November 19, 2013

Developmental and Reproductive Toxicant Identification Committee
California Environmental Protection Agency
Office of Environmental Health Hazard Assessment

via email at P65Public.comments@oehha.ca.gov

re: Agenda Item III: Tabulating Data from Epidemiology Studies in Hazard Identification Documents

Dear DART-IC Members:

I am writing to express concern regarding the proposed table for data from epidemiology studies in hazard identification documents provided on the California Environmental Protection Agency's Office of Environmental Health Hazard Assessment (OEHHA) website (http://www.oehha.ca.gov/prop65/public_meetings/pdf/1013SummMockupTabEpiStudies.pdf). Although this table provides space for the pertinent information from epidemiology studies, as it is laid out, it may encourage selective reporting of data and will make it difficult to compare results across studies.

The need for a systematic way to abstract data and present them for consideration in a full evidence integration process is under much discussion in the hazard and risk assessment arena, as evidenced by such processes as the National Academy of Sciences (NAS) "roadmap," the Cochrane Collaboration, the Grading of Recommendations Assessment, Developmental and Evaluation (GRADE) process, the recently proposed Office of Health Assessment and Translation (OHAT) process, and others (reviewed by Rhomberg *et al*., 2013). It would be good for OEHHA to follow this trend for facilitating systematic review that the hazard and risk assessment process increasingly expects and demands. Based on Gradient's extensive work on systematic reviews and weight-of-evidence (WoE) analyses, below are several suggestions that will improve the utility of epidemiology data tables.

Rather than put all the information from each study in one table, it is useful to have a series of tables that each focus on a specific strand of the pertinent information. For example, the first table should have all the basic information about the studies and the types of data contained in each. The results can then be laid out separately by outcome in subsequent tables. This will help make sure that all data on each outcome is reported. This will also help researchers avoid the tendency to omit null data from the tables, which can make results appear more consistent than they actually are. Examples of these types of tables can be seen in our publications on formaldehyde (Rhomberg *et al*., 2011) and chlorpyrifos (Prueitt *et al*., 2012), attached.

In a similar vein, listing all the results from one study in a single cell makes it difficult to compare results across studies, as it will be difficult to pull out results from similar analyses to compare. One way to address this is to have a column at each dose level or range so one can look across rows for dose-response within studies and down columns for consistency at the same dose levels.

Within the outcome tables, the studies can be organized by factors that may lead to different interpretations of results. For example, if exposure measurement affects study quality, studies can be grouped by the way in which exposure was measured. Or, if several potential confounders are important to consider, studies can be grouped by those that did or did not account for those confounders.

20 University Road, Cambridge, MA 02138 | 617-395-5000 | www.gradientcorp.com

It is critical that tables be constructed in such a way that data are abstracted the same way from each study. Having columns with a lot of information under comments is likely to result in inconsistent reporting across studies. It is much better to have a column for each factor (*e.g.*, selection bias, information bias, exposure measurement error) to be sure it is clear which factors are considered in each study. Gradient recently conducted a survey of best practices in WoE methodology (Rhomberg *et al*., 2013) and applied this to US EPA's framework for casual determinations for the National Ambient Air Quality Standards (Goodman *et al*., 2013). Our evaluations, particularly the latter, describe specific information that should be considered when evaluating study quality and relevance that can be put in study quality tables.

Finally, it may be difficult to have a template table for all epidemiology studies, as different toxicants and outcomes may require different types of information (*e.g.*, if evaluating air toxicants, whether exposure was measured by personal or centrally located air monitors is critical information). Thus, a one-size-fits-all table may not be the best way to move forward. Rather, a template that provides a set of tables that can be adapted based on the causal question is likely to be the most helpful.

In the end, there is no perfect way to tabulate data from many studies with many differences among them. However, it is imperative that all of the data (including null data) be presented in a manner by which the reader will be able to determine consistency and coherence of results, considering study strengths and limitations; this is best accomplished by providing focused tables with more columns and less information in each column. Although compiling tables in this manner initially may be a higher level of effort, it significantly lowers the risk of missing null data and the potential for biased analyses of the WoE for hazard identification.

Thank you for your consideration.

Sincerely,

GRADIENT

Julie E. Goodman, Ph.D., DABT
Principal

email: jgoodman@gradientcorp.com

References

Goodman, JE; Prueitt, RL; Sax, SN; Bailey, LA; Rhomberg, LR. 2013. "Evaluation of the causal framework used for setting National Ambient Air Quality Standards." *Crit. Rev. Toxicol.* doi:10.3109/10408444.2013.837864.

Prueitt, RL; Goodman, JE; Bailey, LA; Rhomberg, LR. 2011. "Hypothesis-based weight-of-evidence evaluation of the neurodevelopmental effects of chlorpyrifos." *Crit. Rev. Toxicol.* 41(10):822-903.

Rhomberg, LR; Bailey, LA; Goodman, JE; Hamade, A; Mayfield, D. 2011. "Is exposure to formaldehyde in air causally associated with leukemia? - A hypothesis-based weight-of-evidence analysis." *Crit. Rev. Toxicol.* 41(7):555-621.

Rhomberg, LR; Goodman, JE; Bailey, LA; Prueitt, RL; Beck, NB; Bevan, C; Honeycutt, M; Kaminski, NE; Paoli, G; Pottenger, LH; Scherer, RW; Wise, KC; Becker, RA. 2013. "A survey of frameworks for best practices in weight-of-evidence analyses." *Crit. Rev. Toxicol.* 43(9):753-784.

Rhomberg *et al*., 2011

Critical Reviews in Toxicology, 2011; 41(7): 555–621
© 2011 Informa Healthcare USA, Inc.
ISSN 1040-8444 print/ISSN 1547-6898 online
DOI: 10.3109/10408444.2011.560140

Ithcare

REVIEW ARTICLE

Is exposure to formaldehyde in air causally associated with leukemia?—A hypothesis-based weight-of-evidence analysis

Lorenz R. Rhomberg[1], Lisa A. Bailey[1], Julie E. Goodman[1], Ali K. Hamade[1], and David Mayfield[2]

[1]*Gradient, Cambridge, Massachusetts, USA, and* [2]*Gradient, Seattle, Washington, USA*

Abstract
Recent scientific debate has focused on the potential for inhaled formaldehyde to cause lymphohematopoietic cancers, particularly leukemias, in humans. The concern stems from certain epidemiology studies reporting an association, although particulars of endpoints and dosimetry are inconsistent across studies and several other studies show no such effects. Animal studies generally report neither hematotoxicity nor leukemia associated with formaldehyde inhalation, and hematotoxicity studies in humans are inconsistent. Formaldehyde's reactivity has been thought to preclude systemic exposure following inhalation, and its apparent inability to reach and affect the target tissues attacked by known leukemogens has, heretofore, led to skepticism regarding its potential to cause human lymphohematopoietic cancers. Recently, however, potential modes of action for formaldehyde leukemogenesis have been hypothesized, and it has been suggested that formaldehyde be identified as a known human leukemogen. In this article, we apply our hypothesis-based weight-of-evidence (HBWoE) approach to evaluate the large body of evidence regarding formaldehyde and leukemogenesis, attending to how human, animal, and mode-of-action results inform one another. We trace the logic of inference within and across all studies, and articulate how one could account for the suite of available observations under the various proposed hypotheses. Upon comparison of alternative proposals regarding what causal processes may have led to the array of observations as we see them, we conclude that the case for a causal association is weak and strains biological plausibility. Instead, apparent association between formaldehyde inhalation and leukemia in some human studies is better interpreted as due to chance or confounding.

