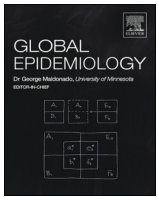
Global Epidemiology 7 (2024) 100143

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology



Triangulation of epidemiological evidence and risk of bias evaluation: A proposed framework and applied example using formaldehyde exposure and risk of myeloid leukemias

Daniel J. Lauer [a,*], Anthony J. Russell [b], Heather N. Lynch [c], William J. Thompson [d], Kenneth A. Mundt [e], Harvey Checkoway [f]

[a] Stantec ChemRisk, Denver, CO, United States of America
[b] Stantec ChemRisk, Cincinnati, OH, United States of America
[c] Stantec ChemRisk, Boston, MA, United States of America
[d] Independent Consultant in Epidemiology, Hume, MO, United States of America
[e] Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, United States of America
[f] Herbert Wertheim School of Public Health, University of California, San Diego, United States of America

ARTICLE INFO

Keywords:
Triangulation
Risk of bias
Evidence synthesis
Formaldehyde
Myeloid leukemia

ABSTRACT

Evidence triangulation may help identify the impact of study design elements on study findings and to tease out biased results when evaluating potential causal relationships; however, methods for triangulating epidemiologic evidence are evolving and have not been standardized. Building upon key principles of epidemiologic evidence triangulation and risk of bias assessment, and responding to the National Academies of Sciences, Engineering, and Medicine (NASEM) call for applied triangulation examples, the objective of this manuscript is to propose a triangulation framework and to apply it as an illustrative example to epidemiologic studies examining the possible relationship between occupational formaldehyde exposure and risk of myeloid leukemias (ML) including acute (AML) and chronic (CML) types.

A nine-component triangulation framework for epidemiological evidence was developed incorporating study quality and ROB guidance from various federal health agencies (i.e., US EPA TSCA and NTP OHAT). Several components of the triangulation framework also drew from widely used epidemiological analytic tools such as stratified meta-analysis and sensitivity analysis. Regarding the applied example, fourteen studies were identified and assessed using the following primary study quality domains to explore potential key sources of bias: 1) study design and analysis; 2) study participation; 3) exposure assessment; 4) outcome assessment; and 5) potential confounding. Across studies, methodological limitations possibly contributing to biased results were observed within most domains. Interestingly, results from one study – often providing the largest and least-precise relative risk estimates, likely reflecting study biases, deviated from most primary study findings indicating no such associations. Triangulation of epidemiological evidence appears to be helpful in exploring inconsistent results for the identification of study results possibly reflecting various biases. Nonetheless, triangulation methodologies require additional development and application to real-world examples to enhance objectivity and reproducibility.

Introduction

Triangulation in epidemiology has been defined as "the practice of strengthening causal inferences by integrating results from several different approaches, where each approach has different (and assumed to be largely unrelated) key sources of potential bias" [1]. The main motivations for advancing triangulation methods are to identify and ultimately minimize risk of bias (ROB) and to enhance study quality assessments in which individual studies may be excluded during evidence synthesis [2]. Triangulation principles maintain that all studies have inherent strengths and weaknesses and, if applied rigorously, triangulation methods may be useful for identifying consistency both

* Corresponding author.
E-mail address: daniel.lauer@stantec.com (D.J. Lauer).

https://doi.org/10.1016/j.gloepi.2024.100143
Received 18 January 2024; Received in revised form 1 April 2024; Accepted 5 April 2024
Available online 9 April 2024

within and across epidemiologic studies [2]. For example, occupational cohort studies often evaluate the association between multiple exposure metrics (i.e. cumulative, duration, peak, average) and disease risk and find inconsistent results across these metrics. Similar interpretational challenges may arise when discordant results for a single exposure variable in relation to disease risk are observed across the relevant body of literature.

Risk of bias (ROB) analyses and broader study quality evaluation tools have been applied in systematic reviews to evaluate a body of literature based on key design and analytical features that may affect the ability to validly identify and quantify possible causal effects. US governmental groups including the National Toxicology Program (NTP) Office of Health Assessment and Translation (OHAT) and the Environmental Protection Agency (US EPA) Integrated Risk Information System (IRIS) and Toxic Substances Control Act (TSCA) programs all routinely employ quality evaluation processes (including, in some cases ROB tools) within chemical hazard and risk assessments [3]. Many of the ROB tools are comprehensive in assessing major study design areas, or domains, that should be considered in evaluating bodies of epidemiologic evidence to elucidate possible causal relations.

Nevertheless, comprehensive criteria and associated guidance for ROB assessments vary, contributing to sometimes contradictory conclusions depending on the specific tool used [4,5]. Current ROB frameworks may be deficient in two key areas: 1) weighing the relative importance of critical quality domains for specific hypotheses and sub-hypotheses; and 2) connecting ROB evaluation findings to the synthesis of the scientific evidence used to reach causal conclusions. Regarding the first limitation, ROB ratings or scores assigned to each domain typically are summed or averaged to derive an overall ROB or quality rating [6]. However, this can be susceptible to overlooking key study-specific details that inform hypotheses relevant to specific biases and obscured when combined. At the extreme, studies may be downgraded or even excluded from systematic reviews because of low quality scores in domains unrelated to the key hypotheses being evaluated.

Like ROB and study quality assessments, meta-analysis is a tool often used in systematic reviews in which precision-weighted averages of reported relative risk estimates can be stratified by study quality features (i.e., study design or exposure assessment approach) to explore heterogeneity across study results. While some of these methods resemble and overlap with triangulation frameworks, key differences emerge upon closer examination of the overriding goals and specific methodologies utilized. For example, meta-analyses often incorporate relatively insensitive study quality tools such as the Newcastle-Ottawa Scale (NOS) that average or aggregate evidence to arrive at semi-quantitative ratings of bias [7]. Moreover, stratification of meta-analytic results may help to identify heterogeneity in study findings without closer consideration of the potential direction, magnitude and relative importance of sources of bias. In contrast, triangulation seeks to isolate and critically evaluate the sources of heterogenous study results through examination of the impact of specific hypothesized biases affecting relative risk estimates across different study designs and populations [8]. Thus, when evaluating inconsistent results across epidemiological studies, incorporating triangulation approaches in tandem with more conventional critical review methods may help overcome limitations associated with solely relying only on meta-analytic (i.e., estimation of an overall or strata-specific risk estimate) or formulaic ROB or quality evaluation processes (i.e., lower quality rating) approaches.

