1972-1987	AML(205.0, 206.0, 207.0), NHL(200, 202)	9	I	A	A
Norway upstream petroleum industry		(77)	Petroleum industry	1981-2003	ALL(*), AML(*), CLL(*), MM(*), NHL(*) HL(*)	7	I	A	D
Paulsboro, New Jersey refinery (Mobil)		(78)	Petroleum industry	1946-1987	NHL(200), HL(201)	8	M	E	D
Petrochemical workers Texas City		(79)	Petroleum industry	1941-1977	NHL(200), HL(201)	7	M	E	D
Petroleum manufacturing plant Illinois, USA (Shell)		(80)	Petroleum industry	1973-1982	ALL(*), AML(*), CLL(*)	8	M	A	D
Petroleum manufacturing plant Illinois, USA (Shell)		(81)	Petroleum industry	1940-1989	NHL(*), HL(*)	9	M	E	D
Pliofilm cohort		(82)	Chemical industry	1940-1987	AML(C)	c	M	A	A
Pliofilm cohort		(53)	Chemical industry	1950-1996	NHL(C), MM(C)	c	M	E	A
Port Arthur refinery workers		(83)	Petroleum industry	1937-1987	ALL(204.0), AML(205.), CLL(204.1), MM(203), NHL(200, 202), HL(201)	8	M	D	D
Richmond and El Segundo refineries		(84)	Petroleum industry	1950-1986	NHL(200), HL(201)	8	M	E	D
Sample of US refineries		(85)	Petroleum industry	1972-1980	MM(*), HL(*)	*	M	E	D
Service station workers in Nordic countries		(86)	Service station workers	1970-1990	AML(*), CLL(*), MM(203), NHL(200, 202), HL(201)	7	I	C	C
Shell Deer Park refinery		(87)	Petroleum industry	1948-1989	NHL(200), HL(201)	8	M	E	D
Shell Louisiana refinery		(88)	Petroleum industry	1973-1999	NHL(200), HL(201)	8	M	E	D
Shoe workers cohort	Italian cohort	(89)	Shoe workers	1950-1990	MM(203), NHL(200, 202)	9	M	E	D

Table 1 continued Overview of publications included in the meta-analyses.

Cohort	Sub-cohort	Reference	Industry	Follow-up period	Included for outcomes (ICD[a])	ICD revision[a,b]	I / M[c]	AML Significance level[d]	Exposure assessment quality[e]
Shoe workers cohort	UK cohort	(89)	Shoe workers	1939-1991	MM(203), NHL(200, 202)	9	M	E	D
Swedish seamen working on product or chemical tankers		(90)	Petroleum tanker workers	1971-1978	MM(203), NHL(200, 202), HL(201)	8	I	E	D
Texaco crude oil workers		(46)	Petroleum workers	1946-1994	ALL(*), AML(*), CLL(*), MM(*), NHL(*), HL(201)	8	M	A	D
Texaco mortality study		(91)	Petroleum industry	1947-1993	HL(201)	8	M	C	D
Texaco mortality study		(92)	Petroleum industry	1947-1993	ALL(*), AML(*), CLL(*), MM(*), NHL(*)	8	M	C	D
Torrance, California petroleum refinery		(93)	Petroleum industry	1959-1997	ALL(204.0), AML(205.0), CLL(204.1), MM(203), NHL(200, 202), HL(*)	8	M	D	D
U.K. oil distribution and oil refinery workers	Refinery	(94)	Petroleum industry	1950-1989	ALL (*)	[a]	M	D	D
U.K. oil distribution and oil refinery workers	Distribution	(94)	Petroleum industry	1950-1989	ALL (*)	[a]	M	C	D
U.K. oil distribution and oil refinery workers		(47)	Petroleum industry	1950-1993	AML(*), CLL (*)	9	M	A	A
U.K. oil distribution and oil refinery workers	Refinery	(95)	Petroleum industry	1951-2003	MM (203), NHL(200, 202), HL(201)	9	M	E	D
U.K. oil distribution and oil refinery workers	Distribution	(95)	Petroleum industry	1951-2003	MM (203), NHL(200, 202), HL(201)	9	M	E	D
Union Oil Company cohort	Oil and gas division	(36)[f]	Petroleum industry	1976-1990	MM(203), NHL(200, 202)	9	M	A	D
Union Oil Company cohort	Refining division	(36)[f]	Petroleum industry	1976-1990	MM(203), NHL(200, 202)	9	M	E	D
Union Oil Company cohort	Oil and gas division	(38)	Petroleum industry	1976-1990	AML(*)	9	M	A	D
U.S. chemical workers		(41)	Chemical industry	1946-1977	HL(201)	8	M	E	A
U.S. chemical workers		(44)	Chemical industry	1946-1977	MM(203)	8	M	E	A
U.S. chemical workers		(43)	Chemical industry	1946-1977	NHL (200, 202)	8	M	E	A

Table 1 continued Overview of publications included in the meta-analyses.

Cohort	Sub-cohort	Reference	Industry	Follow-up period	Included for outcomes (ICD[a])	ICD revision[a,b]	I / M[c]	AML Significance level[d]	Exposure assessment quality[e]
U.S. gasoline distribution employees	Land-based	(42)	Petroleum industry	1946-1986	ALL(*), AML(*), CLL(*)	8	M	B	B
U.S. gasoline distribution employees	Marine	(42)	Petroleum industry	1946-1986	ALL(*), AML(*), CLL(*)	8	M	D	B
U.S. gasoline distribution employees	Land-based and marine	(43)	Petroleum industry	1946-1986	NHL (200, 202)	8	M	C	B

NCI-CAPM, National Cancer Institute-Chinese Academy of Preventive Medicine

[a] (*), ICD revision or specific ICD code was not reported.

[b] A, deaths were coded according to a system developed by Statistics Netherlands (CBS); B, deaths were coded according to the National Institute for Occupational Safety and Health life-table analysis system death categories; C, deaths were coded according to ICD in effect at time of death.

[c] I, Incidence; M, mortality

[d] A, AML RR > 1, $p < 0.1$; B, AML RR > 1, $0.1 \leq p < 0.2$; C, AML RR > 1, $p \geq 0.2$; D, AML RR ≤ 1; E, AML RR not reported.

[e] A, quantitative exposure estimates for benzene; B, semi-quantitative estimates of benzene exposure or quantitative estimates of exposures containing benzene; C, some industrial hygiene sampling results; D, qualitative indication that benzene exposure had occurred.

[f] Non-peer reviewed publication.

Table 2 mRRs (95% CIs) for AML and five lymphoma subtypes in cohort studies of workers exposed to benzene: stratification by *start of follow-up*.

Lymphoma subtype	All studies			Start follow-up before 1970			Start follow-up 1970 and later			Test for difference by follow-up strata (p-value)[c]
	N[a]	n[b]	mRR	N[a]	n[b]	mRR	N[a]	n[b]	mRR	
AML	21	217	1.68 (1.35-2.10)*	12	119	1.43 (1.07-1.92)*	9	98	2.08 (1.59-2.72)	0.06
HL	27	146	0.99 (0.83-1.19)	19	123	1.01 (0.83-1.23)	8	23	0.91 (0.59-1.40)	0.67
NHL[d]	33	647	1.00 (0.89-1.13)*	22	452	0.93 (0.81-1.06)*	11	195	1.21 (0.94-1.55)*	0.07
MM	26	284	1.12 (0.98-1.27)	16	204	1.07 (0.93-1.24)	10	80	1.26 (0.92-1.71)	0.35
ALL	17	47	1.44 (1.03-2.02)	10	30	1.30 (0.88-1.92)	7	17	1.92 (1.00-3.67)	0.31
CLL	18	111	1.14 (0.78-1.67)*	11	69	0.87 (0.50-1.50)*	7	42	1.63 (1.09-2.44)	0.07

[a] Number of studies.

[b] Number of exposed cases.

[c] Test of interaction (29).

[d] NHL or lymphosarcoma/reticulosarcoma (preferred NHL if the study reported both).

* Significant evidence for between study heterogeneity ($p < 0.1$).

Table 3 mRRs (95% CIs) for AML and five lymphoma subtypes in cohort studies of workers exposed to benzene: stratification by *AML significance level.*

Lymphoma subtype	AML significance level[a]	N[b]	n[c]	mRR
AML	A-E (all studies)	21	217	1.68 (1.35-2.10)*
	A-D	21	217	1.68 (1.35-2.10)*
	A-C	16	192	1.88 (1.56-2.27)
	A-B	11	132	2.20 (1.77-2.72)
	A	9	108	2.48 (1.94-3.18)
HL	A-E (all studies)	27	146	0.99 (0.83-1.19)
	A-D	12	69	0.99 (0.77-1.27)
	A-C	9	39	0.82 (0.59-1.15)
	A-B	5	7	0.47 (0.22-0.99)
	A	4	7	0.50 (0.23-1.08)
NHL[d]	A-E (all studies)	33	647	1.00 (0.89-1.13)*
	A-D	15	383	0.97 (0.81-1.16)*
	A-C	13	344	0.99 (0.81-1.21)*
	A-B	7	130	1.21 (0.85-1.72)*
	A	6	101	1.16 (0.77-1.76)*
MM	A-E (all studies)	26	284	1.12 (0.98-1.27)
	A-D	14	160	1.15 (0.95-1.40)
	A-C	12	137	1.19 (0.94-1.49)
	A-B	7	69	1.49 (1.13-1.95)
	A	6	56	1.56 (1.11-2.21)
ALL	A-E (all studies)	17	47	1.44 (1.03-2.02)
	A-D	17	47	1.44 (1.03-2.02)
	A-C	11	29	1.41 (0.90-2.19)
	A-B	7	16	1.74 (0.90-3.36)
	A	5	12	1.74 (0.77-3.90)
CLL	A-E (all studies)	18	111	1.14 (0.78-1.67)*
	A-D	18	111	1.14 (0.78-1.67)*
	A-C	13	93	1.19 (0.74-1.90)*
	A-B	8	57	1.37 (0.73-2.56)*
	A	6	45	1.39 (0.65-2.96)*

[a] A,AML RR > 1, $p < 0.1$; B, AML RR > 1, $0.1 \leq p < 0.2$; C, AML RR > 1, $p \geq 0.2$; D, AML RR ≤ 1; E, AML RR not reported.
[b] Number of studies.
[c] Number of exposed cases.
[d] NHL or lymphosarcoma/reticulosarcoma (preferred NHL if the study reported both).
* Significant evidence for between study heterogeneity ($p < 0.1$).

Table 3 shows mRRs based on random-effects meta-analyses stratified by *AML significance level* for AML, HL, NHL, MM, ALL, and CLL. As could be expected, the lymphoma mRRs based on only the studies that reported a RR for AML (A-D) are largely similar to the mRRs based on all the studies (A-E). These studies provide therefore a relatively unbiased representation of the full set of studies. All outcomes except HL demonstrated an increase in mRRs with

increasing *AML significance level.* However, the 95% confidence intervals successively widened as a result of the reduced number of studies/RRs that were retained with each increase in *AML significance level.* The increase in mRR was most pronounced for MM and ALL, and somewhat weaker for NHL and CLL. In contrast, the mRR for HL dropped with increasing *AML significance level.* We observed significant between-study heterogeneity for NHL and CLL in the subset of studies with *AML significance level* A ($p < 0.10$) (See Supplemental Material, Figure 2). Jackknife analysis eliminating one study at the time demonstrated that, in the NHL analysis of the studies with *AML significance level* A, the RRs from Divine (46) and Delzell (36) had considerable impact on the between-study heterogeneity. Exclusion of both RRs from this analysis resulted in a slight decrease in the mRR from 1.16 (0.77-1.76) to 1.12 (0.77-1.61) with an I^2 (an estimate of the percentage of total variation across studies that was due to heterogeneity rather than chance) of 22.8% (p=0.27). In the CLL analysis of the studies with *AML significance level* A, the RRs provided by Divine and Rushton(46, 47) appeared to be primarily responsible for the observed between-study heterogeneity. Exclusion of both RRs from the meta-analysis resulted in a slight decrease in the mRR from 1.39 (0.65-2.96) to 1.26 (0.65-2.43) with an I^2 of 0% (p = 0.94).

Table 4 shows mRRs based on random-effects meta-analyses and stratified by *exposure assessment quality*. mRRs for NHL, MM and CLL increased with increasing *exposure assessment quality.* The increase in mRR was most pronounced for MM and CLL. Forest plots for AML and the five lymphoma subtypes for all studies with *exposure assessment quality* A and B (A, quantitative exposure estimates for benzene; B, semi-quantitative estimates of benzene exposure or quantitative estimates of exposures containing benzene) are shown in the Supplemental Material (See Supplemental Material, Figure 3). Jackknife analysis eliminating one study at the time demonstrated that in the set of studies with *exposure assessment quality* A and B, the RRs provided by Wong et al. (land-based cohort) (42) and Rushton and Romaniuk (47) had considerable impact on the observed between-study heterogeneity in the CLL analysis. Exclusion of both RRs from the meta-analysis resulted in a slight decrease in the mRR (95% CI) from 1.54 (0.72-3.31) to 1.46 (0.79-2.72), with an I^2 of 0% (p=0.43). The RR provided by Wong (gasoline distribution employees) (43) had a considerable impact on the observed between- study heterogeneity in the NHL analysis of the set of studies with *exposure assessment quality* A and B. Exclusion of this RR resulted in a slight increase in the mRR from 1.04 (0.63-1.72) to 1.27 (0.90-1.79) (I^2 = 0%, p=0.78).

Cross-stratification of *AML significance level* and *exposure assessment quality* with the stratification based on the *start of follow-up*, although limited by a loss of statistical power, showed that mRR patterns with increasing *AML significance level* and *exposure assessment quality* (See Supplemental Material, Tables 1 and 2, respectively) were generally consistent with the patterns observed when meta-analyses were stratified by *start of follow-up* (Table 2).

Egger's test revealed no significant evidence for publication bias in the data available for AML, HL, NHL, ALL, or CLL (See Supplemental Material, Figure 4). We observed evidence for bias for MM (p =0.03), but Egger's test became non-significant after exclusion of all *exposure assessment quality* D studies (p=0.72).

Table 4 mRRs for AML and five lymphoma subtypes in cohort studies of workers exposed to benzene; stratification by *exposure assessment quality.*

Lymphoma subtype	Exposure assessment quality[a]	N[b]	n[c]	mRR
AML	A-D (all studies)	21	217	1.68 (1.35-2.10)*
	A-C	10	108	1.73 (1.26-2.38)
	A-B	9	95	1.82 (1.25-2.66)
	A	6	71	2.32 (1.55-3.47)
HL	A-D (all studies)	27	146	0.99 (0.83-1.19)
	A-C	5	16	0.99 (0.58-1.71)
	A-B	4	6	0.98 (0.36-2.67)
	A	4	6	0.98 (0.36-2.67)
NHL[d]	A-D (all studies)	33	647	1.00 (0.89-1.13)*
	A-C	8	106	1.03 (0.70-1.51)*
	A-B	7	69	1.04 (0.63-1.72)*
	A	6	50	1.27 (0.90-1.79)
MM	A-D (all studies)	26	284	1.12 (0.98-1.27)
	A-C	9	37	1.15 (0.74-1.79)
	A-B	8	28	1.48 (0.96-2.27)
	A	8	28	1.48 (0.96-2.27)
ALL	A-D (all studies)	17	47	1.44 (1.03-2.02)
	A-C	4	11	1.26 (0.50-3.16)
	A-B	4	11	1.26 (0.50-3.16)
	A	1	5	2.80 (0.27-29.23)
CLL	A-D (all studies)	18	111	1.14 (0.78-1.67)*
	A-C	8	61	1.38 (0.71-2.69)*
	A-B	7	53	1.54 (0.72-3.31)*
	A	4	43	2.44 (0.88-6.75)

[a] A, quantitative exposure estimates for benzene; B, semi-quantitative estimates of benzene exposure or quantitative estimates of exposures containing benzene; C, some industrial hygiene sampling results; D, qualitative indication that benzene exposure had occurred.
[b] Number of studies.
[c] Number of exposed cases.
[d] NHL or lymphosarcoma/reticulosarcoma (preferred NHL if the study reported both).
* Significant evidence for between study heterogeneity ($p < 0.1$).

Discussion

We conducted a series of meta-analyses on occupational cohort studies to assess the possible association between benzene and lymphoid neoplasms. Utilizing different dimensions of study quality, we report evidence for an association between occupational benzene exposure and lymphoma subtypes MM, ALL, and CLL. For these subtypes, mRRs increased with increasing study quality, regardless of the strategy that was used to assess study quality. mRRs for NHL also increased with increasing study quality, although this effect was less pronounced. We did not observe an association between occupational benzene exposure and HL. Importantly, with the exception of a chance finding, the increase in mRRs for NHL, MM, ALL, and CLL with increasing study quality most likely reflects an actual underlying association with at least some of these lymphoma subtypes.

Because we observed mRR patterns consistent with a possible association between benzene and all lymphoma subtypes except HL, we formally explored quantitative exposure-response relations for NHL, MM, ALL and CLL, including all studies with *exposure assessment quality* A (studies with quantitative estimates of benzene exposure) based on flexible meta-regression analyses (48). The relatively limited number of studies in category A resulted in uncertain and unstable predictions of the exposure-response curve for NHL, MM, and CLL (data not shown). For ALL only one *exposure assessment quality* A study was available, which precluded conducting a meta-regression for this lymphoma subtype. Therefore, possible exposure-response associations can only be discussed informally on a study-by-study basis.

Assessment of study quality dimensions

We developed three different quality dimensions that reflect the substantial changes in diagnosis and categorization of lymphoid neoplasms over the last half century and the heterogeneity in occupational cohort studies with regard to industry, sample size, and documentation of benzene exposure. The generally higher RRs in the strata with studies that started follow-up in 1970 or later is consistent with better quality of lymphoma diagnosis in more recent years. The higher RRs are particularly noteworthy given that overall benzene exposure was likely reduced in workplaces after 1970-1980. Another secular trend in the quality of cohort studies over time was the greater use of incidence rather than mortality as end point (e.g., 91% of cohorts reporting CLL RRs with *start of follow-up* prior to 1970 used mortality as the end point vs. 43% for studies with start of follow-up in 1970 or later). It is possible that for less aggressive subtypes (e.g., CLL), subjects that died from other causes did not have lymphoma coded on their death certificate (20). However, cross-stratification of results suggested that stratification by period of follow-up explained more of the observed heterogeneity than stratification by mortality/incidence (data not shown). Although it has been suggested that the RR for leukemia subtypes observed in occupational studies might decrease with prolonged follow-up time (49), we found only modest evidence for this

phenomenon for lymphoma subtypes. Substitution of the most recent RRs with those of previous updates did not materially change the results (data not shown).

Because the association between benzene and AML is established, we argue that a well-conducted large epidemiological study on benzene and hemato- and lymphopoietic cancers should find such an association. If at least some evidence of association is not found, one could argue that there must be known or unknown methodological limitations in the study design. Such studies would by extension most likely be noninformative regarding the association between benzene and lymphoid neoplasms. Naturally, one should realize that a failure to find evidence for an association could also be the result of insufficient statistical power. However, in our meta-analyses we observed that the strong increase in mRRs for AML with increasing *AML significance levels* was generally paralleled by increasing RR in lymphoma subtypes. In other words, studies that reported higher (and more significant) RRs for AML generally also reported higher RRs for NHL, MM, ALL, and CLL.
The quality of exposure assessment has a large impact on the ability of an epidemiological study to identify modest increased RRs. The relevance of our *exposure assessment quality* approach was illustrated with the strong increase in mRRs for AML with increasing *exposure assessment quality.* This trend provides support for our assumption that studies that conducted a more detailed benzene exposure assessment likely provide higher overall quality of evidence for the potential association of benzene with adverse health outcomes. Although one would expect that the study quality indicators for *AML significance level* and *exposure assessment quality* to be highly correlated, this is not necessarily the case. For instance, we did observe five studies in the lowest *exposure assessment quality* category (D) that still reported a significant increased RR for AML, and we observed two studies from *exposure assessment quality* category B in the set of studies that reported an AML RR below unity (*AML significance level* category D). Therefore, the two study quality dimensions should be seen as complementary.

Non-Hodgkin Lymphoma
We observed a moderate increased RR of NHL with increasing study quality. However, neither the overall mRR nor any of the strata-specific mRRs reached formal statistical significance. Because our formal meta-regression did not result in robust exposure-response associations, we qualitatively explored exposure-response relations within each *exposure assessment quality* A publication that provided RRs for NHL. Of the six *exposure assessment quality* A studies that reported RRs for NHL only one study reported a significant increased RR (p for trend < 0.02) with increasing cumulative exposure to benzene (50). In contrast, in three of six publications the authors reported that there was no clear trend of RRs for NHL with increasing cumulative exposure to benzene (45, 51, 52), whereas the remaining two publications did not report on the quantitative relation between NHL and cumulative exposure to benzene (41, 53). In addition to these six studies, two publications that included MM in the definition of

NHL did report on the quantitative relation of NHL+MM and cumulative exposure to benzene (25, 44). One of these studies reported an initial increase in RR with increasing exposure to benzene followed by a drop in RR in the upper cumulative exposure group (44), whereas the other study reported no association (25). We note, however, that a recent meta-analysis including both case-control and cohort studies reported a significant elevated mRR for NHL when the analyses were restricted to the higher exposure groups and corrected for the healthy worker (inclusion) effect (13).

Overall the epidemiological evidence for the association between NHL and benzene is conflicting. This is illustrated by three recent meta-analyses that were based on largely the same data but reached a diametrically opposite conclusion on whether exposure to benzene is associated to NHL (4, 7, 13). The inconsistency in findings is partly explained by study quality and failure to correct for biases but might also to a certain extent be explained by the etiological heterogeneity within this group of diseases. If some NHL subtypes (e.g., diffuse large B cell lymphoma (DLBCL) or follicular lymphoma (FL)) are associated with benzene, but others are not, any NHL RR will be attenuated because of the inclusion of non-benzene-associated NHL subtypes. This is even further complicated by the fact that the distribution of NHL subtypes may vary considerably from population to population, which could lead to significant variation in reported associations between potential risk factors and total NHL (54). A series of recent population-based case-control studies provide evidence that the association between some genetic and environmental factors varies between major NHL subtypes like DLBCL and FL (55-59). Another series of recent case-control studies that used relatively high quality retrospective exposure assessment methods have provided evidence that this might also be true for the association between benzene and NHL subtypes (60-62). The studies by Cocco et al. and Wong et al. reported a stronger association with benzene with FL (OR 1.6 (95% CI, 0.9-2.9) and 7.00 (95% CI, 1.45-33.70), respectively) than for DLBCL (OR 0.9 (95% CI, 0.6-1.4) and 0.66 (95% CI, 0.31-1.42), respectively) (60, 62). The study by Miligi et al. did not report a RR for FL (due to the limited number of cases), but reported an OR of 2.4 (95% CI, 1.3-4.5) for DLBCL (61).

Multiple myeloma

Our analyses are supportive of an association of benzene exposure with MM. mRRs increased considerably and reached near statistical significance regardless of the study quality dimension used except for the analyses stratified by *AML significance level* where formal statistical significance was reached for the two highest quality strata. Our results are similar (albeit the point estimates of the mRRs are slightly lower) to the results from a meta-analysis by Infante in which slightly different inclusion criteria were applied (mRR 2.13 (95% CI, 1.31 – 3.46) (6). Further evidence for an association between exposure to benzene and MM have been provided by two recent population-based case-control studies that reported increased MM RRs with increasing exposure to benzene (60, 63). We qualitatively explored the

quantitative exposure-response relation between benzene and MM. Two of eight *exposure assessment quality* A studies reported an increase in RR with increasing cumulative exposure (45, 53); in two studies the authors reported no clear trend of RRs for MM with increasing cumulative exposure to benzene (37, 52); and four studies did not report on the quantitative relation between cumulative exposure to benzene and MM (44, 51, 64, 65). Therefore, although the evidence for an association between "any occupational benzene exposure" versus "background benzene exposure" and the RR of MM appears to be consistent, the evidence for an exposure-response relation between benzene and MM is more ambiguous. This would be explained partly by the much larger statistical power that is required to conduct quantitative exposure-response analysis, often a complication for small-scale occupational cohort studies.

Acute lymphocytic leukemia

The association between exposure to benzene and ALL is difficult to study because the disease is rare in adults (66). It is therefore noteworthy that our analyses do strongly suggest increased RRs for ALL. We were able to identify only two population-based case-control studies that explored benzene-ALL associations in adults (67, 68). One case-control study reported a (nonsignificantly) increased RR for ALL with a suggestion of an exposure-response relation (67), whereas the other study did not observe any cases with ALL (68). Together, the evidence from both cohort and case-control studies are strongly suggestive of a positive association between exposure to benzene and the RR of adult ALL.

Chronic lymphocytic leukemia

Our analyses suggest that exposure to benzene is associated with an increased RR for CLL. This is in line with results from four recent case-control studies that reported RRs ranging from 1.4 to 2.05 (60-63). Two of these case-control studies reported an increase in RR with increasing benzene exposure (60, 63). Of the cohort studies with quantitative exposure assessment, one study reported that the RR for the group with higher cumulative exposure was higher than the RR for the group with lower exposure (25). However, two cohort studies reported no association with cumulative exposure to benzene (45, 47), whereas one study did not report on the quantitative relation between cumulative exposure to benzene and CLL (51).

Conclusion

In line with the recent IARC evaluation of the carcinogenicity of benzene, our meta-analyses provide evidence for the association of occupational benzene exposure to MM, ALL, and CLL (3). Although these findings are suggestive, it is important to realize that most analyses were based on data sets of limited size. The evidence for an association between benzene and NHL

(as defined in ICD-9) is less convincing, but this could be explained by the heterogeneity in the association for particular subgroups of this disease or by not accounting for certain biases. We observed no association between benzene and HL. The discussion on the association between benzene and NHL will likely benefit from NHL subtype-specific analyses. Unfortunately, most current occupational cohort studies lack sufficient statistical power to perform such detailed analyses. Cohort studies with central pathology review and well-designed case-control studies using state-of-the-art retrospective exposure assessment methods will be needed to help evaluate the extent to which occupational benzene exposure is associated with specific subtypes of NHL.

Finally, our overall findings, taken together with the substantial experimental and molecular epidemiologic evidence that benzene exposure alters key components of the immune system relevant for lymphomagenesis (e.g., CD4+ T cell level; CD4+T cell/CD8+ T cell ratio) (69), provide support that benzene is likely to be causally related to one or more subtypes of lymphoma.

Acknowledgements

This work was performed as part of the work package “Integrated Risk Assessment” of the ECNIS Network of Excellence (Environmental Cancer Risk, Nutrition and Individual Susceptibility), operating within the European Union 6th Framework Program, Priority 5: “Food Quality and Safety” (FOOD-CT-2005-513943).

References

1. IARC (International Agency for Research on Cancer). Some industrial chemicals and dyestuffs. IARC Monogr Eval Carcinog Risk Chem Hum.1982;29:93-148.
2. IARC (International Agency for Research on Cancer). Overall evaluations of carcinogenicity: an updating of IARC Monographs volumes 1 to 42. IARC Monogr Eval Carcinog Risk Chem Hum Suppl 7. 1987:1-440.
3. Baan R, Grosse Y, Straif K, et al. A review of human carcinogens–Part F: chemical agents and related occupations. Lancet Oncol. 2009;10(12):1143-4.
4. Alexander DD, Wagner ME. Benzene exposure and non-Hodgkin lymphoma: a meta-analysis of epidemiologic studies. J Occup Environ Med. 2010;52(2):169-89.
5. Bergsagel DE, Wong O, Bergsagel PL, et al. Benzene and multiple myeloma: appraisal of the scientific evidence. Blood. 1999;94(4):1174-82.
6. Infante PF. Benzene exposure and multiple myeloma: a detailed meta-analysis of benzene cohort studies. Ann N Y Acad Sci. 2006;1076:90-109.
7. Kane EV, Newton R. Benzene and the risk of non-Hodgkin lymphoma: a review and meta-analysis of the literature. Cancer Epidemiol. 2010.
8. Lamm SH, Engel A, Byrd DM. Non-Hodgkin lymphoma and benzene exposure: a systematic literature review. Chem Biol Interact. 2005;153-154:231-7.
9. Savitz DA, Andrews KW. Review of epidemiologic evidence on benzene and lymphatic and hematopoietic cancers. Am J Ind Med. 1997;31(3):287-95.
10. Schnatter AR, Rosamilia K, Wojcik NC. Review of the literature on benzene exposure and leukemia subtypes. Chem Biol Interact. 2005;153-154:9-21.

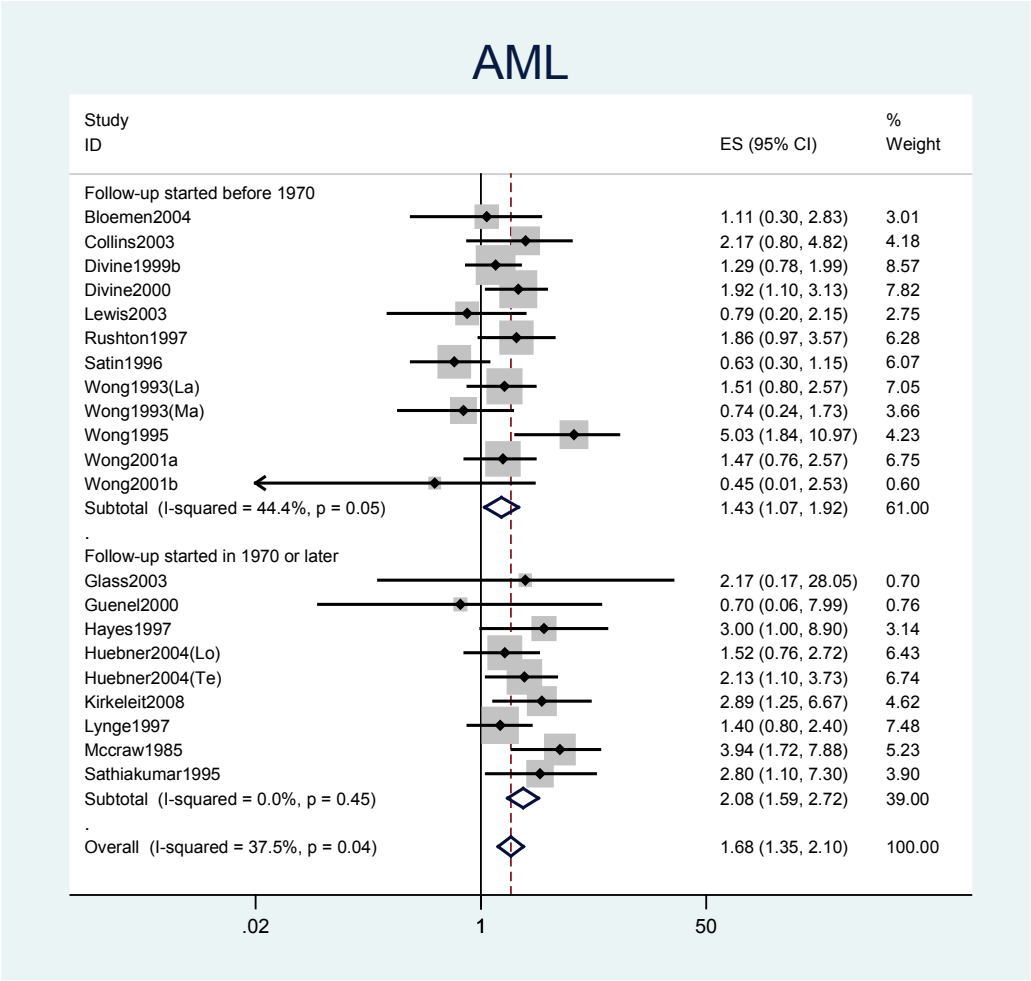
11. Smith MT, Jones RM, Smith AH. Benzene exposure and risk of non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev. 2007;16(3):385-91.
12. Sonoda T, Nagata Y, Mori M, et al. Meta-analysis of multiple myeloma and benzene exposure. J Epidemiol. 2001;11(6):249-54.
13. Steinmaus C, Smith AH, Jones RM, et al. Meta-analysis of benzene exposure and non-Hodgkin lymphoma: biases could mask an important association. Occup Environ Med. 2008;65(6):371-8.
14. Wong O, Raabe GK. 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. Regul Toxicol Pharmacol. 1995;21(2):307-21.
15. Wong O, Raabe GK. Multiple myeloma and benzene exposure in a multinational cohort of more than 250,000 petroleum workers. Regul Toxicol Pharmacol. 1997;26(2):188-99.
16. Wong O, Fu H. Exposure to benzene and non-Hodgkin lymphoma, an epidemiologic overview and an ongoing case-control study in Shanghai. Chem Biol Interact. 2005;153-154:33-41.
17. Wong O, Raabe GK. Non-Hodgkin's lymphoma and exposure to benzene in a multinational cohort of more than 308,000 petroleum workers, 1937 to 1996. J Occup Environ Med. 2000;42(5):554-68.
18. Wong O, Raabe GK. A critical review of cancer epidemiology in the petroleum industry, with a meta-analysis of a combined database of more than 350,000 workers. Regul Toxicol Pharmacol. 2000;32(1):78-98.
19. Aisenberg AC. Historical review of lymphomas. Br J Haematol. 2000;109(3):466-76.
20. Linet MS, Schubauer-Berigan MK, Weisenburger DD, et al. Chronic lymphocytic leukaemia: an overview of aetiology in light of recent developments in classification and pathogenesis. Br J Haematol. 2007;139(5):672-86.
21. Morton LM, Turner JJ, Cerhan JR, et al. Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph). Blood. 2007;110(2):695-708.
22. Swerdlow SH, Campo E, Harris N, et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues : International Agency for Research on Cancer (IARC), 2008.
23. Goldstein BD, Shalat S. Non-Hodgkin's lymphoma and exposure to benzene in petroleum workers. J Occup Environ Med. 2000;42(12):1133-6.
24. Vlaanderen J, Vermeulen R, Heederik D, et al. Guidelines to evaluate human observational studies for quantitative risk assessment. Environ Health Perspect. 2008;116(12):1700-5.
25. Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology. 2003;14(5):569-77.
26. Gun RT, Pratt N, Ryan P, et al. Update of mortality and cancer incidence in the Australian petroleum industry cohort. Occup Environ Med. 2006;63(7):476-81.
27. McElvenny DM, Armstrong BG, Jarup L, et al. Meta-analysis in occupational epidemiology: a review of practice. Occup Med (Lond). 2004;54(5):336-44.
28. Rothman KJ, Boice Jr. JD. Epidemologic analysis with a programmable calculator. Washington DC: U.S. Department of Health, Education and Welfare, 1979.
29. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ. 2003;326(7382):219.
30. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60.
31. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34.
32. Infante PF. Cancer risks in a UK benzene exposed cohort. Occup Environ Med. 2005;62(12):905,6; author reply 905-6.
33. Sorahan T, Kinlen LJ, Doll R. Cancer risks in a historical UK cohort of benzene exposed workers. Occup Environ Med. 2005;62(4):231-6.
34. Ott MG, Teta MJ, Greenberg HL. Lymphatic and hematopoietic tissue cancer in a chemical manufacturing environment. Am J Ind Med. 1989;16(6):631-43.
35. Thomas TL, Waxweiler RJ, Moure-Eraso R, et al. Mortality patterns among workers in three Texas oil refineries. J Occup Med. 1982;24(2):135-41.
36. Delzell E., Sathiakumar N., Cole P, et al. Case-Control study of leukemia, Non-Hodgkin Lymphoma and Multiple Myeloma among employees of Union Oil company of California, 1992.
37. Atkinson S, Coppock J, Fritschi L, et al. Lympho-haematopoietic Cancer and Exposure to Benzene in the Australian Petroleum Industry. Technical Report and Appendices. Melbourne Australia: Monash University and Deakin University, 2001.

38. Sathiakumar N, Delzell E, Cole P, et al. A case-control study of leukemia among petroleum workers. J Occup Environ Med. 1995;37(11):1269-77.
39. Decoufle P, Blattner WA, Blair A. Mortality among chemical workers exposed to benzene and other agents. Environ Res. 1983;30(1):16-25.
40. OSHA (Occupational Safety and Health Administration). Preamble to the final rule for benzene. Washington DC: Occupational Safety and Health Administration, 1987.
41. Wong O. An industry wide mortality study of chemical workers occupationally exposed to benzene. I. General results. Br J Ind Med. 1987;44(6):365-81.
42. Wong O, Harris F, Smith TJ. Health effects of gasoline exposure. II. Mortality patterns of distribution workers in the United States. Environ Health Perspect. 1993;101 Suppl 6:63-76.
43. Wong O. Re: Benzene and the dose-related incidence of hematologic neoplasms in China. J Natl Cancer Inst. 1998;90(6):469-71.
44. Wong O. An industry wide mortality study of chemical workers occupationally exposed to benzene. II. Dose response analyses. Br J Ind Med. 1987;44(6):382-95.
45. Collins JJ, Ireland B, Buckley CF, et al. Lymphohaematopoeitic cancer mortality among workers with benzene exposure. Occup Environ Med. 2003;60(9):676-9.
46. Divine BJ, Hartman CM. Update of a study of crude oil production workers 1946-94. Occup Environ Med. 2000;57(6):411-7.
47. Rushton L, Romaniuk H. 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. Occup Environ Med. 1997;54(3):152-66.
48. Vlaanderen J, Portengen L, Rothman N, et al. Flexible meta-regression to assess the shape of the benzene leukemia exposure-response curve. Environ Health Perspect. 2010;118(4):526-32.
49. Richardson DB. Temporal variation in the association between benzene and leukemia mortality. Environ Health Perspect. 2008;116(3):370-4.
50. Hayes RB, Yin SN, Dosemeci M, et al. Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine-National Cancer Institute Benzene Study Group. J Natl Cancer Inst. 1997;89(14):1065-71.
51. Bloemen LJ, Youk A, Bradley TD, et al. Lymphohaematopoietic cancer risk among chemical workers exposed to benzene. Occup Environ Med. 2004;61(3):270-4.
52. Schnatter AR, Armstrong TW, Nicolich MJ, et al. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med. 1996;53(11):773-81.
53. Rinsky RA, Hornung RW, Silver SR, et al. Benzene exposure and hematopoietic mortality: a long-term epidemiologic risk assessment. Am J Ind Med. 2002;42(6):474-80.
54. Muller AM, Ihorst G, Mertelsmann R, et al. Epidemiology of non-Hodgkin's lymphoma (NHL): trends, geographic distribution, and etiology. Ann Hematol. 2005;84(1):1-12.
55. Lan Q, Morton LM, Armstrong B, et al. Genetic variation in caspase genes and risk of non-Hodgkin lymphoma: a pooled analysis of 3 population-based case-control studies. Blood. 2009;114(2):264-7.
56. Morton LM, Wang SS, Cozen W, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. Blood. 2008;112(13):5150-60.
57. Rothman N, Skibola CF, Wang SS, et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. Lancet Oncol. 2006;7(1):27-38.
58. Skibola CF, Bracci PM, Halperin E, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet. 2009;41(8):873-5.
59. Skibola CF, Bracci PM, Nieters A, et al. Tumor necrosis factor (TNF) and lymphotoxin-alpha (LTA) polymorphisms and risk of non-Hodgkin lymphoma in the InterLymph Consortium. Am J Epidemiol. 2010;171(3):267-76.
60. Cocco P, t'Mannetje A, Fadda D, et al. Occupational exposure to solvents and risk of lymphoma subtypes: results from the Epilymph case-control study. Occup Environ Med. 2010;67(5):341-7.
61. Miligi L, Costantini AS, Benvenuti A, et al. Occupational exposure to solvents and the risk of lymphomas. Epidemiology. 2006;17(5):552-61.
62. Wong O, Harris F, Armstrong TW, et al. A hospital-based case-control study of non-Hodgkin lymphoid neoplasms in Shanghai: Analysis of environmental and occupational risk factors by subtypes of the WHO classification. Chem Biol Interact. 2010;184(1-2):129-46.