Keywords: Epidemiology, formaldehyde, genotoxicity, hazard identification, leukemia, risk assessment

Contents

Abstract 555
1. Introduction and background 557
2. Hypothesis-based weight-of-evidence (HBWoE) evaluation 558
2.1. Overview of approach 558
2.1.1. Hill Criteria and the concept of "accounts" 559
2.2. HBWoE methodology 561
3. Overview of HBWoE as applied to formaldehyde and leukemogenesis 562
4. Weight of epidemiology evidence regarding the association between formaldehyde exposure and leukemia 563
4.1. Overview of epidemiology investigations 563
4.2. Endpoint-by-endpoint analysis 571
4.2.1. All lymphohematopoietic cancers 571
4.2.2. Cancer of lymphoid origin 574
4.2.3. Leukemia 574

Address for Correspondence: Lorenz R. Rhomberg, Gradient, 20 University Road, Cambridge, MA 02138, USA. E-mail: lrhomberg@gradientcorp.com

(Received 29 June 2010; revised 13 September 2010; accepted 13 September 2010)

4.2.4. Lymphatic leukemia ... 578
4.2.5. Hematopoietic cancer of non-lymphoid origin ... 578
4.2.6. Myeloid leukemia ... 581
4.2.7. Other unspecified leukemia ... 582
4.2.8. Hodgkin's lymphoma, non-Hodgkin's lymphoma, and multiple myeloma ... 582
4.3. HBWoE evaluation of epidemiology studies ... 582
4.3.1. Cancer outcome assessments likely lead to disease misclassification ... 585
4.3.2. Exposure assessments likely affected by exposure measurement error or misclassification ... 585
4.3.3. Exposures to other chemicals in the work place may have confounded results ... 586
4.3.4. Exposure-response associations within and among studies are not consistent ... 586
4.3.5. Statistical limitations may have led to spurious associations ... 590
4.3.6. The latency argument proposed by Beane Freeman et al. (2009) appears to be a post hoc explanation for the observed effects. ... 590
4.3.7. Recent formaldehyde meta-analyses do not support an association between formaldehyde exposure and leukemia ... 591
4.4. Summary ... 592
5. Weight of evidence regarding hematotoxicity from formaldehyde exposure ... 593
5.1. Formaldehyde hematotoxicity in animals ... 593
5.1.1. Hematology ... 593
5.1.2. Leukemia ... 594
5.2. Formaldehyde hematotoxicity in humans ... 595
5.3. Hypothesis-based weight-of-evidence evaluation of formaldehyde hematotoxicity studies ... 596
5.3.1. Key animal studies do not provide strong evidence of an association between formaldehyde exposure and hematotoxicity and leukemia ... 599
5.3.1.1. Hematology ... 599
5.3.1.2. Leukemia ... 599
5.3.2. Key human studies do not provide strong evidence of an association between formaldehyde exposure and hematotoxicity ... 601
5.3.3. If formaldehyde causes leukemia in humans, it is likely due to a mechanism that is different from that observed with known leukemogens ... 602
5.3.4. There are alternative explanations for the pancytopenia reported by Zhang et al. (2010b) and the leukopenia reported by other studies ... 603
5.3.4.1. Subjects exposed to formaldehyde share common immunology markers with subjects having dermatitis or other inflammatory conditions ... 603
5.3.4.2. A recent respiratory infection can result in hematological changes—Subjects with exposure to formaldehyde in the study by Zhang et al. (2010b) were more likely than control subjects to have had recent respiratory tract infections ... 604
5.3.4.3. Other unmeasured potential confounders ... 604
5.4. Summary ... 605
6. Weight of evidence regarding a plausible mode of action for formaldehyde leukemogenesis ... 606
6.1. Formaldehyde toxicokinetics ... 606
6.2. Formaldehyde genotoxicity ... 607
6.2.1. DNA adducts and protein cross-links ... 607
6.2.2. Clastogenic and cytogenetic effects ... 608
6.3. HBWoE evaluation of the proposed modes of action for formaldehyde as a leukemogen ... 608
6.3.1. There is no consistent evidence that inhaled formaldehyde induces genotoxicity in bone marrow, NALT, or peripheral HSCs that might lead to leukemia ... 609
6.3.1.1. Bone marrow ... 609
6.3.1.2. Stem cells in the NALT ... 609
6.3.1.3. Circulating peripheral HSCs ... 610
6.3.2. Formaldehyde exposure would have to be very high to induce DNA damage above endogenous levels in the bone marrow, NALT, or circulating HSCs, and would likely be associated with a high degree of irritation ... 612
6.3.3. Circulating HSCs may not readily home back to healthy bone marrow to cause leukemia ... 612
6.4. Summary ... 613
7. Discussion ... 614
Acknowledgments ... 616
Declaration of interest ... 616
References ... 617

1. Introduction and background

Formaldehyde is produced naturally by the human body. It is also a chemical intermediate used in the production of some plywood adhesives, fertilizer, paper, and urea-formaldehyde resins (Agency for Toxic Substances and Disease Registry [ATSDR], 1999). It is found (as a preservative or impurity) in many products around the home, such as antiseptics, medicines, and cosmetics/personal hygiene products (ATSDR, 1999). Formaldehyde is also used for embalming and preserving biological specimens (United States Environmental Protection Agency [US EPA], 2010). Sources of exposure to formaldehyde include occupational exposure during use or production of materials containing formaldehyde; cigarette smoke; off-gassing from manufactured wood products in new mobile homes; and other new products found in homes (e.g., fiberglass, carpets, and paper products) (ATSDR, 1999).