Although the principles and aims of triangulation are intuitive and consistent with the goal of validly interpreting a body of epidemiological evidence, methodologies for applying triangulation in chemical risk assessment lack guidance and standardization. This paper expands on key principles of epidemiological evidence triangulation and proposes a basic framework that builds on recommendations arising from a 2022 workshop entitled Workshops to Support EPA's Development of Human Health Assessments: Triangulation of Evidence in Environmental Epidemiology, hosted by the US National Academies of Science, Engineering, and Medicine (NASEM). Additionally, the proposed framework is applied to an example evaluating occupational formaldehyde exposure and risk of myeloid leukemias (ML) (specifically the etiologically distinct subtypes of acute myeloid leukemia (AML) and chronic myeloid leukemia (CML)).

Methods

A framework for triangulating epidemiological evidence was developed based on guidance from ROB and study quality evaluation methods including those developed by US EPA TSCA [9], US EPA IRIS [10] and NTP OHAT [11]. The proposed framework also incorporates triangulation elements described by Savitz et al. [8] and Lawlor et al. [1] as well as standard meta-analytic techniques for investigating study heterogeneity. Central to the proposed framework were the five key study quality domains (see second column in Table 1) as described in TSCA guidance for evaluating epidemiological studies. Detailed descriptions of the study quality domains and criteria evaluated are presented in the EPA TSCA guidance [12] and in previously published systematic reviews (e.g., [13,14]).

A nine-component triangulation framework was developed in an iterative fashion (see Fig. 1). The first four steps are largely consistent with systematic review principles and include development of a central research question, identification of the relevant body of literature, definition of key study quality domains, and critical study quality review for each study meeting pre-defined inclusion and exclusion criteria. Whereas the first four steps of the proposed framework clearly align with pre-existing systematic review processes, the subsequent five steps reflect more novel triangulation approaches. For example, the fifth component involves a triangulation feasibility assessment, which requires close consideration of the degree of methodological variability for each of the five study quality domains. Assuming a sufficient level of methodological variability for each domain, the next step involves enumeration of specific hypothesized sources of bias relevant to each domain. Upon listing at least two to three testable hypotheses for each domain regarding potential sources of bias, risk visualization can be achieved by creating forest plots for each domain and stratified by the hypothesized sources of bias. Assuming the relevant body of literature is suitable for conducting a sensitivity analysis, step eight seeks to quantify the magnitude and direction of bias by the comparison of estimated summary statistics (in this example pooled meta-risks). Finally, the findings and uncertainties from the triangulation exercise should be integrated and synthesized. To illustrate the proposed framework, the body of epidemiological studies addressing occupational formaldehyde exposure and risk of myeloid leukemias addressed in the EPA IRIS Draft Toxicological Evaluation of Formaldehyde [10] and confirmed by updated searches was assessed.

Applied Example – Inhalation formaldehyde exposure and risk of myeloid leukemia (including AML and CML)

Below, each step of the proposed triangulation framework is described conceptually and applied to the epidemiological literature regarding formaldehyde exposure and risk of myeloid leukemia.

Research question generation

IARC [15] and the NRC Report on Carcinogens [16] concluded that occupational formaldehyde exposure causes leukemia and specifically myeloid leukemia. Furthermore, the EPA IRIS Draft Toxicological Evaluation of Formaldehyde [10] states that "the [epidemiological] evidence demonstrates that formaldehyde inhalation causes myeloid leukemia in humans given appropriate exposure circumstances". Most recently, the French Agency for Food, Environmental and Occupational Health & Safety (ANSES) published an Opinion stating that, based on their scientific assessment of the epidemiological literature on

Table 1
Defined Key Domains based on TSCA criteria and Domain specific hypotheses.

TSCA Domain	Quality Metrics Evaluated	Domain Specific Hypotheses
Study Design/ Analysis	Strength of study design and associated statistical assessment including modeling approaches	1. PMR and case-control studies demonstrating increased risks of ML or ML subtypes may be more susceptible to various biases due to various study design elements including deficiencies in exposure assessment, recruitment strategies, etc. 2. Studies with fewer than five cases or deaths from ML may provide unstable and therefore unreliable RR estimates.
Study Participation	Participant selection methods and completeness; approaches for addressing differences in baseline characteristics of groups	1. Individual studies with poorly defined source populations or high rates of refusal to participate may be more susceptible to selection bias. 2. Studies with significant attrition or missing exposure or outcome data may be more susceptible to selection bias. 3. Case-control and PMR studies based on convenience samples of cases or deaths may be more susceptible to selection bias associated with recruitment strategies, especially when differences in the distribution of baseline characteristics across comparison groups are observed.
Exposure Characterization	Validity of exposure surrogates or measurements; exposure levels and temporality	1. Studies that estimate exposure at the ecological level are likely more susceptible to exposure misclassification than studies with individual exposure measurements or estimates. 2. Among studies incorporating multiple exposure metrics, exposure variables based on modeled or other non-empirical data sources may be more susceptible to misclassification of exposure than those based on personal or environmental testing data.
Outcome Assessment	Reliability of data sources and methods for ascertaining specific outcome (e.g., pathological reports vs. death certificates)	1. Given the differences in survival rates of ML and subtypes, studies solely relying on mortality estimates may generate biased risk estimates relative to studies analyzing disease incidence or mortality by ML subtype. 2. If associations are observed between formaldehyde exposure and ML overall or CML specifically, and not AML, then bias may be present because of the lack of biological plausibility given the potential association between other chemical exposures (i.e. high levels of benzene) and AML. 3. Given significant improvements in ML diagnosis and classification over the last 20 years, is there evidence that study time period (e.g., decade of diagnosis) could contribute to outcome misclassification?
Potential Confounding / Variability Control	Confounder characterization and adjustment; measurement of potential co-exposures	1. Studies that do not adjust or account for concomitant exposures to leukemogenic agents or confounders (i.e., smoking and AML) may be more susceptible to bias away from the null. 2. Studies that do not adequately adjust for non-modifiable confounding factors may be more susceptible to bias in either direction. 3. Is the variability in methods across this body of literature sufficient to identify whether uncontrolled confounding could influence results?

occupational exposure to formaldehyde, formaldehyde causes myeloid leukemias ([17]). While these evaluations largely acknowledged the lack of an animal model for leukemogenesis and the lack of mode of action (MOA) evidence, their conclusions were based on the body of epidemiological studies that largely fails to demonstrate strong or consistent associations [18,19,20]. On the other hand, the conclusion that formaldehyde unlikely causes leukemias has been advanced by European governmental agencies (e.g., [21,22]) and published critical reviews (e.g., [23,24]). Thus, the following research question was identified as a potential case study to illustrate the proposed triangulation framework: Can methods for triangulating the epidemiological evidence help elucidate specific sources of bias within and across the epidemiological literature, which may provide important contextualization for the debate over whether inhaled formaldehyde causes human myeloid leukemias.