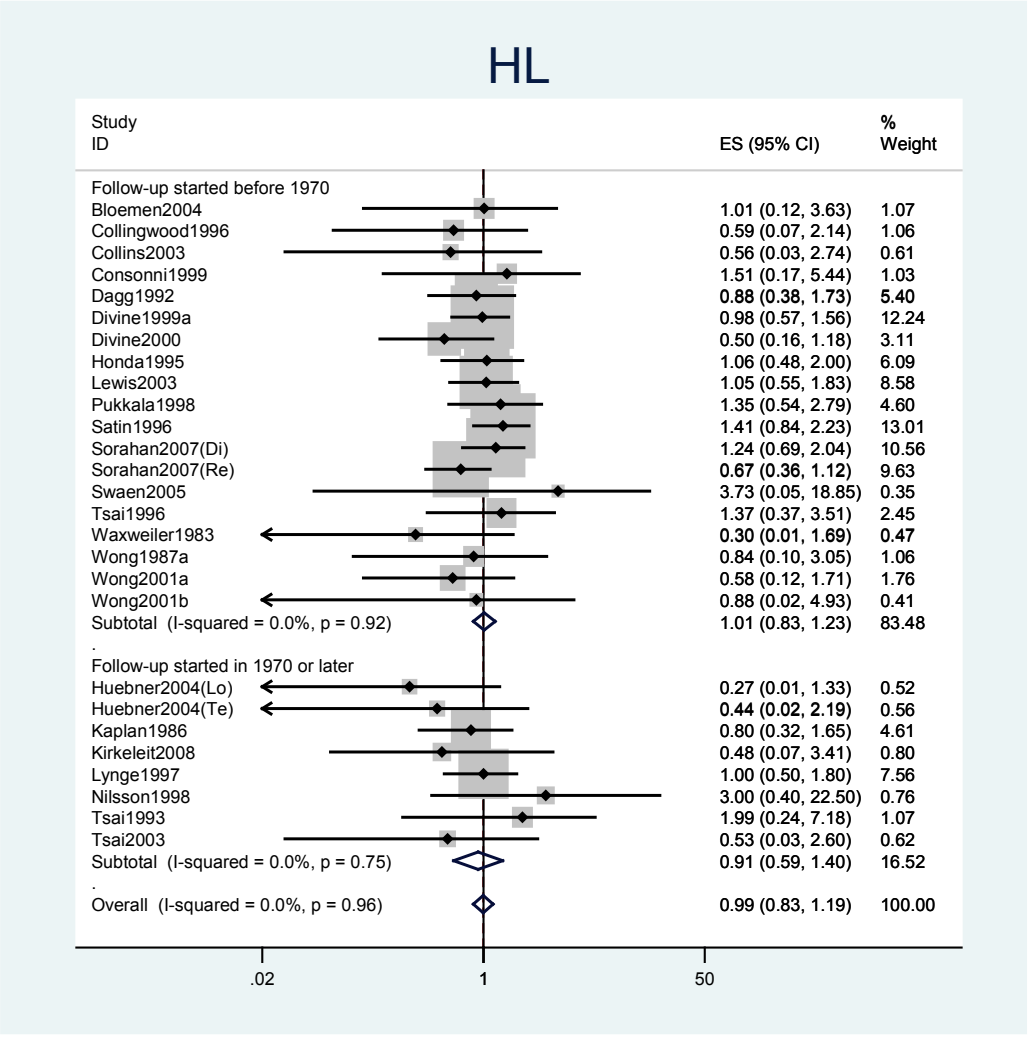
63. Seniori Costantini A, Benvenuti A, Vineis P, et al. Risk of leukemia and multiple myeloma associated with exposure to benzene and other organic solvents: evidence from the Italian Multicenter Case-control study. Am J Ind Med. 2008;51(11):803-11.
64. Swaen GM, Scheffers T, de Cock J, et al. Leukemia risk in caprolactam workers exposed to benzene. Ann Epidemiol. 2005;15(1):21-8.
65. Yin SN, Hayes RB, Linet MS, et al. A cohort study of cancer among benzene-exposed workers in China: overall results. Am J Ind Med. 1996;29(3):227-35.
66. Faderl S, O'Brien S, Pui CH, et al. Adult acute lymphoblastic leukemia: concepts and strategies. Cancer. 2010.
67. Adegoke OJ, Blair A, Shu XO, et al. Occupational history and exposure and the risk of adult leukemia in Shanghai. Ann Epidemiol. 2003;13(7):485-94.
68. Richardson S, Zittoun R, Bastuji-Garin S, et al. Occupational risk factors for acute leukaemia: a case-control study. Int J Epidemiol. 1992;21(6):1063-73.
69. Lan Q, Zhang L, Li G, et al. Hematotoxicity in workers exposed to low levels of benzene. Science. 2004;306(5702):1774-6.
70. Wong O, Harris F, Rosamilia K, et al. An updated mortality study of workers at a petroleum refinery in Beaumont, Texas, 1945 to 1996. J Occup Environ Med. 2001;43(4):384-401.
71. Lewis RJ, Schnatter AR, Drummond I, et al. Mortality and cancer morbidity in a cohort of Canadian petroleum workers. Occup Environ Med. 2003;60(12):918-28.
72. Huebner WW, Wojcik NC, Rosamilia K, et al. Mortality updates (1970-1997) of two refinery/petrochemical plant cohorts at Baton Rouge, Louisiana, and Baytown, Texas. J Occup Environ Med. 2004;46(12):1229-45.
73. Pukkala E. Cancer incidence among Finnish oil refinery workers, 1971-1994. J Occup Environ Med. 1998;40(8):675-9.
74. Guenel P, Imbernon E, Chevalier A, et al. Leukemia in relation to occupational exposures to benzene and other agents: a case-control study nested in a cohort of gas and electric utility workers. Am J Ind Med. 2002;42(2):87-97.
75. Consonni D, Pesatori AC, Tironi A, et al. Mortality study in an Italian oil refinery: extension of the follow-up. Am J Ind Med. 1999;35(3):287-94.
76. Tsai SP, Gilstrap EL, Cowles SR, et al. A cohort mortality study of two California refinery and petrochemical plants. J Occup Med. 1993;35(4):415-21.
77. Kirkeleit J, Riise T, Bratveit M, et al. Increased risk of acute myelogenous leukemia and multiple myeloma in a historical cohort of upstream petroleum workers exposed to crude oil. Cancer Causes Control. 2008;19(1):13-23.
78. Collingwood KW, Raabe GK, Wong O. An updated cohort mortality study of workers at a northeastern United States petroleum refinery. Int Arch Occup Environ Health. 1996;68(5):277-88.
79. Waxweiler RJ, Alexander V, Leffingwell SS, et al. Mortality from brain tumor and other causes in a cohort of petrochemical workers. J Natl Cancer Inst. 1983;70(1):75-81.
80. McCraw DS, Joyner RE, Cole P. Excess leukemia in a refinery population. J Occup Med. 1985;27(3):220-2.
81. Honda Y, Delzell E, Cole P. An updated study of mortality among workers at a petroleum manufacturing plant. J Occup Environ Med. 1995;37(2):194-200.
82. Wong O. Risk of acute myeloid leukaemia and multiple myeloma in workers exposed to benzene. Occup Environ Med. 1995;52(6):380-4.
83. Satin KP, Wong O, Yuan LA, et al. A 50-year mortality follow-up of a large cohort of oil refinery workers in Texas. J Occup Environ Med. 1996;38(5):492-506.
84. Dagg TG, Satin KP, Bailey WJ, et al. An updated cause specific mortality study of petroleum refinery workers. Br J Ind Med. 1992;49(3):203-12.
85. Kaplan SD. Update of a mortality study of workers in petroleum refineries. J Occup Med. 1986;28(7):514-6.
86. Lynge E, Anttila A, Hemminki K. Organic solvents and cancer. Cancer Causes Control. 1997;8(3):406-19.
87. Tsai SP, Gilstrap EL, Cowles SR, et al. Long-term follow-up mortality study of petroleum refinery and chemical plant employees. Am J Ind Med. 1996;29(1):75-87.
88. Tsai SP, Wendt JK, Cardarelli KM, et al. A mortality and morbidity study of refinery and petrochemical employees in Louisiana. Occup Environ Med. 2003;60(9):627-33.
89. Fu H, Demers PA, Costantini AS, et al. Cancer mortality among shoe manufacturing workers: an analysis of two cohorts. Occup Environ Med. 1996;53(6):394-8.

90. Nilsson RI, Nordlinder R, Horte LG, et al. Leukaemia, lymphoma, and multiple myeloma in seamen on tankers. Occup Environ Med. 1998;55(8):517-21.
91. Divine BJ, Hartman CM, Wendt JK. Update of the Texaco mortality study 1947-93: Part I. Analysis of overall patterns of mortality among refining, research, and petrochemical workers. Occup Environ Med. 1999;56(3):167-73.
92. Divine BJ, Hartman CM, Wendt JK. Update of the Texaco mortality study 1947-93: Part II. Analyses of specific causes of death for white men employed in refining, research, and petrochemicals. Occup Environ Med. 1999;56(3):174-80.
93. Wong O, Harris F, Rosamilia K, et al. Updated mortality study of workers at a petroleum refinery in Torrance, California, 1959 to 1997. J Occup Environ Med. 2001;43(12):1089-102.
94. Rushton L. A 39-year follow-up of the U.K. oil refinery and distribution center studies: results for kidney cancer and leukemia. Environ Health Perspect. 1993;101 Suppl 6:77-84.
95. Sorahan T. Mortality of UK oil refinery and petroleum distribution workers, 1951-2003. Occup Med (Lond). 2007;57(3):177-85.

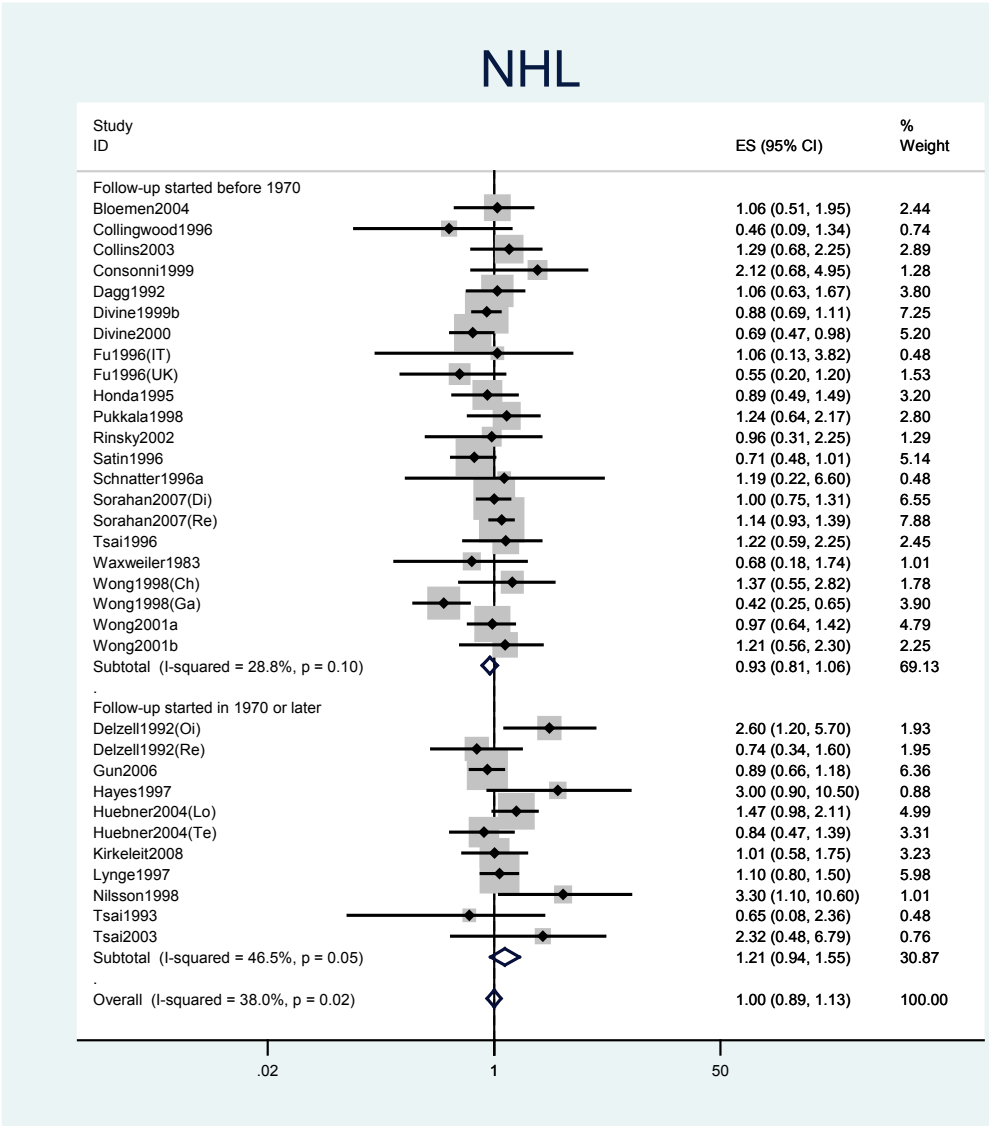
Supplemental Material, Figure 1 Forest plots of all studies for AML and five lymphoma subtypes in cohort studies of workers exposed to benzene, stratified by *start of follow-up*.



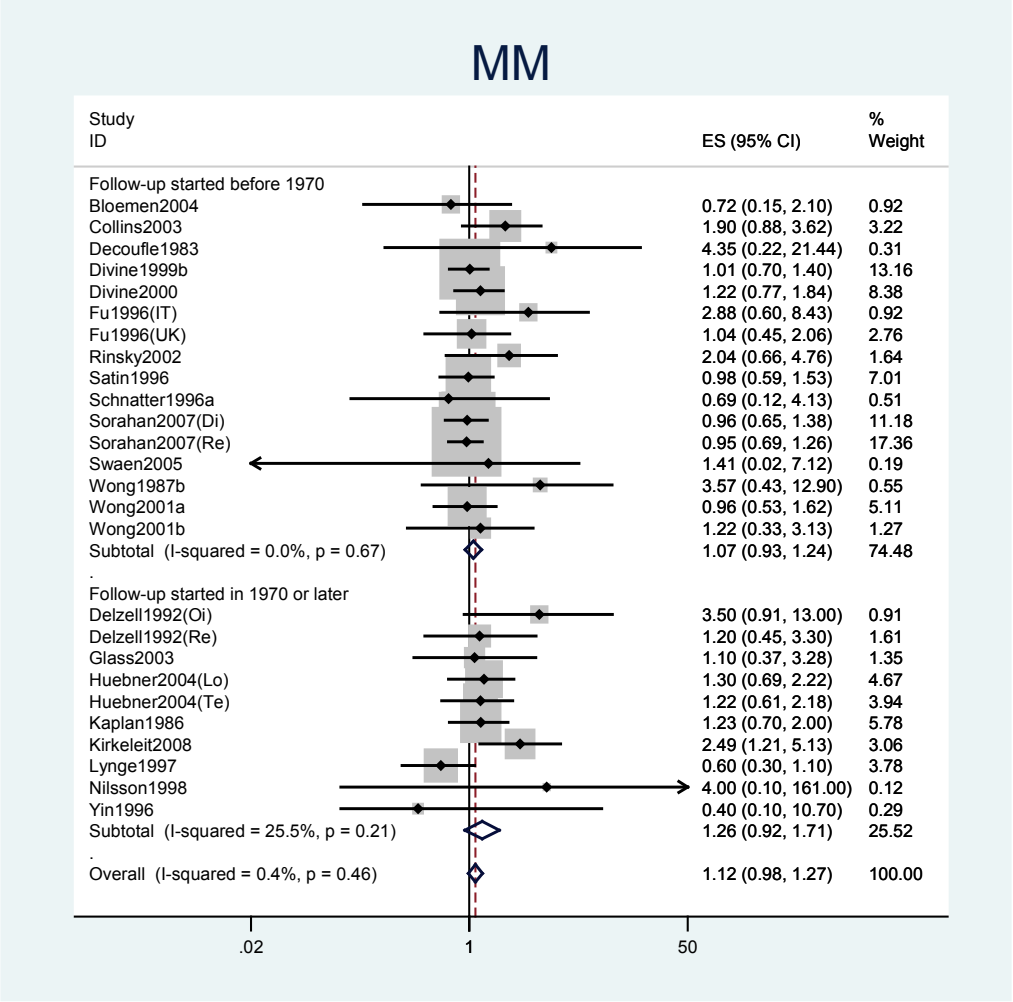
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



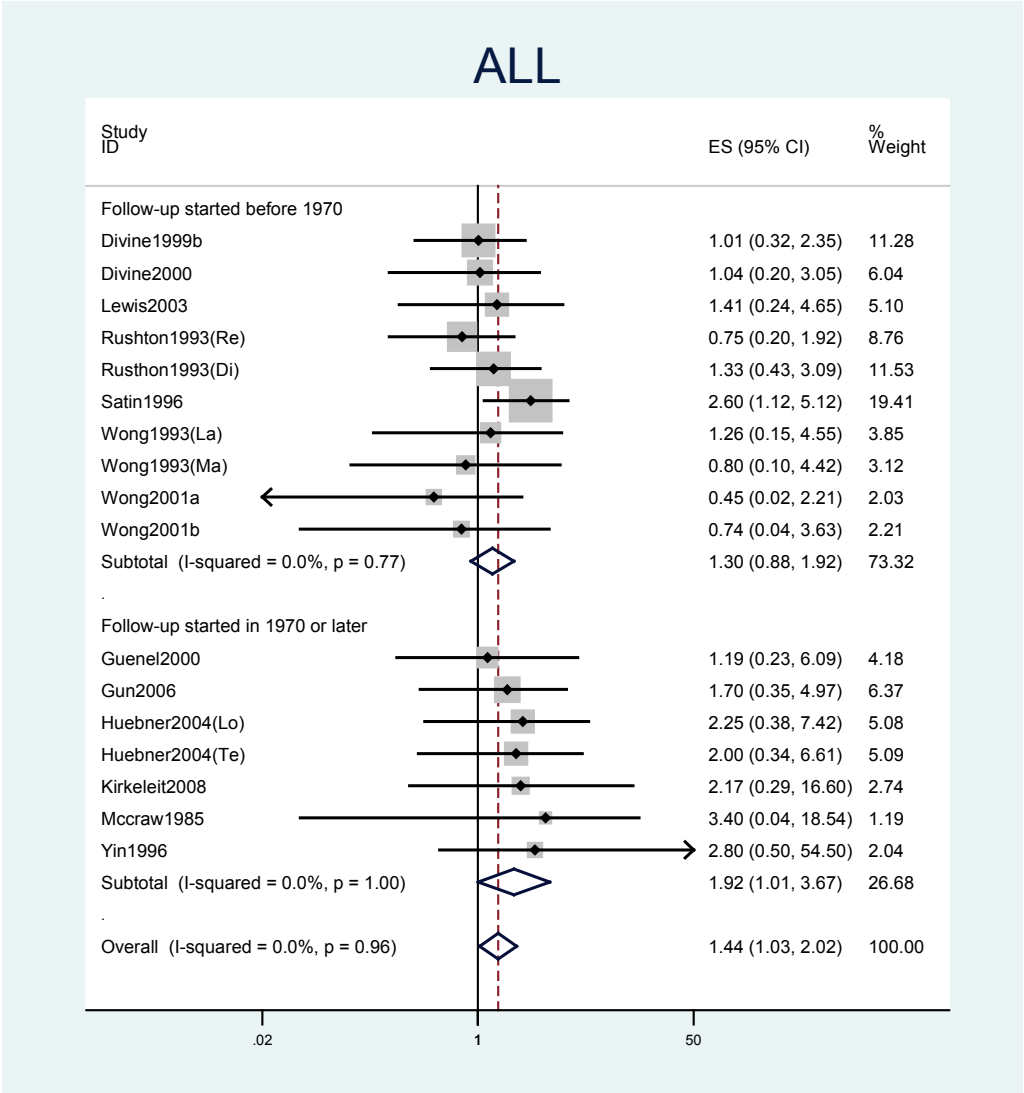
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



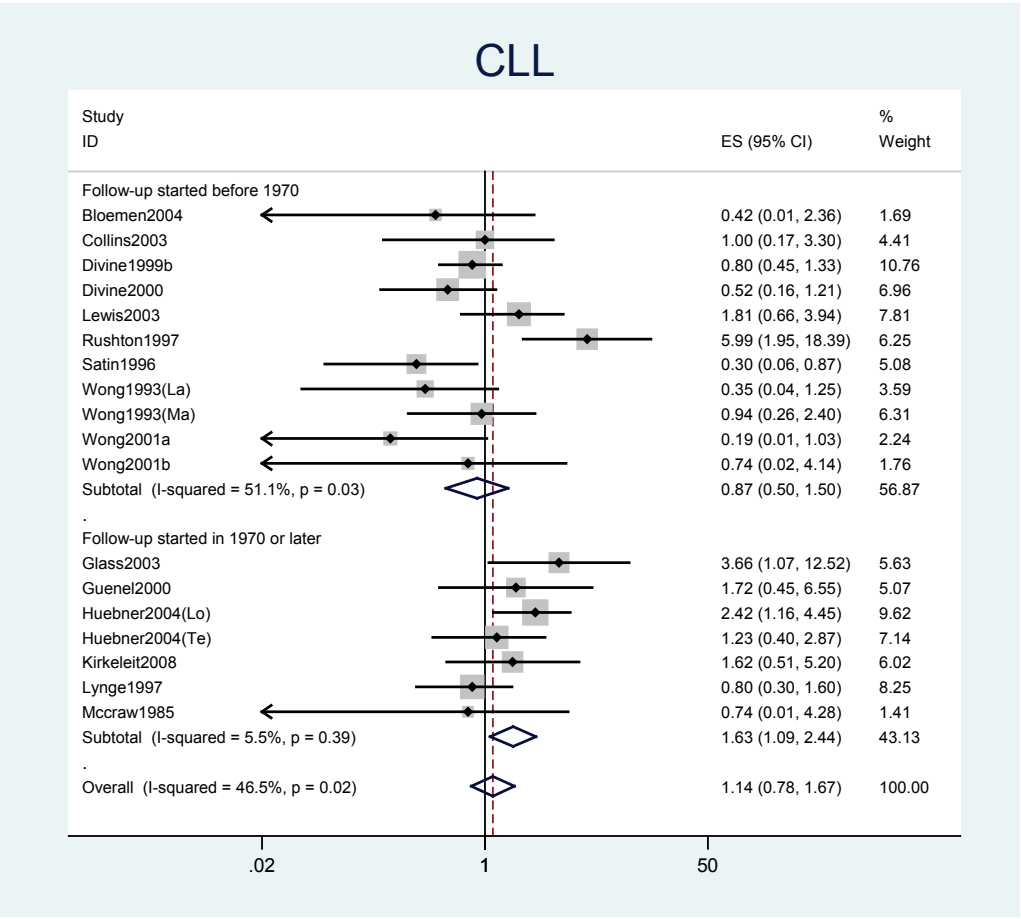
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).

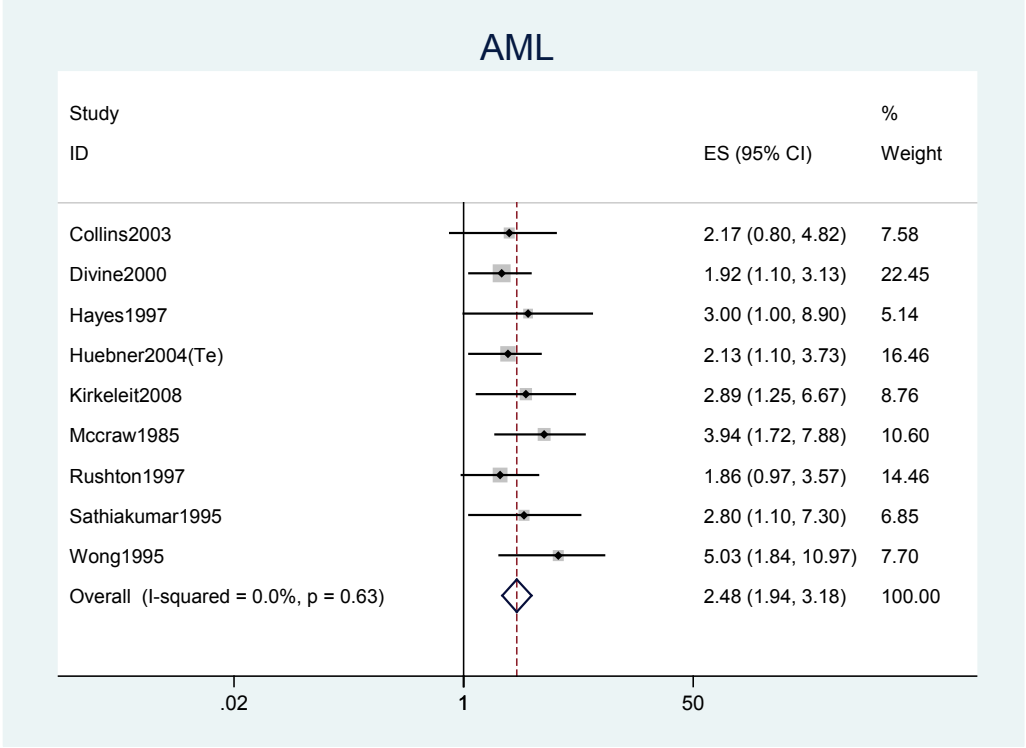


Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



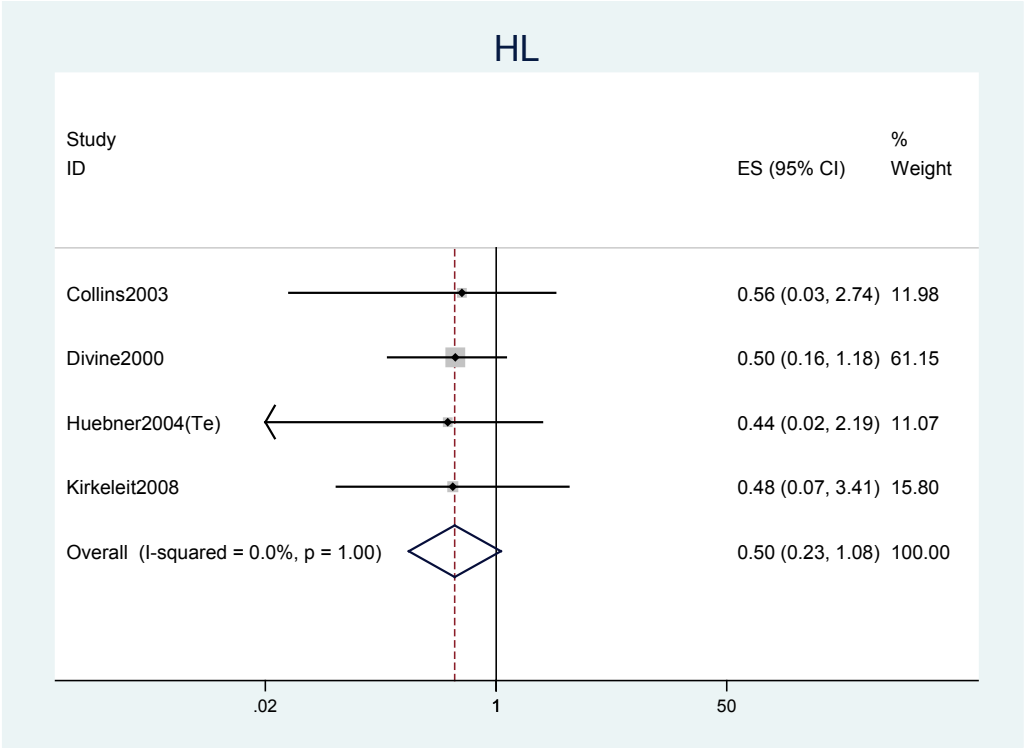
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).

Supplemental Material, Figure 2 Forest plots of studies with *AML significance level* A[a] for AML and five lymphoma subtypes in cohort studies of workers exposed to benzene.

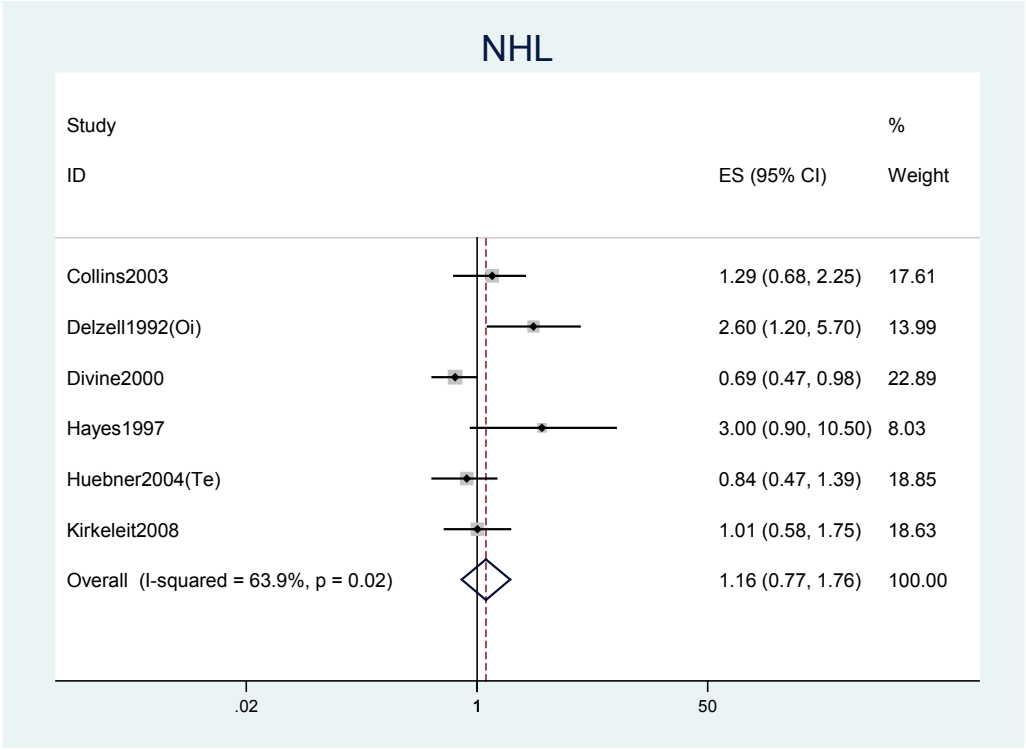


Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).

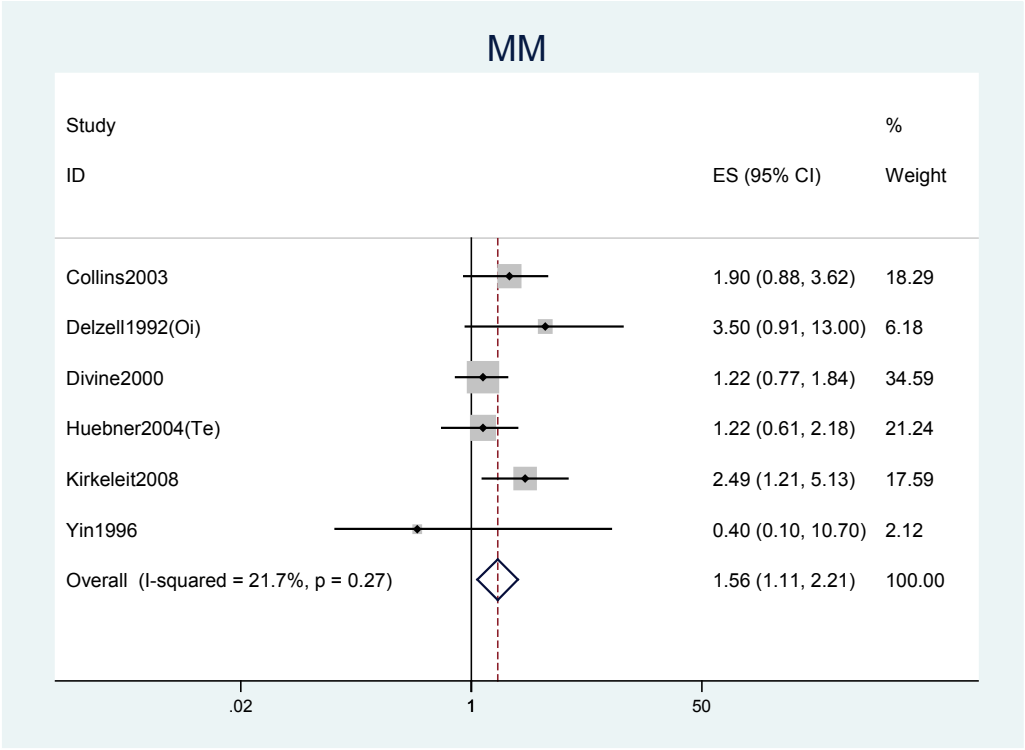
[a] AML RR >1, $p < 0.1$



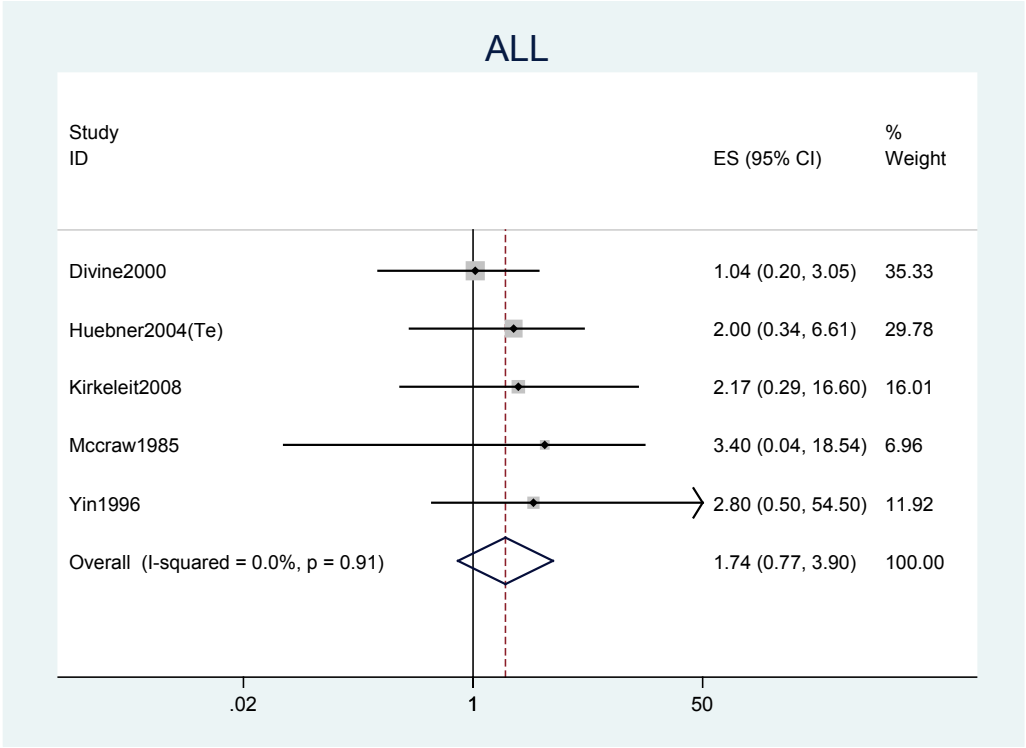
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



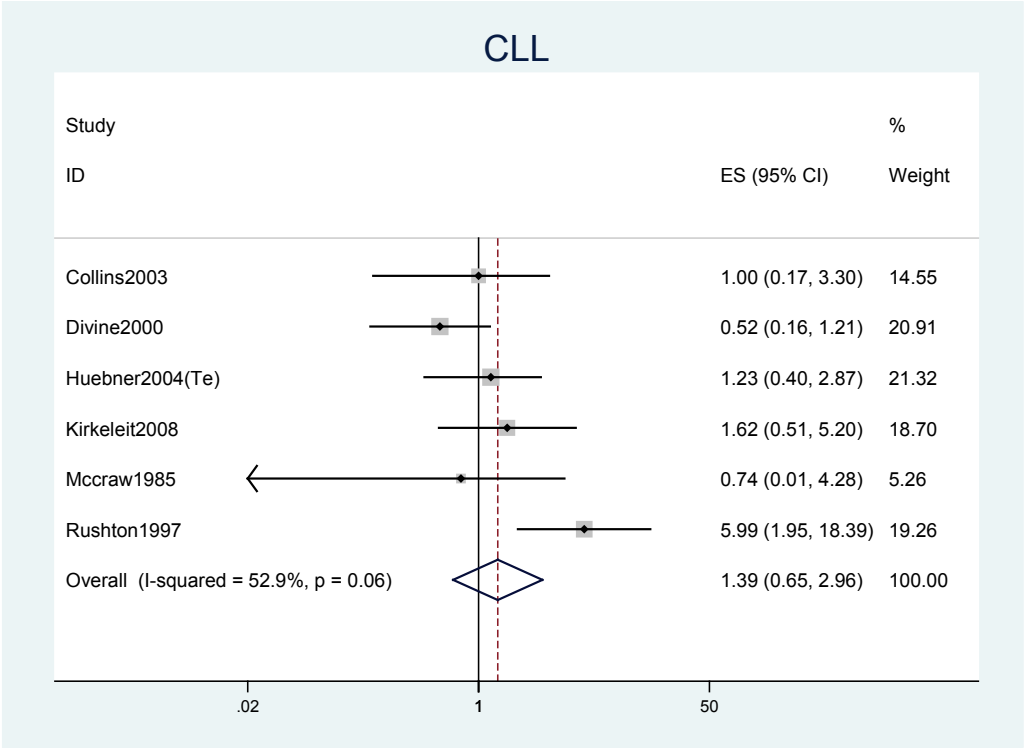
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).