Studies have shown that exposure to high concentrations of formaldehyde in air results in nasal cancer in rats. Some studies in humans exposed to lower concentrations of formaldehyde in air in the workplace found increased incidence of nasopharyngeal cancer, but other studies have not found an increased risk of these cancers in formaldehyde-exposed workers (ATSDR, 1999; Marsh and Youk, 2005; Marsh, 2007a, 2007b; Bachand et al., 2010; US EPA, 2010). More recently, there has been increased concern and scientific debate regarding the potential for exposure to formaldehyde in air to cause lymphohematopoietic cancers in humans, particularly leukemias (US EPA, 2010; Bachand et al., 2010; Beane Freeman et al., 2009; Hauptmann et al., 2009; Zhang et al., 2009, 2010a, 2010b; Pyatt et al., 2008; Golden et al., 2006; Heck and Casanova, 2004).

The concern for formaldehyde-induced leukemogenesis stems from a few epidemiology studies reporting an association between formaldehyde exposure and increased mortality from leukemia (e.g., Beane Freeman et al., 2009; Hauptmann et al., 2009), although other studies show no such effects (e.g., Bachand et al., 2010; Pinkerton et al., 2004). The studies reporting associations have shortcomings, including poor disease classification and unverified estimates of exposure. Studies have been conducted to examine the potential for formaldehyde in air to induce hematotoxicity in animals and humans and leukemia in animals. The animal studies generally reported neither hematotoxicity (Monticello et al., 1989; Appelman et al., 1988; Holmstrom et al., 1989; Kerns et al., 1983; Kamata et al., 1997; Woutersen et al., 1987; Til et al., 1988, 1989; Johannsen et al., 1986) nor leukemia (Albert et al., 1982; Kerns et al., 1983; Sellakumar et al., 1985; Kamata et al., 1997; Feron et al., 1988; Til et al., 1989; Tobe et al., 1989; Takahashi et al., 1986) associated with formaldehyde exposure. Although a few animal studies reported changes in one or more hematology parameters (Dean et al., 1984; Tobe et al., 1989; Vargova et al., 1993), two animal studies reported leukemias (Soffritti et al., 1989, 2002), and a few human study findings were consistent with hematotoxicity from exposure to formaldehyde (Tang et al., 2009; Zhang et al., 2010b), these studies were inconsistent with other study findings and/or plagued by possible confounding.

Despite the lack of substantial and consistent epidemiological and toxicological evidence for formaldehyde leukemogenesis, US EPA has concluded that formaldehyde should be deemed a known human leukemogen (US EPA, 2010), citing possible modes of action put forth by Zhang et al. (2009, 2010a). The three proposed modes of action involve formaldehyde: (1) migrating to and directly targeting bone marrow hematopoietic stem cells; (2) targeting nasal stem cells (nasal-associated lymphoid tissue, or NALT) which then are released from the nasal passage, circulate in the blood, and are eventually incorporated into bone marrow, leading to leukemia; or (3) targeting circulating hematopoietic stem cells, which then migrate back to bone marrow, eventually leading to leukemia. The proposed modes of action, however, find little support in the current literature; there is a large body of evidence indicating that inhaled formaldehyde (at reasonably high exposure levels in humans, 2 ppm) does not move beyond the nasal respiratory mucosa to increase levels in the blood and does not cause DNA damage or cellular transformation (in the bone marrow, circulating hematopoietic stem cells, or the NALT) beyond the portal of entry (Lu et al., 2010, 2011; Moeller et al., 2011; Andersen et al., 2010). These results suggest strongly that if formaldehyde is not getting beyond the nasal respiratory mucosa (as indicated by its lack of genotoxicity and cellular transformation beyond the nasal epithelial cells), it is not likely to induce leukemogenesis (either via genotoxicity or another carcinogenic mode of action).

Acceptance of formaldehyde as a human leukemogen on the strength of observed associations of exposure and effect seen in the epidemiology studies requires accepting the existence of underlying biological processes that embody the causal forces, whether or not these underlying causal processes are identified. This is true of any epidemiological association that is deemed causal, but what is notable about formaldehyde and leukemia is that current understanding both of leukemogenesis by other agents (entailing toxicity to the marrow and genotoxic attack on hematopoietic precursor cells found there) and of formaldehyde kinetics (which appear to preclude such effects distal to the respiratory tract) raises the issue of whether the phenomena observed in the human studies can be interpreted as causal and consistent with known biology. It is not simply that the underlying biological causal processes are unproven—or even hypothetical—but rather, at least at first view, there seems to be no scientifically plausible means for sufficient causal processes to operate based on what is believed to be true about formaldehyde and hematopoiesis.

In the present paper, we evaluate the scientific data relevant to the potential causal association between exposure to formaldehyde in air and leukemia in

© 2011 Informa Healthcare USA, Inc.

humans using the structured hypothesis-based weight-of-evidence (HBWoE) approach we have developed and applied elsewhere (Rhomberg et al., 2010). The HBWoE methodology is described below.

2. Hypothesis-based weight-of-evidence (HBWoE) evaluation

2.1. Overview of approach

Before discussing the evidence regarding formaldehyde's potential leukemogenicity, it is useful to address our overall approach to the weight-of-evidence question by outlining our method, explaining how it differs from other approaches, and setting out why we feel our chosen approach has value. Weed (2005) points out that the term "weight of evidence" is often used loosely; he calls on practitioners to articulate what they mean by the phrase and to specify their approach. Analyses of various technical approaches to weight of evidence have been offered by Krimsky (2005) and Linkov et al. (2009). Clearly, professional judgment is involved, but it is not enough simply to name the evidence at hand and then announce one's conclusion. Our method aims to make the reasoning process and bases for judgments explicit and transparent so that, even if other observers differ with our conclusions, debate can focus on the soundness of the inferences and their connections to study results, rather than devolve into ad hominem arguments about the identity and perspectives of the judges. That is, we seek to make expert judgment a public process by focusing on the *logic* of the process—not just the *outcome*. Ideally, rational evaluation of objective evidence and scientific scrutiny of such evaluation should be the criterion for knowledge, not simple authority of the interpreter.

For some, weight of evidence may connote a process for coming to a yes/no decision in the face of incomplete or contradictory evidence—to agree on a conclusion despite lack of definitive proof—but we seek a method, rather, that arrives at a useful and reasoned characterization of the relative scientific credence that should be placed in alternative interpretations of the data at hand in view of the arguments for and against each alternative. That is, we aim to communicate uncertainty about conclusions so as to enable productive discussion about subsequent decisions.