Study identification

In total, 18 publications that reported results for exposure to formaldehyde and risk of ML, AML, or CML were included in the triangulation exercise. Of these 18 studies, four studies investigated the mortality experience of garment industry workers [25,26] and NCI cohort members [27,28]; however, results for these study populations were updated in more recent analyses [29,23,30]. This resulted in 14 publications included in the final analysis. Supplemental Table 1 provides select characteristics of these fourteen publications, five among funeral industry workers and anatomists [31,32,33,34,35]; six among industrial workers exposed to formaldehyde [29,23,36,30,37,38], and three

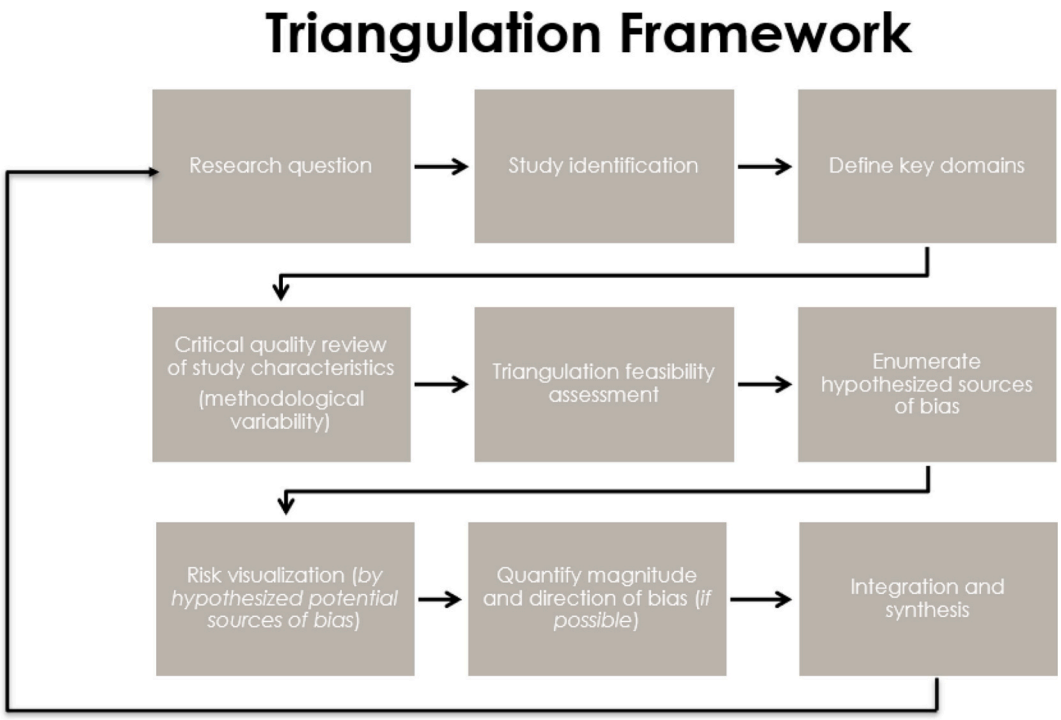


Fig. 1. A proposed Triangulation Framework incorporating elements from ROB guidance, US EPA TSCA guidance, and [8].

population-based studies [39,40,41].

Defining key domain evaluation criteria, critical quality review and methodological variability

Because the TSCA epidemiological study evaluation criteria are non-specific, they were refined to address methodological aspects unique to the research question of whether occupational formaldehyde exposure increases the risk of or possibly causes ML combined, AML and CML. Each eligible study was evaluated according to the modified TSCA quality criteria as described in Table 1 below. The study quality assessment was conducted in an iterative fashion to generate a preliminary list of hypothesized critical potential biases that guided more in-depth investigation of the direction and magnitude of possible biases by domain. Central to the generation of hypothesized sources of bias was characterizing the methodological variability within and between studies for determining the impact of particular study characteristics in relation to risk of bias.

Triangulation feasibility assessment

After the definition and evaluation of the key domains, the critical quality review and determination of methodological variability, a qualitative assessment was conducted to determine whether sufficient methodological variability was present within at least one domain to compare and contrast the reported risk estimates across the studies.

Hypothesized Sources of Bias

Following the critical review of study quality domains, a primary list of hypothesized biases was generated based on initial quality ratings and synthesis of the key strengths and limitations of the analyzed body of literature. Further, attempts were made to identify whether specific biases primarily driven by a single domain (i.e., exposure assessment) could be disentangled from more general study features (i.e., case-control vs. cohort). The 14 publications were evaluated to determine whether there was sufficient variability in methods across the body of literature to reasonably determine the impact of potential biases, and whether individual sources of bias (i.e., by domain) could be disentangled from other sources biases.

Across the study/design analysis domain, study design approaches included proportionate mortality studies, prospective and retrospective cohorts, and case-control studies. While each design is susceptible to distinct sources of potential bias, it was hypothesized that this domain may serve as a proxy for other more informative domains where bias is more easily identified (such as outcome and exposure assessment). For instance, we detected minimal variation in reproducibility of analyses (i.e., sufficient description of methodology to conceptually understand and reproduce analysis) and statistical modeling across studies.

Within the study participation domain, occupational epidemiological cohort studies and cancer-registry-based case-control studies with fully enumerated cohorts and adequately long follow-up time were generally considered to be of higher quality. These studies followed well-defined cohorts which minimized selection bias and attrition (due to censoring of those lost-to-follow-up) when compared with the general population-based case-control or proportionate mortality studies, which were generally considered to be of lower quality. Overall, the study participation domain was less informative, and therefore less capable to identify possible biases in the results of this literature; however, this domain reinforced that the PMR studies and the related case-control study [31] are more susceptible to selection biases due to the lack of a well-defined source cohort and incomplete study population enumeration (see supplemental fig. 1.).

For the potential/confounding domain, given the relatively few established environmental risk factors for myeloid leukemias, and the lack of information on potentially relevant co-exposures (as indicated by historical industrial hygiene measurements) and especially including tobacco smoking histories, this domain proved to be the least informative and therefore least explored, despite the potential importance of smoking as a relatively strong risk factor for AML.

As depicted in Table 1, three aspects of outcome assessment were considered for evaluating potential biases: incidence vs. mortality, outcome specificity (all ML vs AML and CML), and potential for reporting bias. Among cancers with relatively high survival rates, e.g.,

chronic leukemias, disease incidence measures are more likely to provide more complete ascertainment of cases relative to mortality-based estimates. As such, studies that reported incident AML and CML (or other ML subtypes) were considered a priori to be more informative. Additionally, while death certificates for leukemias have been reported generally to agree with tumor registry data, information regarding leukemia subtypes may be missing and contribute to diminished outcome specificity and outcome misclassification [42,43]. Therefore, studies that relied on identifying ML as the cause of death reported on death certificates – especially if these diagnoses were rendered decades before diagnostic and classification improvements occurred – were hypothesized to be less accurate than those relying on registered incident ML cases, determination of which typically is based on pathological confirmation. Studies that did not report findings specifically for ML subtype were considered to be less informative within the outcome assessment domain. Further, studies that reported findings for ML subtype with high precision (indicating less variability within study measures) were considered more informative.