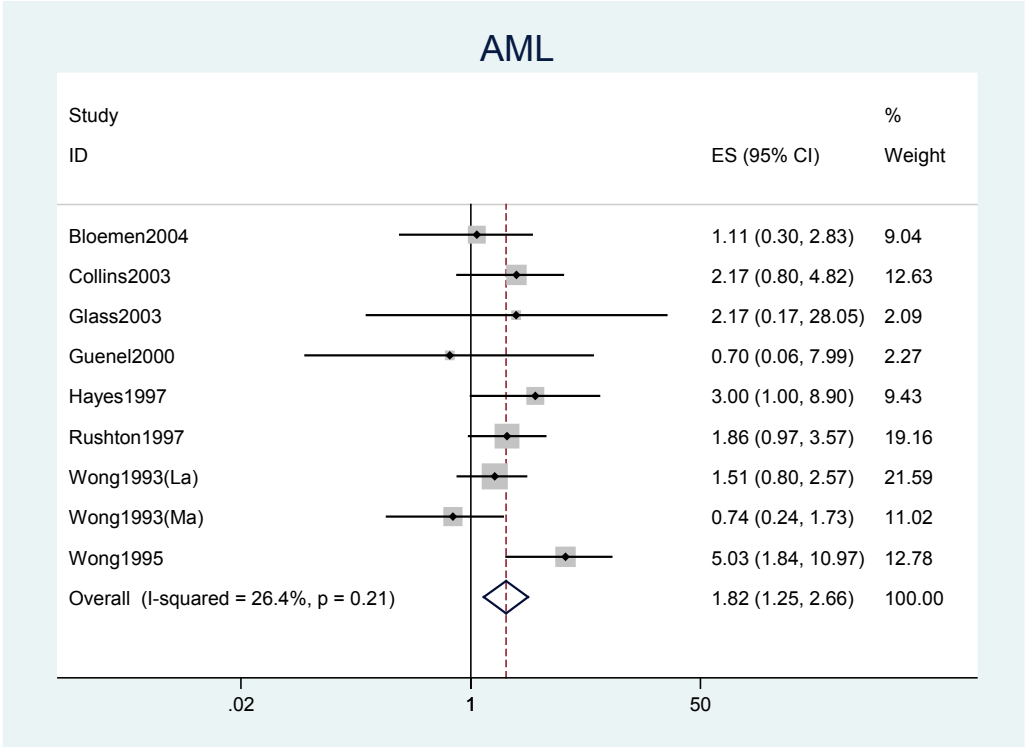


Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



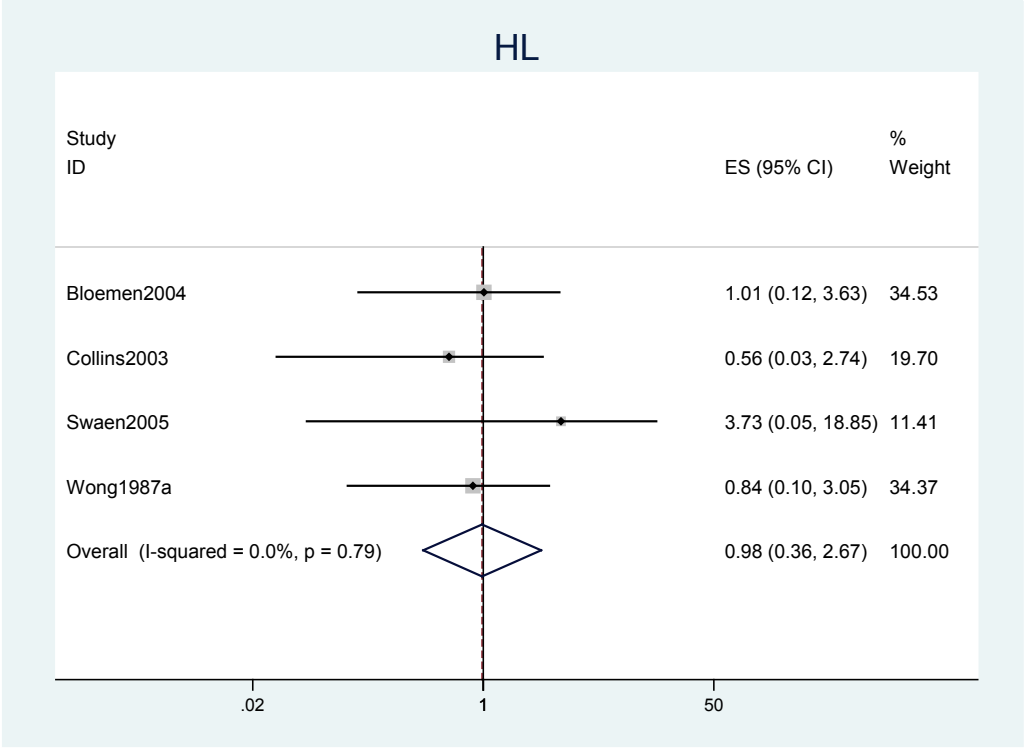
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).

Supplemental Material, Figure 3 Forest plots of studies with *exposure assessment quality* A-B[a] for AML and five lymphoma subtypes in cohort studies of workers exposed to benzene.

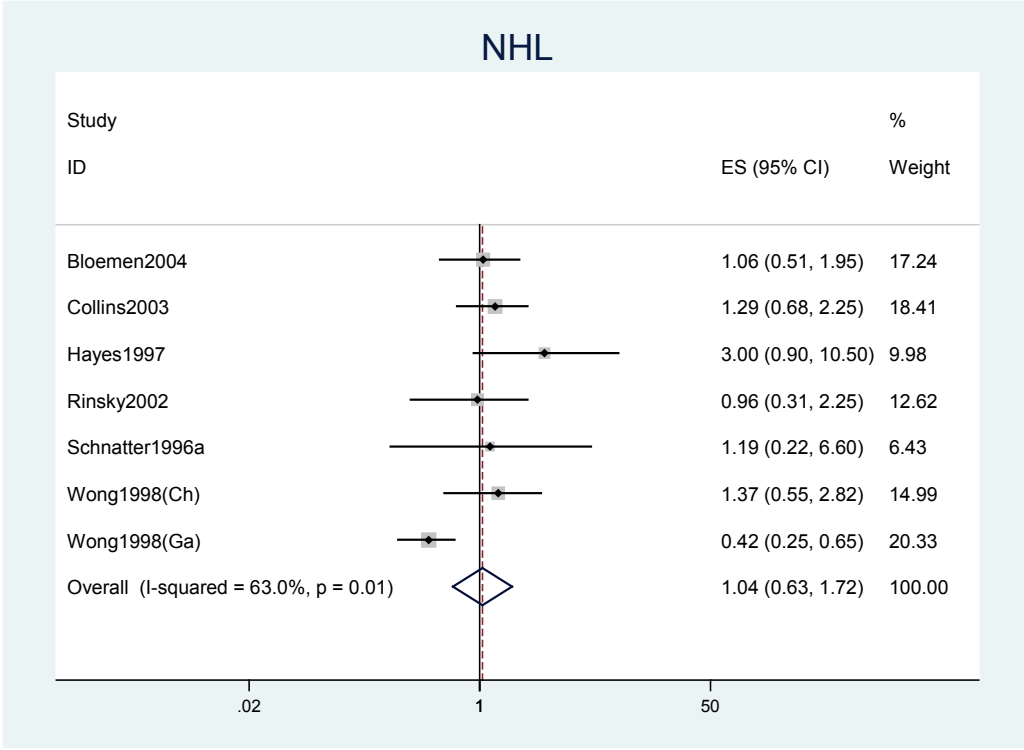


Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).

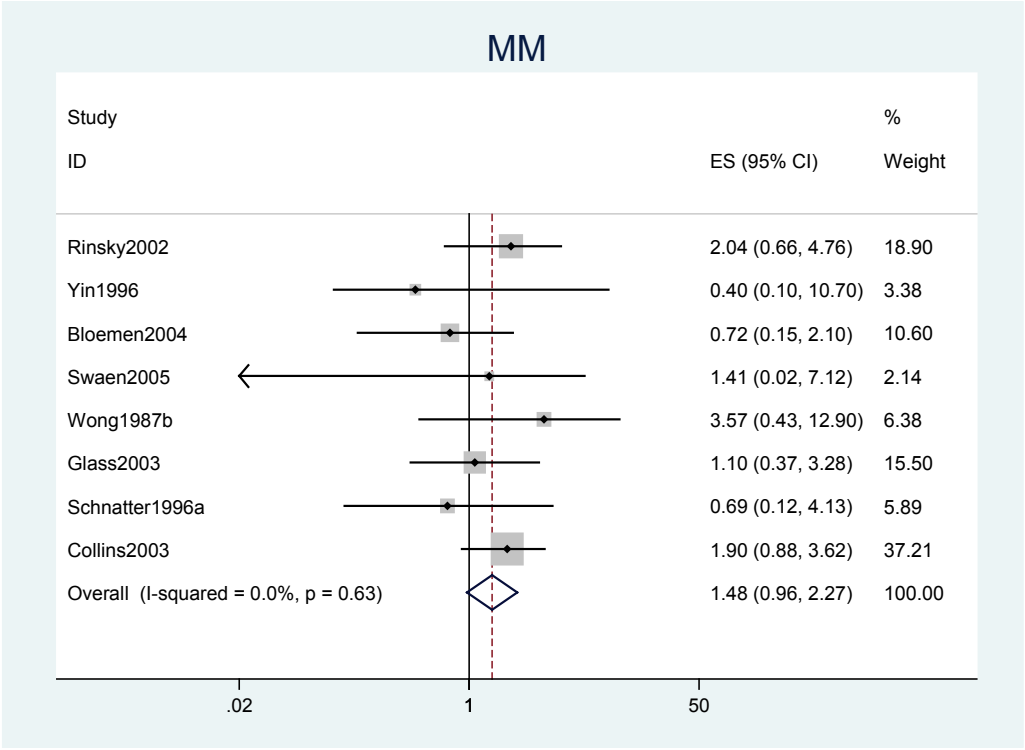
[a] Quantitative exposure estimates for benzene (A), semi-quantitative estimates of benzene exposure or quantitative estimates of exposures containing benzene (B).



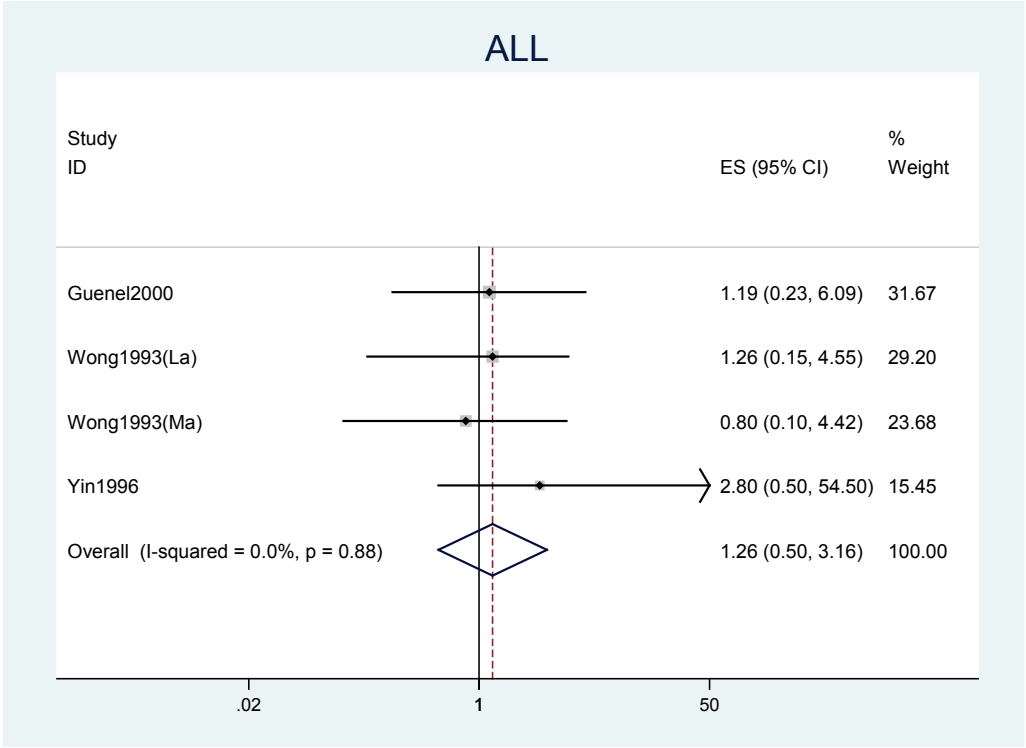
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



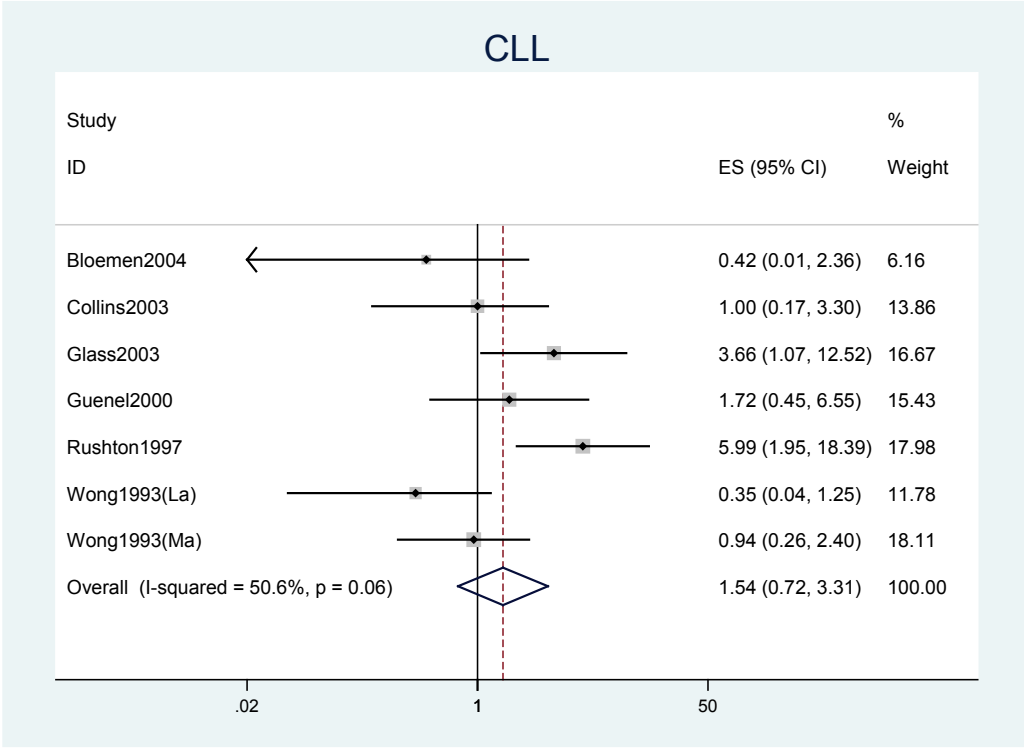
Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).



Sub-cohorts were identified with the following codes: Chemical workers cohort (Ch), Distribution cohort (Di), Gasoline workers cohort (Ga), Italian cohort (It), Land-based workers cohort (La), Louisiana cohort (Lo), Marine workers cohort (Ma), Oil and Gas workers cohort (Oi), Refinery cohort (Re), Texas cohort (Te), UK cohort (UK).

Supplemental Material, Table 1 Pooled relative risks[a] for AML and five lymphoma subtypes; stratification by *start of follow-up* and *AML significance level.*

		All studies			Start follow-up before 1970			Start follow-up 1970 and later		
Lymphoma subtype	AML significance level[b]	N[c]	n[d]	mRR	N[c]	n[d]	mRR	N[c]	n[d]	mRR
	A-E *(all studies)*	**21**	**217**	**1.68 (1.35-2.10)***	**12**	**119**	**1.43 (1.07-1.92)***	**9**	**98**	**2.08 (1.59-2.72)**
	A-D	21	217	1.68 (1.35-2.10)*	12	119	1.43 (1.07-1.92)*	9	98	2.08 (1.59-2.72)
AML	A-C	16	192	1.88 (1.56-2.27)	8	100	1.72 (1.34-2.23)	8	92	2.11 (1.61-2.77)
	A-B	11	132	2.20 (1.77-2.72)	5	64	2.06 (1.47-2.89)	6	68	2.41 (1.77-3.29)
	A	9	108	2.48 (1.94-3.18)	4	51	2.29 (1.54-3.40)	5	57	2.88 (1.95-3.99)
	A-E *(all studies)*	**27**	**146**	**0.99 (0.83-1.19)**	**19**	**123**	**1.01 (0.83-1.23)**	**8**	**23**	**0.91 (0.59-1.40)**
	A-D	12	69	0.99 (0.77-1.27)	8	58	1.03 (0.78-1.36)	4	11	0.83 (0.47-1.48)
HL	A-C	9	39	0.82 (0.59-1.15)	5	28	0.82 (0.55-1.23)	4	11	0.83 (0.47-1.48)
	A-B	5	7	0.47 (0.22-0.99)	2	6	0.51 (0.20-1.27)	3	1[f]	0.40 (0.11-1.44)
	A	4	7	0.50 (0.23-1.08)	2	6	0.51 (0.20-1.27)	2	1[g]	0.46 (0.10-2.09)
	A-E *(all studies)*	**33**	**647**	**1.00 (0.89-1.13)***	**22**	**452**	**0.93 (0.81-1.06)***	**11**	**195**	**1.21 (0.94-1.55)***
	A-D	15	383	0.97 (0.81-1.16)*	8	208	0.82 (0.66-1.02)*	7	175	1.18 (0.91-1.53)*
NHL[e]	A-C	13	344	0.99 (0.81-1.21)*	6	169	0.81 (0.62-1.07)*	7	175	1.18 (0.91-1.53)*
	A-B	7	130	1.21 (0.85-1.72)*	2	40	0.90 (0.49-1.66)*	5	90	1.38 (0.92-2.06)*
	A	6	101	1.16 (0.77-1.76)*	2	40	0.90 (0.49-1.66)*	4	61	1.40 (0.79-2.51)*
	A-E *(all studies)*	**26**	**284**	**1.12 (0.98-1.27)**	**16**	**204**	**1.07 (0.93-1.24)**	**10**	**80**	**1.26 (0.92-1.71)**
	A-D	14	160	1.15 (0.95-1.40)	7	105	1.09 (0.89-1.33)	7	55	1.27 (0.81-2.00)*
MM	A-C	12	137	1.19 (0.94-1.49)	5	82	1.12 (0.89-1.40)	7	55	1.27 (0.81-2.00)*
	A-B	7	69	1.49 (1.13-1.95)	2	29	1.39 (0.94-2.08)	5	40	1.58 (1.03-2.44)
	A	6	56	1.56 (1.11-2.21)	2	29	1.39 (0.94-2.08)	4	27	1.75 (0.94-3.26)
	A-E *(all studies)*	**17**	**47**	**1.44 (1.03-2.02)**	**10**	**30**	**1.30 (0.88-1.92)**	**7**	**17**	**1.92 (1.00-3.67)**
	A-D	17	47	1.44 (1.03-2.02)	10	30	1.30 (0.88-1.92)	7	17	1.92 (1.00-3.67)
ALL	A-C	11	29	1.41 (0.90-2.19)	5	15	1.09 (0.62-1.92)	6	14	2.10 (1.04-4.25)
	A-B	7	16	1.74 (0.90-3.36)	2	5	1.12 (0.39-3.25)	5	11	2.28 (0.99-5.26)
	A	5	12	1.74 (0.77-3.90)	1	3	1.04 (0.27-4.06)	4	9	2.30 (0.84-6.29)

	A-E *(all studies)*	**18**	**111**	**1.14 (0.78-1.67)***	**11**	**69**	**0.87 (0.50-1.50)***	**7**	**42**	**1.63 (1.09-2.44)**
	A-D	18	111	1.14 (0.78-1.67)*	11	69	0.87 (0.50-1.50)*	7	42	1.63 (1.09-2.44)
CLL	A-C	13	93	1.19 (0.74-1.90)*	7	55	0.84 (0.38-1.84)*	6	38	1.61 (1.00-2.59)
	A-B	8	57	1.37 (0.73-2.56)*	4	38	1.08 (0.29-4.06)*	4	19	1.84 (1.12-3.02)
	A	6	45	1.39 (0.65-2.96)*	3	36	1.47 (0.31-7.00)*	3	9	1.33 (0.64-2.76)

[a] The term relative risk (RR) is used to refer to either the risk ratio, the odds ratio (OR), or the standardized mortality ratio (SMR).

[b] AML RR > 1, $p < 0.1$ (A), AML RR > 1, $p < 0.2$ (B), AML RR > 1, $P > 0.2$ (C), AML RR reported (D), AML RR not reported (E)

[c] Number of studies

[d] Number of exposed cases

[e] NHL or Lymphosarcoma/Reticulosarcoma (preferred NHL if the study reported both)

[f] Two out of three studies reported null cases (continuity correction was applied in the meta-analysis)

[g] One out of two studies reported null cases (continuity correction was applied in the meta-analysis)

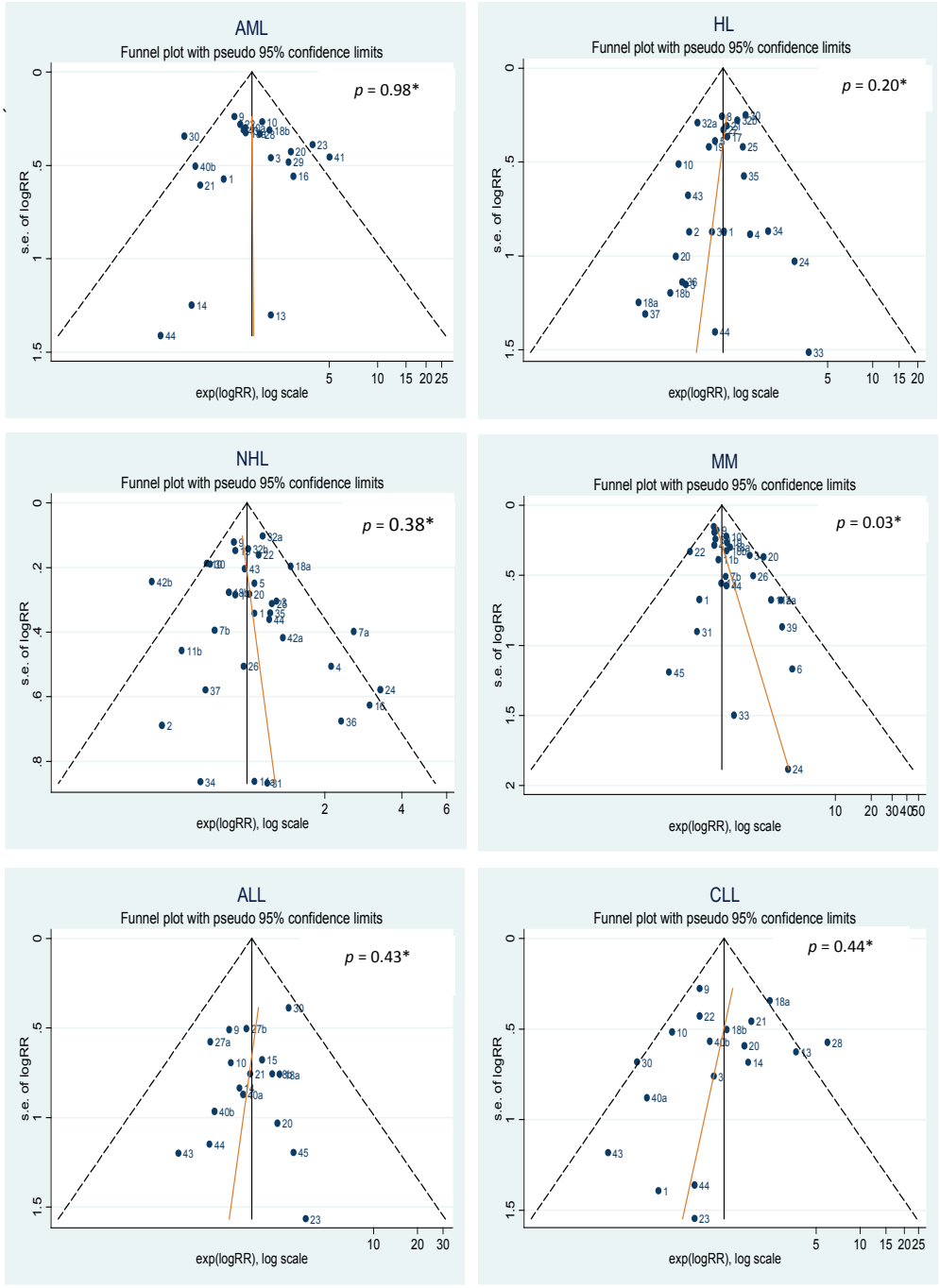
* Significant evidence for between study heterogeneity ($p < 0.1$)

Supplemental Material, Table 2 Pooled relative risks[a] for AML and five lymphoma subtypes; stratification by *start of follow-up* and *exposure assessment quality.*

		All studies			Start follow-up before 1970			Start follow-up 1970 and later		
Lymphoma subtype	Exposure assessment quality[b]	N[c]	n[d]	mRR	N[c]	n[d]	mRR	N[c]	n[d]	mRR
AML	**A-D *(all studies)***	**21**	**217**	**1.68 (1.35-2.10)***	**12**	**119**	**1.43 (1.07-1.92)***	**9**	**98**	**2.08 (1.59-2.72)**
	A-C	10	108	1.73 (1.26-2.38)	6	57	1.76 (1.11-2.79)*	4	51	1.60 (1.00-2.56)
	A-B	9	95	1.82 (1.25-2.66)	6	57	1.76 (1.11-2.79)*	3	38	2.33 (0.92-5.90)
	A	6	71	2.32 (1.55-3.47)	4	39	2.24 (1.28-3.92)	2	32	2.85 (1.05-7.79)
HL	**A-D *(all studies)***	**27**	**146**	**0.99 (0.83-1.19)**	**19**	**123**	**1.01 (0.83-1.23)**	**8**	**23**	**0.91 (0.59-1.40)**
	A-C	5	16	0.99 (0.58-1.71)	4	6	0.98 (0.36-2.67)	1	10	1.00 (0.53-1.90)
	A-B	4	6	0.98 (0.36-2.67)	4	6	0.98 (0.36-2.67)	0	0	--
	A	4	6	0.98 (0.36-2.67)	4	6	0.98 (0.36-2.67)	0	0	--
NHL[e]	**A-D *(all studies)***	**33**	**647**	**1.00 (0.89-1.13)***	**22**	**452**	**0.93 (0.81-1.06)***	**11**	**195**	**1.21 (0.94-1.55)***
	A-C	8	106	1.03 (0.70-1.51)*	6	53	0.92 (0.57-1.49)*	2	53	1.51 (0.61-3.78)
	A-B	7	69	1.04 (0.63-1.72)*	6	53	0.92 (0.57-1.49)*	1	16	3.00 (0.88-10.25)
	A	6	50	1.27 (0.90-1.79)	5	34	1.19 (0.83-1.69)	1	16	3.00 (0.88-10.25)
MM	**A-D *(all studies)***	**26**	**284**	**1.12 (0.98-1.27)**	**16**	**204**	**1.07 (0.93-1.24)**	**10**	**80**	**1.26 (0.92-1.71)**
	A-C	9	37	1.15 (0.74-1.79)	6	21	1.65 (1.02-2.66)	3	16	0.68 (0.40-1.17)
	A-B	8	28	1.48 (0.96-2.27)	6	21	1.65 (1.02-2.66)	2	7	0.92 (0.34-2.47)
	A	8	28	1.48 (0.96-2.27)	6	21	1.65 (1.02-2.66)	2	7	0.92 (0.34-2.47)
ALL	**A-D *(all studies)***	**17**	**47**	**1.44 (1.03-2.02)**	**10**	**30**	**1.30 (0.88-1.92)**	**7**	**17**	**1.92 (1.00-3.67)**
	A-C	4	11	1.26 (0.5-3.16)	2	3	1.03 (0.29-3.65)	2	8	1.58 (0.41-6.04)
	A-B	4	11	1.26 (0.5-3.16)	2	3	1.03 (0.29-3.65)	2	8	1.58 (0.41-6.04)
	A	1	5	2.80 (0.27-29.23)	0	0	--	1	5	2.80 (0.27-29.23)
CLL	**A-D *(all studies)***	**18**	**111**	**1.14 (0.78-1.67)***	**11**	**69**	**0.87 (0.50-1.50)***	**7**	**42**	**1.63 (1.09-2.44)**
	A-C	8	61	1.38 (0.71-2.69)*	5	38	1.16 (0.39-3.41)*	3	23	1.56 (0.62-3.97)
	A-B	7	53	1.54 (0.72-3.31)*	5	38	1.16 (0.39-3.41)*	2	15	2.59 (1.05-6.41)
	A	4	43	2.44 (0.88-6.75)	3	32	1.80 (0.38-8.63)*	1	11	3.66 (1.07-12.52)

[a] The term relative risk (RR) is used to refer to either the risk ratio, the odds ratio (OR), or the standardized mortality ratio (SMR).
[b] Quantitative exposure estimates for benzene (A), semi-quantitative estimates of benzene exposure or quantitative estimates of exposures containing benzene (B), some industrial hygiene sampling results (C), qualitative indication that benzene exposure had occurred (D).
[c] Number of studies
[d] Number of exposed cases
[e] NHL or Lymphosarcoma/Reticulosarcoma (preferred NHL if the study reported both)
* Significant evidence for between study heterogeneity ($p < 0.1$)

Supplemental material, Figure 4 Funnel plots for AML and five lymphoma subtypes with pseudo 95% confidence limits. Ids in the plot refer to the reference list[a].



* Egger's test for bias.

[a] 7a = oil and gas division cohort, 7b = refining division cohort; 11a = Italian cohort, 11b = UK cohort; 18a = Louisiana cohort, 18b = Texas cohort; 27a = refinery workers cohort, 27b = distribution workers cohort; 32a = refinery workers cohort, 32b = distribution workers cohort; 40a = marine workers cohort, 40b = land based workers cohort; 42a = chemical workers cohort (38 and 39), 42b = gasoline distribution employees cohort (see 40).

References Supplemental Material

1. Bloemen LJ, Youk A, Bradley TD, et al. Lymphohaematopoietic cancer risk among chemical workers exposed to benzene. Occup Environ Med. 2004;61(3):270-4.
2. Collingwood KW, Raabe GK, Wong O. An updated cohort mortality study of workers at a northeastern United States petroleum refinery. Int Arch Occup Environ Health. 1996;68(5):277-88.
3. Collins JJ, Ireland B, Buckley CF, et al. Lymphohaematopoeitic cancer mortality among workers with benzene exposure. Occup Environ Med. 2003;60(9):676-9.
4. Consonni D, Pesatori AC, Tironi A, et al. Mortality study in an Italian oil refinery: extension of the follow-up. Am J Ind Med. 1999;35(3):287-94.
5. Dagg TG, Satin KP, Bailey WJ, et al. An updated cause specific mortality study of petroleum refinery workers. Br J Ind Med. 1992;49(3):203-12.
6. Decoufle P, Blattner WA, Blair A. Mortality among chemical workers exposed to benzene and other agents. Environ Res. 1983;30(1):16-25.
7. Delzell E., Sathiakumar N., Cole P, et al. A. Case-control study of leukemia, Non-Hodgkin lymphoma and multiple myeloma among employees of union oil company of California. 1992.
8. Divine BJ, Hartman CM, Wendt JK. Update of the Texaco mortality study 1947-93: Part I. Analysis of overall patterns of mortality among refining, research, and petrochemical workers. Occup Environ Med. 1999;56(3):167-73.
9. Divine BJ, Hartman CM, Wendt JK. Update of the Texaco mortality study 1947-93: Part II. Analyses of specific causes of death for white men employed in refining, research, and petrochemicals. Occup Environ Med. 1999;56(3):174-80.
10. Divine BJ, Hartman CM. Update of a study of crude oil production workers 1946-94. Occup Environ Med. 2000 Jun;57(6):411-7.
11. Fu H, Demers PA, Costantini AS, et al. Cancer mortality among shoe manufacturing workers: An analysis of two cohorts. Occup Environ Med. 1996;53(6):394-8.
12. Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology. 2003;14(5):569-77.
13. Atkinson S, Coppock J, Fritschi L, et al. Lympho-haematopoietic cancer and exposure to benzene in the australian petroleum industry. Technical Report and Appendices, Victoria, Australia: Monash University and Deakin University, 2001.
14. Guenel P, Imbernon E, Chevalier A, et al. Leukemia in relation to occupational exposures to benzene and other agents: A case-control study nested in a cohort of gas and electric utility workers. Am J Ind Med. 2002;42(2):87-97.
15. Gun RT, Pratt N, Ryan P, et al. Update of mortality and cancer incidence in the Australian petroleum industry cohort. Occup Environ Med. 2006;63(7):476-81.
16. Hayes RB, Yin SN, Dosemeci M, et al. Benzene and the dose-related incidence of hematologic neoplasms in china. Chinese academy of preventive medicine-national cancer institute benzene study group. J Natl Cancer Inst. 1997;89(14):1065-71.
17. Honda Y, Delzell E, Cole P. An updated study of mortality among workers at a petroleum manufacturing plant. J Occup Environ Med. 1995;37(2):194-200.
18. Huebner WW, Wojcik NC, Rosamilia K, et al. Mortality updates (1970-1997) of two refinery/petrochemical plant cohorts at Baton Rouge, Louisiana, and Baytown, Texas. J Occup Environ Med. 2004;46(12):1229-45.
19. Kaplan SD. Update of a mortality study of workers in petroleum refineries. J Occup Med. 1986;28(7):514-6.
20. Kirkeleit J, Riise T, Bratveit M, et al. Increased risk of acute myelogenous leukemia and multiple myeloma in a historical cohort of upstream petroleum workers exposed to crude oil. Cancer Causes Control. 2008;19(1):13-23.
21. Lewis RJ, Schnatter AR, Drummond I, et al. Mortality and cancer morbidity in a cohort of Canadian petroleum workers. Occup Environ Med. 2003;60(12):918-28.
22. Lynge E, Anttila A, Hemminki K. Organic solvents and cancer. Cancer Causes Control. 1997;8(3):406-19.
23. McCraw DS, Joyner RE, Cole P. Excess leukemia in a refinery population. J Occup Med. 1985;27(3):220-2.
24. Nilsson RI, Nordlinder R, Horte LG, et al. Leukaemia, lymphoma, and multiple myeloma in seamen on tankers. Occup Environ Med. 1998;55(8):517-21.

25. Pukkala E. Cancer incidence among Finnish oil refinery workers, 1971-1994. J Occup Environ Med. 1998;40(8):675-9.
26. Rinsky RA, Hornung RW, Silver SR, et al. Benzene exposure and hematopoietic mortality: a long-term epidemiologic risk assessment. Am J Ind Med. 2002;42(6):474-80.
27. Rushton L. A 39-year follow-up of the U.K. oil refinery and distribution center studies: results for kidney cancer and leukemia. Environ Health Perspect. 1993;101 Suppl 6:77-84.
28. Rushton L, Romaniuk H. 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. Occup Environ Med. 1997;54(3):152-66.
29. Sathiakumar N, Delzell E, Cole P, et al. A case-control study of leukemia among petroleum workers. J Occup Environ Med. 1995;37(11):1269-77.
30. Satin KP, Wong O, Yuan LA, et al. A 50-year mortality follow-up of a large cohort of oil refinery workers in Texas. J Occup Environ Med. 1996;38(5):492-506.
31. Schnatter AR, Armstrong TW, Nicolich MJ, et al. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med. 1996;53(11):773-81.
32. Sorahan T. Mortality of UK oil refinery and petroleum distribution workers, 1951-2003. Occup Med (Lond). 2007;57(3):177-85.
33. Swaen GM, Scheffers T, de Cock J, et al. Leukemia risk in caprolactam workers exposed to benzene. Ann Epidemiol. 2005;15(1):21-8.
34. Tsai SP, Gilstrap EL, Cowles SR, et al. A cohort mortality study of two California refinery and petrochemical plants. J Occup Med. 1993;35(4):415-21.
35. Tsai SP, Gilstrap EL, Cowles SR, et al. Long-term follow-up mortality study of petroleum refinery and chemical plant employees. Am J Ind Med. 1996;29(1):75-87.
36. Tsai SP, Wendt JK, Cardarelli KM, et al. A mortality and morbidity study of refinery and petrochemical employees in Louisiana. Occup Environ Med. 2003;60(9):627-33.
37. Waxweiler RJ, Alexander V, Leffingwell SS, et al. Mortality from brain tumor and other causes in a cohort of petrochemical workers. J Natl Cancer Inst. 1983;70(1):75-81.
38. Wong O. An industry wide mortality study of chemical workers occupationally exposed to benzene. I. General results. Br J Ind Med. 1987;44(6):365-81.
39. Wong O. An industry wide mortality study of chemical workers occupationally exposed to benzene. II. Dose response analyses. Br J Ind Med. 1987;44(6):382-95.
40. Wong O, Harris F, Smith TJ. Health effects of gasoline exposure. II. Mortality patterns of distribution workers in the united states. Environ Health Perspect. 1993;101 Suppl 6:63-76.
41. Wong O. Risk of acute myeloid leukaemia and multiple myeloma in workers exposed to benzene. Occup Environ Med. 1995;52(6):380-4.
42. Wong O. Re: Benzene and the dose-related incidence of hematologic neoplasms in china. J Natl Cancer Inst. 1998;90(6):469-71.
43. Wong O, Harris F, Rosamilia K, et al. An updated mortality study of workers at a petroleum refinery in Beaumont, Texas, 1945 to 1996. J Occup Environ Med. 2001;43(4):384-401.
44. Wong O, Harris F, Rosamilia K, Raabe GK. Updated mortality study of workers at a petroleum refinery in Torrance, California, 1959 to 1997. J Occup Environ Med. 2001;43(12):1089-102.
45. Yin SN, Hayes RB, Linet MS, et al. A cohort study of cancer among benzene-exposed workers in China: Overall results. Am J Ind Med. 1996;29(3):227-35.