A good weight-of-evidence analysis should attend to all the relevant data, and not simply cite studies (or particular outcomes within studies) that tend to support or refute a conclusion. The frequent practice of reviewing literature by naming the positive or otherwise notable outcomes of the included studies, emphasizing findings by the studies' authors, and leaving the negative results for other endpoints or measures of effect implicit can bias evaluations when studies are positive and negative for different endpoints. The analysis should entail an endpoint-by-endpoint comparative approach, on the grounds that true causal effects should be specific (particular endpoints, not one or another of a set of arguably related endpoints) and repeatable (within the limits of study uncertainty and power). Although study quality and design strengths and shortcomings should be noted, we favor an approach that does not reject outright less-than-ideal studies (the outcomes of which may be informative nonetheless) but, rather, tempers the conclusions drawn. What makes poorer studies less informative is a decreased ability to distinguish between the causative, face-value interpretation of outcomes and the alternative interpretation that the results are spurious because of intrusion of factors not adequately eliminated as possible influences. Thus, the rational and transparent way to down-weight poorer studies is to consider the impact of this ambiguity as one evaluates alternative interpretations of the data, using the patterns of concordance or lack thereof with other studies as part of the evaluation of the likelihood that the study in question has misled us or informed us.

We also seek an approach that integrates inferences across different and diverse kinds of data that can tie together inference based on epidemiology, animal testing, and mode-of-action and pharmacokinetic data. Too often, in our view, these different realms of inquiry are approached separately—each subset of data evaluated within its own realm and according to its own standards—and only then the conclusions are brought together for synthesis. This approach fails to take advantage of the ways in which information from one realm can and should affect interpretation of data within another. For instance, judgments about whether patterns of association seen in human studies represent a causal connection of chemical exposure and disease ought to be based not only on the concordance and repeatability of such patterns among human studies, they also should consider whether animal studies show signs of the operation of the underlying biological processes. Human data have the advantage of greater relevance to the immediate question at hand, but they suffer characteristically from imprecise measures of exposure and effect, and, being uncontrolled and observational, from the difficulty of eliminating possible extraneous influential factors. Animal studies can be controlled more precisely and the underlying biology can be probed more thoroughly, but the relevance of these studies is indirect and only useful to the degree that the animals share underlying causative processes with humans. Since species-specific effects are known in both humans and particular species or strains of experimental animals, lack of concordance of effect across human and animal studies is not a definitive refutation of the proposed causative process, but the reasons for and plausibility of such species differences or other non-concordant outcomes becomes part of the evaluation of correspondence of hypotheses.

An often-overlooked aspect of weight-of-evidence evaluation is the importance of noting when causative explanations have been accommodated to account for results already in hand and when post hoc additions or modifications to hypotheses have been constructed to

explain what might otherwise be contradictory findings. Such modifications of explanatory models as a result of new data are valid parts of scientific discovery as we seek explanations and insights into possible underlying causes through the examination of the patterns of phenomena, but one needs to distinguish such a creative, hypothesis-generating process from the subsequent testing of those hypotheses with results that were not used in formulating the proposed model of causes. To the extent that hypotheses are supportable only with such added assumptions and interpretations, even if these additions are plausible and even if the data are then fully in accord with the hypothesized explanations, this constitutes weaker support than if the tentative explanations preceded, and were only later confirmed by, the data.

We have developed an approach to the above questions that we term "hypothesis-based weight of evidence" (or HBWoE). It is hypothesis based in the sense that its critical aspect is to specify the hypothesized basis for using information at hand to infer the existence of the ability of an agent to cause human health impact. The "hypothesis" referred to in the name "hypothesis-based weight of evidence" consists of the proposed basis for using the cited study results as evidence of human risk. That is, one names the study observations that are being proposed as giving insights into human risk and also names the proposed basis for how those observations could be interpreted as informative about human risk potential. This hypothesized basis can be specific in its biological mode-of-action underpinnings, but it can also be more general. For instance, one might base the proposal that an agent is a human carcinogen on observations of its carcinogenicity in animal studies on the grounds that rodents and humans share a good deal of common mammalian biology and the body of observations about how frequently positive animal tests are found for agents with direct human evidence for carcinogenicity. The strength of such an inference would be judged in view of our experience from other agents regarding how often common biology indeed seems to be operating in human and animal disease, the frequency of concordant and discordant results, and the consistency of animal tests observed for the particular chemical at hand.

The hypothesized basis for inference about human risk from particular data should be seen not just as an extrapolation, but as a generalization—it is a proposal about something in common regarding the causal processes in the study situation and the human population of interest. As a generalization, it ought to apply to other situations as well—or at least have reasons why it does not—and one can evaluate the success of the hypothesis at being in accord with the whole suite of relevant observations at hand. If there are limits to the generalization—it applies to one species but not another, to males but not females, at this dose but not that dose—then the plausibility of such exceptions in view of available evidence and broader knowledge becomes part of the evaluation of the hypothesis against available data. (Such inferences and evaluations are particularly susceptible to the kind of post hoc modification of hypotheses mentioned above, and care must be taken to account for after-the-fact adjustments of the hypothesis in evaluating its strength.)

2.1.1. Hill Criteria and the concept of "accounts"

Whenever a causal hypothesis is proposed, there is always (at least implicitly) a counter-hypothesis that the common link does not exist, and the array of outcomes we see among the studies at hand have other explanations that do not bear the same implications about potential risk in human target populations. When evaluating hypotheses, we suggest that it is important to make these counter-hypotheses explicit as well, including as much specificity about the nature of these "other explanations" as can usefully be provided, so that the alternatives can also be evaluated against all the data. In the end, compelling hypotheses are ones that not only are in accord with and serve to explain patterns and concordances among the data, but also have few ad hoc adjustments to account for observations that do not fit; moreover, they provide markedly more plausible explanations of the array of results on hand than can be provided by the counter-hypotheses. Evaluating explicit hypotheses and their alternatives against all the data provides transparency about the basis for expert professional judgment and communicates how scientifically compelling alternative explanations, with different consequences for human risk potential, ought to be deemed.

The question of evaluating causality in epidemiological data is often approached by applying the so-called "Hill Criteria" developed by Sir Austen Bradford Hill (Hill, 1965). A similar or "extended Hill-Criteria" approach has often been applied beyond the realm of epidemiology. In view of this established practice, the question may arise: What does HBWoE provide that is not already provided by the Hill Criteria? First, one should note that the Hill Criteria were developed for application to epidemiology data, which by nature are more observational than experimental. The criteria relate to the patterns among observational studies that one ought to expect if a common causal effect were operating but, independently, do not demonstrate causation. At most, adherence of data to the criteria constrains the scope for alternative, non-causal explanations. Epidemiology rarely has the ability to put causal explanations to the test (other than by evaluating consistency with further studies), and the kind of critical tests that can be constructed in experimental studies, with alternative influential factors controlled, is rarely available. Our goal of furthering the integration of epidemiological and toxicological inference is aided by an approach that gives experimentation, and the kind of critical tests that it can provide, a central role.