Most studies reported diagnoses of myeloid leukemia based on information recorded on death certificates or from tumor registries using pathological confirmation. Overall, the outcome classifications reported were based on the ICD code designation in place at the time of death or time of diagnosis and often confirmed by expert (i.e., nosologist) review. However, most studies did not present results by time period or even decade of diagnosis.

Significant variability was observed across studies employing different formaldehyde exposure characterization approaches, each with specific limitations and potential sources of bias, as none directly measured exposure at the individual level for all or even most workers. Frequently used exposure surrogates included the mapping of work histories to job exposure matrices (JEMs) based on industrial hygiene measurements and job title and tasks, summed over the duration of employment (see supplemental fig. 2). A few studies reported results for "peak" exposures, although the definition of peak exposure varied by study, and none directly measured peak formaldehyde exposures for individual cohort members [23]. Overall, the domains of outcome assessment and exposure assessment were identified as potentially important for identifying influential sources of bias and are explored in more detail below.

Risk Visualization: An example using the Outcome and Exposure Assessment Domains

Forest plots (with risk estimates stratified by domain-specific study features) and summary tables visually explored possible bias scenarios including potential direction and magnitude of biases for the two domains which were previously identified as containing the most variability. These helped identify key study features (i.e., precision of reported risk estimates, exposure assessment based on post-diagnosis or informant recall vs. job-specific exposure matrices) that could influence apparent associations between exposure indicators and disease risk, and the potential magnitude of the bias. Forest plots were generated with R version 4.2.2. Results for each of these two domains are further described below.

Regarding outcome assessment, Fig. 2 presents study results by ML subtype. Overall, three studies [39,40,41] reported incident AML with pathological verification. Risk estimates for AML and CML generally clustered around the null and when elevated were not statistically significant. As demonstrated by the consistency and relative precision of the risk estimates for AML reported in Fig. 2, our investigation of disease specific associations proved unremarkable, suggesting no clear bias, although the results for CML were generally less precise due to smaller numbers. The single result from Hauptmann et al. [31] appears to be an outlier, likely due to its high imprecision.

* Case-control studies **Cohort based studies.

While few studies reported results by time period of diagnosis, the NCI cohort study [29,23] reported results over time within the study population for ML associated with comparable levels of peak formaldehyde exposure. This provides some indication that the year of death or diagnosis may influence risk estimates due to changing diagnostic criteria, although attributing the observed trend to any one hypothesis is difficult given the instability of the risk estimates (see Table 2).

In contrast to self-reported occupational information (such as that used in Saberi Hosnijeh et al. [40]) and the low prevalence of substantial exposure to formaldehyde in the population-based studies, exposure metrics derived from occupational-based JEMs [29,23] will likely allow for more reliable exposure estimates given the use of detailed employment records combined with some empirical industrial hygiene measurements to inform historical exposure estimates. As shown in Fig. 3, only three studies reported ML results by cumulative or peak exposure [29,23,31]. Reported associations varied slightly across exposure metrics (e.g., cumulative, average intensity, peak) [29,23,28,31]. Nevertheless, regardless of exposure metric used, results were strikingly consistent within the NCI industrial cohort workers cohort (i.e., Beane Freeman et al. [29] and the extended analyses by Checkoway et al. [23], indicating no clear associations, whereas Hauptmann et al. [31] reported remarkably high risks across all exposure groups.

Table 2
Relative risk (RR) of ML stratified by year of diagnosis from [23].

Year of death	RR Peak 2.0 - < 4.0 ppm	RR Peak > 4.0 ppm
Before 1981	3.71 (0.71–19.36)	3.92 (0.78–19.67)
1981–1994	1.23 (0.27–5.56)	2.70 (0.79–9.17)
1995–2004	0.67 (0.17–2.67)	0.71 (0.20–2.51)

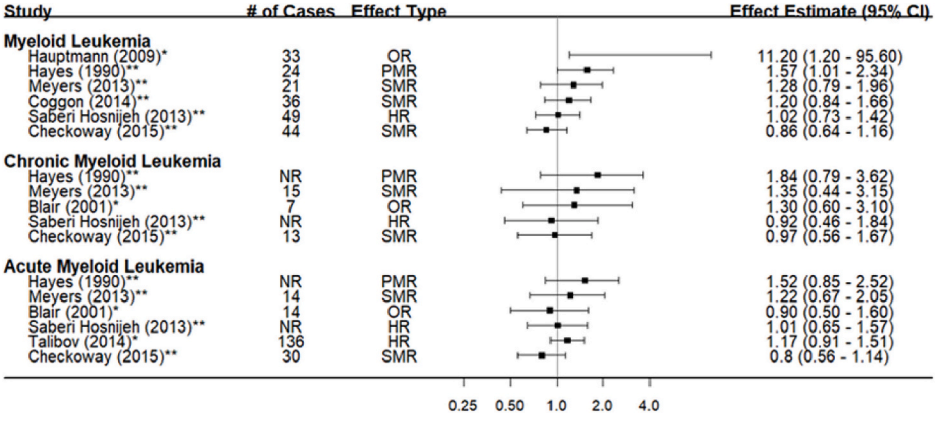


Fig. 2. Forest plot depicting overall results of formaldehyde exposure (ever vs. never exposed) and risk of myeloid leukemia, CML and AML.

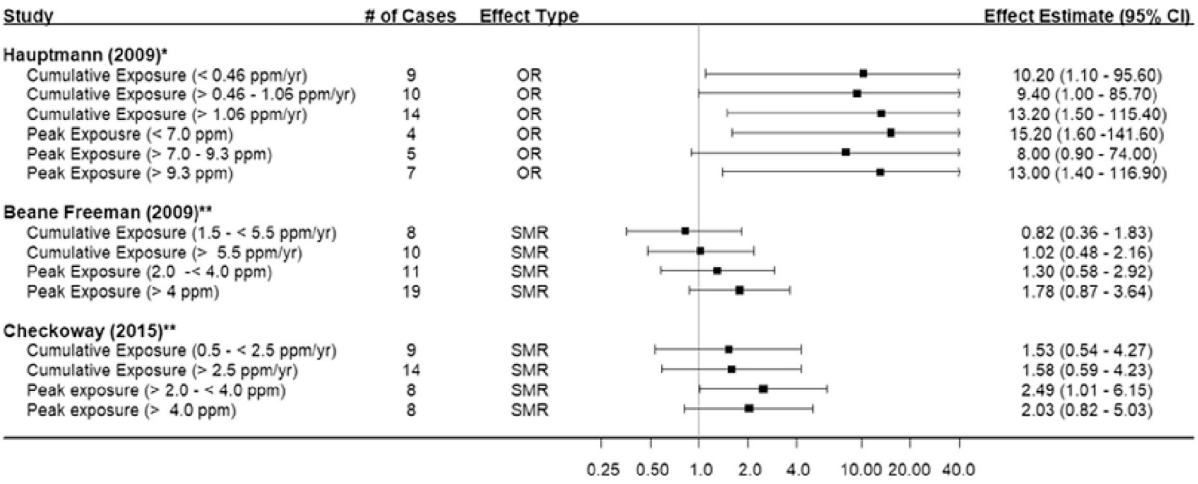


Fig. 3. Forest plot for formaldehyde exposure metric and risk of ML by study and exposure metric.
* Case-control study with upper CI's truncated for best graphical fit **Cohort based studies.