Chapter 5

Flexible Meta-Regression to Assess the Shape of the Benzene-Leukemia Exposure-Response Curve

Jelle Vlaanderen
Lützen Portengen
Nathaniel Rothman
Qing Lan
Hans Kromhout
Roel Vermeulen

Environmental Health Perspectives. 2010;118(4):526-32.

Abstract

Background Previous evaluations of the shape of the benzene-leukemia exposure-response curve (ERC) were based on single or small sets of human occupational studies. Integrating evidence from all available studies that are of sufficient quality combined with flexible meta-regression models is likely to provide better insight into the functional relation between benzene exposure and risk of leukemia.

Objectives We used natural splines in a flexible meta-regression method to assess the shape of the benzene-leukemia ERC.

Methods We fitted meta-regression models to 30 aggregated risk estimates extracted from nine human observational studies and performed sensitivity analyses to assess the impact of *a priori* assessed study characteristics on the predicted ERC.

Results The natural spline showed a supralinear shape at cumulative exposures less than 100 ppm-years, although this model fitted the data only marginally better than a linear model (p = 0.06). Stratification based on study design and jackknifing indicated that the cohort studies had a considerable impact on the shape of the ERC at high exposure levels (> 100 ppm-years), but that predicted risks for the low exposure range (< 50 ppm-years) were robust.

Conclusions Although limited by the small number of studies and the large heterogeneity between studies, the inclusion of all studies of sufficient quality combined with a flexible meta-regression method provides the most comprehensive evaluation of the benzene-leukemia ERC to date. The natural spline based on all data indicates a significantly increased risk of leukemia (relative risk (RR) = 1.14; 95% confidence interval (CI), 1.04–1.26) at an exposure level as low as 10 ppm-years.

Introduction

Although broad consensus exists in the scientific community that benzene is a leukemogen, there is considerable uncertainty regarding the actual shape of the exposure-response curve (ERC) (1, 2). Most of the current epidemiological evidence for an increased leukemia risk stems from studies among workers exposed to relatively high levels of benzene (3, 4). The risks of leukemia at lower benzene exposures, however, remain largely unclear. Importantly, exposure to benzene in occupational settings has dropped considerably during the last three decades (5). Furthermore, a large proportion of the general population is exposed to low levels of benzene (< 0.2 ppm) through car exhaust, cigarette smoke and other sources (6). Therefore, the primary interest of current benzene risk assessment is in risks associated with exposure to low levels of benzene.

In existing risk assessments of benzene, the ERC was assessed based on evidence from one 'best' study or, alternatively, from a limited set of 'best' studies (3, 4). These investigations were mostly conducted among relatively highly exposed workers, so the derived ERC might not be directly applicable to workers exposed to lower levels or the general population. Furthermore, these risk assessments used linear models to describe the ERC, but increasing evidence from molecular epidemiologic studies of workers exposed to a wide range of benzene levels indicates that the shape of the ERC for benzene and its toxic effects may be nonlinear. This hypothesis is based on the observation that the dose-related production of urinary metabolites of benzene, which include the toxic metabolites muconic acid and hydroquinone and the less toxic metabolites phenol and catechol, decreases with increasing benzene exposure (7-9). Furthermore, some evidence shows that benzene metabolism favors the production of the toxic metabolites hydroquinone and muconic acid at low exposures (7). This is especially important because hydroquinone is the pre-cursor of 1,4-benzoquinone, which is generally regarded as the most hematotoxic metabolite of benzene (7). The non-linear production of benzene's toxic metabolites would have important consequences for risk assessment because one would expect this to result in a nonlinear relationship between benzene exposure and health outcomes as well. Indeed, in a study that looked at the shape of the dose-response curve of benzene-related hematological effects, a sharper drop in the peripheral white blood cell count was observed at low levels of exposure (< 1 ppm) than at higher levels of exposure (10, 11).

To explore the shape of the benzene-leukemia ERC, we performed flexible meta-regressions on a set of studies that reported results from quantitative exposure-response analysis for benzene and leukemia. For our analyses, we used a modified version of the approach proposed by Bagnardi et al. that was applied to studies on alcohol and mortality and on silica and lung cancer (12, 13). It consists of fitting a set of regression models that includes (flexible) regression splines and linear models to aggregated data, adjusting for the expected

correlation of estimated (relative) risks within studies. An improvement of this approach over existing meta-regression methods is that the use of regression splines eliminates the need to make strict *a priori* assumptions regarding the shape of the ERC, which allows for a more objective evaluation of its actual shape.

Methodology

Identification of studies and evaluation of study quality

Publications eligible for the meta-regression were identified by a PubMed search that included the MESH keywords 'benzene', 'humans', and 'leukemia' in combination with either 'cohort studies' or 'case-control studies'. Other publications were added by scrutinizing references included in a literature review by Schnatter et al. that was identified in the original PubMed search and in regulatory risk assessments by the Canadian Centre for Occupational Health and Safety, the U.S. National Institute for Occupational Safety and Health, the U.S. Agency for Toxic Substances and Disease Registry and the U.S. Environmental Protection Agency (U.S. EPA) (4, 14-17). The quality of the 11 studies that reported results from quantitative exposure-response analysis for benzene and leukemia (mortality or incidence) was evaluated using a previously developed evaluation framework (18). The first tier of the framework consists of six criteria that are related to crucial aspects of the quality of the design, the quality of conduct, and the quality of the reporting of human observational studies (see Supplemental Material). A study was excluded from the meta-regression if it did not meet all of the six criteria. Nine studies were of sufficient quality to be included in the meta-regression (Table 1). Two studies that reported results from quantitative exposure response analysis were excluded: one expressed exposure in undefined units (19), and the other provided insufficient details that resulted in a lack of insight regarding the decisions made in the statistical analysis (20).

Extraction of data from the incorporated studies

A database was constructed based on published data available for the studies incorporated in the meta-regression. We extracted only risk estimates reported for cumulative exposure to benzene (expressed in ppm-years or ppm-months). The database contained the following fields: study identifier, study design, exposure category, risk estimate, confidence interval for the risk estimate, number of cases and controls for each exposure category (nested case-control studies) or the size of the study population for the exposure category (cohort studies). Three different epidemiologic study designs contributed to the current meta-regression: the nested case-control design (n = 3), the cohort design with an internal reference population (n = 1), and the cohort design with an external reference population (n = 5) (Table 1). Reported odds ratios (ORs), relative risks (RRs) and standardized mortality ratios (SMRs) were combined and interpreted as estimates of the RR for the purpose of this meta-regression (21). The

studies selected for the meta-regression were also different with regard to the definition of the reference population that was used. The cohort studies assumed 'background (environmental) exposure' in their reference populations (22-27). Typical daily environmental exposure to benzene can range up to 0.2 ppm which, over a seventy year live span, accumulates to a maximum of 14 ppm-years of cumulative exposure (6). In the nested case-control studies individuals in the lowest exposure category were used as reference population (ranging from < 0.17 ppm-years to < 1 ppm-years occupational exposure) (28-30).

Preparation of the data extracted from the publications

Three steps were necessary to prepare the extracted data for use in the meta-regression models. In the first step, we assigned a specific cumulative exposure estimate to each risk estimate. It is common practice to report only the boundaries of the exposure categories used in an exposure-response analysis, and this was the case for all included studies except the Swaen study, which reported an average mean exposure (26, 31). To estimate the mean exposure for the assigned exposure categories, we assumed a log-normal distribution for the cumulative exposure for each study. Any data providing information on the exposure distribution within a study were collected from the publication (i.e., the number of person-years per exposure category, the number of controls per exposure category, or the number of expected cases per exposure category). Maximum likelihood estimation was used to fit a Probability Density Function (PDF) to the available data. Using the PDF, we assigned an average cumulative exposure to each exposure category based on its respective boundaries. To avoid unreasonably high estimates of the average exposure in the highest exposure group, we truncated the exposure distribution at the maximum reported cumulative exposure level. This method of assigning specific cumulative exposures is similar to the approach that was proposed by Hartemink et al. (32).

In the second step, we estimated the variance of each specific risk estimate to allow weighting based on the precision of the risk estimates in the meta-regression (33). The variance of RRs and ORs was estimated using the reported confidence intervals following a method discussed by Rothman et al. (34). Estimated variances for studies that reported asymmetrical confidence intervals on the log scale were based on the upper confidence limit only (35, 36).

In the third step, we estimated the covariance between the different risk estimates within a study by applying the approach advocated by Shi and Copas (37). This approach is necessary because risk estimates of a study based on a common internal reference group will be correlated. Ignoring this correlation in the meta-regression underestimates the variance of the risk estimates from the study that results in an overestimation of its weight (12, 31). The Canada-Petrol study lacked the information necessary to estimate the covariance matrix; therefore, uncorrected variances were used for this study. For studies that reported SMRs, we did not estimate covariance because SMRs within a single study can be assumed to be largely

independent when the expected number of deaths used to calculate the SMRs is based on a sufficiently large population.

Application of the (regression) models to describe the exposure-response relation
Natural spline models (with knots at the 20^{th}, 50^{th} and 80^{th} percentiles) as well as linear models were fitted to the data to investigate the shape of the exposure-response relation. To improve the statistical properties of the regression models, we fitted all models to the natural logarithm of the reported risk estimates (38). Regression models were fitted to the data using a modified version of a macro developed by Bagnardi et al. (12). All regression models allowed for (random) study-specific intercepts and exposure effects to accommodate potential between-study heterogeneity (33). Model deviance was used to compare goodness of fit between (nested) models.

Sensitivity analyses
For two cohort studies (Pliofilm and Dow) multiple updates were available. For these cohorts the most recent update was included in the meta-regression. To assess the impact of varying follow-up times in our analysis, we also conducted the meta-regression with risk estimates that were abstracted from earlier updates of the Pliofilm and Dow cohorts (39, 40). Substitution of the risk estimates did not have a substantial effect on the shape of the predicted ERC or on model fit to the data (data not shown). Sensitivity of the predicted ERC to the inclusion of specific studies was assessed with a jackknifing analysis, excluding one study at a time before (re)predicting the exposure-response relation. In addition, we analyzed the cohort studies (including studies with an external reference group as well as the single study with an internal reference group) and nested case-control studies separately and compared their ERC predictions. To allow flexible prediction of the ERC, all sensitivity analyses were done using natural splines.

Prediction of risk estimates
Benzene-leukemia ERC RRs were estimated for four plausible scenarios at three different levels of cumulative exposure (10, 20, and 40 ppm-years; corresponding to 0.25, 0.5 and 1 ppm intensity of exposure over a tenure of 40 years). We used fitted regression models to predict the risk estimates with associated confidence intervals. In addition, corrected risk estimates and confidence intervals were calculated by subtracting the intercept at zero exposure predicted by the regression model from the risk estimates.

Software
Average exposure levels for reported exposure categories were estimated using R v2.7 (R Core Development Group, Vienna, Austria). All other statistical analyses were performed using SAS software for Windows (version 9.1, SAS Institute Inc., Cary, NC, USA).

Results

Nine studies had sufficient quality to be included in the meta-regression. All included studies were performed in the occupational setting. Together these studies provided 30 risk estimates over a range of 0.32–554.3 (assigned) ppm-years (Figure 1A). Nineteen (63%) of the risk estimates were assigned a cumulative exposure < 50 ppm-years (Figure 1B). Most of the risk estimates for the lower exposure range were provided by nested case-control studies. The differences in exposure levels between studies can be largely attributed to the different industries in which the studies were performed. The nested case-control studies were all performed in the petroleum industry, whereas the cohort studies were performed in the chemical industry (Pliofilm, Dow, Wong, and Swaen studies), in a shoe factory (Costantini study), or a wide range of different industries (Chinese Academy of Preventive Medicine-National Cancer Institute (CAPM-NCI) study) (Table 1).

Predictions of the ERC based on a natural spline model and a linear model are presented in Figure 2. The lower deviance of the natural spline (deviance = 25.84, 27 df) compared to that of the linear model (deviance = 29.25, 28 df) suggests a slightly better fit (chi square test (1 df), $p = 0.06$). The natural spline model also indicates a strong supralinear shape of the ERC in the low-exposure region, resulting in a considerable lower intercept than the linear model (RR = 1.33 vs. 1.65).

Results from a jackknifing analysis (Figure 3) suggest that the Pliofilm and CAPM-NCI studies were particularly influential for the (high-exposure region of the) predicted ERC. Exclusion of the Plioflim study from the meta-regression resulted in a strong reduction of risks predicted for cumulative exposures > 100 ppm-years, whereas exclusion of the CAPM-NCI study had the opposite effect. Exclusion of other studies had little impact on the predicted ERC.
Stratified analyses showed that study design had a considerable impact on the predicted ERC (Figure 4). The ERC based on the cohort studies had a similar shape compared with the ERC based on the full data. However, the supralinear shape was somewhat less pronounced and the predicted intercept slightly lower (RR = 1.13 vs. 1.33). The deviance of the natural spline model fitted to the cohort studies was smaller than the deviance of the corresponding linear model (deviance = 8.43 and 11.97 respectively; chi-square test (1 df), p = 0.06). The analysis based on the three nested case-control studies resulted in extremely wide confidence intervals around the predicted ERC and was essentially uninformative (Figure 4).

Table 1 Details of the studies included in the meta-regression.

Study	Year	Study design	Risk estimates	Country	Industry	Reference category	Lowest exposure category[a]	Upper exposure category	Study outcome	ICD code (revision)[b]	Size of the study population	Reference
Wong	1987	Cohort	SMR	USA	Chemical industry	National population death rates	< 15 ppm-yrs	≥60 ppm-yrs	Mortality	204-207 (8)	7676 individuals, 6 cases	(27)
CAPM-NCI	1997	Cohort	RR	China	Variety of industries	Workers employed in work units or factories where benzene was not used	< 40 ppm-yrs	> 100 ppm-yrs	Incidence	204-208 (9)	74828 exposed, 35805 unexposed, 47 cases	(24)
Pliofilm	2002	Cohort	SMR	USA	Chemical industry	National population death rates	0.01–40 ppm-yrs	> 400 ppm-yrs	Mortality	204-208[c]	1291 individuals, 15 cases	(25)
Costantini	2003	Cohort	SMR	Italy	Shoe factory	National and regional specific death rates	< 40 ppm-yrs	> 200 ppm-yrs	Mortality	204-207 (8)	1687 individuals, 11 cases	(23)
DOW	2004	Cohort	SMR	USA	Chemical industry	National and regional specific death rates	< 28.3 ppm-yrs	> 79.1 ppm-yrs	Mortality	204-208 (9)	2266 individuals, 12 cases	(22)
Swaen[d]	2005	Cohort	SMR	The Netherlands	Chemical industry	National population death rates	3.4 ppm-yrs	401.5 ppm-yrs	Mortality	Not available[e]	311 individuals, 1 case	(26)
Canada petrol	1996	Nested case-control	OR	Canada	Petroleum industry	Workers exposed to < 0.17 ppm-yrs	0.18-0.49 ppm-yrs	8.0-219.8 ppm-yrs	Incidence	204-207 (8)	14 cases, 55 controls	(30)

UK-Petrol	1997	Nested case-control	OR	UK	Petroleum industry	Workers exposed to < 0.26 ppm-yrs	0.26-0.59 ppm-yrs	> 4.79 ppm-yrs	Mortality / Incidence	204-208 (9)	90 cases, 354 controls	(29)
AHW	2003	Nested case-control	OR	Australia	Petroleum industry	Workers exposed to ≤1 ppm-yrs	1-2 ppm-yrs	> 16 ppm-yrs	Incidence	204-208 (9)	33 cases and 165 controls	(28, 50)

[a] Lowest exposure category for which a risk estimate was reported (excluding reference category).
[b] International Classification of Diseases (ICD) code and revision that was used to assign disease outcomes to 'Leukemia' category. ICD 204: lymphoid leukemia; 205: myeloid leukemia; 206: monocytic leukemia; 207: other specified leukemia; 208: leukemia of unspecified cell type.
[c] ICD code for 'Leukemia' category in effect at time of death of the cases.
[d] Average mean exposure for tertiles of the exposure distribution; because of a lack of observed cases, a risk estimate was reported only for the middle tertile.
[e] Disease categorization was based on the Dutch electronic causes of death file.

Table 2 Comparison of predicted relative risks for three cumulative exposure levels.

			RR (95% CI)		
Model	Deviance (df)	Intercept	10 ppm-yrs	20 ppm-yrs	40 ppm-yrs
Predictions meta regression - all studies					
Scenario A: Natural spline	25.84 (27)	1.33 (0.87-2.05)	1.52 (1.08-2.15)	1.73 (1.27-2.34)	2.11 (1.51-2.96)
Scenario A corrected for intercept			1.14 (1.04-1.26)	1.29 (1.07-1.56)	1.59 (1.15-2.19)
Scenario B: Natural spline without intercept	28.39 (28)	NA	1.22 (1.11-1.34)	1.46 (1.22-1.75)	1.96 (1.44-2.68)
Scenario D1: Linear model without intercept	38.67 (29)	NA	1.05 (1.02-1.07)	1.10 (1.05-1.15)	1.20 (1.09-1.32)
Prediction meta regression - cohort studies					
Scenario C: Natural spline	8.43 (15)	1.13 (0.71-1.81)	1.25 (0.83-1.88)	1.38 (0.96-1.97)	1.67 (1.22-2.27)
Scenario C corrected for intercept			1.10 (1.04-1.17)	1.22 (1.09-1.36)	1.48 (1.19-1.83)
Scenario D2: Linear model without intercept	15.95 (17)	NA	1.04 (1.02-1.07)	1.09 (1.04-1.14)	1.19 (1.09-1.31)

NA, not applicable.

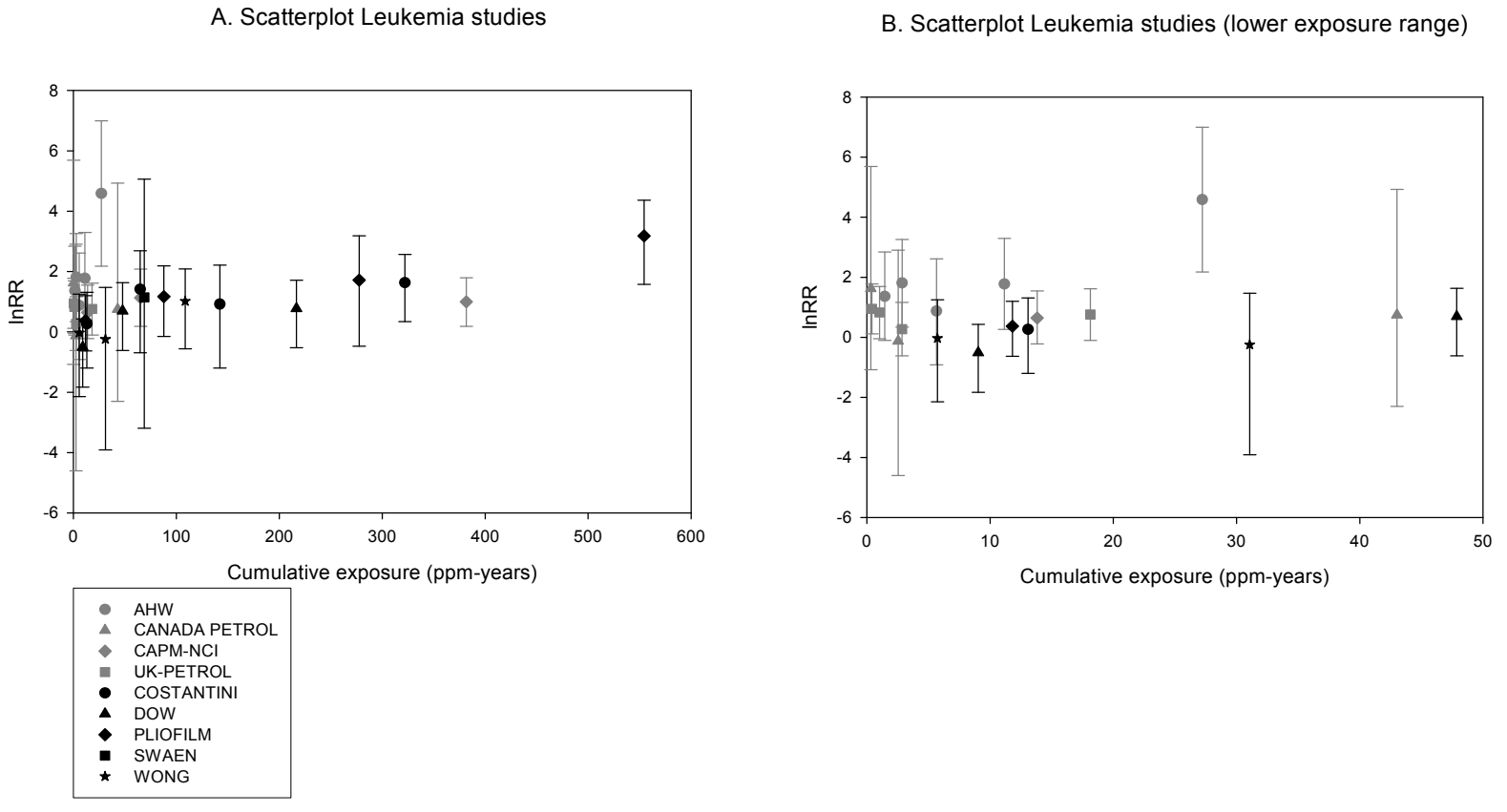


Figure 1 Scatter plot of the risk estimates extracted from the nine studies included in the meta-regression based on the assigned average cumulative exposure: full range of cumulative exposures (A) and cumulative exposures < 50 ppm-years (B).

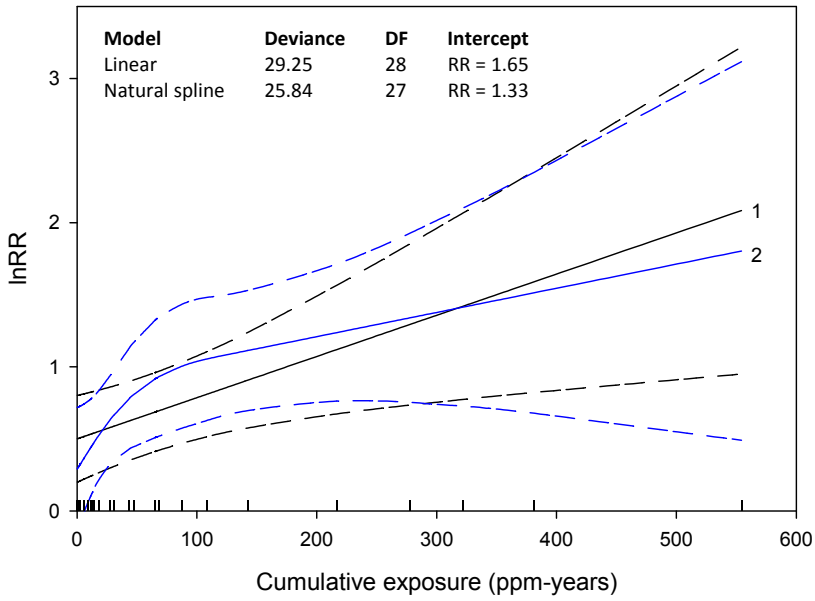


Figure 2 Predicted exposure-response curve using all risk estimates from the nine included studies based on a natural spline and linear regression model. Plot 1 is the predicted ERC based on a linear model. Plot 2 is the predicted ERC based on a natural spline model (knots are located at 2.9, 22.7, and 125.5 ppm-years). Dashed lines represent the 95% CIs of the predictions. Rug plot indicates the distribution of the estimated cumulative exposure for each risk estimate included in the analyses.

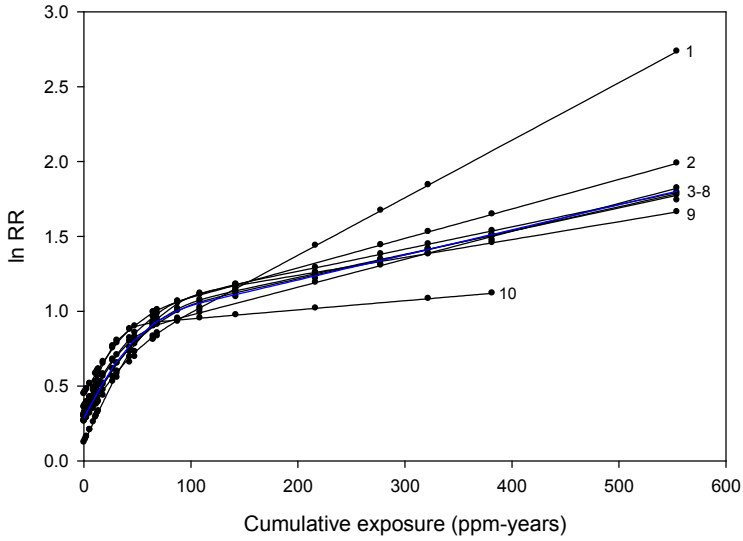


Figure 3 Sensitivity analysis of studies on the prediction of the exposure-response curve based on a natural spline. Graph represents nine plots of the predicted ERC based on all studies minus one. The plots are identified by the study that was excluded: 1, CAPM-NCI; 2, DOW; 3, Costantini; 5, U.S.-Chemical; 6, Swaen; 7, Canada-Petrol; 8, AHW; 9, UK-Petrol; and 10, Pliofilm. Plot 4 is the predicted ERC based on all available studies.

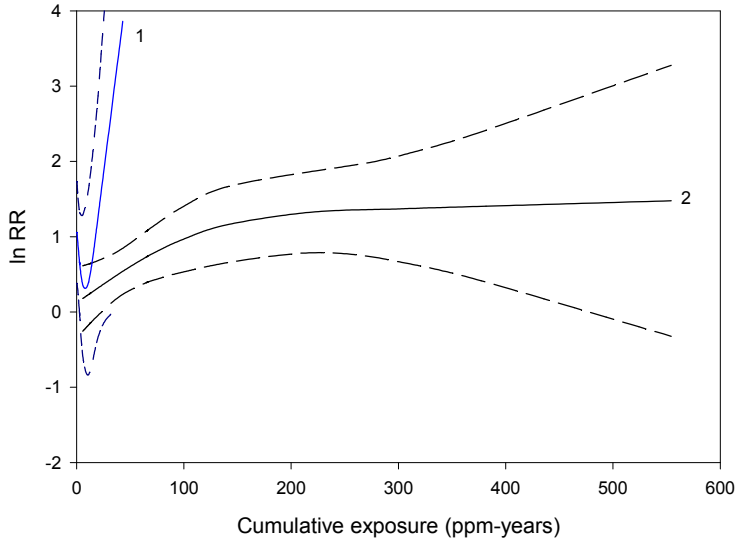


Figure 4 Predicted exposure-response curve stratified by study design based on a natural spline. Plot 1 is the predicted ERC based on only the nested case-control studies (knots are located at 1.0, 2.9, and 18.1 ppm-years). Plot 2 is the predicted ERC based on only the cohort studies (knots are located at 13.1, 67.1, and 277.6 ppm-years). Dashed lines represent the 95% CIs of the predictions.

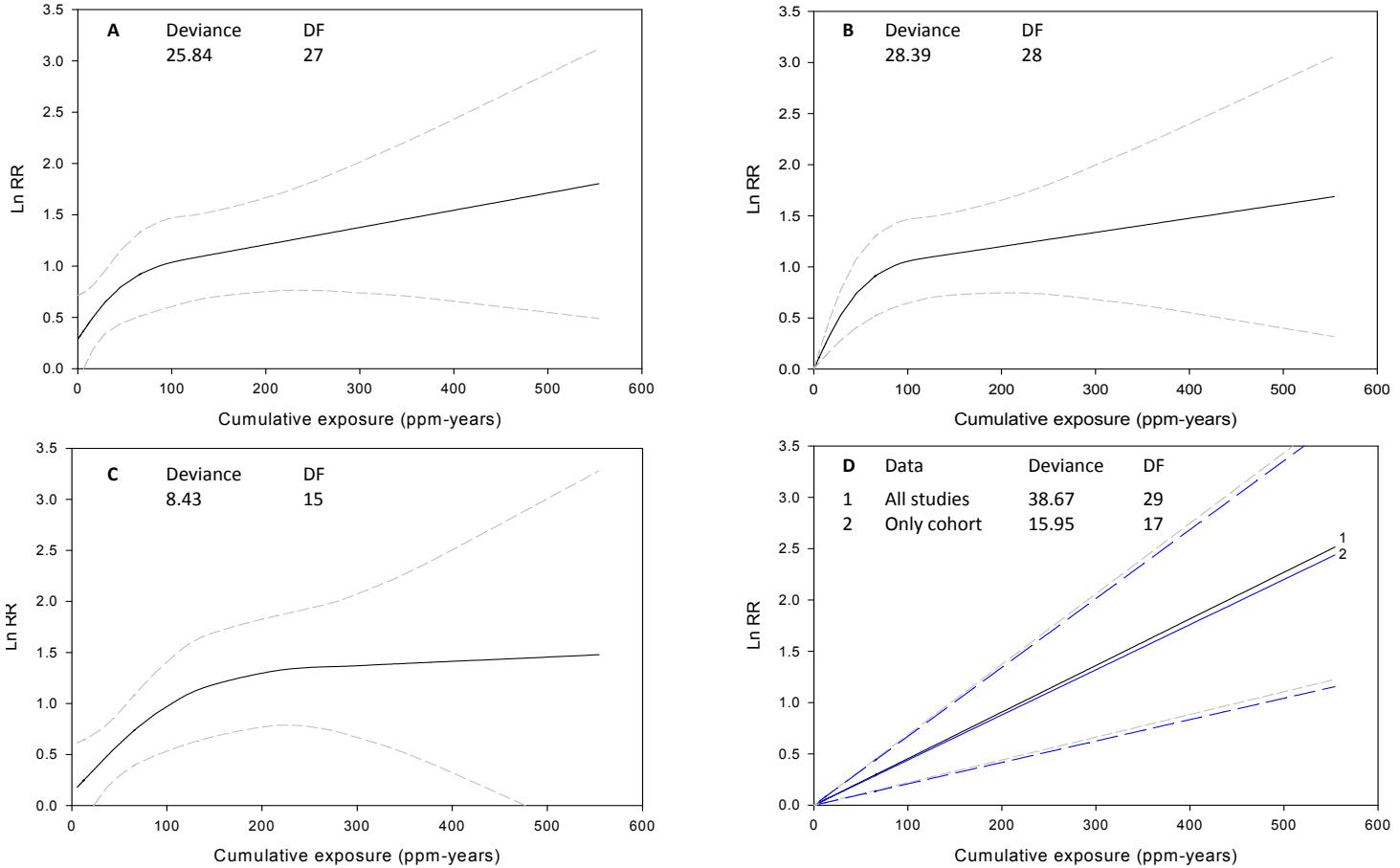


Figure 5 Four different scenarios for the shape of the benzene leukemia exposure-response curve The four scenarios are: (A) Natural spline with intercept fitted to all studies (best fitting model), (B) Natural spline without intercept fitted to all studies, (C) Natural spline with intercept fitted to the cohort studies, (D) Linear model without intercept fitted to the all the studies (1) and only the cohort studies (2). Dashed lines are 95% CIs.

We predicted the RRs for leukemia for three cumulative exposure levels (10, 20, 40 ppm-years) based on four different modeling scenarios for the shape of the benzene leukemia ERC (Table 2). The four scenarios were a natural spline with intercept fitted to all studies (scenario A), a natural spline without intercept fitted to all studies (scenario B), a natural spline with intercept fitted to the cohort studies (scenario C), and a linear model without intercept fitted to all studies (scenario D1) or only the cohort studies (scenario D2) (Figure 5).
Scenario A predicted the highest RRs (RR = 1.52, 1.73, and 2.11 for cumulative exposures of 10, 20 and 40 ppm-years respectively), although these dropped considerably (RR = 1.14, 1.29, and 1.59) after correction for the predicted intercept. RRs predicted in scenario B (RR = 1.22, 1.46, and 1.96) were somewhat lower than the uncorrected RRs from scenario A. Predictions using data from the cohort studies only (scenario C) were also lower than those predicted by scenario A (all studies) (RR = 1.25, 1.38, 1.67), although these differences largely disappeared after we corrected for the intercept. Finally, predictions of the RRs based on scenario D1 (RR = 1.04, 1.09, 1.19) and D2 (RR = 1.05, 1.10, 1.20) were very similar and considerably lower than the predictions based on the other scenarios.

Discussion

Interpretation of the predicted exposure-response curve
We presented estimates of the benzene-leukemia ERC based on predictions from two regression models. Both the natural spline and linear regression models indicated a positive relation between cumulative exposure to benzene and leukemia risk, although risk appeared to increase more strongly at low exposures in the natural spline model. This supralinear shape of the natural spline model at low exposures is consistent with the increasing evidence that saturable metabolism plays an important role in the low-dose carcinogenicity of benzene (7-10, 41-44).

Alternative explanantions for the non-linear relation between (inhalatory) benzene exposure and leukemia that we found are depletion of susceptible individuals at high benzene exposure levels and bias due to attenuation of the exposure-response relation within or between studies (45). Attenuation of exposure-response relations is commonly observed in occupational studies and can be caused by several factors, including the healthy worker survivor effect, high disease background rates, exposure measurement error, and confounding and effect modification (45). However, not all of these factors are equally likely to have played a role in occupational studies on benzene and leukemia. Confounding should be considered a potential factor that might have introduced attenuation, but none of the included studies demonstrated (or was able to demonstrate) a confounding effect from potential confounders such as ionizing radiation, smoking and family history of leukemia (46). In addition, it is unlikely that these factors could have caused serious distortion of the study

findings considering the general lack of association with exposure to benzene across the assessed industries (ionizing radiation), weak association with leukemia (smoking), and rare occurrence (family history of leukemia) (46).
Factors that contributed to the heterogeneity observed between studies were differences in study design and exposure assessment. Although all studies were comparable regarding the context of exposure (occupational exposure), considerable differences exist in the geographical location, type of industry, and intensity and frequency of exposure to benzene (Table 1). Differences in study design resulted in different types of risk estimates that were reported: RRs, ORs and SMRs (Table 1). However, ORs and SMRs can be interpreted as reasonable approximations of the RR when the disease is rare, and these measures have been pooled with RRs for meta-analysis in previous analyses (21, 47). According to our evaluation, the quality of exposure assessment was sufficient in all included studies (18). However, systematic differences in exposure assessment strategies between studies might have contributed to the between-study heterogeneity. Because all included studies assessed exposure retrospectively based on a relatively limited set of exposure measurements, exposure estimation in these studies was based partly on decision rules to extrapolate exposure measurements to time periods and exposure circumstances for which no measurements were available (22-30). The significant amount of expert judgment that goes into these decision rules makes it conceivable that systematic differences in exposure assessment may exist between studies. This situation is illustrated by the exposure assessment for the Pliofilm cohort where three groups of authors have published three different sets of exposure estimates, based on the same exposure measurement data (1, 25, 40, 48). In contrast, the three nested case-control studies attempted to limit systematic error in exposure assessment by applying similar exposure assessment approaches (28-30).

We tried to assess the potential impact of between-study heterogeneity in sensitivity analyses. Visual inspection of the results from a jackknifing analysis (Figure 3) showed that two studies that provided risk estimates for the highest assigned cumulative exposures had a considerable impact on the ERC for the higher exposure range. The impact on the lower exposure range was less pronounced. Exclusion of the Australian Health Watch (AHW) study, which reported relatively high risks for the low exposure range (Figure 1), had little impact on the shape of the ERC. Results from the sensitivity analysis stratified by study design indicated considerable differences between the ERC based on the nested case-control studies from the petroleum industry and the ERC based on the cohort studies. The shape of the ERC based on the cohort studies only was similar to the shape of the ERC based on all studies. Although the shape of the ERC for the nested case-control studies could be estimated only very imprecisely, the results indicated that these studies were largely responsible for the rather high intercepts that were predicted for the ERCs based on the full data.

Assessing publication bias in a flexibe meta-regression is complicated because no standard statistical approaches are available to deal with the correlated effect estimates and nonlinear exposure-response relations. However, in our opinion it is unlikely that studies with quantitative benzene exposure estimates would not have reported risk estimates for leukemia even if these had been negative because this is one of the major cancer outcomes associated with the exposure. Also, considering the large effort that is required to generate quantitative benzene exposure estimates, it is improbable that any study that had quantitative exposure estimates available would not have been published at all.

Considering that risk estimates in the included studies were calculated in reference to populations with assumed no or negligible occupational exposure to benzene, one might have expected a predicted marginal intercept (lnRR) of approximately 0 (RR = 1) at 0 ppm-years. Although intercepts above 0 are frequently observed in exposure-response studies based on epidemiologic data, most other meta-regression studies have avoided the issue by forcing their regression models to fit through the origin (12, 31). We did include intercepts in our regression models to attain the best possible fit to the data, which resulted in intercepts of RRs of 1.33 and 1.65 in the natural spline and linear regression models, respectively. An explanation for these nonzero intercepts may be the lack of risk estimates for very low exposure levels (< 0.32 ppm-year). Risk estimates at slightly higher, but still very low, exposure levels (ppm-years) already indicate a strong increase in risk. A natural spline, being linear in its tails, may be unable to track the curvature of the ERC at these low levels, resulting in a nonzero intercept. The same reasoning would apply to intercepts from the linear model, although the effect may be even more extreme. However we cannot exclude the effect of lower-than-expected leukemia risk in the reference populations or (conversely) higher non-benzene-related leukemia risk in the exposed populations. Finally, attenuation of the ERC due to random and systematic error in the exposure assessment might also have forced the intercept up (45, 49).