Second, as often applied, the Criteria become something of a checklist or a set of headings for citation of outcomes favorable or opposed to a causal hypothesis, but each evaluation is often not done very rigorously or transparently and suffers from the criticism we mentioned

© 2011 Informa Healthcare USA, Inc.

above—simply citing the studies that fit and announcing a professional judgment conclusion. Hypothesis-based weight of evidence can be seen as a process for encouraging rigorous and transparent evaluation of the criteria, particularly those referring to consistency, specificity, repeatability, and biological plausibility. In keeping with the theme of not simply making judgments, but rather showing the proposed basis for those judgments, HBWoE emphasizes not just the conclusions about each criterion, but also a transparent and articulated examination of its logical and evidentiary basis. To rigorously address the question of biological plausibility, one needs to follow a method similar to what we propose.

Finally, as Bradford Hill originally intended, his criteria (which he called "postulates") were designed to articulate the basis for judgments and facilitate the integration of evaluations across criteria, not simply as a checklist for which, if enough features of the array of data seemed to fit, causality could be concluded. Hill saw the postulates as guides to thinking rather than as measures of evidence. In our reading of Hill's original paper, his intent for the application is along precisely the lines we propose—the evaluation of a specific causal hypothesis against alternative non-causal explanations. Bradford Hill makes explicit the importance of considering alternative "accounts" of the observations at hand in stating:

> None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non. What they can do, with greater or less strength, is to help us to make up our minds on *the fundamental question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?* (Hill, 1965) [emphasis added]

The essence of the "accounts" (which we put forth in this context as a technical term) is that they constitute being explicit about Bradford Hill's "ways of explaining the set of facts before us." They are not conclusions or findings but, rather, provisional proposals for the reasons behind the set of observations at hand.

Hypothesis-based weight of evidence comes down to evaluation of alternative accounts. An account is a set of proposed explanations and hypotheses that could be put forth to explain all of the observed data at hand. The array of all observations among all relevant studies comprises the fixed set of available facts; the challenge of scientific investigation is to discern what causes and processes account for those facts having come out as they did. Among the explanations that could be tentatively proposed are causal underlying processes that, if true, would lead to observed patterns and apparent connections within and among studies, but one could also entertain explanations that attribute particular outcomes to chance fluctuations, biases in measurement or reporting, confounding factors, operation of case-specific influences of unknown nature, or other such reasons. In the end, all the facts have to be accounted for by some combination of these, since the study outcomes came out as they did for some reason, even if we do not have clear ideas of what those reasons are. Any one proposed set of such reasons constitutes an account—a tentative "story" as to why the facts are as they are.

Clearly, there could be an infinite set of different accounts, but, in practice, there will be a few major contenders. Since the purpose of the weight-of-evidence evaluation is to identify underlying causal factors of relevance to our larger question, the key account will be one that proposes such an underlying causal factor. Such an account is centered on the proposed ability of a chemical to cause and increase the frequency of appearance of a particular toxic effect, put forward as a reason behind the existence of much of the apparent patterns and connections within and among studies. But there may be some facts on hand that are not readily attributed to such a factor, either ones that appear to contradict the general operation of the hypothesized cause or ones that, although not overtly contradicting, nonetheless are not explained by the key causal hypothesis. These facts need tentative explanations as well, from which subsidiary explanations also become part of the account.

There is always an important second account—one that denies the existence of the key causal factor and instead attributes the facts that appear to be explained by such a factor to other causes, either an alternative causal principle or simply a set of case-specific reasons under which any appearance of patterns within and across studies is mere happenstance. When one doubts the outcomes of a poor-quality study, one is in effect entertaining the possibility that some array of other factors or reasons (beside the one the study aimed at characterizing) has accounted for the outcomes, and the study's design does not allow one to attribute the outcomes confidently to the nominally tested influence.

When the "causal" account's plausibility overwhelms the alternative's, which by comparison seems to lack non-arbitrary reasons to deny the apparent patterns of causation, then we can feel confident that we have characterized a truly causal factor. But we undertake weight-of-evidence evaluations precisely when the case is not so clear—when the causal account itself has many facts that require modification or assumed special conditions of the causal hypothesis, or when there are apparently refuting facts that must be explained away as potential counterexamples. In short, weight of evidence is applied when the data at hand have contradictions and limitations such that even the optimal account requires ad hoc elements and assumptions to account for at least some of the problematic facts. The weight of evidence for the existence of the key causal factor consists of the comparative plausibility of the alternative accounts—the one that invokes it and the one that denies it. The credence we should give to an account and its implications for human health risk assessment depends on the degree to which it

provides a more satisfactory and plausible accounting of the array of observations at hand than do any competing accounts. That is, we see the metaphor of "weight" of evidence as being evaluated with a two-pan balance—the relative plausibility of competing accounts—rather than as a single scale showing how much evidence in accord with a conclusion can be accumulated. Our approach to revealing and characterizing the plausibility of each account is to "unpack" the set of explanations they invoke, noting how much each strains credulity in view of the data at hand and wider knowledge of the relevant science. The explanations in each account need not be proven—what is important is that one set out the following questions:

- What is being proposed as causal and generalizable phenomena (i.e., what constitutes the basis for applying observations of biological perturbations or realized risks in other contexts to project potential risks to humans as they are exposed)?
- What is being proposed as the basis for deviations that lead to observations that do not fit the hypothesized causal model (i.e., that would otherwise be counterexamples or refutations)?
- What assumptions are made that are ad hoc (i.e., to explain particulars, but for which the evidence consists of their plausibility and the observations they are adduced to explain)?
- What further auxiliary assumptions have to be made, and how reasonable are they in view of our wider knowledge and understanding?
- What is relegated to error, happenstance, or other causes not relevant to the question at hand?
- For those events or processes proposed as critical for a given account, what other observable manifestations should they have? Are these other manifestations indeed found?
- If either the operation or necessity of the proposed critical events for a given account were disproven, how else would one explain the array of outcomes?

2.2. HBWoE methodology

Although HBWoE is intended to be flexible in its application, the approach generally consists of the following steps, which are not intended to be a checklist and may involve an approach that is not necessarily in this order.