Quantify magnitude and direction of bias

In order to infer the magnitude and direction of bias potentially associated with domain-specific features identified in the previous steps of the proposed framework, principles of meta-analysis and sensitivity analysis were used. Specifically, Fig. 4 provides pooled meta-risks of the most recent cohort updates among the available literature. Pooled meta-relative risks were calculated assuming a Poisson distribution with 95% CI's using the online software SISA.[1] The pooled meta-risks for all myeloid leukemias, chronic myeloid leukemia and acute myeloid leukemia were 1.41 (95% CI; 1.20–1.64, based on 164 observed cases and 116.78 expected), 1.17 (95% CI; 0.82–1.63, based on 35 observed cases and 29.89 expected), and 0.99 (95% CI; 0.77–1.26, based on 58 observed cases and 64.54 expected), respectively. Studies that reported hazard ratios [40,41] were excluded from the pooled meta-risk calculation, as they estimated risk by incorporating a time dimension.

Based on the critical study quality evaluation results, two studies [31,32] had suggestive evidence of sources of bias in multiple domains and, in particular, outcome and exposure assessment. Given the multiple sources of potential bias, as indicated by the triangulation exercise, associated with these two studies, a sensitivity analysis was conducted to estimate the quantitative impact of these risk estimates on the overall meta-risk estimates for the entire body of literature. Integrating each of these five various triangulations repeatedly resulted in two studies with consistently high risk of bias across each domain [31,32]. Removing those risk estimates from our initial results (Fig. 4) produces Fig. 5 below. As shown, the pooled meta-risk for ML overall was attenuated (1.04, 95% CI: 0.84–1.26 vs. 1.41, 95% CI: 1.20–1.64 respectively) with an approximate 38% reduction in observed cases and a 26% reduction in the overall meta-risk. A similar, but less dramatic movement towards the null can also be seen for AML (0.90, 95% CI: 0.68–1.16 vs. 0.99, 95% CI: 0.77–1.20 respectively). This may, in part, demonstrate the observed magnitude and direction (away from the null) of bias from these two studies. The remaining body of epidemiologic literature reviewed consistently reported myeloid leukemia estimates clustered around the null.

Integration and synthesis

The final step was to integrate and synthesize the evidence on the potential role of each of the biases within the domains and their potential impact on possible causal conclusions [8]. Overall, and not uncommon among occupational epidemiological studies, the risk of bias arising from potential exposure misclassification remains substantial. The variability in exposure assessment provides some evidence that these methodological differences and underlying assumptions influenced results. In particular, some evidence of bias away from the null was possible given the attenuated risk estimates when individuals with missing exposure data were excluded from the case-control analysis of embalmers and funeral directors [31]. Interestingly, shortly after it was published, Cole et al. [44] critically assessed the Hauptmann et al. [31] study, reporting that the despite the reported statistical associations (specifically exposure odds ratios) reported, there was evidence of no excess of ML in the underlying cohort.

Within the outcome assessment this triangulation evaluation indicates a lack of increased risk among studies that evaluated AML and CML incident cases with pathological confirmation or cause of death coded by an independent trained nosologist. Further, studies relying solely on deaths due to myeloid leukemia as reported on death certificates such as in the early PMR studies (e.g., [32,34,35]), especially if the majority of death records predate diagnostic and classification refinements developed in the 1990's and 2000's, may be susceptible to outcome ascertainment error leading to bias away from the null. See supplemental fig. 3 for additional analysis.

Within the exposure domain, likely due to the various methodological approaches employed for exposure assessment across studies, some positive associations were reported, although inconsistently and likely subject to exposure estimation error and other sources of bias, or chance. Higher confidence was placed in findings from studies that relied upon detailed employment records and some degree of historical monitoring. Within studies that used multiple exposure metrics to explore the relationship between formaldehyde exposure and ML, some exposure variables likely were more susceptible to misclassification than others e.g., peak versus cumulative or average exposure.

Overall, and as illustrated in figs. 1–5, most reported epidemiological results do not demonstrate statistically significant associations between occupational exposure to formaldehyde and risk of ML, AML or CML. One study presented several large statistically significant associations, but these were based on small numbers of ML deaths and highly imprecisely estimated [31]. Sporadic positive results were reported in three additional studies, but these were inconsistent across studies [29,23,32]. Further, triangulating the results of all studies by the selected domains and study characteristics using forest plots to visualize the contrasts repeatedly identified two studies that both reported anomalously large relative risk estimates (mostly PMRs and ORs) and that were more likely to be influenced by multiple biases, especially arising from exposure assessment approaches and small numbers of ML-