Implication of the findings for quantitative risk assessment

To facilitate the use of our meta-regression results in Quantitative Risk Assessment (QRA) we provided the benzene leukemia ERC for three plausible scenarios (Figure 5, Table 2). In scenario A, the natural spline model was used, which fitted the data slightly better than a linear model. However, the estimate of leukemia risk at 0 ppm-years (the intercept) for this model was much lower than that for the linear model. Because application of these models for risk assessment purposes will most likely entail subtraction of the intercept from all predictions (effectively lowering the predicted (increased) risks for benzene at each exposure level), we favor scenario A (Figure 5A) for risk prediction over the alternatives (Table 2). If one believes that the intercept is due to the natural spline failing to track the shape of the ERC at very low exposures, one may prefer predictions from a spline model without an intercept, thus forcing the predicted ERC through the origin (scenario B, Figure 5B). This model fitted the

data only slightly worse than scenario A and may therefore be considered a plausible alternative (chi-square test (1 df), p = 0.11). If one believes that the nested case-control studies should not be used for prediction of the ERC, scenario C (Figure 5C) could be used, which was based on a natural spline model with an intercept fitted to data from the cohort studies only. This scenario resulted in slightly lower predicted risks for exposures < 100 ppm-years. Finally, scenarios D1 (all studies, Figure 5D) and D2 (cohort studies only) were based on linear models without an intercept and are therefore similar in spirit to models commonly used in QRA (3, 4). Clearly, these models fitted the data worse than the relevant alternative models in scenarios A (chi-square test (2 df), p = 0.002) and C (chi-square test (2 df), p = 0.02). To compare predictions from scenarios A-D with predictions from existing QRAs, we estimated RRs for three cumulative exposures in the low exposure range (10, 20, and 40 ppm-years) (Table 2). This showed that risk estimates from our models at these exposures are very similar to those based on the (multiplicative) models used by the U.S. EPA and the California Environmental Protection Agency in their QRA of benzene (3, 4). However, while these QRAs were based on data from either the Pliofilm study or the CAPM-NCI study, our approach allowed us to use all available epidemiological evidence to date and should therefore be more robust. In addition, our approach allowed for a nonlinear shape of the ERC to be used in QRA, which appears to be more appropriate. It is important to note that all analyses were performed on the overarching disease outcome 'leukemia'. Slight differences in the definition of this disease existed between studies (Table 1). Unfortunately, analyses for specific subtypes of leukemia will be hampered by a lack of data.

Conclusion

Flexible meta-regression of the aggregated risk estimates from a set of occupational human observational studies offers an efficient approach to acquiring more insight into the functional relation between exposure to benzene and leukemia. The flexible meta-regression model predicted a supralinear shape of the ERC. Although the limited number of available studies and the large heterogeneity between studies were considerable limitations, sensitivity analyses demonstrated that results were not strongly affected. Our application of a flexible meta-regression method provides the most comprehensive evaluation of the benzene-leukemia ERC to date.

Acknowledgements

This work was performed as part of the work package "Integrated Risk Assessment" of the ECNIS Network of Excellence (Environmental Cancer Risk, Nutrition and Individual Susceptibility), operating within the European Union 6th Framework Program, Priority 5:

"Food Quality and Safety" (FOOD-CT-2005-513943). The authors thank Dr. Bagnardi for providing his macro "flexible". The macro used in our study is available on request by the author.

References

1. Crump KS. Risk of benzene-induced leukemia predicted from the Pliofilm cohort. Environ Health Perspect. 1996;104 Suppl 6:1437-41.
2. Paustenbach DJ, Bass RD, Price P. Benzene toxicity and risk assessment, 1972-1992: implications for future regulation. Environ Health Perspect. 1993;101 Suppl 6:177-200.
3. California EPA. Public Health Goal for Benzene In Drinking Water. Scramento, California: Office of Environmental Health Hazard Assessment, 2001.
4. U.S. EPA. Carcinogenic Effects of Benzene: An Update. EPA/600/P-97/001F. Washington, DC: United States Environmental Protection Agency, 1998.
5. Capleton AC, Levy LS. An overview of occupational benzene exposures and occupational exposure limits in Europe and North America. Chem Biol Interact. 2005;153-154:43-53.
6. Johnson ES, Langard S, Lin YS. A critique of benzene exposure in the general population. Sci Total Environ. 2007;374(2-3):183-98.
7. Kim S, Vermeulen R, Waidyanatha S, et al. Using urinary biomarkers to elucidate dose-related patterns of human benzene metabolism. Carcinogenesis. 2006;27(4):772-81.
8. Kim S, Vermeulen R, Waidyanatha S, et al. Modeling human metabolism of benzene following occupational and environmental exposures. Cancer Epidemiol Biomarkers Prev. 2006;15(11):2246-52.
9. Rothman N, Bechtold WE, Yin SN, et al. Urinary excretion of phenol, catechol, hydroquinone, and muconic acid by workers occupationally exposed to benzene. Occup Environ Med. 1998;55(10):705-11.
10. Lan Q, Vermeulen R, Zhang L, et al. Benzene exposure and hematotoxicity [letter]. Science. 2006;312(5776):998-9.
11. Lan Q, Zhang L, Li G, et al. Hematotoxicity in workers exposed to low levels of benzene. Science. 2004;306(5702):1774-6.
12. Bagnardi V, Zambon A, Quatto P, et al. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. Am J Epidemiol. 2004;159(11):1077-86.
13. Lacasse Y, Martin S, Gagne D, et al. Dose–response meta-analysis of silica and lung cancer. Cancer Causes Control. 2009;20(6):925-33.
14. ATSDR. Toxicological Profile for Benzene. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, 2007.
15. NIOSH. Revised Recommendation for an Occupational Exposure Standard for Benzene. Washington DC: National Institute for Occupational Safety and Health, 1976.
16. Penney J. Report to the Occupational Disease Panel (Industrial Disease Standards Panel) on Occupational Exposure to Benzene and Leukaemia. Hamilton, ON: Canadian Centre for Occupational Health and Safety:1995.
17. Schnatter AR, Rosamilia K, Wojcik NC. Review of the literature on benzene exposure and leukemia subtypes. Chem Biol Interact. 2005;153-154:9-21.
18. Vlaanderen J, Vermeulen R, Heederik D, et al. Guidelines to evaluate human observational studies for quantitative risk assessment. Environ Health Perspect. 2008;116(12):1700-5.
19. Guenel P, Imbernon E, Chevalier A, et al. Leukemia in relation to occupational exposures to benzene and other agents: a case-control study nested in a cohort of gas and electric utility workers. Am J Ind Med. 2002;42(2):87-97.
20. Collins JJ, Ireland B, Buckley CF, et al. Lymphohaematopoeitic cancer mortality among workers with benzene exposure. Occup Environ Med. 2003;60(9):676-9.
21. McElvenny DM, Armstrong BG, Jarup L, et al. Meta-analysis in occupational epidemiology: a review of practice. Occup Med (Lond). 2004;54(5):336-44.
22. Bloemen LJ, Youk A, Bradley TD, et al. Lymphohaematopoietic cancer risk among chemical workers exposed to benzene. Occup Environ Med. 2004;61(3):270-4.

23. Seniori Costantini A, Quinn M, Consonni D, et al. Exposure to benzene and risk of leukemia among shoe factory workers. Scand J Work Environ Health. 2003;29(1):51-9.
24. Hayes RB, Yin SN, Dosemeci M, et al. Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine—National Cancer Institute Benzene Study Group. J Natl Cancer Inst. 1997;89(14):1065-71.
25. Rinsky RA, Hornung RW, Silver SR, et al. Benzene exposure and hematopoietic mortality: A long-term epidemiologic risk assessment. Am J Ind Med. 2002;42(6):474-80.
26. Swaen GM, Scheffers T, de Cock J, et al. Leukemia risk in caprolactam workers exposed to benzene. Ann Epidemiol. 2005;15(1):21-8.
27. Wong O. An industry wide mortality study of chemical workers occupationally exposed to benzene. II. Dose response analyses. Br J Ind Med. 1987;44(6):382-95.
28. Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology. 2003;14(5):569-77.
29. Rushton L, Romaniuk H. 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. Occup Environ Med. 1997;54(3):152-66.
30. Schnatter AR, Armstrong TW, Nicolich MJ, et al. Lymphohaematopoietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. Occup Environ Med. 1996;53(11):773-81.
31. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. Am J Epidemiol. 1992;135(11):1301-9.
32. Hartemink N, Boshuizen HC, Nagelkerke NJ, et al. Combining risk estimates from observational studies with different exposure cutpoints: a meta-analysis on body mass index and diabetes type 2. Am J Epidemiol. 2006;163(11):1042-52.
33. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88.
34. Rothman KJ, Boice JD, Austin H. Epidemiologic analysis with a programmable calculator. New ed. Boston, Mass.: Epidemiology Resources, 1982.
35. D. BJ,Jr, McLaughlin JK. Errors in TCE analysis. Environ Health Perspect. 2001;109(3):A108-9.
36. Wartenberg D, Reyner D, Scott CS. Trichloroethylene and cancer: epidemiologic evidence. Environ Health Perspect. 2000;108 Suppl 2:161-76.
37. Shi JQ, Copas JB. Meta-analysis for trend estimation. Stat Med. 2004;23(1):3,19; discussion 159-62.
38. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. Epidemiology. 1993;4(3):218-28.
39. Bond GG, McLaren EA, Baldwin CL, et al. An update of mortality among chemical workers exposed to benzene. Br J Ind Med. 1986;43(10):685-91.
40. Rinsky RA, Smith AB, Hornung R, et al. Benzene and leukemia. An epidemiologic risk assessment. N Engl J Med. 1987;316(17):1044-50.
41. Rappaport SM, Kim S, Lan Q, et al. Evidence that humans metabolize benzene via two pathways. Environ Health Perspect. 2009;117(6):946-52.
42. Rappaport SM, Waidyanatha S, Qu Q, et al. Albumin adducts of benzene oxide and 1,4-benzoquinone as measures of human benzene metabolism. Cancer Res. 2002;62(5):1330-7.
43. Rappaport SM, Yeowell-O'Connell K, Smith MT, et al. Non-linear production of benzene oxide-albumin adducts with human exposure to benzene. J Chromatogr B Analyt Technol Biomed Life Sci. 2002;778(1-2):367-74.
44. Rappaport SM, Waidyanatha S, Yeowell-O'Connell K, et al. Protein adducts as biomarkers of human benzene metabolism. Chem Biol Interact. 2005;153-154:103-9.
45. Stayner L, Steenland K, Dosemeci M, et al. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environ Health. 2003;29(4):317-24.
46. Zeeb H, Blettner M. Adult leukaemia: what is the role of currently known risk factors? Radiat Environ Biophys. 1998;36(4):217-28.
47. Steinmaus C, Smith AH, Jones RM, et al. Meta-analysis of benzene exposure and non-Hodgkin lymphoma: biases could mask an important association. Occup Environ Med. 2008;65(6):371-8.
48. Paustenbach DJ, Price PS, Ollison W, et al. Reevaluation of benzene exposure for the Pliofilm (rubberworker) cohort (1936-1976). J Toxicol Environ Health. 1992;36(3):177-231.

49. Berman DW, Crump KS. Update of potency factors for asbestos-related lung cancer and mesothelioma. Crit Rev Toxicol. 2008;38 Suppl 1:1-47.
50. Atkinson S, Coppock J, Fritschi L, et al. Lympho-haematopoietic Cancer and Exposure to Benzene in the Australian Petroleum Industry. Technical Report and Appendices. Melbourne Australia: Monash University and Deakin University, 2001.

Supplemental Material: Criteria used to evaluate eligibility of studies for inclusion in the meta-regression

Table 1 includes six evaluation criteria that were used to assess the quality of the studies that were potentially eligible for inclusion in the benzene-leukemia meta-regression. Justification and application of the criteria were discussed in detail in Chapter 2.

Supplemental Material, Table 1 Evaluation criteria.

Criteria	Possible outcome	Required outcome
1 Is the study design case-control, cohort or cross-sectional?	Yes / No	Yes
2 Is exposure expressed on a ratio scale and specific for the agent of interest?	Yes / No	Yes
3 Is a detailed description of the statistical analysis provided?	Yes / No	Yes
4 Are criteria for inclusion of subjects into the study described with sufficient detail?	Yes / No	Yes
5 Is the assessment of the health effect performed according to recognized norms?	Yes / No	Yes
6 Are all known potential strong confounding factors considered in the study design?	Yes / No	Yes

Chapter 6

Application of OMICS Technologies in Occupational and Environmental Health Research; Current Status and Projections

Jelle Vlaanderen
Lee E. Moore
Martyn T. Smith
Qing Lan
Luoping Zhang
Christine F. Skibola
Nathaniel Rothman
Roel Vermeulen

Occupational and Environmental Medicine. 2010;67(2):136-43.

Abstract

OMICS technologies are relatively new biomarker discovery tools that can be applied to study large sets of biological molecules. Their application in human observational studies (HOS) has become feasible in recent years due to a spectacular increase in the sensitivity, resolution and throughput of OMICS-based assays. Although, the number of OMIC techniques is ever expanding, the five most developed OMICS technologies are genotyping, transcriptomics, epigenomics, proteomics and metabolomics. These techniques have been applied in HOS to various extents. However, their application in occupational environmental health (OEH) research has been limited. Here, we will discuss the opportunities these new techniques provide for OEH research. In addition we will address difficulties and limitations to the interpretation of the data that is generated by OMICS technologies. To illustrate the current status of the application of OMICS in OEH research, we will provide examples of studies that used OMICS technologies to investigate human health effects of two well-known toxicants, benzene and arsenic.

Introduction to OMICS technologies

In the biological sciences the suffix –omics is used to refer to the study of large sets of biological molecules (1). The idea that the field of molecular biology needed to move from studying isolated biological molecules towards a broad analysis of large sets of biological molecules was underscored with the completion of human genome project (HGP) in 2001 (2, 3). The HGP demonstrated that a relatively limited number of genes could be identified in the human genome, which substantiated the theory that complex biological processes were regulated on other levels than DNA sequence alone. This realization triggered the rapid development of several fields in molecular biology that together are described with the term 'OMICS'. The OMICS field ranges from genomics (focused on the genome) to proteomics (focused on large sets of proteins, the proteome) and metabolomics (focused on large sets of small molecules, the metabolome). We divide the field of genomics into genotyping (focused on the genome sequence), transcriptomics (focused on genomic expression) and epigenomics (focused on epigenetic regulation of genome expression). An overview of the different OMICS fields that will be discussed in this paper is presented in Table 1. In this review we define the field of occupational and environmental health (OEH) research as the study of interactions between the following domains: environment (the exposome) (4), individual (genetic) susceptibility (the (epi)genome), and biological outcomes (the responsome) (5) (Figure 1). In this context, biological outcomes can be defined as clinical diseases as well as relevant (pre-clinical) intermediate endpoints. In theory, OMICS technologies have a large potential value for OEH research because the environment is known to influence many of the described processes and therefore OMICS technologies are likely to provide valuable information especially where the three domains overlap. Although the field of OMICS is ever expanding (e.g., see http://omics.org), currently five different OMICS fields are well established: genotyping, gene expression profiling, epigenomics, proteomics and metabolomics. In this paper, we will address the spectacular increase in sensitivity, resolution and throughput of OMICS-based techniques in recent years, and we will discuss the difficulties regarding the interpretation of data generated by these techniques. To illustrate the current status of the application of OMICS in OEH research and the progress that has been made in recent years, we will provide examples of studies that have used OMICS technologies to investigate human health effects of two well-known environmental/occupational toxicants, benzene and arsenic.

Table 1 Overview of the different OMICS technologies.

Technology	Molecules of interest	Definition	Temporal variance	Influence by disease status
Genotyping	DNA	Assessment of variability in DNA sequence in the genome	None	No
Epigenomics	Epigenetic modifications of DNA	Assessment of factors that regulate gene expression without changing DNA sequence of the genome	Low / Moderate	Probable
Gene expression profiling	RNA	Assessment of variability in composition and abundance of the transcriptome	High	Yes
Proteomics	Proteins	Assessment of variability in composition and abundance of the proteome	High	Yes
Metabolomics	Small molecules	Assessment of variability in composition and abundance of the metabolome	High	Yes

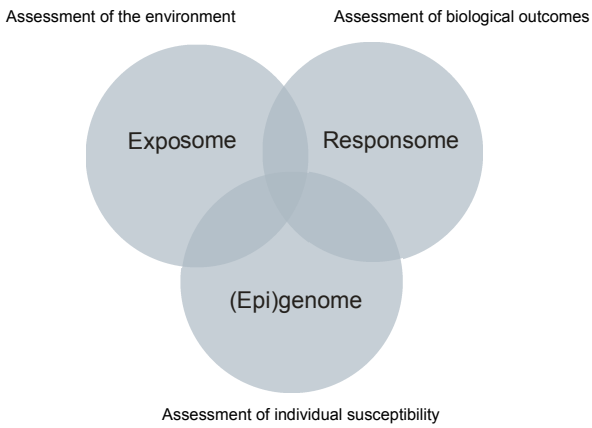


Figure 1 OMICS within the domain of OEH. Genotyping operates completely within the domain of genomics. The other OMICS technologies operate in the intersection between the exposome (assessment of the environment), the responsome (assessment of health effects) and the (epi)genome (assessment of individual genetic susceptibility).

Overview OMICS technologies

Genomics

We divide the field of genomics into genotyping, transcriptomics, and epigenomics.

Genotyping

Genotyping is focused on the identification of the physiological function of genes and the elucidation of the role of specific genes in disease susceptibility (6). The HGP has provided insight into the number of genes and their location in the human genome (2, 3, 7). This knowledge in combination with major technological improvements resulted in the development of assays that are able to assess variability in the DNA sequence of many thousands of genes in a single experiment. This development has opened the possibility to study the combined effect of variability in multiple genes on the development of complex diseases. While several types of genetic variation exist (e.g., insertions and deletions of nucleotide base pairs and copy number variation (CNVs)), single nucleotide polymorphisms (SNPs) are the most commonly investigated (2). At this moment over 9 million detected SNPs are available in public databases (8, 9). Because SNPs are highly abundant in the human genome, they are commonly used as markers for genetic variation in disease-gene association studies (10). Due to limited genetic variation and haplotype structure and a high level of linkage disequilibrium within small regions of the genome, a subset of informative SNPs, called tag SNPs, can be genotyped as proxies for haplotype blocks to identify regional associations that influence disease or phenotypes of interest (11). Fine mapping (e.g., sequencing) can further narrow the associated region in the search for the true causal variant(s). However, functional studies are needed to test whether associated SNPs alter the structure or function of DNA, RNA or proteins and influence phenotypes. Among others, functional SNPs might alter peptide sequences, transcription factor binding sites and exonic splicing enhancer/suppressor sites.

The first SNP-based studies focused on one or more SNPs per gene in a limited set of candidate genes. However, since the introduction of array-based genotyping techniques, allowing the simultaneous assessment of up to one million SNPs in a single assay, it has become possible to cover, with varying resolution, the entire genome in what are now commonly referred to as genome-wide association studies (GWAS). These GWAS have uncovered, and will continue to uncover, interesting and previously unknown polymorphic variants that are associated with a variety of chronic diseases. The effect sizes of these findings have in general been small (OR 1.2-1.5) fuelling debates on positive interactions between one or more common variants and the environment (12). Yet, identifying these gene-environment interactions will be difficult in ongoing GWAS given the low prevalence of exposures and/or the poor characterization of environmental exposures in these large, often multi-center/country studies. As such, OEH research can play an important role in the

identification of gene-environment interactions as the exposure is more prevalent and assessed with greater accuracy than in population- or hospital-based case-control studies that have provided most GWAS to date. Of course, sample sizes will likely be much smaller in these studies limiting the statistical power, and therefore the number of SNPs that can be tested simultaneously. Until recently most OEH studies on gene-environment have been focused on candidate genes, where the success depends on previous knowledge and ability for selection of candidate genes (13). Application of GWAS has been limited except in a study on exposure to environmental tobacco smoke (14). The application of GWAS to OEH studies will, however, result in some computational challenges as the number of genes that have a possible interaction with the exposure are large. Recently, several papers have proposed new statistical approaches for gene-environment-wide-interaction studies which minimize the type 1 error (i.e., false positives) while gaining efficiency and power (15-17).

Although they occur less frequently than SNPs, CNVs play an important role in genetic variation (18). CNVs are caused by genomic structural variations such as insertions, deletions, and duplications and have been defined as 'segments of DNA that are 1 kb or larger and present at variable copy number in comparison with a reference genome' (19). CNVs located in gene promoter regions can influence gene expression, and might influence the development of complex disease traits where gene dosage is altered but not abolished (19). CNVs proximal to genes but not in promoter sequences could perturb the 'histone code' and also influence gene expression. Further, CNVs located in exons could result in mis-spliced mRNA with detrimental effects on protein expression. Techniques that have been used to assess CNVs in the genome include comparative genomic hybridization (CGH), a technique that compares labeled DNA from individuals in a study population with differently labeled reference genomic DNA (20), and SNP-based platforms that use allele intensity ratios to make inferences about CNVs (19). CNV has been frequently assessed in studies that investigated the effects of the *glutathione S-transferase M1* (*GSTM1*) gene on environment-cancer associations (21, 22). To date most studies assessed the effect of having the null genotype (deletion) of *GSTM1* gene versus having at least one copy of the gene. Recent studies were also able to assess gene dosage effects (i.e., does having two copies of the GSTM1 gene result in stronger associations with cancer than having one copy?) (23, 24).

Transcriptomics

The abundance of specific mRNA transcripts in a biological sample is a reflection of the expression levels of the corresponding genes (25). Gene expression profiling is the identification and characterization of the mixture of mRNA that is present in a specific sample. An important application of gene expression profiling is to associate differences in mRNA mixtures originating from different groups of individuals to phenotypic differences between the groups (26). In contrast to genotyping, gene expression profiling allows characterization of the level of gene expression. Both the presence of specific forms of mRNA

and the levels in which these forms occur are parameters that provide information on gene expression (27). The transcriptome in contrast to the genome is highly variable over time, between cell types and will change in response to environmental changes (Table 1). A gene expression profile provides a quantitative overview of the mRNA transcripts that were present in a sample at the time of collection. Therefore, gene expression profiling can be used to determine which genes are differently expressed as result of changes in environmental conditions. A typical gene expression profiling study includes a group of individuals with similar phenotype (e.g., exposure level, disease status) and compares the gene expression profile of this group to the profile of a reference group matched on selected factors such as age and sex to the group of interest. Studies of this type usually report a set of genes that are differently expressed between the groups.

Epigenomics

The focus of epigenomics is to study epigenetic processes on a large (ultimately genome-wide) scale (28, 29). Epigenetic processes are mechanisms other than changes in DNA sequence that are involved in local activity states such as gene transcription and gene silencing (30-32). Although the range of epigenetic mechanisms that are discovered is expanding, epigenomics is mainly based on two most comprehensively studied mechanisms, DNA methylation and histone modification (28, 33-39). However, in recent years RNA interference of gene expression by non-coding RNAs such as microRNA and siRNA has acquired considerable attention (30, 40, 41). Changes in DNA methylation, histone modification and RNA interference are often associated and it is believed that interaction exists between these epigenetic processes (30). Here, the focus will be on DNA methylation and histone modification. DNA methylation is the addition of a methyl group to cytosine in a CpG dinucleotide. A distinction is made between global methylation and CpG island specific methylation. About 70% of the CpG dinucleotides in the human genome are methylated. However, CpG dinucleotides in CpG islands are predominantly unmethylated (34). Hypermethylation of CpG islands located in promoter regions of genes is related to gene silencing. Under normal conditions gene silencing is related to phenomena such as genomic imprinting, x-chromosome inactivation and tissue specific gene expression (28, 36). Altered gene silencing plays a causal role in human disease (30, 34, 37, 38, 42). The effect of hypomethylation of the genome outside CpG islands is less well understood but may be involved in chromosomal instability (31, 34). Histone proteins are involved in the structural packaging of DNA in the chromatin complex. Post translational histone modifications such as acetylation and methylation are believed to regulate chromatin structure and therefore gene expression (37, 38).

Proteomics

In general the function of cells can be described by the proteins that are present in the intra- and inter-cellular space and the abundance of these proteins (43). Although all proteins are

based on mRNA precursors, post translational modifications (PTMs) and environmental interactions make it impossible to predict abundance of specific proteins based on gene expression analysis alone. The proteome consists of all proteins present in specific cell types or tissue. In contrast to the genome, the proteome is highly variable over time, between cell types and will change in response to changes in its environment (44). Proteomics provides insights into the role proteins have in biological systems. A major challenge is the high variability in proteins and protein abundance in certain types of biologic samples (e.g., the concentration of proteins in plasma ranges up to nine orders of magnitude) (45). This requires the development of technologies that can detect a wide range of proteins in samples from different origins (46). Many proteomic technologies are currently available but broadly a distinction can be made between approaches that are based on detection by mass spectrometry and protein microarrays using capturing agents such as antibodies. An important focus is the identification of proteins including the presence of PTMs of proteins and identification of proteins interacting in protein-complexes (43, 44). Another focus of proteomics is quantification of the protein abundance. Protein expression levels represent the balance between translation and degradation of proteins in cells. It is therefore assumed that the abundance of a specific protein is related to its role in cell function. However, the high dynamic range (i.e., the ratio between the smallest and largest concentration and/or mass value) of proteins complicates this type of proteomic analysis (43, 44).

Metabolomics

Metabolic phenotypes are the by-products that result from the interaction between genetic, environmental, lifestyle and other factors (47). The metabolome consists of small molecules (e.g., lipids or vitamins) that are also known as metabolites (48). Metabolites are involved in the energy transmission in cells (metabolism) by interacting with other biological molecules following metabolic pathways. Metabolomics is defined as the study of metabolic profiles in easily collected biological samples such as urine, saliva or plasma (48). The metabolome is highly variable and time dependent, and it consists of a wide range of chemical structures (Table 1). An important challenge of metabolomics is to acquire qualitative and quantitative information concerning the metabolites that occur under normal circumstances in order to be able to detect perturbations in the complement of metabolites as result of changes in environmental factors.

Challenges for the application of OMICS in OEH

The development of new OMICS technologies is an important first step towards implementation of OMICS markers in OEH. However, similar to other (bio)markers of exposure, susceptibility and effect, the successful implementation of OMICS markers in OEH

requires appropriate study designs, thorough validation of markers, and careful interpretation of study results (49-51).

Study design

As indicated in Table 1, the transcriptome, proteome and metabolome are highly variable over time and are likely to be influenced by the disease process. This indicates that great care should be given to the timing of biological sample collection and adequate processing (e.g., field stabilization of mRNA) of the sample to minimize measurement error and to avoid potential differential misclassification biases. In Table 2 the advantages and disadvantages of the different human observational study (HOS) designs with regard to the collection and use of biological markers are given. In general, it can be stated that hospital-based case-control studies are the least suitable for the application of these technologies in HOS research, as they are more prone to selection and differential bias, while prospective studies or cross-sectional studies seem most suitable for such approaches. Moreover, hospital case-control studies are problematic as it is impossible to determine if changes in biomarkers are the cause or consequence of a disease. Semi-longitudinal studies might be extremely powerful for some OMICS technologies such as transcriptomics, proteomics, and metabolomics where biological measures are taken before and after exposure or change in disease status. In these study designs each individual serves as their own control eliminating the influence of population variance.

Validation of biomarkers

The value of an OMICS-based biomarker in OEH depends on the reliability of an assay to qualitatively and quantitatively assess the biomarker and on the association between the biomarker and the biological endpoint of interest (exposure, susceptibility or health effect). The reliability of an assay can be tested by investigating the variability of an assay within and between laboratories and comparing results to the variability of existing assays (standards). A necessary step towards an increase in the reliability of OMICS assays is standardization. Several initiatives have developed standards for new OMICS assays with regards to comparison to existing techniques (microarray quality control (MAQC)), data formats to describe experimental details (minimum information about a microarray experiment (MIAME)) and assessment of sample quality (external RNA controls consortium (ERCC)) (52, 53). Once the reliability of assays has been established in the laboratory transitional studies that assess the association between biomarkers and biological endpoints in humans are needed (49). To achieve an accurate estimate of the association between a biomarker and a biological endpoint reliable and valid measurements of exposure and covariates are needed as well.

Table 2 Comparison of advantages and limitations relevant to the collection of biological specimens and data interpretation in molecular epidemiology study designs (adapted from Garcia-Closas et al.(49)).

Study Design	Advantages	Limitations
Cross-sectional	• Facilitates intense collection and timely processing of specimens (e.g., freshly frozen samples, cryopreserved lymphocytes) • Allows detailed collection of exposure and confounder information	• Relevance of intermediate endpoints altered by current exposures in healthy individuals not always clear
Hospital-based case-control	• Facilitates intense collection and timely processing of specimens (e.g., freshly frozen samples, cryopreserved lymphocytes) • Participation rates for biological collections might be enhanced • Facilitates follow-up of cases for treatment response and survival	• More prone to selection and differential biases than other designs • Some biomarkers might be affected by disease process or hospital stay
Population-based case-control	• Less subject to biases (e.g., selection, exposure misclassification) than hospital-based studies	• Some biomarkers might be affected by disease process • May be more difficult to obtain high participation rates for biological collection than hospital based designs • Implementation of intense, specialized blood and tumor collection and processing protocols can be challenging
Prospective cohort	• Allows study of multiple disease end points • Allows study of transient biomarkers and biomarkers affected by disease status • Selection bias and differential misclassification are avoided: non-differential misclassification might be reduced for some exposures • Nested case-control or case-cohort studies can be used to improve efficiency of the design	• Implementation of intense, specialized collection and processing protocols for the entire cohort can be challenging • Obtaining tissue samples and following cases for treatment response and survival can be challenging in many cohorts

A true association between a biomarker and a biological endpoint can be obscured by measurement error. To acquire insight into the impact of measurement error on the observed association between a biomarker and a biological endpoint a repeated sampling design, at least on part of the population, is necessary. Repeated sampling on individuals will allow researchers to compare biomarker variability within individuals to biomarker variability between individuals. One measure that can be used to assess the variability of biomarkers within and between individuals is the intraclass correlation coefficient, which represents the proportion of the total variance that can be attributed to the between-individual variance (49). The level of measurement error that is acceptable for a biomarker depends on the magnitude of the true association between the biomarker and the biological endpoint of interest. For biomarkers with a dichotomous outcome (e.g., genotyping) the accuracy of the biomarker is based on the sensitivity (e.g., probability of correctly identifying a SNP) and the specificity (e.g., probability of incorrectly identifying a SNP) of the biomarker.

Interpretation of study results

In recent years technological developments have had a major impact on the development of new types of study designs of OMICS-based studies. One trend that has been seen consistently within the different OMICS fields is the enormous increase in resolution of the assays (the number of 'endpoints' that can be assessed in a single assay) and throughput of the assays (the number of samples that can be analyzed per time period). Many of the improvements are based on the introduction of chip-based assays such as DNA-microarrays. A major implication of the possibility to investigate multiple endpoints (e.g., up to 1,000,000 SNPs in a single assay) in large populations is the possibility for researchers to move away from hypothesis-based studies (focused on a limited set of endpoints) towards hypothesis-free (agnostic) types of study designs (including much larger sets of endpoints). Although the hypothesis-free studies might contribute considerably to the elucidation of the complex biological processes that underlie clinically manifested health effects, it is important to realize that the interpretation of data generated by these types of studies requires a different approach than the interpretation of data generated by more traditional hypothesis-based studies. In hypothesis-based study designs 'frequentist' measures such as 95% confidence intervals or p-values provide a reasonably good measure to assess the statistical significance of the study's finding. However, the interpretation of such measures is based on the inclusion of a limited number of hypotheses for which the researchers assume that there is a good possibility that the null-hypothesis might be rejected (i.e., there is a high prior probability of a true positive finding). In a hypothesis-free analytic approach, a study is initiated without a well-defined hypothesis for each included endpoint investigated (i.e., a flat prior probability for each finding). However, as a result of chance, the increased number of possible endpoints in a study is accompanied by higher probability of the possibility of a detecting statistically significant false-positive results (54). Therefore, the traditional statistical approaches that are commonly used in epidemiology are of less value in hypothesis-free studies. A current

challenge for the OMICS field is the development of (statistical) approaches that can be used for the interpretation of the high-dimensional data generated by these high-throughput techniques. Several statistical strategies (and also approaches in study designs) have been developed to reduce the probability of false positives results. Examples are the Bonferroni adjustment for multiple significance testing or more sophisticated Bayesian approaches which include estimation of the false positive report probability (15-17, 54, 55). However, replication of the initial findings in follow-up studies remains the strongest safeguard against false-positive results. Studies that incorporate thousands of biological endpoints should therefore primarily be seen as discovery studies that can aid to the generation of new hypotheses. Therefore, new OMICS studies should incorporate strategies for built in replication of the study findings. Application of a different analytical technique to test the hypothesis *a priori* in a second/validation set of samples will reduce the possibility that the initial finding was an artifact of the technology used. A potential strategy for built-in replication is to perform the initial analysis on a subset of well-characterized samples matched on potential confounders and effect modifiers and confirm the findings by using alternative analysis methods on the remaining often larger sample set. A potential problem in OEH research is, however, that replication is often complicated as there are often only a limited number of relatively small studies on a single exposure. Even if another large study can be found on a single exposure replication might still be complicated by the fact that the populations are exposed to different levels.

In addition to aspects that contribute to random error, systematic error (bias) is also a potential threat to the validity of HOS utilizing OMICS technologies (56-58). The types of bias that might occur will be largely similar to types of bias that might occur in all HOS. However, issues such as sample collection, handling and storage of samples, and analysis technique-specific biases might be especially relevant for studies applying OMICS technologies (57, 59, 60). Very recently guidelines for the reporting of genetic association studies (STREGA) have been published (61). These guidelines underline the necessity of detailed reporting in publications on genetic association studies to allow scientist to assess the potential of bias in study outcomes. Development of similar guidelines for the other OMICS fields will contribute to the identification of relevant types of bias.

Pathway analysis and systems biology

OMICS technologies will enable researchers to look at the complete complement, expression, and regulation of genes, proteins and metabolites. However, at the present time, most statistical analyses are often based on a (simplistic) one-by-one comparison of markers between exposure and/or disease groups. Recently, analytical tools/databases have become available to perform more integrated analyses of biological functions and changes in biological functions as a result of environmental factors. Examples of such approaches are gene ontology (GO), pathway analysis and Structural Equation Modeling (SEM) (62-65). GO is

based on a library that consists of gene profiles that are associated with biological processes (66). Gene sets that are identified in microarray experiments as differently expressed are tested for their association with a profile in the GO library (64). In pathway analysis, not only the profile of genes associated with a specific biological process is tested, but also the functional interactions between genes in a profile (63). While still large gaps in the knowledge of biological pathways exist, each new study will contribute to build a base of knowledge necessary for these types of analyses. SEM is a statistical approach that can be used to simultaneously model multiple genes and multiple SNPs within a gene in a hierarchical manner that reflects their underlying role in a biological system (62).
The increasing knowledge of biological pathways will facilitate the integration of the separate OMICS fields into systems biology approaches. System biology has been described as a global quantitative analysis of the interaction of all components in a biological system to determine its phenotype (67-69). This integration is facilitated by a continuous increase in computing power and possibilities for data sharing.

Examples of the use of OMICS in occupational and environmental health research

In Table 3 a number of studies are listed to illustrate the current application of OMICS technologies in OEH research. Benzene and arsenic were chosen as examples because of the large populations with potential exposure to these agents in both the occupational and environmental setting and the relatively large number of studies on these agents that have applied OMICS technologies. It should be noted that inclusion of the example studies was not intended as a systematic overview of studies applying OMICS in OEH research in these specific areas but merely to provide a resource of studies that are indicative of the potential of these new technologies. We highlight three studies from Table 3 in some more detail to illustrate the progress in the OMICS field that has been made in recent years. A nice illustration of the progress of the use of genotyping methods in OEH research is a study on hematological effects among a cohort of 250 workers exposed to benzene and 140 controls (70-72). Initial gene-environment analyses in this study were based on candidate gene-approaches focusing on genes involved in the metabolism of benzene (4 genes, 4 SNPs) (71), DNA double strand break repair (7 genes, 24 SNPs) (70), and cytokine and cellular adhesion molecule pathways (20 genes, 40 SNPs) (72). In a more recent analysis of the same study population, Lan et al. used a chip-based assay (GoldenGate assay) for genotyping which allowed for a larger number of SNPs to be assessed (414 genes, 1433 SNPs) (73) . These SNPs were selected from the SNP500Cancer database, and were, therefore, hypothesized to be involved in the development of cancer. However, the influence of these SNPs on benzene-induced hematotoxicity was largely unknown for most SNPs. This study should therefore primarily be seen as hypothesis-generating and indeed has provided information on several putative genes involved in benzene hematotoxity that went well beyond the more classical focus in OEH research on metabolic genes.

Table 3 Examples of the use of OMICS technologies in occupational and environmental studies that investigate health effects in human populations exposed to benzene or arsenic.

OMICS field	Exposure	Topic	References
Genotyping	Benzene	Interaction between SNPs and benzene-induced toxicity.	(70, 71, 73, 77-80)
Genotyping	Arsenic	Interaction between SNPs and arsenic-induced skin lesions.	(81, 82)
Genotyping	Arsenic	Interaction between SNPs and arsenic metabolism.	(83, 84)
Genotyping	Arsenic	Interaction between SNPs and exposure to arsenic in relation to non-melanoma skin cancer.	(85)
CNV	Arsenic	Interaction between DNA CNV and exposure to arsenic in relation to transitional cell carcinoma.	(86)
CNV	Arsenic	Interaction between DNA CNV and exposure to arsenic in relation to bladder tumors.	(87)
Epigenomics	Benzene	Relation between gene specific hypermethylation and exposure to benzene.	(88)
Epigenomics	Arsenic	Relation between epigenetic silencing of tumor suppressor genes and exposure to both tobacco and arsenic.	(89)
Epigenomics	Arsenic	Relation between genomic methylation and exposure to arsenic.	(90, 91)
Transcriptomics	Benzene	Relation between gene expression and exposure to benzene.	(92)
Transcriptomics	Arsenic	Interaction between exposure to arsenic and arsenical skin lesions in relation to genome-wide gene expression.	(74)
Transcriptomics	Arsenic	Relation between gene expression and exposure to arsenic.	(93, 94)
Proteomics	Benzene	Impact of exposure to benzene on the composition of the proteome.	(95, 96)
Proteomics	Arsenic	Impact of exposure to arsenic on the composition of the proteome.	(97, 98)

Although the authors addressed issues of multiple comparisons to reduce the chance of false positive findings due to the large number SNPs included in the analysis, it is still critical that the results are replicated in subsequent independent studies.

An example of a hypothesis-free approach towards the assessment of the transcriptome comes from a study by Argos et al. (74). In this micro-array based study, ~22,000 genome wide gene transcripts were measured in 25 subjects with arsenic induced skin-lesions and 15 controls. A false discovery rate of 1% was defined a priori to reduce the risk of chance findings. A set of 486 genes that were differentially expressed between cases and controls was reported. The gene transcripts were also analyzed with the use of gene ontology and pathway analysis approaches to elucidate the biological pathways that are involved in arsenic induced skin-lesions. Similar to the genotyping results of the studies discussed above, results

from the genome-wide assessment of the transcriptome should be interpreted with great care and require replication in independent studies before they can be used as valid exposure or effect markers (75, 76).