- Systematically review all studies that are potentially relevant to the causal question at hand (i.e., epidemiology, mode of action, pharmacokinetic, toxicology) and summarize the results without regard to whether they tend to support or undermine particular interpretations. All potentially relevant data and modes of analysis, not only those featured or noted as significant by the studies' authors, should be included. The aim is to specify the set of relevant observations that can be brought to bear. Ask further questions about the data within these studies—specifically, think about the quality of the individual studies (strengths and weaknesses of study design, potential for ambiguity of interpretation of outcomes). Note the interpretation of data by the authors and how well those conclusions are supported by the reported observations. Note instances where evidence of associations depends on choosing the most significant among a set of parallel analyses of the same data (e.g., with different category cut-offs or different dose measures) and note whether there is any a priori reason to favor one mode of analysis over others. Note instances where the interpretation of proposed causes may have been accommodated to account for patterns in the data after the fact (e.g., preferring one dose measure over another because it provides a more interpretable pattern to dose-response data). The aim is to provide the basis for a critical review of the available studies, rather than simply collecting the findings noted and conclusions drawn by study authors.
- Within a realm of investigation (e.g., epidemiology, animal toxicology studies), examine the data for particular endpoints across studies. The aim is to evaluate consistency, specificity of apparent effects, and repeatability of outcomes. Note instances of similar patterns across studies, species, sexes, strains, etc., and also instances of apparent discordance among these. The aim is to provide the basis for judging the apparent limitations or exceptions to proposals about generally operating causal effects.
- Identify and articulate lines of argument by which results from available studies could be used to infer the existence, nature, or magnitude of human risk. These could be newly proposed or they could be proposals already put forth within the scientific community that one seeks to evaluate. Each line of argument should specify the data on which the inference would be based and also the reasoning for why those data are informative about the human risk question. Typically, the reasoning would entail a generalization about causal forces such that some commonality is proposed between the causal forces seen in the study data and those that would be presumed to operate in the human target population. It is important to specify how widely the invoked commonality is proposed to apply (e.g., just to humans but not experimental animals, or just to one sex, or just to humans and a particular strain of animals). The proposed reasons for why the limits to generalization exist should also be specified, to the degree possible (so one can evaluate whether they have an evidentiary basis or are simply ad hoc). These lines of argument are the "hypotheses" of HBWoE, and they are articulated so that one can evaluate how well they are in agreement with all of the data, how well they would explain patterns in the data if they were true, what other observable consequences the invoked causal principles should have, and whether in fact these consequences are observed.

© 2011 Informa Healthcare USA, Inc.

- Trace through the logic within each line of evidence. That is, think about how all of the relevant studies within each line of evidence support each other, considering consistencies and inconsistencies across studies. For example, one would do this for all of the epidemiology studies together (i.e., apply Bradford Hill Criteria), all of the mode-of-action and pharmacokinetic data together, and all of the toxicology data together. The aim is to establish how well the hypotheses being examined comport with and help explain common patterns in the data, what data seem to constitute exceptions or contrary outcomes to the hypothesized causal principles, and what reasons for such exceptions might be proposed.
- Trace through the logic regarding all lines of evidence as a whole and how they inform interpretation of each other. Specifically, how the epidemiology studies as a whole, mode-of-action studies as a whole, and toxicology data as a whole (that we have articulated as part of Step 4) inform interpretation of one another. The question is whether explanations or hypothesized causal factors proposed in one realm (e.g., epidemiology) have aspects that should be observable in others (e.g., mode-of-action studies), enabling evaluation of whether signs of those causal processes do or do not appear where expected.
- Next, one needs to formulate alternative accounts. Each account comprises a set of proposals, hypotheses, assertions, and assumptions that together should provide a tentative story for why all of the relevant observations came out as they did. Each of the causal hypotheses identified in Step 5 would constitute the core of an account, but the same account should also include the proposed reasons why facts that do not fit or are deemed to be outside the span of generalization should not be taken as disproofs because their non-concordance is explicable. An account that denies a central causal hypothesis as an explanation for an apparent association needs to provide an alternative proposed explanation for the observed patterns.
- Finally, evaluate alternative, and competing, accounts. Now that one has worked carefully through not only each study and each individual line of evidence but, importantly, considered how each line of evidence informs the other, it is at this point that one asks how well each hypothesis is supported by the data and how many ad hoc assumptions are required to support each hypothesis. The rationale and reasoning for how the data support (or do not support) each account's hypotheses, together with the plausibility of subsidiary explanations or assumptions in view of wider biological knowledge, constitute the basis for evaluating the scientific support each account gets from available data. The comparative support constitutes the basis for judging the relative credence that alternative accounts should be given.
- The goal in the end is to present the lines of reasoning for (not to prove or disprove) each account, based on the science and integration of the lines of evidence, so that the data will speak for themselves in supporting (or not supporting) the overarching hypotheses that have been put forth.
- By comparison of the various accounts, one may be left with a variety of outcomes or proposed next steps. The results may suggest sharpening a proposed hypothesis, or there may be obvious data gaps that can now be pursued more clearly so that each account can be defined more clearly, or one account may be more clearly supported by the data than other accounts. An advantage of the HBWoE approach is that it can help identify research that would be most able to inform outstanding questions and resolve ambiguous interpretations.

In this article, we first describe an overview of the HBWoE evaluation of formaldehyde and leukemogenesis by describing the various accounts that must be considered before concluding whether a possible causal association exists between formaldehyde exposure and leukemogenesis. We then describe the details of our analysis for each of the lines of evidence (epidemiology, toxicology, pharmacokinetic, and mode of action) that form the bases of these accounts, individually and in terms of how each inform each other.

3. Overview of HBWoE as applied to formaldehyde and leukemogenesis

The HBWoE evaluation for human leukemogenesis from inhaled formaldehyde comes down to evaluating the comparative degree to which each of the alternative accounts is supported by reference to scientific evidence. In short, one is faced with a contradiction between the apparent (though not certainly causal) association of leukemia with formaldehyde exposure in at least some human studies and the apparent implausibility of such a causal effect in view of current biological understanding. The apparent contradiction can be reconciled in one of two ways: (1) by accepting that human risks are actually increased and positing that the biological impossibility of such increases is somehow mistaken—that is, since the effect appears, it must have a possible causal explanation; or (2) by concluding that doubts about possible mechanisms have merit, and the apparent association of formaldehyde and leukemia seen in some human studies does not in fact indicate a causal connection (and that those studies showing lack of effect are indeed the ones to be taken at face value)—that is, the appearance of some apparent associations is in fact accounted for by chance or by shortcomings in the ostensibly positive human studies, which, according to this view, should be deemed false-positive results.