[1] www.quantitativeskills.com/sisa/statistics/smr.htm

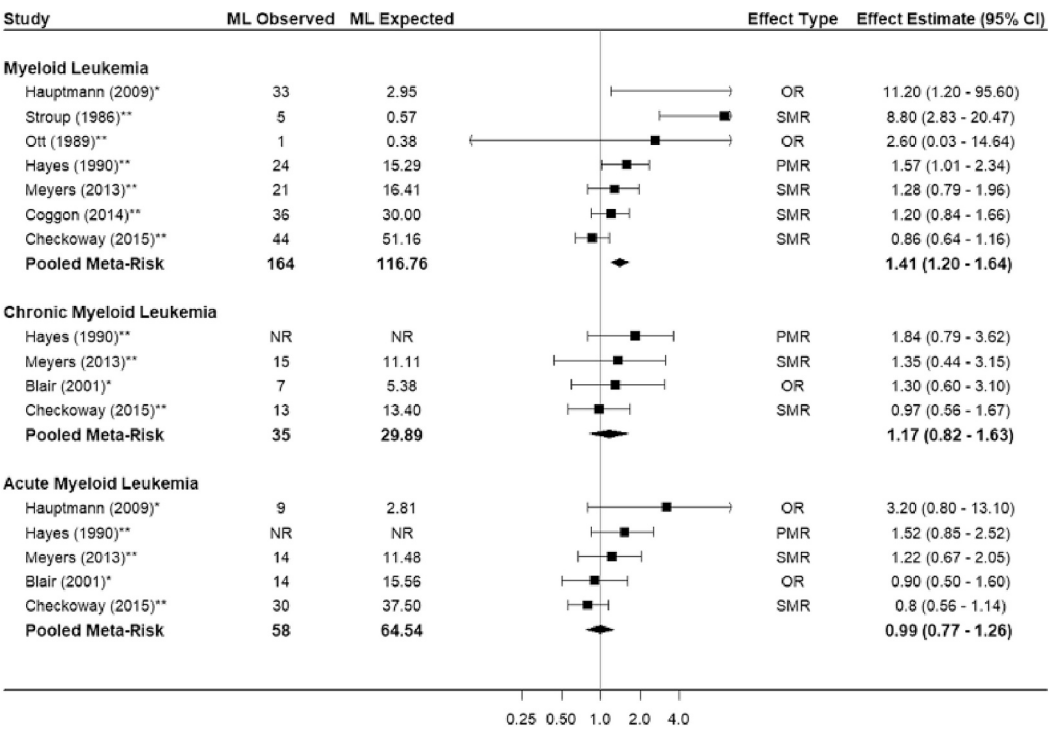


Fig. 4. Overall Results including pooled meta-risk estimates.
*Case-control studies **Cohort based studies.

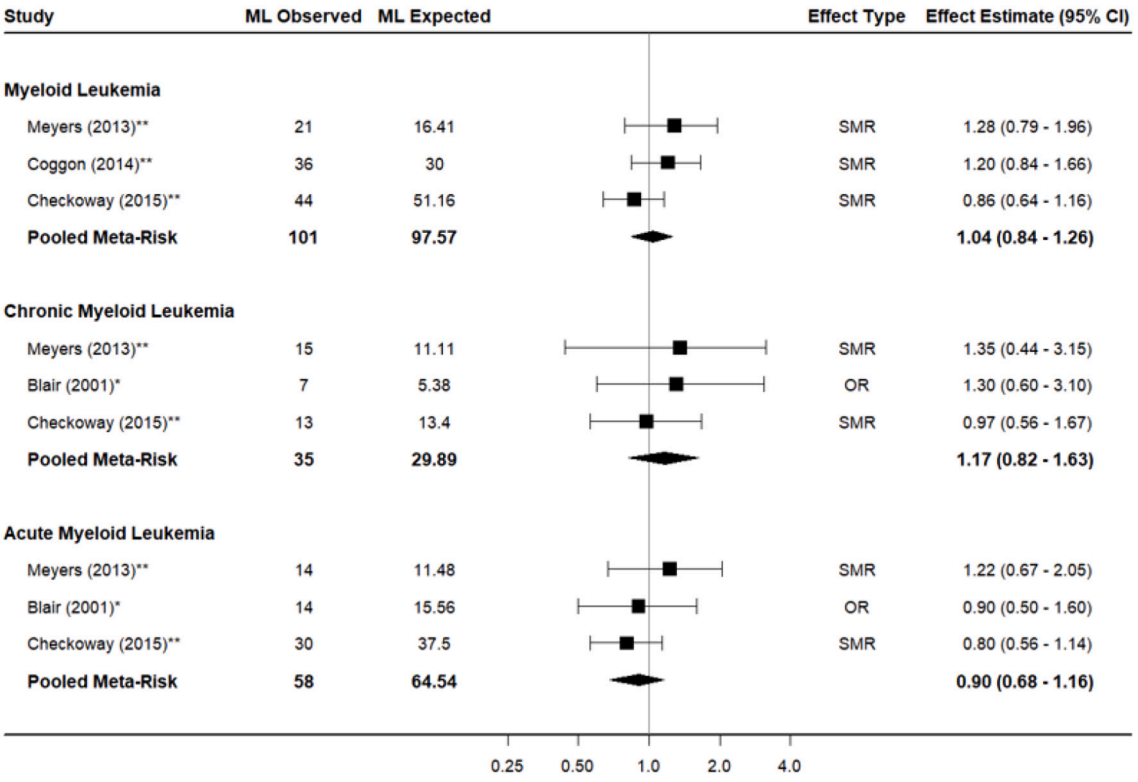


Fig. 5. Overall Results including pooled meta -risk estimates with studies considered biased removed.
*Case-control studies **Cohort based studies.

specific deaths that likely biased results away from the null ([31,32]). The same studies reporting the largest relative risk estimates consistently were identified among those most susceptible to multiple hypothesized biases, whereas most of the epidemiological evidence, and especially the studies considered to have lower ROB, does not indicate any clear or consistent association between various metrics of occupational formaldehyde exposure and the risk of AML and CML.

Discussion

Triangulation methods draw on common practices used throughout epidemiological systematic reviews and meta-analyses (e.g., forest plots used to visualize heterogeneity in reported study results) and therefore are intuitive. The proposed triangulation framework and applied example specific to occupational formaldehyde exposure and risk of myeloid leukemia provide, to the authors knowledge, one of the first attempts to develop and implement an adaptable triangulation framework. The study results suggest that epidemiological evidence triangulation offers several possible advantages when evaluating and synthesizing the findings from a body of studies to address sufficiently focused hypotheses regarding specific biases. For example, Schwilk et al. [45] performed "an updated meta-analysis focusing on high-exposure groups" and evaluation of bias of the epidemiological literature on formaldehyde and leukemia. However, with the exception of Egger's and Begg's test to investigate publication bias, a methodology for identifying sources of bias was not provided. Despite the lack of a reproducible or transparent methodology for carefully evaluating potential sources of bias, the authors reported that "[a]lthough confounding, publication bias, diagnostic bias, or substantial exposure or outcome misclassification cannot be completely ruled, [the] evaluations suggest that these biases are unlikely causes of the associations identified" ([45]). Importantly, the investigators acknowledged that "some inconsistencies remain" in the epidemiological literature and that future research was necessary to "explain these inconsistencies". The proposed triangulation framework therefore demonstrates the vital importance of developing methodologies to closely evaluate potential sources of bias both within and across studies that may influence causal determinations.

The proposed triangulation framework also has several limitations, not the least of which is the lack of information provided by some publications that may help elucidate biases and their potential sources. For example, exposure and confounder data generally were insufficient to readily assess the direction and especially magnitude of any biases and disentangling individual sources of potential biases. Nonetheless, in the absence of more detailed information, the ability to apportion the contribution of any of these potential biases specifically to any one source is difficult, and it remains likely that the combination of biases leads to the more extreme reported results for formaldehyde and ML, AML and CML relative to the more consistent body of null epidemiological results. It should also be noted that, while the exposure and outcome domains were investigated in detail for potential sources of bias as part of the applied example, future triangulation exercises might benefit from risk visualization of all domains. One additional source of bias which bears mentioning is publication bias. As the goal of triangulation is to assess the available epidemiological evidence, evaluating the potential bias from unpublished or unreported studies may fall outside the scope of triangulation. However, it appears that publication bias is unlikely for this body of literature; a systematic review and meta-analysis on environmental carcinogens (including formaldehyde) and MLs also included the occupational cohort studies summarized here as well as two population-based control studies. Based on visual inspection of funnel plots and Egger's test the authors concluded that for this body of literature publication bias appears unlikely [24].