Way forward

It is clear that there have been great technological advances in the different OMICS fields. Some of these technologies have and are starting to be applied in OEH research and will undoubtedly lead to numerous new insights in the near future. With the development of validated technologies, appropriate study designs, better sample handling and advanced statistical methods for data interpretation, OMICS techniques will eventually contribute significantly to OEH and will help the field progress towards an integrated view of the interaction between environment and human health. To achieve this integrated view it will be important to not only focus on genetic variants but also on more functional measures of the phenotype and accurate assessment of exposure. The challenge in this effort will be that the closer one gets to a functional measure of the phenotype (i.e., proteomics, metabolomics) the more complex it will be to capture physiologically relevant variability and the more crucial the development of advanced study designs, sampling collection procedures, measurement techniques, and methods for statistical analysis will be to allow interpretation of these parameters.

Acknowledgments

This work was performed as part of the work package “Integrated Risk Assessment” of the ECNIS Network of Excellence (Environmental Cancer Risk, Nutrition and Individual Susceptibility), operating within the European Union 6th Framework Program, Priority 5: “Food Quality and Safety” (FOOD-CT-2005-513943).

References

1. Smith MT, Vermeulen R, Li G, et al. Use of 'Omic' technologies to study humans exposed to benzene. Chem Biol Interact. 2005;153-154:123-7.
2. Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001;409(6822):928-33.
3. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.
4. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005;14(8):1847-50.
5. Smith MT. Future perspectives of molecular cancer epidemiology. EJC supplements 2008;6:188.
6. Syvänen A. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature reviews Genetics. 2001;2(12):930-42.
7. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

8. Eichler EE, Nickerson DA, Altshuler D, et al. Completing the map of human genetic variation. Nature. 2007;447(7141):161-5.
9. Rocha D, Gut I, Jeffreys AJ, et al. Seventh international meeting on single nucleotide polymorphism and complex genome analysis: 'ever bigger scans and an increasingly variable genome'. Hum Genet. 2006;119(4):451-6.
10. Engle LJ, Simpson CL, Landers JE. Using high-throughput SNP technologies to study cancer. Oncogene. 2006;25(11):1594-601.
11. Patil N, Berno AJ, Hinds DA, et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science. 2001;294(5547):1719-23.
12. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40(6):695-701.
13. Castro-Giner F, Kauffmann F, de Cid R, et al. Gene-environment interactions in asthma. Occup Environ Med. 2006;63(11):776,86, 761.
14. Colilla S, Kantoff PW, Neuhausen SL, et al. The joint effect of smoking and AIB1 on breast cancer risk in BRCA1 mutation carriers. Carcinogenesis. 2006;27(3):599-605.
15. Mukherjee B, Chatterjee N. Exploiting gene-environment independence for analysis of case-control studies: an empirical Bayes-type shrinkage estimator to trade-off between bias and efficiency. Biometrics. 2008;64(3):685-94.
16. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. Am J Epidemiol. 2009;169(2):219-26.
17. Wakefield J, De Vocht F, Hung RJ. Bayesian mixture modeling of gene-environment and gene-gene interactions. Genet Epidemiol. 2010;34(1):16-25.
18. Shlien A, Malkin D. Copy number variations and cancer. Genome Med. 2009;1(6):62.
19. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54.
20. Costa JL, Meijer G, Ylstra B, et al. Array comparative genomic hybridization copy number profiling: a new tool for translational research in solid malignancies. Semin Radiat Oncol. 2008;18(2):98-104.
21. Benhamou S, Lee WJ, Alexandrie AK, et al. Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. Carcinogenesis. 2002;23(8):1343-50.
22. Carlsten C, Sagoo GS, Frodsham AJ, et al. Glutathione S-transferase M1 (GSTM1) polymorphisms and lung cancer: a literature-based systematic HuGE review and meta-analysis. Am J Epidemiol. 2008;167(7):759-74.
23. Crosbie PA, Barber PV, Harrison KL, et al. GSTM1 copy number and lung cancer risk. Mutat Res. 2009;664(1-2):1-5.
24. Sorensen M, Raaschou-Nielsen O, Brasch-Andersen C, et al. Interactions between GSTM1, GSTT1 and GSTP1 polymorphisms and smoking and intake of fruit and vegetables in relation to lung cancer. Lung Cancer. 2007;55(2):137-44.
25. Manning AT, Garvin JT, Shahbazi RI, et al. Molecular profiling techniques and bioinformatics in cancer research. Eur J Surg Oncol. 2007;33(3):255-65.
26. Nachtomy O, Shavit A, Yakhini Z. Gene expression and the concept of the phenotype. Stud Hist Philos Biol Biomed Sci. 2007;38(1):238-54.
27. Celis JE, Kruhoffer M, Gromova I, et al. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. FEBS Lett. 2000;480(1):2-16.
28. Callinan PA, Feinberg AP. The emerging science of epigenomics. Hum Mol Genet. 2006;15 Spec No 1:R95-101.
29. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. Nature. 2007;447(7143):433-40.
30. Egger G, Liang G, Aparicio A, et al. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429(6990):457-63.
31. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465-76.
32. Bird A. Perceptions of epigenetics. Nature. 2007;447(7143):396-8.
33. Higgs DR, Vernimmen D, Hughes J, et al. Using Genomics to Study How Chromatin Influences Gene Expression. Annual Review of Genomics and Human Genetics. 2007;8:299-325.
34. Liu L, Wylie RC, Andrews LG, et al. Aging, cancer and nutrition: the DNA methylation connection. Mech Ageing Dev. 2003;124(10-12):989-98.
35. Bibikova M, Lin Z, Zhou L, et al. High-throughput DNA methylation profiling using universal bead arrays. Genome Res. 2006;16(3):383-93.

36. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. Nat Rev Genet. 2007;8(4):253-62.
37. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007;8(4):286-98.
38. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128(4):683-92.
39. Moore LE, Huang WY, Chung J, et al. Epidemiologic considerations to assess altered DNA methylation from environmental exposures in cancer. Ann N Y Acad Sci. 2003;983:181-96.
40. Costa FF. Non-coding RNAs: lost in translation? Gene. 2007;386(1-2):1-10.
41. de Fougerolles A, Vornlocher HP, Maraganore J, et al. Interfering with disease: a progress report on siRNA-based therapeutics. Nat Rev Drug Discov. 2007;6(6):443-53.
42. Moss TJ, Wallrath LL. Connections between epigenetic gene silencing and human disease. Mutat Res. 2007;618(1-2):163-74.
43. Sellers TA, Yates JR. Review of proteomics with applications to genetic epidemiology. Genet Epidemiol. 2003;24(2):83-98.
44. Fliser D, Novak J, Thongboonkerd V, et al. Advances in urinary proteome analysis and biomarker discovery. J Am Soc Nephrol. 2007;18(4):1057-71.
45. Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. Nature. 2008;452(7187):571-9.
46. Wingren C, Borrebaeck CA. Antibody microarrays: current status and key technological advances. Omics. 2006;10(3):411-27.
47. Holmes E, Loo RL, Stamler J, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature. 2008;453(7193):396-400.
48. Claudino WM, Quattrone A, Biganzoli L, et al. Metabolomics: available results, current research projects in breast cancer, and future applications. J Clin Oncol. 2007;25(19):2840-6.
49. Garcia-Closas M, Vermeulen R, Sherman M, et al. Application of biomarkers in cancer epidemiology. In: Schottenfeld D, F FJ,Jr, eds. Cancer Epidemiology and Prevention. New York, NY: Oxford University Press, 2006:70-88.
50. Vineis P, Gallo V. The epidemiological theory: principles of biomarker validation. In: Vineis P, Gallo V, eds. Epidemiological concepts of validation of biomarkers for the identification/quantification of environmental carcinogenic exposures. Lodz: Nofer Institute of Occupational Medicine, 2007.
51. Duncan MW. Omics and its 15 minutes. Exp Biol Med (Maywood). 2007;232(4):471-2.
52. Morrison N, Wood AJ, Hancock D, et al. Annotation of environmental OMICS data: application to the transcriptomics domain. Omics. 2006;10(2):172-8.
53. Wilkes T, Laux H, Foy CA. Microarray data quality - review of current developments. Omics. 2007;11(1):1-13.
54. Wacholder S, Chanock S, Garcia-Closas M, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst. 2004;96(6):434-42.
55. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. 1995;310(6973):170.
56. Semmes OJ. The "omics" haystack: defining sources of sample bias in expression profiling. Clin Chem. 2005;51(9):1571-2.
57. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer. 2005;5(2):142-9.
58. Dumeaux V, Borresen-Dale AL, Frantzen JO, et al. Gene expression analyses in breast cancer epidemiology: the Norwegian Women and Cancer postgenome cohort study. Breast Cancer Res. 2008;10(1):R13.
59. Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: B: Are the results of the study valid? JAMA. 2009;301(2):191-7.
60. Yesupriya A, Evangelou E, Kavvoura FK, et al. Reporting of human genome epidemiology (HuGE) association studies: an empirical assessment. BMC Med Res Methodol. 2008;8:31.
61. Little J, Higgins JP, Ioannidis JP, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. Eur J Epidemiol. 2009;24(1):37-55.
62. Nock NL, Larkin EK, Morris NJ, et al. Modeling the complex gene x environment interplay in the simulated rheumatoid arthritis GAW15 data using latent variable structural equation modeling. BMC Proc. 2007;1 Suppl 1:S118.
63. Draghici S, Khatri P, Tarca AL, et al. A systems biology approach for pathway level analysis. Genome Res. 2007.
64. Khatri P, Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics. 2005;21(18):3587-95.

65. Weniger M, Engelmann JC, Schultz J. Genome Expression Pathway Analysis Tool—analysis and visualization of microarray gene expression data under genomic, proteomic and metabolic context. BMC Bioinformatics. 2007;8:179.
66. The Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. Genome Res. 2001;11(8):1425-33.
67. Lund E, Dumeaux V. Systems epidemiology in cancer. Cancer Epidemiol Biomarkers Prev. 2008;17(11):2954-7.
68. Lemberger T. Systems biology in human health and disease. Mol Syst Biol. 2007;3:136.
69. Studer SM, Kaminski N. Towards systems biology of human pulmonary fibrosis. Proc Am Thorac Soc. 2007;4(1):85-91.
70. Shen M, Lan Q, Zhang L, et al. Polymorphisms in genes involved in DNA double-strand break repair pathway and susceptibility to benzene-induced hematotoxicity. Carcinogenesis. 2006;27(10):2083-9.
71. Lan Q, Zhang L, Li G, et al. Hematotoxicity in workers exposed to low levels of benzene. Science. 2004;306(5702):1774-6.
72. Lan Q, Zhang L, Shen M, et al. Polymorphisms in cytokine and cellular adhesion molecule genes and susceptibility to hematotoxicity among workers exposed to benzene. Cancer Res. 2005;65(20):9574-81.
73. Lan Q, Zhang L, Shen M, et al. Large-scale evaluation of candidate genes identifies associations between DNA repair and genomic maintenance and development of benzene hematotoxicity. Carcinogenesis. 2009;30(1):50-8.
74. Argos M, Kibriya MG, Parvez F, et al. Gene expression profiles in peripheral lymphocytes by arsenic exposure and skin lesion status in a Bangladeshi population. Cancer Epidemiol Biomarkers Prev. 2006;15(7):1367-75.
75. Gillis B, Gavin IM, Arbieva Z, et al. Identification of human cell responses to benzene and benzene metabolites. Genomics. 2007;90(3):324-33.
76. Smith MT. Misuse of genomics in assigning causation in relation to benzene exposure. Int J Occup Environ Health. 2008;14(2):144-6.
77. Chen Y, Li G, Yin S, et al. Genetic polymorphisms involved in toxicant-metabolizing enzymes and the risk of chronic benzene poisoning in Chinese occupationally exposed populations. Xenobiotica. 2007;37(1):103-12.
78. Gu SY, Zhang ZB, Wan JX, et al. Genetic polymorphisms in CYP1A1, CYP2D6, UGT1A6, UGT1A7, and SULT1A1 genes and correlation with benzene exposure in a Chinese occupational population. J Toxicol Environ Health A. 2007;70(11):916-24.
79. Kim YJ, Choi JY, Paek D, et al. Association of the NQO1, MPO, and XRCC1 polymorphisms and chromosome damage among workers at a petroleum refinery. J Toxicol Environ Health A. 2008;71(5):333-41.
80. Zhang Z, Wan J, Jin X, et al. Genetic polymorphisms in XRCC1, APE1, ADPRT, XRCC2, and XRCC3 and risk of chronic benzene poisoning in a Chinese occupational population. Cancer Epidemiol Biomarkers Prev. 2005;14(11 Pt 1):2614-9.
81. Breton CV, Zhou W, Kile ML, et al. Susceptibility to arsenic-induced skin lesions from polymorphisms in base excision repair genes. Carcinogenesis. 2007;28(7):1520-5.
82. McCarty KM, Chen YC, Quamruzzaman Q, et al. Arsenic methylation, GSTT1, GSTM1, GSTP1 polymorphisms, and skin lesions. Environ Health Perspect. 2007;115(3):341-5.
83. Schlawicke Engstrom K, Broberg K, Concha G, et al. Genetic polymorphisms influencing arsenic metabolism: evidence from Argentina. Environ Health Perspect. 2007;115(4):599-605.
84. Steinmaus C, Moore LE, Shipp M, et al. Genetic polymorphisms in MTHFR 677 and 1298, GSTM1 and T1, and metabolism of arsenic. J Toxicol Environ Health A. 2007;70(2):159-70.
85. Applebaum KM, Karagas MR, Hunter DJ, et al. Polymorphisms in nucleotide excision repair genes, arsenic exposure, and non-melanoma skin cancer in New Hampshire. Environ Health Perspect. 2007;115(8):1231-6.
86. Hsu LI, Chiu AW, Pu YS, et al. Comparative genomic hybridization study of arsenic-exposed and non-arsenic-exposed urinary transitional cell carcinoma. Toxicol Appl Pharmacol. 2008;227(2):229-38.
87. Moore LE, Smith AH, Eng C, et al. Arsenic-related chromosomal alterations in bladder cancer. J Natl Cancer Inst. 2002;94(22):1688-96.
88. Bollati V, Baccarelli A, Hou L, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res. 2007;67(3):876-80.
89. Marsit CJ, Karagas MR, Schned A, et al. Carcinogen exposure and epigenetic silencing in bladder cancer. Ann N Y Acad Sci. 2006;1076:810-21.
90. Chanda S, Dasgupta UB, Guhamazumder D, et al. DNA hypermethylation of promoter of gene p53 and p16 in arsenic-exposed people with and without malignancy. Toxicol Sci. 2006;89(2):431-7.

91. Pilsner JR, Liu X, Ahsan H, et al. Genomic methylation of peripheral blood leukocyte DNA: influences of arsenic and folate in Bangladeshi adults. Am J Clin Nutr. 2007;86(4):1179-86.
92. Forrest MS, Lan Q, Hubbard AE, et al. Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. Environ Health Perspect. 2005;113(6):801-7.
93. Fry RC, Navasumrit P, Valiathan C, et al. Activation of inflammation/NF-kappaB signaling in infants born to arsenic-exposed mothers. PLoS Genet. 2007;3(11):e207.
94. Wu MM, Chiou HY, Ho IC, et al. Gene expression of inflammatory molecules in circulating lymphocytes from arsenic-exposed human subjects. Environ Health Perspect. 2003;111(11):1429-38.
95. Joo WA, Sul D, Lee DY, et al. Proteomic analysis of plasma proteins of workers exposed to benzene. Mutat Res. 2004;558(1-2):35-44.
96. Vermeulen R, Lan Q, Zhang L, et al. Decreased levels of CXC-chemokines in serum of benzene-exposed workers identified by array-based proteomics. Proc Natl Acad Sci U S A. 2005;102(47):17041-6.
97. Zhai R, Su S, Lu X, et al. Proteomic profiling in the sera of workers occupationally exposed to arsenic and lead: identification of potential biomarkers. Biometals. 2005;18(6):603-13.
98. Hegedus CM, Skibola CF, Warner M, et al. Decreased urinary beta-defensin-1 expression as a biomarker of response to arsenic. Toxicol Sci. 2008;106(1):74-82.

Chapter 7

The Impact of Saturable Metabolism on Exposure-Response Relations in Two Studies of Benzene Induced Leukemia

Jelle Vlaanderen
Lützen Portengen
Stephen M. Rappaport
Deborah C. Glass
Hans Kromhout
Roel Vermeulen

American Journal of Epidemiology (in press)

Abstract

Enzymatic saturation of metabolic pathways is one factor that potentially contributes to the nonlinear exposure-response relations that are frequently reported in occupational epidemiological studies. The authors propose an approach to explore the contribution of saturable metabolism to previously reported exposure-response relations by integrating predictive models of relevant biomarkers of exposure into the epidemiological analysis. The approach is demonstrated with two studies of leukemia in benzene-exposed workers which differed greatly in their magnitudes and durations of exposure. Substitution of biomarker levels for external estimates of benzene exposure reduced the fold difference of the log relative risk (RR) of leukemia per unit of cumulative exposure between the two studies by 11% to 44%. Nevertheless, a considerable difference in the log RR per unit of cumulative exposure remained between the two studies suggesting that exposure misclassification, differences in study design, and potential confounding factors also contributed to the heterogeneity in risk estimates.

Introduction

Cumulative exposure (the product of intensity and duration of exposure) is a standard exposure metric used in many occupational studies of chronic health effects (1). The frequent use of cumulative exposure as index of the target tissue dose is based on three broad assumptions: (1) the cumulative probability of developing a disease is proportional to the sum of the daily probabilities of developing a disease; (2) the daily probability of developing a disease increases monotonically with the concentration in the target tissue; and (3) the concentration in the target tissue is linearly related to the external exposure (2). Several authors have demonstrated that these assumptions are not always valid (3-5). For example, Doll and Peto demonstrated that lung cancer incidence was related to at least the fourth power of smoking duration and only to the second power of smoking intensity (5). Further, in a study of respiratory cancer and arsenic exposure, Lubin et al. reported that a higher relative risk (RR) was predicted for cumulative exposure delivered at high intensity for shorter duration than for cumulative exposure delivered at lower intensity for longer duration (3). In attempting to explain their findings, Lubin et al. speculated that the impact of 'exposure delivery' might be explained by a non-linear relation between exposure intensity (air concentration of arsenic) and the concentration of arsenic metabolites in the target tissue, due to saturation of arsenic metabolism (3). Effects of saturable metabolism on delivered dose have been shown to affect outcomes of toxicology studies (6) and have been suspected of being important in occupational epidemiological studies (7).

Alteration of exposure-response relations due to saturable metabolism could affect the interpretation of occupational epidemiological studies in risk assessment since the common practice of linear extrapolation of exposure-response curves (ERCs) derived at high exposure levels to lower levels relevant to the general public could result in highly biased RRs. The direction of such biases could be either positive or negative depending on whether the saturating pathway was activating or deactivating in terms of the ultimate toxicant(s). In either case, estimates of the RR per unit of exposure would no longer be comparable at different ends of the exposure spectrum.

Here we describe applications of a physiologically based pharmacokinetic (PBPK) model (8) and empirical models based on biomarker data (9, 10) to quantitatively relate levels of airborne exposure to internal levels of toxic metabolites in occupational epidemiological studies. Our approach is based on the conjecture that if a health outcome is not proportional to the air concentration of an agent because of saturable metabolism, then the exposure-response relation would become linear (i.e. outcome proportional to exposure) when air concentrations are substituted by predictions of relevant biomarker levels from PBPK or empirical models. As a result, RRs that are expressed per unit of metabolite or biomarker level

for a given toxicant should be more comparable across studies than RRs expressed per unit of airborne exposure.

In a recent meta-regression analysis we observed that RR estimates from occupational epidemiological studies suggested a supralinear shape of the benzene-leukemia ERC (i.e. greater than proportional RR at lower exposure levels) (11). Such nonlinear behavior is supported by several lines of evidence suggesting that pathways leading to production of benzene metabolites saturate with increasing exposure to airborne benzene (9, 10, 12-18). The major metabolic pathways for benzene are shown in Figure 1. Benzene is metabolized by cytochrome P450 (CYP) enzymes (primarily CYP2E1) to benzene oxide (BO), which is in equilibrium with its valence tautomer oxepin, and is the source of all other metabolites. For airborne exposure levels between 0.1 and 10 ppm, phenol (PH) represents 70-85% of the urinary metabolites, hydroquinone (HQ), *t,t*-muconic acid (MA), and catechol (CA) each represent 5-10% and *S*-phenylmercapturic acid (SPMA) represents < 1% (19).

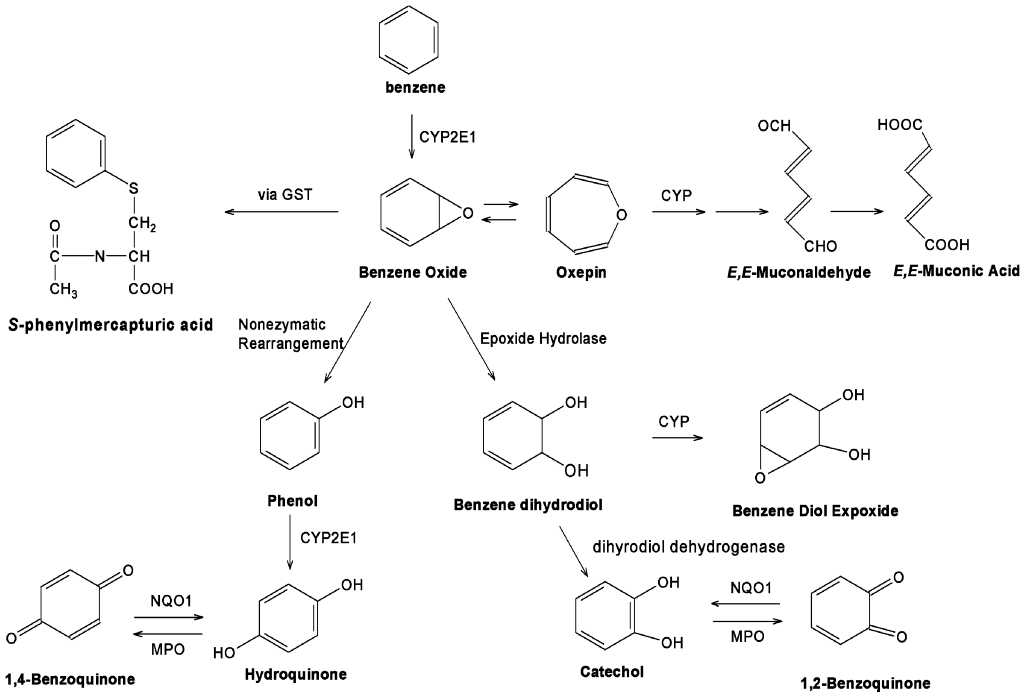


Figure 1 Simplified metabolic scheme for benzene showing major pathways and metabolizing genes.

To illustrate our approach, we used data from two prominent occupational studies of benzene-induced leukemia, namely, the Health Watch (HW) study (20), and the Pliofilm (PL) study (21). The more recent HW study was performed in the petroleum industry where workers were generally exposed for many years to relatively low air concentrations of benzene (seldom above 5 ppm) (20). In contrast, the older PL study was performed in a chemical factory where workers experienced much higher benzene exposures (up to 60 ppm), but for short periods of time (21). Interestingly, the HW study reported much higher RRs for leukemia by cumulative exposure category than the PL study even though HW included lower cumulative exposures than PL (20, 21). Here, we examine the difference in log RR of leukemia per unit of cumulative benzene exposure in the HW and PL studies and explore the extent to which this difference might be explained by saturable benzene metabolism.

Materials and methods

In Figure 2 we present a scheme of our approach for integrating the predictions of benzene metabolite levels into the epidemiological analyses of benzene exposure and leukemia risk. The scheme consists of the following steps:

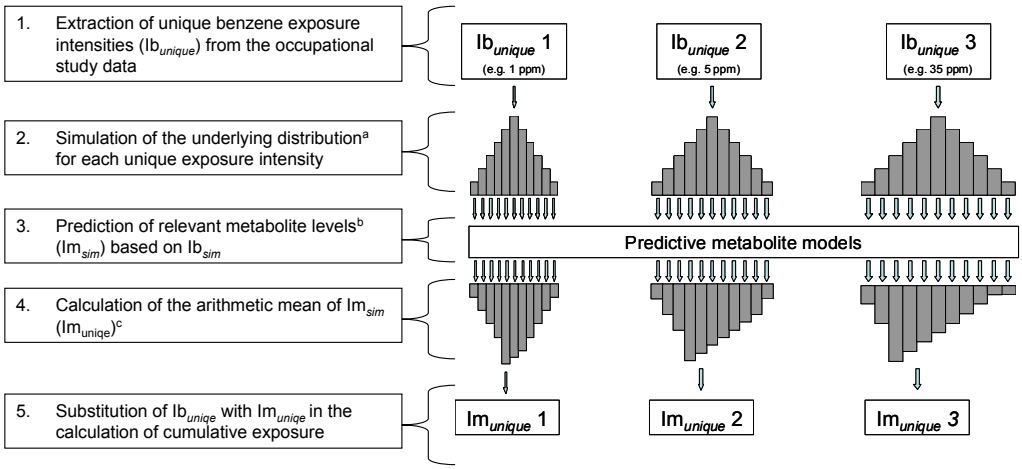


Figure 2 Approach to substitute benzene with measures of relevant benzene metabolite biomarkers.

[a] For each Ib$_{unique}$ an underlying distribution consisting of one thousand measurements was simulated. The simulation was based on fifty (hypothetical) workers per Ib$_{unique}$ and twenty (hypothetical) exposure measurements per worker.
[b] Sum of catechol, hydroquinone, *t,t*-muconic acid, *S*-phenyl mercapturic acid, and phenol.
[c] The geometric mean of Im$_{sim}$ was calculated for each simulated worker (Im$_{worker}$), the median of Im$_{worker}$ was calculated for each unique exposure level (Im$_{unique\text{-}gm}$), and the arithmetic mean (Im$_{unique}$) was estimated based on Im$_{unique\text{-}gm}$ and its variance.

Extraction of unique benzene exposure intensities from the study data

The PL and HW studies were selected based on their quality (22) and the large contrast between them in terms of the duration and intensity of benzene exposure. All unique intensities of benzene exposure (Ib_{unique}) assigned to individual subjects in the original epidemiological analysis were extracted. We interpret Ib_{unique} as the arithmetic mean of an underlying distribution of eight-hour average exposure intensity levels experienced by a subject.

Simulation of the underlying distribution for each unique exposure intensity

To reconstruct the full distribution of exposure intensities that each study subject is likely to have experienced, we simulated the underlying exposure distribution of each Ib_{unique} as follows. First, within-worker variance components (δ_{ww}) and between-worker variance components (δ_{bw}) were assumed for the HW and PL studies using the estimated median values of δ_{ww} and δ_{bw} reported by Kromhout et al. (23) for the chemical industry (PL: δ_{ww} = 2.05, δ_{bw} = 1.49) and the petroleum-refining industry (HW: δ_{ww} = 3.35, δ_{bw} = 1.43). Then, the natural logarithm of the geometric mean ($Ib_{unique\text{-}gm}$) corresponding to each Ib_{unique} was calculated as $\ln(Ib_{unique\text{-}gm}) = [\ln(Ib_{unique}) - \ln(\delta_{ww}+\delta_{bw})/2]$ (24). Next, a geometric mean exposure intensity ($Ib_{worker\text{-}gm}$) was simulated for fifty virtual workers based on $Ib_{uniqe\text{-}gm}$ and δ_{bw}. Finally, twenty exposure intensity levels (Ib_{sim}) were simulated for each simulated worker based on $Ib_{worker\text{-}gm}$ and δ_{ww}. In total, one thousand values of Ib_{sim} exposure were simulated for each value of Ib_{unique}.

Prediction of urinary metabolite levels

The sum of urinary levels of CAT, HQ, MA, PH, and SPMA (hereafter *sum of metabolites*) was used as the indicator of benzene metabolism. These *sums of metabolites* were estimated from the corresponding values of Ib_{sim} by three different methods.

1) *Physiologically based pharmacokinetic model.* The PBPK model reported by Yokley et al. (16) was used to predict the amount of CAT, HQ conjugates, MA, PH conjugates, and SPMA excreted in urine after 8 hours of continuous exposure to benzene. We constructed the PBPK model using Berkeley Madonna software (version 8.3.14, Berkeley, CA, USA). We assumed an 8-hour urinary output of 0.4 L (25) to convert the amount of the *sum of metabolites* that was predicted by the PBPK model into urinary concentrations of the *sum of metabolites*.

2) *Michaelis-Menten-like model.* The Michaelis-Menten-like (MML) model reported by Rappaport et al. (10) was used to predict the *sum of metabolites*. This MML model assumes two metabolic pathways that compete for access to benzene plus an intercept which accounts for a background level of benzene metabolites from dietary and endogenous sources (26). Parameters for the MML model had been estimated by fitting the model to data from 263 nonsmoking females from two studies of Chinese workers for whom individual levels of airborne benzene (0.001-299 ppm) and urinary metabolite levels had been documented (10, 27, 28).

3) *Regression splines model.* The 5 regression splines reported by Kim et al. (19) were used to relate airborne benzene exposure to urinary levels of CAT, HQ, MA, PH, and SPMA. The predictions of the individual regression splines were summed to calculate the *sum of metabolites.* The splines had been derived from a cross-sectional study that included airborne benzene (0.03-88.9 ppm) and urinary metabolite measurements of 326 exposed workers in Tianjin, China (9, 19).

Note that there is a considerable overlap (240 nonsmoking females) in the data that were used to derive the regression splines and the parameters for the MML model. For the empirical models (the MML model and the regression splines model) the urinary level of metabolites predicted for a benzene exposure of zero ppm was subtracted from the model prediction to isolate the contribution of benzene exposure to the *sum of metabolites*. Because the regression splines were fitted using log-transformed airborne benzene exposure levels, we used the second derivative of the splines at the first knot (0.004 ppm for HQ, MA, PH, SPMA and 0.04 ppm for CAT) to predict the shape of the curve for exposures below the first knot. The intercepts of the second derivatives at zero occupational exposure (CAT = 11.93 µmol/l, HQ = 5.79 µmol/l, MA = 0.79 µmol/l, PH = 63.23 µmol/l, SPMA = 0.003 µmol/l) were summed to estimate the background level of the *sum of metabolites* for the regression splines model.

The three models were used to predict Im_{sim} for each level of Ib_{sim}. Subsequently, the geometric mean of Im_{sim} was calculated for each simulated worker (Im_{worker}), the median of Im_{worker} was calculated for each unique exposure level ($Im_{unique\text{-}gm}$), and the arithmetic mean (Im_{unique}) was estimated based on $Im_{unique\text{-}gm}$ and its variance (Figure 2).

Epidemiological analyses

Epidemiological analyses were conducted using two measures of exposure intensity, namely, each subject's arithmetic mean air concentration (Ib_{unique}) and the corresponding arithmetic mean metabolite level (Im_{unique}) based on the PBPK, MML, and regression splines models. Ib_{unique} and Im_{unique} levels were combined with the duration of exposure data from the original studies to calculate cumulative and average intensity of exposure to benzene or the *sum of metabolites*. The log RR per unit of exposure to benzene or to the *sum of metabolites* was estimated with conditional logistic regression for the HW data (similar to the analysis of Glass et al. (20)) and with Cox proportional hazards regression for the PL data (similar to the analysis of Rinsky et al. (21)). Akaike's Information Criterion (AIC) (29) was used to compare the fit of the epidemiological models to the data. Both types of models were fitted with the PHREG procedure in SAS (version 9.2, Cary, NC, USA). All models were run five times to account for the potential influence of random variation in the simulation of the underlying distribution of unique exposure intensities. The MIANALYZE procedure in SAS was used to combine the results of the five runs. To explore whether there was evidence for a non-linear

effect of exposure to benzene or to the *sum of metabolites* we repeated the epidemiological analyses with the clogit (HW) and coxph (PL) procedures from the survival package for R (version 2.9.1, Vienna, Austria) and included a penalized spline (degrees of freedom based on the AIC) for exposure. A likelihood ratio test demonstrated that allowing more flexibility for the effect of exposure within studies did not significantly increase model fit in any of the analyses (results not shown). Consequently, only the linear estimate of the log RR per unit of exposure was used in the current analysis.

Results

Health Watch and Pliofilm studies

In Figure 3 the characteristics of the exposure metrics used in HW and PL are compared by plotting the cumulative distribution of cumulative exposure (Figure 3a), duration of exposure (Figure 3b), and average intensity of exposure to benzene (Figure 3c). The plots for the PL study are based on exposed workers only (70% of the study population). Figure 3a illustrates that the HW study population had consistently lower cumulative exposures than the PL study population. More extreme differences between HW and PL were observed for duration of exposure (the full population of HW was exposed for 5-10 years, while the majority of the PL study population was exposed < 1 year) and average intensity of exposure (calculated as cumulative exposure divided by duration of exposure). The average intensity of exposure in the HW study ranged from 0.001-1 ppm, while in the PL study the average intensity of exposure ranged from 1-40 ppm. The median assigned exposure intensity was 0.1 ppm in the HW study and 23 ppm in the PL study (not shown in Figure 3).

In Table 1 the estimate for the log RR per unit of cumulative exposure, duration of exposure, and average intensity of exposure to benzene are reported for HW and PL. Relative risks with 95% confidence intervals were calculated for four different exposure levels based on the log RR per unit of exposure and the associated standard error. The estimate of the log RR per unit of cumulative exposure was significant in both HW ($p = 0.0005$) and PL ($p < 0.0001$). However the slope-estimate for HW was ~20-fold higher than the estimate for PL. Interestingly, no formal statistical significant effect for duration of exposure was observed in HW ($p = 0.26$), while the effect in PL was significant ($p = 0.001$). Furthermore, the effect of average intensity of exposure observed in HW was ~100 fold higher than the effect of average intensity observed in PL. The comparison of AIC values (29) across the different models in Table 1 showed that the models based on cumulative exposure and average intensity of exposure to benzene fit the data equally well in the HW study, but not in the PL study, where cumulative exposure to benzene provided a much better fit than average intensity of exposure.

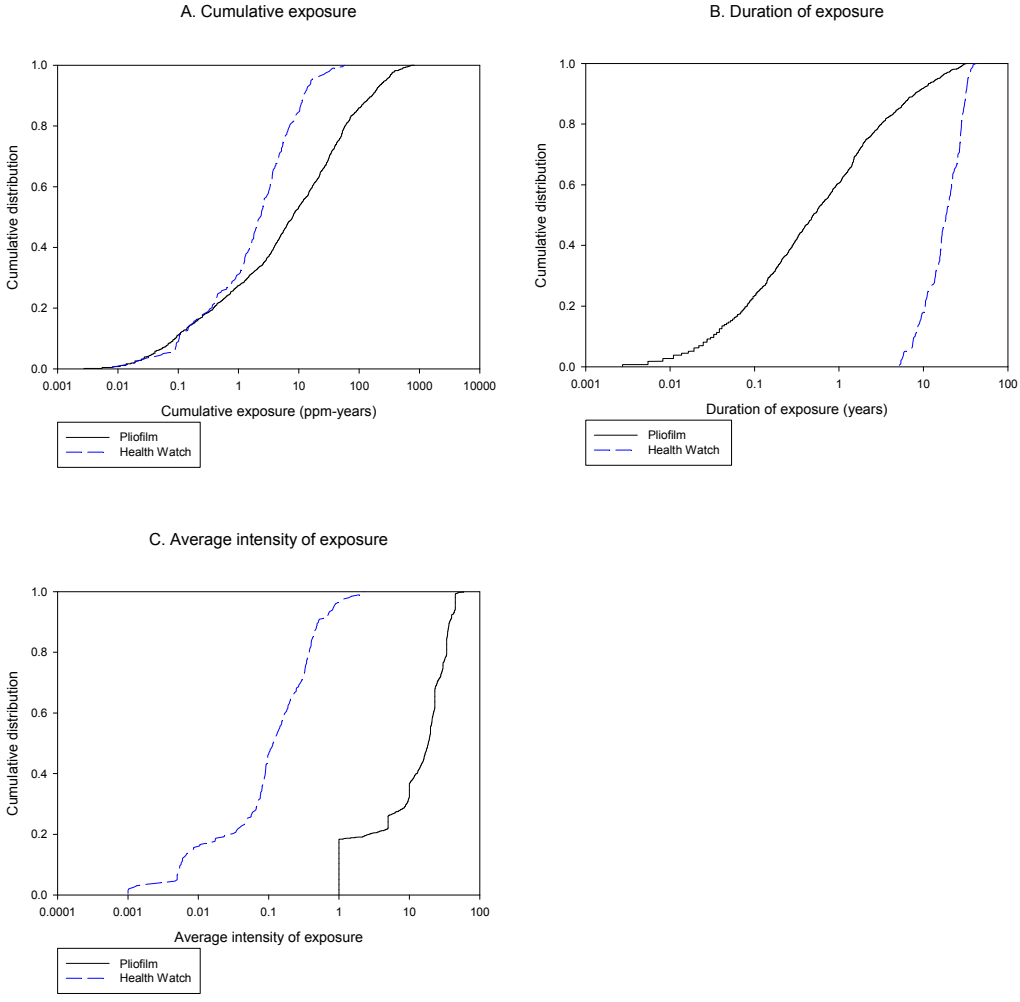


Figure 3 Cumulative distribution of cumulative exposure, duration of exposure, and average intensity of exposure to benzene in the Pliofilm[a] and Health Watch studies.

[a] Plots are based on occupationally benzene exposed individuals only; non-exposed individuals in the PL study were excluded (30% of the study population).

Table 1 Estimate of the risk of leukemia per unit cumulative, duration or average intensity of exposure and relative risks with 95% confidence intervals for four different exposure levels.