In pursuit of the first account that suggests a causal mechanism must exist between formaldehyde exposure

and leukemia because their effects are seen, several candidate causal mechanisms have been hypothesized (Zhang et al., 2009, 2010a). As these mechanisms are evaluated, it is important to consider their ad hoc nature; rather than being suggested a priori because of plausibly relevant observed properties, they are constructed after the fact specifically to propose a remedy to the fatal shortcoming of impossibility. Furthermore, they are constrained by the need to offer a possible causal connection between leukemia and formaldehyde inhalation without producing observable effects that contradict currently accepted knowledge and observations. This ad hoc nature does not make the hypothesized mechanisms false, but it does put a premium on finding some independent, positive evidence of their operation and role rather than simply relying on their ability, if true, to furnish the needed mechanisms or apparent consistencies with observations, since they were chosen in part as support of these observations and proposed mechanisms.

An alternative, and contrasting, account is that it is not possible for formaldehyde to move beyond the nasal respiratory mucosa to cause systemic DNA damage and cellular transformation (in the bone marrow, circulating hematopoietic stem cells, or the NALT), and therefore there is no biologically plausible mechanism for formaldehyde leukemogenesis. This account is supported by a large body of hematotoxicity studies (in animals and humans); toxicokinetic, genotoxocity, and mechanistic data in animals, humans, and in vitro; and a large body of null epidemiology findings. Under this account, the significant number of null epidemiology findings are considered true results, and the few positive findings in the epidemiology studies (which have shortcomings, including poor disease classification and poor estimates of exposure), are likely attributable to confounding by other exposures or to chance. If this account is true, an association between inhalation of formaldehyde and leukemia would be understood as not plausible for humans.

Our HBWoE evaluation compares these two accounts by first describing what is known and what has been interpreted from the formaldehyde epidemiology, toxicology, and mode-of-action data, pointing out questions that arise from within and across these studies and their interpretation, the answers to (or at least discussions of) which provide the bases for tracing the logic for each alternative hypothesis.

4. Weight of epidemiology evidence regarding the association between formaldehyde exposure and leukemia

To conduct the HBWoE analysis of the epidemiology data regarding the association between formaldehyde exposure and leukemia, we first conducted a literature search, using PubMed and TOXLINE, for all human studies measuring or estimating formaldehyde exposure and the incidence of or mortality from any lymphohematopoietic cancer. Search terms included "leukemia," "lymphoma," "Hodgkin," "non-Hodgkin," "hematologic neoplasm," "myeloma," "hematopoietic," "lymphatic," "formaldehyde," "epidemiol*," "occupation*," "cohort*," and "worker*." We also relied on the reference lists of several review articles and meta-analyses (e.g., Bachand et al., 2010; Zhang et al., 2010a; Bosetti et al., 2008; Collins and Lineker, 2004). We critically reviewed each relevant study and focused particularly on two cohorts that have received much recent attention: the National Cancer Institute (NCI) industrial worker and embalmer cohorts. The former was analyzed in several studies using traditional cohort study designs, whereas individuals were drawn from the latter to conduct case-control analyses.

After providing a brief overview of the epidemiology literature below, we describe an endpoint-by-endpoint analysis of each lymphohematopoietic cancer and groups of cancers that have been investigated. This is followed by an HBWoE evaluation of the epidemiology evidence with respect to the hypothesis that formaldehyde causes leukemia.

4.1. Overview of epidemiology investigations

Several cohort and case-control studies have been conducted on formaldehyde exposure and lymphohematopoietic cancers (Tables 1 and 2). The first study published was of pathologists and medical laboratory technicians in the United Kingdom (UK) who were followed through 1973 (Harrington and Shannon, 1975). Since that time, studies of embalmers, undertakers, funeral directors, radiologists, pathologists, anatomists, leather tannery workers, iron foundry workers, plastics manufacturing workers, wood industry workers, garment workers, pest-control workers, and workers at formaldehyde production or usage plants have been conducted in the United States, the UK, France, Sweden, Italy, Denmark, Finland, and Canada. Cohort studies ranged in size from 154 to 126,347 subjects with follow-up beginning as early as 1925 and up through 2004. Among the eight case-control studies we identified, the largest included 1511 cases, and follow-up periods among the studies ranged from 1940 to 2000 (Table 2). Formaldehyde exposure was rarely measured in any study and, when it was, concentration information was not available for the entire period of employment. Owing to the limited concentration data, exposure was typically estimated based on job descriptions. Formaldehyde risks were then calculated based on the date of hire/first exposure, minimum employment duration, duration of employment/exposure, time since first exposure, cumulative exposure, average exposure, average intensity of exposure, peak exposure, and number of peak exposures. Health outcomes were coded according to the International Classification of Diseases (ICD) 7th, 8th, or 9th revision (Table 3). Because the majority were coded using the 8th revision (ICD-8) and there are few differences between the 8th and 9th revisions, classifications in the following sections and the tables refer to the 8th revision unless otherwise noted. The health outcomes assessed included mortality from

© 2011 Informa Healthcare USA, Inc.

Table 1. Formaldehyde cohort studies.

Reference	Study population	Subjects (n)	Job/Exposure Category	Period of Employment	Period of Follow-up	Total Follow-Up (person-years)	Minimum Employment (years)	Mean Time-Weighted Average Exposure (ppm)	Peak Exposure (ppm)	Cumulative Number of Peaks ≥4.0 ppm
Harrington and Shannon, 1975	UK Pathologists and medical laboratory technicians	156 154	Pathologists Medical laboratory technicians	1955–1973 1963–1973	1955–1973 1963–1973	24,119.7 73,025.6				
Walrath and Fraumeni, 1983	New York State embalmers	1,132	Embalmers (length of time from first license to death was used to approximate exposure)	1902–1980	1925–1980					
Wong et al., 1983	United States Formaldehyde plant workers	2,026	White male chemical workers	1940–1977	1940–1977	32,514.3				
Levine et al., 1984	Ontario, Canada undertakers	1,477	Undertakers exposed to formaldehyde	1928–1957	1950–1977	34,774				
Walrath and Fraumeni, 1984	California embalmers	1,007	Embalmers (length of time from first license to death was used to approximate exposure)	1916–1978	1925–1980					
Bertazzi et al., 1986, 1989	Italian male resin producers	1,332	Workers exposed to formaldehyde, exposed to other compounds or exposure unknown	1959–1980	1959–1986	5,731	≥1 month			
Logue et al., 1986	US radiologists and pathologists	785 455	Radiologists Pathologists (based on entrance into professional society)	1962–1977	1962–1977					
Stroup et al., 1986	US anatomists	2,317	Anatomists	1889–1969	1925–1979					
Edling et al., 1987	Swedish abrasive manufacturing workers	521	Abrasives industry workers	1958–1981	1958–1981		≥5			
Robinson et al., 1987	US plywood mill workers	2,283	Plywood mill workers	1945–1955	1945–1977	57,588	≥1			
Stern et al., 1987	Minnesota and Wisconsin leather tannery workers	9,365	Tannery A Tannery B Department (finishing 0.5–7 ppm formaldehyde)	1940–1979	1940–1982					
Matanoski et al., 1991	US pathologists	6,411	Pathologists	1912–1950	1925–1978					
Hayes et al., 1990	US embalmers and funeral directors	4,046	Embalmers and funeral directors exposed to formaldehyde (measured average 0.98–3.99 ppm and peak 20 ppm)	NR	1975–1985					
Hall et al., 1991	UK pathologists	3,872	Pathologists	1974–1987	1974–1987					

Table 1. continued on next page

© 2011 Informa Healthcare USA, Inc.