Interestingly, our evaluations found no clear indication suggesting bias towards the null for any study domain. However, bias towards the null was frequently postulated in several of the studies reviewed [29,39,40,41] due to non-differential misclassification of either exposure or outcome. The default hypothesis of non-differential misclassification contributing to bias towards the null may not be supported and therefore cannot be claimed unless specific assumptions are met [46]. In theory, where no true underlying association exists, the best representation is the null, and potential error (e.g., random or measurement error) may not be detectable.

Accordingly, triangulation also may facilitate differentiating apparently "mixed" bodies of evidence for which conflicting results may be related to possible biases arising from variability in study design and quality from "truly conflicting" evidence, i.e., studies of similar exposure potential, study design quality reporting different results unlikely due to identifiable biases. Triangulation may be particularly useful in the context of regulatory reviews and risk assessments, in which voluminous bodies of epidemiological evidence must be carefully evaluated to inform regulations. EPA emphasized the importance of carefully evaluating the effects of bias when synthesizing apparently conflicting bodies of evidence in its recently finalized IRIS Handbook [10]. Other lines of scientific inquiry also can inform the interpretation of evidence triangulation. For example, for formaldehyde and risk of AML, the strong biological evidence of formaldehyde's inability to reach the bone marrow must inform the triangulator that observed associations likely have other explanations including study biases. Therefore, the potential is great for refining critical assessments and evidence syntheses including systematic reviews, specifically through more purposeful evidence triangulation to provide insights regarding the epidemiological evidence as it might or might not align with impressions drawn from animal and mechanistic evaluations.

Additionally, while it may seem counterintuitive, the triangulation exercises above that found little difference across some study characteristics in fact demonstrate the importance of this somewhat novel approach for evaluating biases and their impact on evidence synthesis. In standard ROB approaches that average study quality scores across domains, key criteria or domains can be diluted by less informative ones. Furthermore, demonstrating that there is little or no variability across categories of study characteristics provides information that allows sharpening the focus on areas where reported study results more likely may have been influenced by bias or some other methodological difference. Nonetheless, our approach and application of triangulation principles and methods will require further development and refinements to increase objectivity and reproducibility, and additional pragmatic examples on different topics and bodies of epidemiological studies are needed. For example, within this proposed framework, further development is needed for generating testable hypotheses, appropriately visualizing risk estimates and quantifying the magnitude and direction of bias. Ideally, triangulation frameworks will be refined and applied to complement– and further strengthen – evaluations and conclusions based on formal systematic review frameworks. The approach outlined and applied above is intended to provide a pragmatic starting point and applied example that can be advanced, modified and refined. Its attractiveness in part arises from the series of ROB assessments by domains and study characteristics and the repeated visualization of study findings in the context of all others similarly examined. In other words, the study results begin to speak for themselves, and possible interpretations and conclusions become more tangible.

CRediT authorship contribution statement

Daniel J. Lauer: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Anthony J. Russell:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Heather N. Lynch:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **William J. Thompson:** Writing – review & editing, Formal analysis, Conceptualization. **Kenneth A. Mundt:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Harvey Checkoway:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

DJL, AJR, and HNL are employed by Stantec ChemRisk, a consulting firm that provides scientific support to the government, corporations, law firms, and various scientific/professional organizations. KAM and

WJT are independent consultants who have provided scientific evaluation on behalf of clients in litigation and regulatory settings in the United States and Europe. HC has received funding for consulting projects related to formaldehyde. The content and the conclusions of the manuscript are exclusively those of the authors. This work was supported, in part, by funding from the Foundation for Chemistry Research and initiatives, a 501(c)(3) tax-exempt organization established by the American Chemistry Council.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2024.100143.