A. cumulative exposure

Study	Estimate (s.e.)[a]	AIC	RR at 1 ppm-yr[b]	RR at 10 ppm-yrs[b]	RR at 20 ppm-yrs[b]	RR at 40 ppm-yrs[b]
Health Watch	$9.40*10^{-2}$ ($2.71*10^{-2}$) $p = 0.0005$	103.5	1.10 (1.04-1.16)	2.57 (1.51-4.37)	6.58 (2.27-19.07)	43.33 (5.16-363.79)
Pliofilm	$5.10*10^{-3}$ ($1.11*10^{-3}$) $p < 0.0001$	206.9	1.01 (1.00-1.01)	1.05 (1.03-1.08)	1.11 (1.06-1.16)	1.23 (1.12-1.34)

B. Duration of exposure

Study	Estimate (s.e.)[a]	AIC	RR at 1 yr[b]	RR at 5 yrs[b]	RR at 10 yrs[b]	RR at 20 yrs[b]
Health Watch	$3.39*10^{-2}$ ($3.02*10^{-2}$) $p = 0.26$	119.0	1.03 (0.62-1.72)	1.18 (0.09-15.22)	1.40 (0.01-231.60)	1.97 (0-53637)
Pliofilm	$7.17*10^{-2}$ ($2.15*10^{-2}$) $p = 0.001$	214.3	1.07(1.03-1.12)	1.43 (1.16-1.77)	2.05 (1.34-3.12)	4.20 (1.81-9.76)

C. Average intensity of exposure[c]

Study	Estimate (s.e.)[a]	AIC	RR at 1 ppm[b]	RR at 2 ppm[b]	RR at 5 ppm[b]	RR at 10 ppm[b]
Health Watch	2.06 ($5.86*10^{-1}$) $p = 0.0005$	103.5	7.81 (2.48-24.66)	61.04 (6.13-608)	n.c.[d]	n.c.[d]
Pliofilm	$2.50*10^{-2}$ ($1.27*10^{-2}$) $p = 0.04$	216.3	1.03 (0.94-1.12)	1.05 (0.88-1.25)	1.14 (0.74-1.75)	1.29 (0.54-3.07)

[a] Estimate of the log relative risk of leukemia per unit of exposure. Estimate and associated standard error are from conditional logistic regression for the HW data and from Cox proportional hazards regression for the PL data.
[b] Relative risks and 95% confidence intervals were predicted with the log relative risk per unit of exposure and the associated standard error.
[c] Calculated as cumulative exposure divided by duration of exposure.
[d] Not calculated (outside the range of study exposure levels).

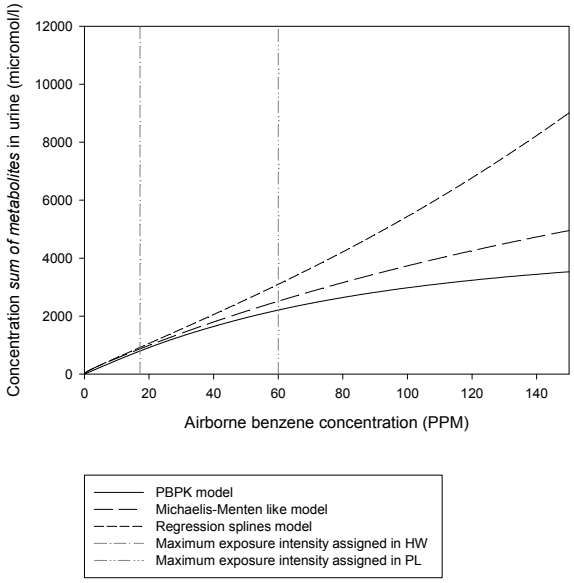


Figure 4a Predictions of the *sum of metabolites* by three different models (exposure range 0 – 150 ppm).

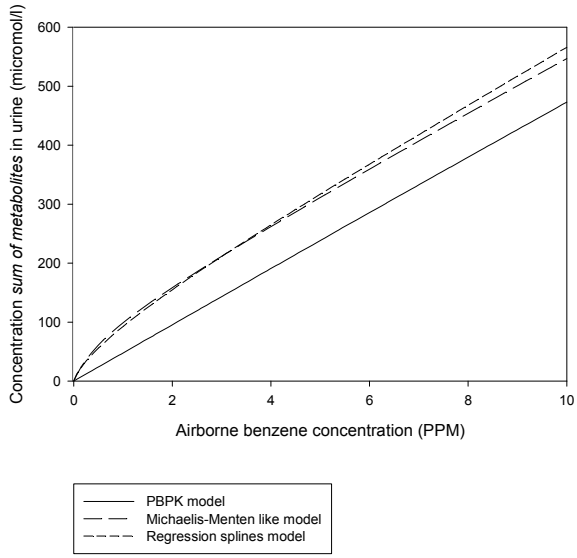


Figure 4b Predictions of the *sum of metabolites* by three different models (exposure range 0 – 10 ppm).

Predicted levels of benzene metabolites

Figures 4a (exposure range 0 – 150 ppm) and 4b (exposure range 0 – 10 ppm) show the *sum of metabolites* predicted by the PBPK, MML, and regression splines models. The two reference lines in Figure 4a indicate the maximum exposure intensity levels that were assigned in HW (~17 ppm) and PL (~60 ppm). At benzene concentrations above about 50 ppm the metabolite levels predicted from the regression splines were substantially higher than those from the other two models (Figure 4a).

For the exposure range below 10 ppm, the MML and regression splines models predicted greater-than-proportional production of the *sum of metabolites*, while the prediction of the PBPK model was proportional to the airborne benzene concentration (Figure 4b).

Effect of modeling saturable metabolism on risk estimates

In Table 2 the slope-estimates are given for the log RR of leukemia per unit of either cumulative exposure to benzene or the *sum of metabolites*. Based on cumulative exposure to benzene, the unit risk estimate for the PL study was 18-fold higher than that for the HW study. When cumulative *sum of metabolites* was used as the predictor variable instead of cumulative benzene exposure, the difference in unit risks between the two studies decreased to 16-fold (11% decrease) for the PBPK model, 12-fold (33% decrease) for the regression splines model, and 10-fold (44% decrease) for the MML model. Evaluation of AIC values indicated that modeling saturable metabolism did not improve the fits of models for either HW or PL. Analyses based on average intensity of benzene exposure and average intensity of the *sum of metabolites* resulted in a similar (albeit stronger) impact of modeling saturable metabolism on the leukemia risk estimates in HW and PL ranging from a 29% to 51% decrease in fold difference.

Discussion

By considering possible saturation of metabolism in two prominent epidemiological studies of benzene-induced leukemia (PL and HW) we were able to narrow the difference in estimates of the log RR per unit of cumulative exposure by about a factor of two (PL/HW: decreased from 18 for external exposure to 10 for the MML model of metabolite levels). We therefore provided some evidence that the heterogeneity in RR estimates between these two studies might be explained, in part, by saturable metabolism.

Table 2 Log relative risk of leukemia per unit cumulative exposure to benzene or *sum of metabolites*[a].

Exposure metric	HW[b] Estimate (s.e.)[c]	AIC[d]	PL[e] Estimate (s.e.)[c]	AIC[d]	Fold difference[f]
Benzene	$98.9*10^{-3}$ ($2.93*10^{-2}$) *p = 0.0007*	103.7	$5.35*10^{-3}$ ($1.23*10^{-3}$) *p < 0.0001*	207.0	18
Sum of metabolites[a] PBPK model	$2.09*10^{-3}$ ($6.16*10^{-4}$) *p = 0.0007*	104.4	$0.13*10^{-3}$ ($3.03*10^{-5}$) *p < 0.0001*	207.5	16
Sum of metabolites[a] MML model	$1.22*10^{-3}$ ($3.54*10^{-4}$) *p = 0.0006*	103.9	$0.12*10^{-3}$ ($2.79*10^{-5}$) *p < 0.0001*	207.6	10
Sum of metabolites[a] Regression splines model	$1.32*10^{-3}$ ($3.78*10^{-4}$) *p = 0.0005*	103.3	$0.11*10^{-3}$ ($2.45*10^{-5}$) *P < 0.0001*	207.2	12

[a] Sum of catechol, hydroquinone, *t,t*-muconic acid, *S*-phenyl mercapturic acid, and phenol.
[b] Health Watch study.
[c] Estimates and associated standard errors (s.e.) are from conditional logistic regression for the HW data and from Cox proportional hazards regression for the PL data. All models included a simulation step to estimate the underlying distribution of unique exposure intensities used in the original epidemiological analysis. Models were run five times to account for the potential influence of random variation in the simulation of the underlying distribution of unique exposure intensities. The MIANALYZE procedure in SAS (version 9.2, Cary, NC, USA) was used to combine the results of the five runs.
[d] Calculated as the average of the AIC values reported for the five runs.
[e] Pliofilm study
[f] Fold difference is defined as the log relative risk per unit of cumulative exposure derived in the HW study divided by the log relative risk per unit of cumulative exposure derived in the PL study.

The interpretation of our findings is complicated by the fact that saturable metabolism was probably not the only factor contributing to the heterogeneity of the leukemia risk estimates reported in the HW and PL studies. Exposure misclassification is a common problem in retrospective occupational cohort studies (30) that could have affected the estimated log RR per unit of cumulative exposure in the two studies. Systematic differences in exposure misclassification between the studies might have resulted from the considerable degree of extrapolation and expert judgment, particularly in the PL study which had access to far fewer benzene measurements than the HW study (~750 measurements in the PL study (31)) compared to > 3870 measurements in the HW study (32)). Indeed, the limited quality of the exposure data in the PL study resulted in three groups of authors publishing three different sets of exposure estimates, based on the same data (21, 33-35). Other potential explanations for the heterogeneity in risk estimates between HW and PL could be related to differences between the study populations in their susceptibility to benzene induced health effects, differences in background rates for leukemia (it has been suggested that background rates were too low in the HW study (36)), differences in study design (nested case-control vs. cohort), or differences in confounding factors (7, 37). In addition, leukemia is not a single disease entity. Stronger associations have been shown for some leukemia subtypes (e.g., acute myeloid leukemia) than for other subtypes. Differences between the study populations in the distribution of subtypes that contributed to the overarching disease outcome 'leukemia' might have contributed to heterogeneity in the risk estimates as well.

Our decision to use the sum of the major urinary benzene metabolites to reflect saturable metabolism was motivated by the readily availability of models that quantitatively described the relation between airborne exposure to benzene and these biomarkers of exposure and the fact that they collectively account for virtually all of the metabolized benzene dose (10, 16, 19). Furthermore although some individual metabolites (i.e., HQ and 1,4-benzoquinone) are frequently mentioned as being the most toxic benzene metabolites, there is still considerable uncertainty regarding exactly which metabolites are causally related to leukemia (15, 19, 26). When we applied our approach using the predictions of the PBPK and regression splines model for only urinary HQ concentrations, the results were essentially the same as for the *sum of metabolites* (see Supplemental Material). Although the biomarkers were not measured in the target tissue (the bone marrow, where benzene induced toxicological effects are primarily thought to take place), they do reflect saturation of the CYP enzymes which are primarily active in the liver (10, 15).

We should consider the possibility that the predictive models included in our approach did not fully reflect the extent to which saturable metabolism actually occurs in humans. In the PBPK model, error might have been introduced by invalid specification of the model or its parameters, and measurement errors of benzene and its metabolites might have played a role in the empirical spline and MML models. It is also important to mention that the

empirical models were derived in populations of Chinese predominantly female workers that may not be directly comparable to the predominantly white male workers that were included in the study populations of HW and PL.

Our approach is a form of dosimetric modeling which constructs measures of exposure using explicit hypotheses about the exposure-dose and/or dose-risk relation (38). Extending the approach by incorporating explicit hypotheses about the interaction between duration of exposure and the concentration in the target tissue, clearance from the target tissue, repair processes, etc. could provide further insight into the exposure disease relations. Connolly et al. presented an example of such an extended model applied to formaldehyde and cancer of the respiratory tract (39). However, extension of our approach in the current example of benzene and leukemia would likely be limited by the statistical power and the levels of detail that are available in the exposure data in the HW and PL studies.

As an alternative to constructing hypothesis-based dosimetric models, flexible approaches that allow more independent weighting of duration, intensity, and timing of exposure can also be used to better predict cancer incidence in occupational studies and to evaluate heterogeneity in risk estimates across studies. Several approaches have been proposed for doing this (e.g., (3, 4, 40)), but are generally difficult to apply in occupational studies that often lack sufficient statistical power for such complex models. Knowledge about saturable metabolism can provide clues as to where to expect non-linearity in flexible models and will therefore be helpful in the interpretation of the outcomes of flexible approaches.

Our approach provides a general framework for estimating internal exposure levels from external measurements and then using these internal levels to explore the effect of saturation on exposure-response relationships. Our approach can be applied to studies of any toxicant provided that substantial data are available for external exposure levels, and the metabolic pathway supports a hypothesis regarding the impact of saturable metabolism on the exposure-response relation.

Acknowledgements

This work was performed as part of the work package “Integrated Risk Assessment” of the ECNIS Network of Excellence (Environmental Cancer Risk, Nutrition and Individual Susceptibility), operating within the European Union 6th Framework Program, Priority 5: “Food Quality and Safety” (FOOD-CT-2005-513943).

References

1. Rappaport SM. Assessing workplace exposures: turning to the past for guidance. Occup Environ Med. 2009;66(7):429-30.
2. Smith TJ. Occupational exposure and dose over time: limitations of cumulative exposure. Am J Ind Med. 1992;21(1):35-51.
3. Lubin JH, Moore LE, Fraumeni JF, Jr, et al. Respiratory cancer and inhaled inorganic arsenic in copper smelters workers: a linear relationship with cumulative exposure that increases with concentration. Environ Health Perspect. 2008;116(12):1661-5.
4. Seixas NS, Robins TG, Becker M. A novel approach to the characterization of cumulative exposure for the study of chronic occupational disease. Am J Epidemiol. 1993;137(4):463-71.
5. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. J Epidemiol Community Health. 1978;32(4):303-13.
6. Andersen ME. Saturable metabolism and its relationship to toxicity. Crit Rev Toxicol. 1981;9(2):105-50.
7. Stayner L, Steenland K, Dosemeci M, et al. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environ Health. 2003;29(4):317-24.
8. Clewell HJ, 3rd, Andersen ME, Barton HA. A consistent approach for the application of pharmacokinetic modeling in cancer and noncancer risk assessment. Environ Health Perspect. 2002;110(1):85-93.
9. Kim S, Vermeulen R, Waidyanatha S, et al. Modeling human metabolism of benzene following occupational and environmental exposures. Cancer Epidemiol Biomarkers Prev. 2006;15(11):2246-52.
10. Rappaport SM, Kim S, Lan Q, et al. Evidence that humans metabolize benzene via two pathways. Environ Health Perspect. 2009;117(6):946-52.
11. Vlaanderen J, Portengen L, Rothman N, et al. Flexible meta-regression to assess the shape of the benzene-leukemia exposure-response curve. Environ Health Perspect. 2010;118(4):526-32.
12. Henderson RF, Sabourin PJ, Bechtold WE, et al. The effect of dose, dose rate, route of administration, and species on tissue and blood levels of benzene metabolites. Environ Health Perspect. 1989;82:9-17.
13. Lin YS, Vermeulen R, Tsai CH, et al. Albumin adducts of electrophilic benzene metabolites in benzene-exposed and control workers. Environ Health Perspect. 2007;115(1):28-34.
14. Rothman N, Li GL, Dosemeci M, et al. Hematotoxicity among Chinese workers heavily exposed to benzene. Am J Ind Med. 1996;29(3):236-46.
15. Snyder R. Benzene's toxicity: a consolidated short review of human and animal studies by HA Khan. Hum Exp Toxicol. 2007;26(9):687-96.
16. Yokley K, Tran HT, Pekari K, et al. Physiologically-based pharmacokinetic modeling of benzene in humans: a Bayesian approach. Risk Anal. 2006;26(4):925-43.
17. Rappaport SM, Waidyanatha S, Qu Q, et al. Albumin adducts of benzene oxide and 1,4-benzoquinone as measures of human benzene metabolism. Cancer Res. 2002;62(5):1330-7.
18. Rappaport SM, Waidyanatha S, Yeowell-O'Connell K, et al. Protein adducts as biomarkers of human benzene metabolism. Chem Biol Interact. 2005;153-154:103-9.
19. Kim S, Vermeulen R, Waidyanatha S, et al. Modeling human metabolism of benzene following occupational and environmental exposures. Cancer Epidemiol Biomarkers Prev. 2006;15(11):2246-52.
20. Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology. 2003;14(5):569-77.
21. Rinsky RA, Hornung RW, Silver SR, et al. Benzene exposure and hematopoietic mortality: A long-term epidemiologic risk assessment. Am J Ind Med. 2002;42(6):474-80.
22. Vlaanderen J, Vermeulen R, Heederik D, et al. Guidelines to evaluate human observational studies for quantitative risk assessment. Environ Health Perspect. 2008;116(12):1700-5.
23. Kromhout H, Symanski E, Rappaport SM. A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. Ann Occup Hyg. 1993;37(3):253-70.
24. Rappaport SM, Selvin S. A method for evaluating the mean exposure from a lognormal distribution. Am Ind Hyg Assoc J. 1987;48(4):374-9.
25. Leung HW. Use of physiologically based pharmacokinetic models to establish biological exposure indexes. Am Ind Hyg Assoc J. 1992;53(6):369-74.
26. McDonald TA, Holland NT, Skibola C, et al. Hypothesis: phenol and hydroquinone derived mainly from diet and gastrointestinal flora activity are causal factors in leukemia. Leukemia. 2001;15(1):10-20.

27. Waidyanatha S, Rothman N, Li G, et al. Rapid determination of six urinary benzene metabolites in occupationally exposed and unexposed subjects. Anal Biochem. 2004;327(2):184-99.
28. Kim S, Vermeulen R, Waidyanatha S, et al. Using urinary biomarkers to elucidate dose-related patterns of human benzene metabolism. Carcinogenesis. 2006;27(4):772-81.
29. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, eds. Second international symposium on information theory. Budapest: Academiai Kiado, 1973:267-281.
30. Blair A, Stewart P, Lubin JH, et al. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med. 2007;50(3):199-207.
31. Rinsky RA, Young RJ, Smith AB. Leukemia in benzene workers. Am J Ind Med. 1981;2(3):217-45.
32. Vlaanderen J, Fransman W, Miller B, et al. A graphical tool to evaluate temporal coverage of occupational history by exposure measurements. Occup Environ Med. 2010;67(9):636-8.
33. Crump KS. Risk of benzene-induced leukemia predicted from the Pliofilm cohort. Environ Health Perspect. 1996;104 Suppl 6:1437-41.
34. Paustenbach DJ, Price PS, Ollison W, et al. Reevaluation of benzene exposure for the Pliofilm (rubberworker) cohort (1936-1976). J Toxicol Environ Health. 1992;36(3):177-231.
35. Rinsky RA, Smith AB, Hornung R, et al. Benzene and leukemia. An epidemiologic risk assessment. N Engl J Med. 1987;316(17):1044-50.
36. Schnatter AR. Benzene exposure and leukemia. Epidemiology. 2004;15(4):509; author reply 510-1.
37. Hertz-Picciotto I, Smith AH. Observations on the dose-response curve for arsenic exposure and lung cancer. Scand J Work Environ Health. 1993;19(4):217-26.
38. Kriebel D, Checkoway H, Pearce N. Exposure and dose modelling in occupational epidemiology. Occup Environ Med. 2007;64(7):492-8.
39. Conolly RB, Kimbell JS, Janszen D, et al. Human respiratory tract cancer risks of inhaled formaldehyde: dose-response predictions derived from biologically-motivated computational modeling of a combined rodent and human dataset. Toxicol Sci. 2004;82(1):279-96.
40. Richardson DB. Multistage modeling of leukemia in benzene workers: a simple approach to fitting the 2-stage clonal expansion model. Am J Epidemiol. 2009;169(1):78-85.

Supplemental Material, Table 1 Log relative risk of leukemia per unit of cumulative urinary hydroquinone concentration.

Exposure metric	HW[a] Estimate (s.e.)[b]	AIC[c]	PL[d] Estimate (s.e.)[b]	AIC[c]	Fold difference[e]
Benzene	$98.9*10^{-3}$ ($2.93*10^{-2}$) *p = 0.0007*	103.7	$5.35*10^{-3}$ ($1.23*10^{-3}$) *p < 0.0001*	207.0	18
Urinary hydroquinone *PBPK model*	$19.6*10^{-3}$ ($5.01*10^{-3}$) *p = 0.0007*	104.1	$1.38*10^{-3}$ ($3.10*10^{-4}$) *P < 0.0001*	207.5	12
Urinary hydroquinone *Regression splines model*	$14.2*10^{-3}$ ($4.08*10^{-3}$) *p = 0.0005*	104.0	$1.52*10^{-3}$ ($3.47*10^{-4}$) *p < 0.0001*	207.8	9

[a] Health Watch study

[b] Estimates and associated standard error (s.e.) are from conditional logistic regression for the HW data and from Cox proportional hazards regression for the PL data. All models included a simulation step to estimate the underlying distribution of unique exposure intensities used in the original epidemiological analysis. Models were run five times to account for the potential influence of random variation in the simulation of the underlying distribution of unique exposure intensities. The MIANALYZE procedure in SAS (version 9.2, Cary, NC, USA) was used to combine the results of the five runs.

[c] Calculated as the average of the AIC values reported for the five runs.

[d] Pliofilm study

[e] Fold difference is defined as the log relative risk per unit of cumulative exposure derived in the HW study divided by the log relative risk per unit of cumulative exposure derived in the PL study.

Chapter 8

General Discussion

In this thesis a set of approaches has been presented to advance the use of occupational epidemiological studies in risk assessment. The approaches emphasized the importance of study quality, introduced transparent methods for evidence synthesis, and explored the potential for risk assessment of integrating biomarkers into occupational epidemiological studies.

Quality of occupational epidemiological studies

The quality of the design and the conduct of an occupational epidemiological study affect the potential for bias in study results and thus its value for risk assessment. Therefore, the evaluation of study quality should be a key aspect of risk assessment when it is based on occupational epidemiological studies. A comprehensive evaluation of study quality will facilitate the use of all available evidence but will acknowledge potential differences in the weight of evidence for risk assessment. The guidelines that were discussed in Chapter 2 provide a framework for evaluation of occupational epidemiological studies for risk assessment. In addition to more standard 'study quality' parameters such as evaluation of selection bias or potential impact of confounding factors, the guidelines focus on the quality of exposure assessment. Exposure assessment is a crucial aspect of occupational epidemiological studies, especially in studies that define exposure-response relations in quantitative terms (1).
In this thesis the guidelines were successfully applied in multiple chapters to evaluate the quality of epidemiological studies on occupational exposure to benzene and associated health outcomes. In Chapter 2 a set of epidemiological studies reporting on the association between occupational exposure to benzene and the risk of acute myeloid leukemia (AML) was evaluated to demonstrate the usefulness and practical implications of the application of the guidelines. In Chapter 4 the guidelines were used to explore the impact of study quality on the outcome of meta-analyses of occupational benzene exposure and lymphoma subtypes. In Chapter 5 the guidelines were used to select a set of occupational epidemiological studies that were of sufficient quality to be included in a meta-regression of benzene and leukemia.
In addition to the application in this thesis, the guidelines were also already used as the basis of quality evaluations of occupational epidemiological studies for meta-analyses of exposures to endotoxin and asbestos and the risk of lung cancer (2, 3).

The application of the evaluation guidelines in different chapters in this thesis highlighted two issues that are likely widespread in occupational epidemiology literature and are highly relevant for risk assessment: only a limited number of studies conducted quantitative exposure assessment; and most evaluated publications provided only limited details on the design and the conduct of the exposure assessment and assignment.

High quality quantitative exposure assessment is essential for risk assessment
In Chapter 2, only 5 out of 116 publications that were initially identified in the peer reviewed literature were found to be informative for a dose-response assessment for benzene and AML. Most studies were excluded because exposure levels were not quantified. A similar observation comes from a study by Sutedja et al. (4), where in a systematic review of the association of amyotrophic lateral sclerosis with exposure to chemical agents or metals, the authors excluded 30 out of 37 studies focused on exposure to chemical agents and excluded 47 out of 50 studies on exposure to metals because of the poor quality of exposure assessment in these studies. These two examples illustrate that the number of occupational epidemiological studies that is potentially informative for risk assessment is dramatically reduced when the quality of exposure assessment is taken into account. This is important because while studies with a qualitative exposure component (i.e., exposed versus non-exposed) might contribute to hazard identification, studies with a quantitative exposure component can be much more informative in discussions of causality of associations and in dose-response assessment. Risk assessment would therefore undoubtedly benefit from a further emphasis on and development of quantitative exposure assessment in occupational epidemiological studies.

As a result of differences from study to study in availability of exposure data, methods to generate quantitative exposure estimates are still far from standardized (5, 6). However, several aspects are central to the quality of quantitative exposure estimates and are discussed below.

Insight into the variability of exposure is an important aspect of a quantitative exposure assessment strategy. This variability can be attributed to a combination of measurement error and variation in exposure levels over time and between individuals. Classical measurement error, by which analytical and sampling error is covered, usually only plays a marginal role because its magnitude is often orders of magnitude smaller than variability in exposure over time and between individuals. The influence of exposure variability is dependent on the exposure assessment strategy. Often, exposure assessment at the individual level is considered to be the gold standard. However, this strategy is most sensitive to intra-individual variability of the exposure (the variability from day to day). If intra-individual variability is not correctly addressed in an individual-based exposure assessment strategy, severe underestimation of the exposure-response relationship might occur when this variability is large relative to the inter-individual variability of the exposure (the variability between individuals) in the population. Categorization of the population in *a priori* defined exposure groups, and use of measured average exposure for each of the *a priori* defined exposure groups in an exposure-response relationship is less sensitive to intra-individual variability (7). In most cases this strategy, leading to Berkson error, is known to lead to unbiased estimates of the association between exposure and response, however, unexplained differences in

health risks within exposure groups and unaddressed differences in exposure levels within exposure groups will lead to less precision and consequently a reduction of power in comparison with the individual exposure assessment strategy (8).

For most occupational epidemiological studies, the optimal exposure assessment strategy would involve distributing the sampling effort to achieve a balance between maximizing the number of workers that is monitored to account for between worker variability and maximizing the number of exposure measurements collected for each worker to capture within-worker variability (9, 10). Furthermore, group-based exposure assessment strategies should seek to minimize the variability within exposure groups, maximize the contrast between exposure groups, and minimize the standard error of the average exposure in each exposure group (8-10).

The quality of the exposure assessment in occupational epidemiological studies also depends on the quality of available exposure data. The quality of exposure data is determined by the analytical techniques that were used to measure and analyze exposure levels, but more so by the sampling strategy that was used to collect measurements (11). Before exposure data is used in occupational epidemiological studies the analytical techniques that were used to generate the exposure data should be evaluated (e.g., by side-by-side comparisons of the analytical technique employed with the techniques that are considered as best practice at the time of the study (12, 13)). Furthermore, researchers should evaluate whether the available exposure data is representative for the study population and not, for example, representing only worst-case scenarios (which could occur when relying only on exposure data collected for regulatory purposes) (11).

The biological relevance of an exposure metric used in an occupational epidemiological study is crucial for the quality of quantitative exposure estimates. The biological relevance of an exposure metric is determined by the ability to capture the duration, timing, and concentration of an agent in the target tissue (14, 15). Currently cumulative (airborne) exposure is often used as a standard exposure metric to reflect the target tissue dose (concentration in target tissue times duration of exposure). However, the assumptions that underlie the use of cumulative exposure as a proxy for the target tissue dose are frequently violated (14). Therefore, insight into biological considerations such as the time window of exposure that is relevant to the health effect of interest, and the fate of an agent once entered in the human body, would certainly further increase the quality of quantitative exposure assessment in occupational epidemiological studies (14, 15).

Increased transparency in the reporting of studies will facilitate the evaluation of occupational epidemiological studies for risk assessment

The application of the evaluation guidelines presented in this thesis illustrated that evaluation of the quality of occupational epidemiological studies was frequently hampered by a lack of detail or transparency regarding design, conduct, and analysis of studies in the evaluated publications. Currently many journals allow authors of published manuscripts to provide additional study details in online stored material. These facilities offer researchers the possibility to provide extensive details regarding design, conduct, and analysis that are relevant for the evaluation of occupational epidemiological studies for risk assessment and, if used to its full extent, will contribute significantly to an increase in transparency of the reporting of study results.

Because the quality of exposure assessment has a large impact on the validity of risk estimates that are provided by an occupational epidemiological study, extra attention should be paid to the reporting of methods and underlying assumptions used to generate the exposure estimates. A separate publication focused specifically on the exposure assessment part of an occupational epidemiological study that accompanies the publication in which main results are presented would be very beneficial for risk assessment purposes. Although examples exist in the literature (e.g., (16-20)), this is not yet standard practice and should be applied more broadly in occupational epidemiological studies. One approach to stimulate researchers to provide more detail on exposure assessment could be the requirement of providing sufficient information on the exposure assessment through online materials, or a separately published exposure assessment paper or report before results from an epidemiological analysis are considered for risk assessment.

An important aspect that should be addressed in the reporting of exposure assessment is the uncertainty of quantitative exposure estimates. Because exposures are generally assessed retrospectively, direct validation of estimated exposure levels is often difficult and exposure misclassification occurs frequently. Exposure misclassification has an impact on the internal validity of study findings (e.g., attenuation of exposure-response relations), but can also have an impact on the comparability of studies in risk assessment (1). Risk assessment would therefore greatly benefit if occupational epidemiological studies would characterize the uncertainty of quantitative exposure estimates by conducting sensitivity analyses. Examples of sensitivity analysis of the exposure assessment in occupational epidemiological studies can be found in publications by Loomis et al. (21) and Bhatti et al. (22) where the authors reported that the sensitivity analysis contributed considerably to their confidence in the study findings. The incorporation of probabilistic or Bayesian methods is another viable approach to gain further insight into the uncertainty surrounding exposure estimates in occupational epidemiological studies. While these approaches have been applied to a limited extent in

specific exposure assessment publications (23, 24), their application in occupational epidemiological studies has not yet matured.

Chapter 3 is an example of how increased transparency in the reporting of exposure assessment could facilitate the evaluation of the quality of occupational epidemiological studies. In this example, one specific aspect which significantly contributed to the quality of retrospective exposure assessment (the temporal coverage of occupational history by exposure measurements) was assessed with a graphical tool. The graphical tool is a straightforward method to provide insight into potential differences in temporal coverage of occupational history by exposure measurements between occupational epidemiological studies. Application of the tool to three nested case-control studies on occupational exposure to benzene and leukemia illustrated that there were considerable differences in the temporal coverage of occupational history with exposure measurement data between these studies. These differences would have remained unnoticed had these studies been evaluated based on peer-reviewed publications only. Illustratively, two of the three nested case-control studies to which the graphical tool was applied in Chapter 3 were also included in the demonstration of the application of the evaluation guidelines in Chapter 2. In that example a study from the UK (25) was ranked higher than a study from Australia (26), mainly because the UK study reported sensitivity analyses in the peer-reviewed publication. In hindsight, if the information on the difference in temporal coverage between these studies (45% for the Australian study and 9% for the UK study) would have been available during the application of the evaluation guidelines in Chapter 2, the Australian study would have likely ranked higher than the UK study. The inclusion of the proposed graphical tool in the reporting of occupational studies will facilitate evaluation of the quality of exposure assessment and is therefore recommended as standard information in occupational epidemiological publications.

Approaches for transparent evidence synthesis

Chapter 4 is an illustration of how differences in the quality of occupational epidemiological studies can be acknowledged in a transparent approach for evidence synthesis. In this chapter the focus was on the association between exposure to benzene and lymphoma subtypes. The quality of exposure assessment and the year-of-start of follow-up (a proxy for the quality of the outcome classification) were used to assess the quality of the studies included in a meta-analysis. Furthermore, arguing that any study that was not able to detect at least suggestive evidence of the established strong association between benzene and AML most likely had serious methodological limitations in one or more aspects of the study design, the strength of the reported AML association was used as proxy for the overall study quality as well. For several lymphoma subtypes meta-relative risks increased with increasing study quality,

regardless of the strategy used to assess study quality. With the exception of a chance finding, these trends most likely reflect an underlying association of benzene with at least some of these lymphoma subtypes. As such this approach demonstrated how a transparent integration of detailed evaluation of study quality into data synthesis contributed to identification of lymphoma subtypes with a potential causal relation to occupational exposure to benzene.
The work in Chapter 5 demonstrated how methods for quality evaluation and evidence synthesis can be applied in exposure-response analysis as well. A flexible meta-regression of aggregated risk estimates from a set of occupational epidemiological studies offered more insight into the functional relation between benzene exposure and leukemia. The meta-regression approach provided a comprehensive and quantitative assessment of the exposure-response relation and sensitivity analyses offered insight into the uncertainty of the predicted exposure-response curve (ERC). The predicted ERC could be applied directly in dose-response assessment and could therefore provide a considerable improvement over the approaches that have frequently been used in the benzene-leukemia dose-response assessment (such as using a single 'best' study) (27-30).

The application of the approaches described in this thesis to occupational epidemiological studies on benzene illustrated that the successful application of these types of approaches requires considerable amounts of (high quality) data. For example, in Chapter 4 the number of studies that was available for meta-analysis dropped considerably with increasing study quality resulting in an increased uncertainty surrounding the summary estimates in the high quality strata. Furthermore, the application of the meta-regression approach to lymphoma subtypes in Chapter 4 resulted in predictions that were not robust due to a lack of data. Taking into consideration that benzene is arguably one of the most studied occupational exposures it is uncertain to which extent these approaches could currently be applied successfully to other occupational exposures. However, the meta-regression approach was recently successfully applied to assess the exposure-response relationship for occupational asbestos exposure (another well studied occupational exposure) and lung cancer (31). Increasing the number of occupational epidemiological studies with a strong focus on the quality of exposure assessment would facilitate the application of the evidence synthesis approaches discussed in this thesis.

In addition to a lack of data (statistical power), publication bias is also a threat to the interpretability of evidence synthesis. Publication bias occurs if results from studies that are not in the public domain are systematically different from the results from studies that are in the public domain. The classic example of such a situation is when 'negative' small studies are less likely to be accepted for publication in the scientific literature. However, the opposite situation in which 'positive' studies are withheld from publication because they conflict with the interests of researchers or sponsors could also be envisioned.

Recently, the development of a database in which study details are registered in the design stage of a study and that provides unrestricted space to store study details (similar to the clinical trial registries) has been suggested to advance the use of occupational epidemiological studies in risk assessment (32, 33). Such an approach would likely increase the transparency in study outcomes and would diminish the potential impact of publication bias in evidence synthesis. However, opponents of the idea for registration noted that *a priori* registration of study hypotheses would likely limit the number of hypotheses that can be tested in a single study and would therefore limit the exploration of new ideas (34, 35). An approach in which study details are registered, but that does not restrict researchers to explore only *a priori* defined hypotheses would facilitate increased transparency in the evidence synthesis of occupational epidemiological studies without limiting researchers in their ability to explore new ideas.
Naturally, an increase in the quality of design and reporting of occupational epidemiological studies (especially with respect to exposure assessment) would also likely contribute to a diminished impact of publication bias on evidence synthesis as researchers or journals would be less inclined to withhold or reject studies from publication when there are few methodological flaws.

Incorporating biomarkers into occupational epidemiology: potential advantages for risk assessment

The technological advances in the development of biomarkers over the last years have stimulated the use of biomarkers in occupational epidemiological studies (36, 37). From the risk assessment perspective, biomarkers might contribute to an increase in the quality of occupational epidemiological studies and might facilitate formal integration of toxicological data into epidemiological studies (38) (e.g., through the linkage of exposure levels experienced in human populations to outcomes from animal bioassays with the use of a physiologically based pharmacokinetic model (39), or by using markers of intermediate health effects to inform exposure-response analysis (40)).
The potential advantages of using biomarkers in occupational epidemiological studies are commonly summarized as the possibility to improve the accuracy of exposure measurement, the possibility to identify intermediate health effects, the possibility to attain more homogeneous classifications of health effects, and the possibility to identify subpopulations with increased susceptibility to develop health effects in the presence of a toxicant (37). Examples of the potential opportunities for risk assessment of incorporating biomarkers in occupational epidemiological studies are the use of chromosomal aberrations (an intermediate health effect) to predict cancer risk (41), the use of patterns of TP53 mutations to distinguish tobacco-related lung cancer from non tobacco-related lung cancer (a more homogeneous classification of the health outcome) (42), and the use of genetic

polymorphisms to assess differential susceptibility for cancer risk in the presence of an adverse exposure (e.g., (43)). However, the actual value of a biomarker for an occupational epidemiological study largely depends on its validity. While the list of biomarkers is growing fast, the validation of new biomarkers is lagging behind. Another problem is the limited availability of historical measurements of biomarkers of exposure, which complicates the current application in epidemiological studies. Prospectively stored biological samples could create opportunities for further incorporating biomarkers of exposure in future occupational epidemiological studies, although the actual value of this approach will largely depend on the ability of biomarkers and sampling strategies to capture inter- and intra-individual variability in exposure experienced in the study population (44, 45).

In Chapter 6 an overview of a new generation of biomarker technologies (collectively described with the term OMICS) was provided and the challenges surrounding the implementation of these biomarkers in occupational (and environmental) epidemiological studies were discussed. The need for high quality study designs and the validation of biomarkers, and the difficulties with the interpretation of results from biomarker-based studies were emphasized. For most biomarker (especially OMICS) technologies, these issues currently prohibit the incorporation of biomarker-based occupational epidemiological studies in risk assessment. However, with the development of validated technologies, appropriate study designs, better sample handling and advanced statistical methods for data interpretation, biomarkers might eventually contribute significantly to risk assessment and could provide a more integrated view of the interaction between the environment and human health.

Chapter 7 is an example of how biomarker data can be formally integrated into occupational epidemiological studies. An approach was developed to translate airborne exposure estimates, used in the exposure assessment of epidemiological analyses, into relevant biomarkers of exposure with a set of predictive models. The approach was applied in an example of benzene and leukemia to assess the impact of saturation of benzene's enzymatic pathways on heterogeneity in the unit risk estimates that were reported in two occupational epidemiological studies. Application of the approach in the example of benzene and leukemia suggested that the heterogeneity in the unit risk estimates was partly attributable to saturable metabolism. However, considerable heterogeneity remained between the studies after the application of the approach. Differences in exposure (mis)classification, uncontrolled confounding factors, or other systematic differences between the studies are likely explanations for the remaining heterogeneity.
The approach in Chapter 7 can be seen as a form of dosimetric modeling (i.e., constructing measures of exposure using explicit hypotheses about the exposure-dose and/or dose-risk relation) (46). Models that, in addition to dosimetric modeling, attempt to incorporate all key events that lead to an adverse health effect are known as biologically based dose-response

(BBDR) models (47). Examples of BBDR models have been provided by Connolly et al. and Hack et al. (39, 40). The model by Conolly et al. describes the quantitative relation between formaldehyde and respiratory tract cancer (39). This model includes regional dosimetry predictions for the respiratory tract, sub-models that link the regional dosimetry predictions to DNA-protein cross-links (DPX) and regenerative cellular proliferation (CRCP), and a two-stage clonal growth model linking DPX and CRCP to tumor formation (39). Realistically, this type of approach will likely only be available for a small set of occupational exposures in the near future because the full quantitative linkage of all steps from exposure to disease will require vast amounts of research data. Hack et al. dealt with this problem by using a Bayesian network to quantitatively link elements in the exposure-disease continuum (40). In their approach, data on biomarkers spanning the continuum of exposure to disease was used to develop an exposure-response curve for benzene and AML. Interestingly, in this approach the Bayesian network-derived benchmark concentration was an order of magnitude lower than the benchmark concentration derived using the direct relationship between airborne benzene and AML. The authors suggested this was due to the inclusion of 8-OHdG (a marker for oxidative stress) that allowed the detection of intermediate health effects in the low-dose region (40).
It is important to realize that the validity of dosimetric and BBDR models largely depend on the availability, validity and accuracy of biomarker data used to generate and validate the model. Therefore this type of models will currently primarily contribute to generation and evaluation of hypotheses (48). However, further development of techniques to measure and validate biomarkers might contribute to a more central role for dosimetric and BBDR models in the assessment of risks at low exposure levels.

Conclusion

Occupational epidemiological studies can be very informative for risk assessment. However, the evidence that occupational epidemiological studies can generate is currently not used to its full extent. The approaches presented in this thesis can be used as a framework to advance the use of occupational epidemiological studies in risk assessment. Some of the approaches are ready for application in formal risk assessment (such as the evaluation guidelines, the graphical tool, and the methods for evidence synthesis); other approaches (such as incorporation of biomarkers in occupational epidemiological studies) need further development. Further progress in the use of occupational epidemiological studies in risk assessment should come from tailoring study designs to the needs of risk assessment. A strong focus on high quality quantitative exposure assessment in the design of new occupational epidemiological studies would significantly contribute to an increased weight of evidence for risk assessment and would likely improve the overall quality of risk assessment for many exposures.

References

1. Blair A, Stewart P, Lubin JH, et al. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med. 2007;50(3):199-207.
2. Lenters V, Basinas I, Beane-Freeman L, et al. Endotoxin exposure and lung cancer risk: a systematic review and meta-analysis of the published literature on agriculture and cotton textile workers. Cancer Causes Control. 2010;21(4):523-55..
3. Lenters V, Vermeulen R, Dogger S, et al. A meta-analysis of asbestos and lung cancer: is better quality exposure assessment associated with steeper slopes of the exposure-response relationships? Environ Health Perspect. In preperation.
4. Sutedja NA, Veldink JH, Fischer K, et al. Exposure to chemicals and metals and risk of amyotrophic lateral sclerosis: a systematic review. Amyotroph Lateral Scler. 2009;10(5-6):302-9.
5. Stewart PA, Lees PS, Francis M. Quantification of historical exposures in occupational cohort studies. Scand J Work Environ Health. 1996;22(6):405-14.
6. Sahmel J, Devlin K, Paustenbach D, et al. The role of exposure reconstruction in occupational human health risk assessment: Current methods and a recommended framework. Crit Rev Toxicol. 2010.
7. Armstrong BG. The effects of measurement errors on relative risk regressions. Am J Epidemiol. 1990;132(6):1176-84.
8. Tielemans E, Kupper LL, Kromhout H, et al. Individual-based and group-based occupational exposure assessment: some equations to evaluate different strategies. Ann Occup Hyg. 1998;42(2):115-9.
9. Loomis D, Kromhout H. Exposure variability: concepts and applications in occupational epidemiology. Am J Ind Med. 2004;45(1):113-22.
10. Heederik D, Attfield M. Characterization of dust exposure for the study of chronic occupational lung disease: a comparison of different exposure assessment strategies. Am J Epidemiol. 2000;151(10):982-90.
11. Stewart P. Challenges to retrospective exposure assessment. Scand J Work Environ Health. 1999;25(6):505-10.
12. de Vocht F, Huizer D, Prause M, et al. Field comparison of inhalable aerosol samplers applied in the European rubber manufacturing industry. Int Arch Occup Environ Health. 2006;79(8):621-9.
13. Stephenson DJ, Lillquist DR, DeRosso FD, et al. Side-by-side comparison of three sampling methods for aerosolized endotoxin in a wastewater treatment facility. J Environ Health. 2004;67(4):16-9.
14. Smith TJ. Occupational exposure and dose over time: limitations of cumulative exposure. Am J Ind Med. 1992;21(1):35-51.
15. Seixas NS, Robins TG, Becker M. A novel approach to the characterization of cumulative exposure for the study of chronic occupational disease. Am J Epidemiol. 1993;137(4):463-71.
16. Glass DC, Adams GG, Manuell RW, et al. Retrospective exposure assessment for benzene in the Australian petroleum industry. Ann Occup Hyg. 2000;44(4):301-20.
17. Armstrong TW, Pearlman ED, Schnatter AR, et al. Retrospective benzene and total hydrocarbon exposure assessment for a petroleum marketing and distribution worker epidemiology study. Am Ind Hyg Assoc J. 1996;57(4):333-43.
18. Burstyn I, Boffetta P, Kauppinen T, et al. Estimating exposures in the asphalt industry for an international epidemiological cohort study of cancer risk. Am J Ind Med. 2003;43(1):3-17.
19. Agostini M, Ferro G, Olsson A, et al. Exposure assessment for a nested case-control study of lung cancer among European asphalt workers. Ann Occup Hyg. 2010;54(7):813-23.
20. Stewart PA, Coble JB, Vermeulen R, et al. The diesel exhaust in miners study: I. Overview of the exposure assessment process. Ann Occup Hyg. 2010;54(7):728-46.
21. Loomis A, Kromhout H, Kleckner RC, et al. Effects of the analytical treatment of exposure data on associations of cancer and occupational magnetic field exposure. Am J Ind Med. 1998;34(1):49-56.
22. Bhatti P, Stewart PA, Linet MS, et al. Comparison of occupational exposure assessment methods in a case-control study of lead, genetic susceptibility and risk of adult brain tumours. Occup Environ Med. 2011; 68(1):4-9.
23. Xue J, Zartarian V, Wang SW, et al. Probabilistic Modeling of Dietary Arsenic Exposure and Dose and Evaluation with 2003-2004 NHANES Data. Environ Health Perspect. 2010;118(3):345-50.
24. Ramachandran G. Retrospective exposure assessment using Bayesian methods. Ann Occup Hyg. 2001;45(8):651-67.

25. Rushton L, Romaniuk H. 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. Occup Environ Med. 1997;54(3):152-66.
26. Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology. 2003;14(5):569-77.
27. U.S. EPA. Carcinogenic Effects of Benzene: An Update. EPA/600/P-97/001F. Washington, DC: United States Environmental Protection Agency, 1998.
28. California EPA. Public Health Goal for Benzene In Drinking Water. Scramento, California: Office of Environmental Health Hazard Assessment, 2001.
29. Gezondheidsraad. Comissie Beoordeling carcinogeniteit van stoffen. Benzeen. Rijswijk, 1997.
30. Gezondheidsraad. Benzeen. Toetsing van een basisdocument en voorstel risico-evaluatie. Den Haag, 1987.
31. van der Bij S, Koffijberg H, Lenters V, et al. Asbestos and lung cancer: meta-regression of the exposure-response relationship. to be published.
32. Swaen GM, Carmichael N, Doe J. Strengthening the reliability and credibility of observational epidemiology studies by creating an Observational Studies Register. J Clin Epidemiol. 2010.
33. European Centre for Ecotoxicology and Toxicology of Chemicals. Workshop: Enhancement of the Scientific Process and Transparency of Observational Epidemiology Studies. Brussels ed., November 2009.
34. Vandenbroucke JP. Preregistration of epidemiologic studies: an ill-founded mix of ideas. Epidemiology. 2010;21(5):619-20.
35. Editors. The registration of observational studies-when metaphors go bad. Epidemiology. 2010;21(5):607-9.
36. Knudsen LE, Hansen AM. Biomarkers of intermediate endpoints in environmental and occupational health. Int J Hyg Environ Health. 2007;210(3-4):461-70.
37. Boffetta P. Biomarkers in cancer epidemiology: an integrative approach. Carcinogenesis. 2010;31(1):121-6.
38. Schulte PA, Waters M. Using molecular epidemiology in assessing exposure for risk assessment. Ann N Y Acad Sci. 1999;895:101-11.
39. Conolly RB, Kimbell JS, Janszen D, et al. Human respiratory tract cancer risks of inhaled formaldehyde: dose-response predictions derived from biologically-motivated computational modeling of a combined rodent and human dataset. Toxicol Sci. 2004;82(1):279-96.
40. Hack CE, Haber LT, Maier A, et al. A Bayesian network model for biomarker-based dose response. Risk Anal. 2010;30(7):1037-51.
41. Bonassi S, Znaor A, Norppa H, et al. Chromosomal aberrations and risk of cancer in humans: an epidemiologic perspective. Cytogenet Genome Res. 2004;104(1-4):376-82.
42. Hainaut P, Pfeifer GP. Patterns of p53 G-->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. Carcinogenesis. 2001;22(3):367-74.
43. Hein DW, Doll MA, Fretland AJ, et al. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. Cancer Epidemiol Biomarkers Prev. 2000;9(1):29-42.
44. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005;14(8):1847-50.
45. Wild CP. Environmental exposure measurement in cancer epidemiology. Mutagenesis. 2009;24(2):117-25.
46. Kriebel D, Checkoway H, Pearce N. Exposure and dose modelling in occupational epidemiology. Occup Environ Med. 2007;64(7):492-8.
47. Luke NS, Sams R,2nd, DeVito MJ, et al. Development of a quantitative model incorporating key events in a hepatotoxic mode of action to predict tumor incidence. Toxicol Sci. 2010;115(1):253-66.
48. Crump KS, Chen C, Chiu WA, et al. What role for biologically based dose-response models in estimating low-dose risk? Environ Health Perspect. 2010;118(5):585-8.

Affiliation of Contributors

Dick Heederik, Hans Kromhout, Lützen Portengen, Roel Vermeulen
Division of Environmental Epidemiology (EEPI), Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, the Netherlands

Qing Lan, Lee E. Moore, Nathaniel Rothman
Division of Cancer Epidemiology and Genetics, National Cancer Institute, Department of Health and Human Services, National Institutes of Health, Bethesda, Maryland, USA

Stephen M. Rappaport, Christine F. Skibola, Martyn T. Smith, Luoping Zhang
Division of Environmental Health Sciences, University of California, Berkeley, California, USA

Igor Burstyn
Department of Environmental and Occupational Health, School of Public Health, Drexel University, Philadelphia, Pennsylvania, USA

Wouter Fransman
Business Unit Quality & Safety, TNO Quality of Life, Zeist, the Netherlands

Fintan Hurley, Brian Miller
Institute for Occupational Medicine, Edinburgh, UK

Deborah C. Glass
Department of Epidemiology & Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

Summary

Background

The identification and quantification of risk factors that are characterized by low exposure levels, moderately increased risks, and unspecific exposure-disease relations is a major challenge facing risk assessment today. Occupational epidemiological studies can play a role in addressing this challenge. The main advantage of occupational epidemiological studies over other potential sources of information for risk assessment (primarily animal bioassays) is that in these studies humans are being investigated, rendering the extrapolation of study results from animals to humans unnecessary. However this advantage is also a disadvantage because occupational epidemiological studies are mostly observational by nature which makes them prone to bias.

Although some limitations of the use of occupational epidemiological studies in risk assessment are inherent to the discipline, improvements in the design, conduct and interpretation of studies will likely enhance their use in risk assessment. Furthermore, recent developments in the field of molecular biology and the related increase in the understanding of carcinogenesis and other adverse health effects have opened opportunities to further advance the contribution of occupational epidemiological studies to risk assessment.

This thesis consists of a set of approaches to advance the use of occupational epidemiological studies in risk assessment. Some of the approaches are ready to be applied in risk assessment. Other approaches need further development before their actual value for risk assessment can be assessed.

Incorporating the weight of evidence principle in risk assessment based on occupational epidemiological studies

In Chapter 2 a set of guidelines to evaluate the quality of occupational epidemiological studies for (quantitative) risk assessment is presented. Because most occupational epidemiological studies are not conducted with risk assessment in mind, not all studies are equally informative for risk assessment. The quality of an occupational epidemiological study for risk assessment is determined by the quality of design, conduct and reporting of a study and by the relevance of the study hypothesis for the risk assessment in question. Differences in the quality of design and conduct are often not reflected in parameters of statistical uncertainty (e.g. confidence intervals). It is therefore important that a rigorous evaluation of occupational epidemiological studies is performed before study data is used in risk assessment. The guidelines in Chapter 2 consist of 20 evaluation criteria. The quality of (quantitative) exposure assessment receives particular attention in the guidelines because of the often neglected but notable large impact of exposure assessment on the overall quality of occupational epidemiological studies. The usefulness and the practical implications of the guidelines were demonstrated in an example of benzene and acute myeloid leukemia (AML). The focus was to identify studies that are informative for benzene-AML dose-response assessment. Only five out of 116 publications that were initially identified in the peer-reviewed literature were

found to be informative. Most studies that were not informative for dose-response assessment were so because exposure levels were not quantified. The guidelines enable a structured and transparent evaluation of occupational epidemiological studies for risk assessment.

Chapter 3 provides an example of how an increased transparency in the reporting of exposure assessment could facilitate the evaluation of the quality of occupational epidemiological studies. In this example, one specific aspect which significantly contributed to the quality of exposure assessment (the temporal coverage of the occupational history by exposure measurements) was assessed. A graphical tool that provides insight into potential differences in temporal coverage between occupational epidemiological studies was presented. Application of the tool to three nested case-control studies of occupational exposure to benzene and leukemia illustrated that there were considerable differences in the temporal coverage with exposure measurements between studies. It is suggested that the graphical tool be implemented in the reporting of occupational epidemiological studies to facilitate the evaluation of study quality for risk assessment.

Synthesis of the findings of occupational epidemiological studies is essential for risk assessment to generate summary risk estimates and to address potential heterogeneity in the study findings. In Chapter 4 an approach for meta-analysis that acknowledges differences in study quality is presented. The approach was demonstrated in an example of benzene and lymphoma subtypes. The quality of exposure assessment and the year-of-start of follow-up (a proxy for the quality of the outcome classification) were used to assess the quality of the studies included in the meta-analysis. Furthermore, arguing that any study that was not able to detect at least suggestive evidence of the established strong association between benzene and AML most likely had serious methodological limitations in one or more aspects of the study design, the strength of the reported AML association was used as proxy for the overall study quality as well. For several lymphoma subtypes meta-relative risks increased with increasing study quality, regardless of the strategy that was used to assess study quality. With the exception of a chance finding, these trends most likely reflect an underlying association of benzene with some of these lymphoma subtypes. As such this approach demonstrated how a transparent integration of the evaluation of study quality into evidence synthesis contributed to the identification of lymphoma subtypes that are potentially causally related to occupational benzene exposure.

The work in Chapter 5 demonstrates how methods for quality evaluation and evidence synthesis can be applied in exposure-response analysis. A flexible meta-regression of aggregated risk estimates from a set of occupational epidemiological studies offered insight into the functional relation between benzene exposure and leukemia. The meta-regression approach provided a comprehensive and quantitative assessment of the (shape of the)

exposure-response relation and sensitivity analyses offered formal insight into the uncertainty of the predicted exposure response curve (ERC). The predicted ERC can be applied directly in dose-response assessment and will provide a considerable improvement over the approaches that are currently used in benzene-leukemia dose-response assessment (such as using a single 'best' study).

Using biomarkers to advance the contribution of occupational epidemiological studies to risk assessment

Biomarkers can play an important role in advancing the contribution of occupational epidemiological studies to risk assessment. The recent developments in the field of molecular biology have provided epidemiologists with many tools to move beyond the classic study design of correlating external exposure levels to clinically manifested disease rates. In occupational epidemiological studies, new biomarkers might contribute to improved quality of exposure assessment, help to better understand the distribution of susceptibility to health outcomes in study populations, and might increase the ability to observe adverse health effects in the early stages of a disease. Insight into toxicological processes that underlie an exposure-response relation will also allow better use of information that is provided in animal bioassays and might facilitate formal incorporation of data from animal bioassays in human exposure-response analysis.

In Chapter 6 an overview of the potential application of OMICS technologies (biomarker discovery tools that can be applied to study large sets of biological molecules) in occupational (and environmental) epidemiological studies is presented. The application of OMICS technologies in occupational epidemiology has become feasible in recent years due to a spectacular increase in the sensitivity, resolution and throughput of OMICS-based assays. The five most developed OMICS technologies are genotyping, transcriptomics, epigenomics, proteomics and metabolomics. The opportunities of these new techniques for (occupational) epidemiology and the difficulties and limitations in the interpretation of the data generated by OMICS technologies are discussed. With the further development of validated technologies, appropriate study designs and sample handling, and advanced statistical methods for data interpretation, OMICS techniques will eventually contribute significantly to occupational epidemiology and will help the field progress towards an integrated view of the interaction between the environment and human health.

Chapter 7 is an example of how biomarker data can be formally integrated into occupational epidemiological studies. An approach was developed to translate airborne exposure estimates used in the exposure assessment of epidemiological analyses into relevant biomarkers of exposure with a set of empirical and physiologically-based predictive models. The approach was applied in an example of benzene and leukemia to assess the impact of the saturation of benzene's enzymatic pathways on the heterogeneity in the risk estimates

reported in two occupational epidemiological studies. Application of the approach in an example of benzene and leukemia suggested that the heterogeneity in the risk estimates was partly attributable to saturable metabolism. However, considerable heterogeneity remained between the studies after the application of the approach to translate airborne exposure estimates into relevant biomarkers of exposure. Differences in exposure misclassification, uncontrolled confounding factors, or other systematic differences between the studies are likely explanations for the remaining heterogeneity.

Discussion

Occupational epidemiological studies can be very informative for risk assessment. However, the evidence that occupational epidemiological studies can generate is currently not used to its full extent. The approaches presented in this thesis can be used as a framework to advance the use of occupational epidemiological studies in risk assessment. Some of the discussed approaches are ready to be applied in risk assessment (such as the evaluation guidelines, the graphical tool, and the methods for evidence synthesis); other approaches (such as incorporation of biomarkers in occupational epidemiological studies) need further development. Further progress in the use of occupational epidemiological studies in risk assessment should come from tailoring study designs to the needs of risk assessment. A strong focus on high quality quantitative exposure assessment in the design of new occupational epidemiological studies and transparency of the steps undertaken to develop quantitative exposure estimates would significantly contribute to an increased *weight of evidence* for risk assessment and would likely improve the overall quality of risk assessment for many exposures.

Samenvatting

Achtergrond

Het identificeren en kwantificeren van risicofactoren die worden gekarakteriseerd door lage blootstellingniveaus, gematigd verhoogde risico's en een aspecifieke relatie tussen blootstelling en ziekte, is een grote uitdaging voor de risicobeoordeling van chemische stoffen (hierna: risicobeoordeling). Arbeidsepidemiologische studies kunnen een bijdrage leveren aan deze uitdaging. Het voornaamste voordeel van arbeidsepidemiologische studies boven andere potentiële bronnen van informatie voor risicobeoordeling (voornamelijk dierexperimentele studies) is dat in deze studies mensen worden bestudeerd en dus extrapolatie van studieresultaten geobserveerd bij dieren naar de humane populatie niet nodig is. Dit voordeel brengt echter ook een nadeel met zich mee omdat arbeidsepidemiologische studies van nature voornamelijk observationeel zijn, wat de methodologische kwaliteit van een epidemiologisch onderzoeksontwerp beïnvloedt.
Sommige beperkingen van het gebruik van arbeidsepidemiologische studies in risicobeoordeling zijn inherent aan de onderzoeksdiscipline. Echter, verbeteringen in het ontwerp, de uitvoering en de interpretatie zal het gebruik van dit type studies in risicobeoordelingen bevorderen. Bovendien kunnen de recente ontwikkelingen in de moleculaire biologie, en het daaraan gerelateerde verbeterde inzicht in het ontstaan van kanker en andere aandoeningen, de bijdrage van arbeidsepidemiologische studies aan risicobeoordeling verder bevorderen.
In dit proefschrift wordt een aantal strategieën beschreven die het gebruik van arbeidsepidemiologische studies in risicobeoordeling kunnen bevorderen. Sommige van de strategieën kunnen direct worden toegepast. Andere strategieën moeten nog verder worden ontwikkeld voordat hun bijdrage aan risicobeoordeling kan worden beoordeeld.

Gebruik van weight of evidence in de risicobeoordeling van chemische stoffen op basis van epidemiologische studies
In hoofdstuk 2 wordt een aantal richtlijnen voor de evaluatie van de kwaliteit van arbeidsepidemiologische studies voor (kwantitatieve) risicobeoordeling gepresenteerd. De kwaliteit van arbeidsepidemiologische studies voor risicobeoordeling wordt bepaald door de kwaliteit van het ontwerp, de uitvoering en de rapportage van een studie, en door de relevantie van de studiehypothese voor de risicobeoordeling in kwestie. Verschillen tussen studies in de kwaliteit van het ontwerp en de uitvoering worden vaak niet gereflecteerd in parameters van statistische onzekerheid (bijvoorbeeld betrouwbaarheidsintervallen). Het is daarom belangrijk dat arbeidsepidemiologische studies een gedetailleerde evaluatie krijgen voordat ze worden gebruikt voor risicobeoordeling. De richtlijnen in hoofdstuk 2 bestaan uit 20 evaluatiecriteria. De kwaliteit van (kwantitatieve) blootstellingskarakterisering krijgt extra aandacht in de evaluatierichtlijnen omdat deze, hoewel vaak onderschat, een grote impact kan hebben op de kwaliteit van een arbeidsepidemiologische studie. De bruikbaarheid en de praktische implicatie van de toepassing van de richtlijnen zijn gedemonstreerd aan de hand

van een voorbeeld van benzeen en acute myeloïde leukemie (AML). In dit voorbeeld lag de focus op de identificatie van studies die informatief zijn voor de blootstelling-respons relatie voor benzeen en AML. Slechts vijf van de 116 publicaties die oorspronkelijk waren geïdentificeerd in de wetenschappelijke literatuur werden als voldoende informatief beoordeeld. De meeste studies waren niet informatief voor blootstelling-respons karakterisering omdat blootstellingniveaus niet waren gekwantificeerd. De evaluatierichtlijnen bevorderen de gestructureerde en transparante evaluatie van arbeidsepidemiologische studies in risicobeoordeling.

In hoofdstuk 3 wordt gedemonstreerd hoe meer transparantie in de rapportage van de blootstellingskarakterisering van arbeidsepidemiologische studies de evaluatie van studies kan faciliteren. In dit voorbeeld wordt een specifiek aspect besproken dat significant bijdraagt aan de kwaliteit van de blootstellingskarakterisering. Met een figuur wordt inzicht verschaft in potentiële verschillen tussen studies wat betreft de dekking van de beroepshistorie met blootstellingsmetingen over de tijd. Toepassing van de figuur op drie patiëntcontrole-onderzoeken naar de relatie tussen benzeenblootstelling en leukemie (uitgevoerd in cohorten van beroepsmatig blootgestelde personen) illustreerde dat de dekking van de beroepshistorie door blootstellingsmetingen over de tijd behoorlijk verschilde tussen deze studies. Gebruik van de gepresenteerde figuur in de rapportage van arbeidsepidemiologische studies zal de evaluatie van de kwaliteit van deze studies voor risicobeoordeling bevorderen.

Synthese van de bevindingen van arbeidsepidemiologische studies is essentieel voor risicobeoordeling om studiespecifieke risicoschattingen te combineren en om eventuele heterogeniteit tussen studies te identificeren. In hoofdstuk 4 wordt een strategie voor meta-analyse gepresenteerd waarin verschillen in kwaliteit tussen studies kunnen worden erkend. De strategie wordt gedemonstreerd aan de hand van een voorbeeld van benzeen en maligne lymfoom subtypen. De kwaliteit van blootstellingskarakterisering en het jaar waarin gestart werd met de follow-up (een maat voor de kwaliteit van de classificatie van de gezondheidsuitkomst) worden gebuikt om de kwaliteit van de studies in de meta-analyse te beoordelen. Verder wordt de sterkte van de associatie tussen beroepsmatige benzeenblootstelling en AML ook gebruikt als maat voor studiekwaliteit. Deze proxy is gebaseerd op de aanname dat wanneer een studie niet tenminste suggestief bewijs kan vinden voor de bewezen associatie tussen benzeenblootstelling en AML, deze studie waarschijnlijk serieuze beperkingen heeft in één of meerdere aspecten van het studieontwerp. Voor een aantal maligne lymfoom subtypen stegen de meta-relatieve risico's met een stijging in studiekwaliteit, ongeacht de strategie die werd gehanteerd om studiekwaliteit te beoordelen. Hoewel het mogelijk kan zijn dat er sprake is van een toevalsvinding, lijken de geobserveerde trends een onderliggende associatie tussen benzeenblootstelling en sommige maligne lymfoomtypen te reflecteren. Deze strategie laat zien hoe een transparante integratie van de beoordeling van studiekwaliteit in de synthese

van bewijs uit arbeidsepidemiologische studies bijdraagt aan de identificatie van een mogelijk causaal verband tussen beroepsmatige benzeenblootstelling en verscheidene maligne lymfoom subtypen.

Hoofdstuk 5 laat zien hoe methoden voor kwaliteitsevaluatie en de synthese van bewijs kunnen worden gebruikt in blootstelling-respons karakterisering. Een flexibele metaregressie is toegepast op de geaggregeerde risicoschattingen uit een set arbeidsepidemiologische studies. Dit leverde inzicht in de functionele relatie tussen benzeenblootstelling en leukemie. De meta-regressie gaf een kwantitatief inzicht in de blootstelling-respons relatie en sensitiviteitsanalyses gaven inzicht in de onzekerheid van de voorspelde relatie. De voorspelde blootstelling-respons relatie kan direct worden toegepast in dosis-respons karakterisering. Dit is een verbetering van de huidige situatie in de dosis-respons karakterisering voor benzeen en leukemie die vaak gebaseerd is op de selectie van één 'beste' studie.

Het gebruik van biomarkers om de bijdrage van de arbeidsepidemiologie aan de risicobeoordeling van chemische stoffen te bevorderen
Biomarkers kunnen een belangrijke rol spelen in de bevordering van de bijdrage van arbeidsepidemiologische studies aan risicobeoordeling. Recente ontwikkelingen in de moleculaire biologie hebben epidemiologen de mogelijkheid gegeven verder te gaan dan het klassieke studieontwerp waarin blootstellingen in de omgevingslucht worden gecorreleerd aan klinisch manifeste aandoeningen. Nieuwe biomarkers kunnen in arbeidsepidemiologische studies bijdragen aan een verbeterde kwaliteit van de blootstellingskarakterisering, bijdragen aan een beter begrip van de distributie van de ontvankelijkheid voor gezondheidsuitkomsten in studiepopulaties en bevorderen mogelijk het vermogen om nadelige gezondheidsuitkomsten te observeren in een vroeg stadium van een ziekte. Inzicht in de toxicologische processen die bijdragen aan een blootstelling-respons relatie zal ook het gebruik van de informatie die beschikbaar is uit dierexperimentele studies verbeteren en kan mogelijk de formele integratie van data uit dierexperimentele studies in humane blootstelling-respons analyse bevorderen.

In hoofdstuk 6 wordt een overzicht gegeven van de mogelijke toepassing van OMICS technologieën (biomarker technieken die kunnen worden toegepast op grote groepen van biologische moleculen) in arbeidsepidemiologische studies. De toepassing van OMICS technologieën in arbeidsepidemiologische studies is mogelijk geworden door de spectaculaire verbetering in de sensitiviteit, resolutie en verwerkingscapaciteit van op OMICS gebaseerde analysemethoden in de afgelopen jaren. De vijf meest ontwikkelde OMICS technologieën zijn genotypering, transcriptomics, epigenomics, proteomics en metabolomics. De mogelijkheden van deze nieuwe technieken voor de (arbeids-) epidemiologie en de moeilijkheden en beperkingen voor de interpretatie van de door OMICS technologieën gegenereerde data

worden besproken. Met een verdere ontwikkeling van gevalideerde technologieën, geschikte studieontwerpen, behandeling van biologische monsters en geavanceerde statistische methoden kunnen OMICS technologieën uiteindelijk een significante bijdrage leveren aan de arbeidsepidemiologie en zullen een bijdrage leveren aan een completer beeld van de interactie tussen omgevingsfactoren en de humane gezondheid.

In hoofdstuk 7 wordt gedemonstreerd hoe biomarker data formeel in arbeidsepidemiologische studies kan worden geïntegreerd. Een strategie is ontwikkeld om met behulp van empirische en op fysiologische data gebaseerde modellen blootstellingschattingen in de buitenlucht, gebruikt in de blootstellingskarakterisering van epidemiologische studies, om te zetten in relevante biomarkers van blootstelling. Deze strategie is toegepast in een voorbeeld van benzeen en leukemie. In dit voorbeeld is de impact van de verzadiging van de enzymatische paden van benzeen op de heterogeniteit in de risicoschattingen uit twee arbeidsepidemiologische studies bekeken. Toepassing van deze strategie suggereerde dat de heterogeniteit in de risicoschattingen gedeeltelijk was te wijten aan de verzadiging van het metabolisme van benzeen. Echter, significante heterogeniteit bleef bestaan tussen de twee studies, ook na toepassing van de hierboven beschreven strategie. Verschillen in misclassificatie van de blootstelling, ongecontroleerde verstorende factoren, of andere systematische verschillen tussen de studies zijn aannemelijke verklaringen voor de resterende heterogeniteit.

Discussie

Arbeidsepidemiologische studies kunnen een grote bijdrage leveren aan risicobeoordeling. De informatie die potentieel beschikbaar is uit arbeidsepidemiologische studies wordt op dit moment echter nog niet volledig benut voor risicobeoordeling. De strategieën die worden gepresenteerd in dit proefschrift kunnen functioneren als een structuur om het gebruik van arbeidsepidemiologische studies in risicobeoordeling te bevorderen. Sommige strategieën zijn klaar om te worden toegepast in risicobeoordeling (de evaluatierichtlijnen, de figuur dat de dekking van de beroepshistorie met blootstellingmetingen over tijd illustreert en de methoden voor de synthese van bewijs). Andere strategieën behoeven nog verdere ontwikkeling (de integratie van biomarkers in arbeidsepidemiologische studies). Een verdere vooruitgang wat betreft het gebruik van arbeidsepidemiologische studies in risicobeoordeling kan komen uit het beter laten aansluiten van studies op de benodigde informatie voor risicobeoordeling. Een sterke focus op de kwaliteit van de blootstellingskarakterisering tijdens het ontwerpen van nieuwe studies en transparantie in de stappen die zijn genomen om tot een kwantitatieve blootstellingskarakterisering te komen, zullen significant bijdragen aan meer *weight of evidence* voor risicobeoordeling. Waarschijnlijk zal dit bijdragen aan de verbetering van de risicobeoordeling van veel chemische stoffen.

Scientific publications

Vlaanderen J, Portengen L, Rappaport SRM, Glass DC, Kromhout H, Vermeulen R. The Impact of Saturable Metabolism on Exposure-Response Relations in Two Studies of Benzene Induced Leukemia. Am J. Epidemiol. In press.

Vlaanderen J, Lan Q, Kromhout H, Rothman, N, Vermeulen R. Occupational Benzene Exposure and the Risk of Lymphoma Subtypes: A Meta-analysis of Cohort Studies Incorporating Three Study Quality Dimensions. Environ Health Perspect. 2011;119(2):159-67.

Vlaanderen J, Fransman W, Miller B, Bursty I, Heederik D, Hurley F, Vermeulen R, Kromhout H. A graphical tool to evaluate temporal coverage of occupational history by exposure measurements. Occup Environ Med. 2010;67(9):636-8.

Vlaanderen J, Moore LE, Smith MT, Lan Q, Zhang L, Skibola CF, Rothman N, Vermeulen R. Application of OMICS technologies in occupational and environmental health research; current status and projections. Occup Environ Med. 2010;67(2):136-43.

Vlaanderen J, Portengen L, Rothman N, Lan Q, Kromhout H, Vermeulen R. Flexible meta-regression to assess the shape of the benzene-leukemia exposure-response curve. Environ Health Perspect. 2010;118(4):526-32.

Davies HW, Vlaanderen J, Henderson SB, Brauer M. Correlation between co-exposures to noise and air pollution from traffic sources. Occup Environ Med. 2009;66(5):347-50.

Vineis P, Khan AE, Vlaanderen J, Vermeulen R. The impact of new research technologies on our understanding of environmental causes of disease: the concept of clinical vulnerability. Environ Health. 2009;8:54.

Vlaanderen J, Vermeulen R, Heederik D, Kromhout H, ECNIS Integrated Risk Assessment Group. Guidelines to evaluate human observational studies for quantitative risk assessment. Environ Health Perspect. 2008;116(12):1700-5.

Pronk A, Yu F, Vlaanderen J, Tielemans E, Preller L, Bobeldijk I, Deddens JA, Latza U, Baur X, Heederik D. Dermal, inhalation, and internal exposure to 1,6-HDI and its oligomers in car body repair shop workers and industrial spray painters. Occup Environ Med. 2006;63(9):624-31.

Dankwoord

Hoewel ik mezelf beschouw als een man van weinig woorden heb ik het toch voor elkaar gekregen een proefschrift van 224 pagina's te produceren. Dus ga ik alles op alles zetten om in ieder geval deze pagina zo kort mogelijk te houden. Ik ben erg blij met dit proefschrift en ben dankbaar voor de kansen die ik bij het IRAS heb gekregen. Ik verlaat het IRAS met een breed beeld van wat epidemiologisch onderzoek inhoudt en denk dat ik daarmee een prima basis heb gekregen voor een succesvolle carrière in dit vakgebied.

Hans en Roel hartelijk bedankt voor de begeleiding de afgelopen jaren. Ik heb veel van jullie geleerd en denk dat ik nog jaren op vele plekken welkom ben door alleen maar jullie naam te laten vallen! Gelukkig zullen we in de nabije toekomst nog regelmatig samenwerken. Hans ik weet dat je vaak heb moeten lijden om onze Poolse vrienden tevreden te houden, maar wat mij betreft is dat het zeker waard geweest. Lützen bedankt voor je advies en statistische hulp de afgelopen jaren. Zonder jouw bijdrage was het aantal artikelen in dit proefschrift een stuk lager geweest. Alle andere collega's bij het IRAS bedankt voor de samenwerking en gezelligheid de afgelopen jaren. De lijst met AIO's waarmee ik de afgelopen jaren heb gewerkt, gelunchwandeld en geborreld is te lang om jullie allemaal bij naam te noemen, maar jullie hebben het IRAS tot een inspirerende en gezellige werkplek gemaakt. Susan, leuk dat je mijn paranimf wil zijn. We moeten toch een beetje voor elkaar zorgen tussen de EMF'ers, niet?

Gelukkig had ik de afgelopen jaren ook nog een leven naast het IRAS. Arjen, Bas, Joost, Lucas, en Mirko: er is niets beter om even los te komen van je proefschrift, dan een middag Ajax schreeuwen. Analyseren of die ruit nou naar voren of naar achteren moet, schelden op trainers en overbetaalde voetballers die het niet begrijpen en soms (heel soms) juichen om een wereldgoal! Bas, laten we vooral een keer naar Engeland zeilen ergens in de komende tien jaar. Joost, te gek dat je de voorkant van dit proefschrift wilde maken. Hester en Enneke, ik ben heel blij dat we het zo goed met elkaar kunnen vinden en met de interesse die jullie altijd hebben getoond. En, leuk dat je mijn paranimf wil zijn! Virissa, gelukkig hebben we elkaar op de valreep beter leren kennen. Dit had ik zeker niet willen missen! Ik zie erg uit naar onze vakantie.

Ten slotte mijn ouders. Jullie zijn er altijd voor me geweest. Of het nou hulp met huiswerk op school, sjouwen met een drumstel of (wekenlang) klussen aan een oud huis was, jullie stonden altijd voor me klaar. Zonder jullie was dit proefschrift er niet geweest (en ik waarschijnlijk dakloos). Ik hoop dat ik er ooit kan zijn voor jullie zoals jullie er nu voor mij zijn.

Curriculum Vitae

Jelle Vlaanderen werd geboren op 11 januari 1981 in Soest. Na in 1999 zijn VWO diploma behaald te hebben aan het Griftland College te Soest startte hij met de opleiding Farmacie aan de Universiteit Utrecht. In 2003 startte hij met het masterprogramma Toxicology and Environmental Health aan het Institute for Risk Assessment Sciences (IRAS), onderdeel van de Universiteit Utrecht. Tijdens dit programma liep hij stage bij een onderzoek naar de blootstelling aan isocyanaten in autoschadeherstelbedrijven, dat werd uitgevoerd door IRAS en TNO. Ook deed hij 6 maanden onderzoek naar de correlatie tussen verkeersgeluid en luchtverontreiniging aan de Universiteit van British Columbia in Vancouver, Canada. In 2005 behaalde hij zijn doctoraaldiploma Farmacie en zijn masterdiploma Toxicology and Environmental Health. Na zijn afstuderen kwam hij in dienst bij het IRAS waar hij meewerkte aan onderzoeken naar de blootstelling aan silica in de Europese mineraalindustrie en de blootstelling aan dieselmotoremissies in de op- en overslagsector in Nederland. In 2006 begon hij aan het promotieonderzoek dat is beschreven in dit proefschrift. Dit onderzoek werd uitgevoerd als onderdeel van ECNIS (Environmental Cancer Risk, Nutrition and Individual Susceptibility), een netwerk dat opereerde in het zesde kaderprogramma voor onderzoek en technologische ontwikkeling van de Europese Unie. In 2009 bracht hij 3 maanden door aan de Monash Universiteit in Melbourne, Australië dankzij een ECNIS uitwisselingsbeurs.

Jelle Vlaanderen was born January 11, 1981, in Soest, the Netherlands. After completing his secondary school at the Griftland College in Soest in 1999, he started studying Pharmaceutical Sciences at Utrecht University. In 2003 he started the Master of Science program in Toxicology and Environmental Health at the Institute for Risk Assessment Sciences (IRAS), part of Utrecht University. During this Master's program he participated in a research project on exposure to isocyanates in car body repair shops which was conducted by IRAS and TNO. He also conducted 6 months of research at the University of British Columbia in Vancouver, Canada to assess the correlation between traffic noise and air pollution. In 2005 he received MSc degrees in Pharmaceutical sciences and in Toxicology and Environmental Health. After graduating he started working as a junior researcher at IRAS where he participated in projects on silica exposure in the European mineral industry, and exposure to diesel motor emissions in storage and transfer companies in the Netherlands. In 2006 he began the research project presented in this thesis. This project was conducted as part of ECNIS (Environmental Cancer Risk, Nutrition and Individual Susceptibility); a Network of Excellence that operated in the European Union 6th Framework Programme for Research and Development. In 2009 he spent 3 months at Monash University in Melbourne, Australia on an ECNIS exchange fellowship.