References

[1] Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol 2016;45(6):1866–86.
[2] National Academies of sSciences E, mMedicine (NASEM). Triangulation in environmental epidemiology for EPA human health assessments: Proceedings of a workshop. 2022.
[3] Lynch HN, Mundt KA, Pallapies D, Ricci PF. Lost in the woods: finding our way back to the scientific method in systematic review. Global. Epidemiology 2022;4. article 100093.
[4] Lynch HN, Goodman JE, Tabony JA, Rhomberg LR. Systematic comparison of study quality criteria. Regul Toxicol Pharmacol 2016;76:187–98. https://doi.org/10.1016/j.yrtph.2015.12.017.
[5] Pega F, Norris SL, Backes C, Bero LA, Descatha A, Gagliardi D, et al. RoB-SPEO: A tool for assessing risk of bias in studies estimating the prevalence of exposure to occupational risk factors from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. Environment International 2020;135:105039. https://doi.org/10.1016/j.envint.2019.105039. Epub 2019 Dec 18. PMID: 31864023; PMCID: PMC7479507.
[6] Eick SM, Goin DE, Chartres N, Lam J, Woodruff TJ. Assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools. Syst Rev 2020;9(1):1–13.
[7] Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol 2010;25(9): 603–5. https://doi.org/10.1007/s10654-010-9491-z.
[8] Savitz DA, Wellenius GA, Trikalinos TA. The problem with mechanistic risk of bias assessments in evidence synthesis of observational studies and a practical alternative: assessing the impact of specific sources of potential bias. Am J Epidemiol 2019;188(9):1581–5.
[9] U.S. EPA. Application of systematic review in TSCA risk evaluations. In: U.S. EPA Office of chemical safety and pollution prevention. Washington, DC: EPA; 2018. 740-P1–8001.
[10] US EPA. IRIS toxicological review of formaldehyde-inhalation (external review draft, 2022). Washington, DC: U.S. Environmental Protection Agency; 2022. EPA/635/R-22/039, 2022.
[11] NTP (National Toxicology Program). Handbook for Conducting Systematic Reviews for Health Effects Evaluations – March 2019. https://ntp.niehs.nih.gov/whatwestudy/assessments/noncancer/handbook; 2019 [last updated March 4, 2019].
[12] U.S. EPA.. Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies. Washington, DC: U.S. EPA Office of Chemical Safety and Pollution Prevention; 2021. EPA-D-20-031.
[13] Lynch HN, Lauer DJ, Thompson WJ, Leleck O, Freid RD, Collins J, et al. Systematic review of the association between talc and female reproductive tract cancers. Toxicol: Front; 2023.
[14] Lynch HN, Kozal JS, Russell AJ, Thompson WJ, Divis HR, Freid RD, et al. Systematic review of the scientific evidence on ethylene oxide as a human carcinogen. Chem Biol Interact 2022;364:110031. https://doi.org/10.1016/j.cbi.2022.110031.
[15] IARC (International Agency for Research on Cancer). Formaldehyde [IARC monograph]. In: A review of human carcinogens: Chemical agents and related occupations (pp. 401-435). Lyon, France; 2012. http://monographs.iarc.fr/ENG/Monographs/vol100F/index.php.
[16] NRC (National Research Council). Review of the formaldehyde assessment in the National Toxicology Program 12th report on carcinogens. Washington (DC): National Academies Press (US); 2014. https://doi.org/10.17226/18948.
[17] French Agency for Food, Environmental and Occupational Health & Safety (ANSES) (2023). Occupational exposure to formaldehyde: an established link with myeloid leukemia. Maisons-Alfort, France: Retrieved from: Occupational exposure to formaldehyde: an established link with myeloid leukaemia | Anses - Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail.
[18] Gentry R, Thompson CM, Franzen A, Salley J, Albertini R, Lu K, et al. Using mechanistic information to support evidence integration and synthesis: a case study with inhaled formaldehyde and leukemia. Crit Rev Toxicol 2020;50(10): 885–918.
[19] Gentry PR, Rodricks JV, Turnbull D, Bachand A, Van Landingham C, Shipp AM, et al. Formaldehyde exposure and leukemia: critical review and reevaluation of the results from a study that is the focus for evidence of biological plausibility. Crit Rev Toxicol 2013;43(8):661–70.
[20] Mundt KA, Gallagher AE, Dell LD, Natelson EA, Boffetta P, Gentry PR. Does occupational exposure to formaldehyde cause hematotoxicity and leukemia-specific chromosome changes in cultured myeloid progenitor cells? Crit Rev Toxicol 2017;47(7):598–608.
[21] ECHA (European Chemicals Agency). Committee for risk assessment. RAC. Opinion proposing harmonized classification and labeling at EU level of formaldehyde. https://echa.europa.eu/documents/10162/b8dfa022-9544-72e8-dcaa-7491dff3c0d5; 2012.
[22] SCOEL (Scientific Committee on Occupational Exposure Limits). SCOEL/REC/125 formaldehyde: recommendation from the scientific committee on occupational exposure limits. 2017.
[23] Checkoway H, Dell LD, Boffetta P, et al. Formaldehyde exposure and mortality risks from acute myeloid leukemia and other Lymphohematopoietic malignancies in the US National Cancer Institute cohort study of Workers in Formaldehyde Industries. J Occup Environ Med 2015;57(7):785.
[24] Mundt KA, Dell LD, Boffetta P, Beckett EM, Lynch HN, Desai VJ, et al. The importance of evaluating specific myeloid malignancies in epidemiological studies of environmental carcinogens. BMC Cancer 2021;21(1):227. https://doi.org/10.1186/s12885-021-07908-3.
[25] Pinkerton L, Hein M, Stayner L. Mortality among a cohort of garment workers exposed to formaldehyde: an update. Occup Environ Med 2004;61(3):193–200.
[26] Stayner LT, Elliott L, Blade L, Keenlyside R, Halperin W. A retrospective cohort mortality study of workers exposed to formaldehyde in the garment industry. Am J Ind Med 1988;13(6):667–81.
[27] Blair A, Stewart P, O'Berg M, et al. Mortality among industrial workers exposed to formaldehyde. J Natl Cancer Inst 1986;76(6):1071–84.
[28] Hauptmann M, Lubin JH, Stewart PA, Hayes RB, Blair A. Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries. J Natl Cancer Inst 2003;95(21):1615–23.
[29] Beane Freeman LE, Blair A, Lubin JH, et al. Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: the National Cancer Institute Cohort. JNCI 2009;101(10):751–61.
[30] Meyers AR, Pinkerton LE, Hein MJ. Cohort mortality study of garment industry workers exposed to formaldehyde: update and internal comparisons. Am J Ind Med 2013;56(9):1027–39.
[31] Hauptmann M, Stewart PA, Lubin JH, et al. Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. J Natl Cancer Inst 2009;101(24):1696–708.
[32] Hayes RB, Blair A, Stewart PA, Herrick RF, Mahar H. Mortality of US embalmers and funeral directors. Am J Ind Med 1990;18(6):641–52.
[33] Stroup NE, Blair A, Erikson G. Brain cancer and other causes of death in anatomists. J Natl Cancer Inst 1986;77(6):1217–24.
[34] Walrath J, Fraumeni Jr JF. Mortality patterns among embalmers. Int J Cancer 1983;31(4):407–11.
[35] Walrath J, Fraumeni Jr JF. Cancer and other causes of death among embalmers. Cancer Res 1984;44(10):4638–41.
[36] Coggon D, Ntani G, Harris EC, Palmer KT. Upper airway cancer, myeloid leukemia, and other cancers in a cohort of British chemical workers exposed to formaldehyde. Am J Epidemiol 2014;179(11):1301–11.
[37] 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.
[38] Pira E, Romano C, Verga F, La Vecchia C. Mortality from lymphohematopoietic neoplasms and other causes in a cohort of laminated plastic workers exposed to formaldehyde. Cancer Causes Control 2014;25:1343–9.
[39] Blair A, Zheng T, Linos A, Stewart P, Zhang Y, Cantor K. Occupation and leukemia: a population-based case—control study in Iowa and Minnesota. Am J Ind Med 2001;40(1):3–14.
[40] Saberi Hosnijeh FS, Christopher Y, Peeters P, et al. Occupation and risk of lymphoid and myeloid leukaemia in the European prospective investigation into Cancer and nutrition (EPIC). Occup Environ Med 2013;70(7):464–70.
[41] Talibov M, Lehtinen-Jacks S, Martinsen JI, et al. Occupational exposure to solvents and acute myeloid leukemia: a population-based, case-control study in four Nordic countries. Scand J Work Environ Health 2014:511–7.
[42] Percy CL, Miller BA, Ries LAG. Effect of changes in cancer classification and the accuracy of cancer death certificates on trends in cancer mortality. Ann N Y Acad Sci 1990;609:87–99.
[43] Percy CL, Stanek E, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. Am J Public Health 1981;71:242–50.
[44] Cole P, Adami HO, Trichopoulos D, Mandel J. Formaldehyde and lymphohematopoietic cancers: a review of two recent studies. Regul Toxicol Pharmacol 2010;58(2):161–6. https://doi.org/10.1016/j.yrtph.2010.08.013. Epub 2010 Aug 22. PMID: 20736040.
[45] Schwilk E, Zhang L, Smith M, et al. Formaldehyde and Leukemia: An Updated Meta-Analysis and Evaluation of Bias. Journal of Occupational and Environmental Medicine. 2010; 52 (9): 878-886.
[46] Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. Int J Epidemiol 2005;34(3):680–7. https://doi.org/10.1093/ije/dyi060.