Table 1. continued.

Reference	Study population	Subjects (n)	Job/Exposure Category	Period of Employment	Period of Follow-up	Total Follow-Up (person-years)	Minimum Employment (years)	Mean Time-Weighted Average Exposure (ppm)	Peak Exposure (ppm)	Cumulative Number of Peaks ≥4.0 ppm
Andjelkovich et al., 1995	US iron foundry workers	3,929	Iron foundry workers (formaldehyde exposed or unexposed)	1960–1987	1960–1989	83,064	≥6 months	Low 0.05 Medium 0.55 High 1.5		
Dell and Teta, 1995	New Jersey workers at plastics manufacturing and R&D facility	5,932	Hourly and salaried employees	1946–1967	1946–1988		≥7 months			
Hansen and Olsen, 1995	Denmark formaldehyde male workers	126,347	Working for company making or importing formaldehyde at least 10 years before diagnosis	1970–1984	1970–1984					
Chiazze et al., 1997	South Carolina fiberglass workers	4,631	Cumulative exposure to formaldehyde	1951–1991	1951–1991	73,259				
Stellman et al., 1998	US wood industry workers	45,399	Woord workers Wood dust exposed workers (asbestos and formaldehyde exposure)	1982–1988	1982–1988	2,101,145				
Marsh et al., 2001	US fiberglass workers	32,110	Workers exposed to formaldehyde in ten fiberglass plants	1945–1978	1946–1992	209,726	≥1			
Coggon et al., 2003	UK factory workers where formaldehyde was used or produced	14,014	Formaldehyde production workers	1941–1989	1941–2000			<0.1 0.1–0.5 0.6–2.0 >2.0		
Pinkerton et al., 2004	Georgia and Pennsylvania garment workers	11,039	Garment workers	1955–1982	1955–1998		≥3 months			
Ambroise et al., 2005	French pest-control workers	181	Pest-control workers (ever employed)	1979–2000	1979–2000	3107				
Beane Freeman et al., 2009 (update of Hauptmann et al., 2003)	US workers at formaldehyde production or usage plants	25,619	Formaldehyde production workers (exposed or unexposed)	1934–1966	1934–2004	998,106			0 0.1–1.9 2.0–3.9≥4.0	Data not shown

Table 1. continued on next page

Table 1. continued.

Reference	Average intensity (ppm)	Cumulative exposure (ppm)	Duration of exposure or length of employment (years)	Time since first exposure (years)	Latency period (years)	Hire date or year of first exposure (year)	ICD code	Observed (total)	Expected (total)	Possible co-exposures discussed
Harrington and Shannon, 1975							ICD-8, 200–209 201 204–207	8/3 1/01/1	4.0/5.5 0.7/1.6 1.6/2.2	
Walrath and Fraumeni, 1983					<35 ≥35		ICD-8, 200–209 200 201 202, 203 204–207	25 5 2 6 12	20.6 4.7 2.3 4.9 8.5	Embalming fluids that contain other chemicals (e.g., tissue moisturizers, antiseptic solutions, dyes, and deodorizers)
Wong et al., 1983			<5 5–9 10–14 15–19 20+		10 20	Before 1961 After 1961	ICD-8, 200–209 201 204–207	6 2 2	4.42 0.83 1.70	Formaldehyde, oxygenated hydrocarbons, benzene, asbestos, pigments
Levine et al., 1984							ICD-8, 200–209 204–207	8 4	6.5 2.5	Methanol, phenol, and dyes
Walrath and Fraumeni, 1984			<20 ≥20				ICD-8, 200–209 200 201 202, 203, 208, 209 204–207	19 3 0 4 12	15.6 3.1 2.5 3.0 6.9	Embalming fluids containing coloring and modifying agents, anticoagulants, surfactants, deodorants, and vehicles
Bertazzi et al., 1986, 1989							ICD-8, 200–209	3	1.11	Styrene, polystyrene
Logue et al., 1986						Before 1962 After 1962	ICD-7, 200–203, 205 204	Not reported	Not reported	Radiation
Stroup et al., 1986							ICD-8, 200–209 200 201 204–207 202–203, 208–209	18 2 0 10 6	14.6 2.9 1.9 6.8 3.0	Solvents, methyl alcohol, phenol, and biological agents
Edling et al., 1987							ICD-8, 200–202 203	2 2	1.0 0.5	Aluminum oxide, silicon carbide, clay, phenol, silica, total dust
Robinson et al., 1987			<20 ≥20		<20 ≥20		ICD-7, 200 201 204 202, 205	4 2 3 3	3.9 1.8 0.9 1.1	Wood dust, pentachlorophenol, carbon disulfide, and volatiles
Stern et al., 1987			<1 1–9 ≥10		≥15		ICD-7, 200–205 204 200–203, 205	8/14 4/6 4/8	12.3/19.4 5.2/8.0 7.0/11.4	Cu, Cr, Mn, Co, n-butyl acetate, MEK, MIK, toluene, xylene, acetone, dust, and butyl cellosolve

Table 1. continued on next page

© 2011 Informa Healthcare USA, Inc.

Table 1. continued.

Reference	Average intensity (ppm)	Cumulative exposure (ppm)	Duration of exposure or length of employment (years)	Time since first exposure (years)	Latency period (years)	Hire date or year of first exposure (year)	ICD code	Observed (total)	Expected (total)	Possible co-exposures discussed
Matanoski et al., 1991							ICD-8, 200–209	115	82.7	Phenol, methyl alcohol, glutaraldehyde, and biologic materials, and in the past were exposed to mercury, arsenic, and zinc
							201204–207	3	4.2	
								34	27.1	
								12	10.7	
								20	14.6	
								22	16.3	
								7	9.4	
								24	15.3	
								20	8.8	
								3	0.8	
								4	2.6	
Hayes et al., 1990							ICD-8, 200–209	1104	0.83	
							201		6.93	
							200, 202		2.63	
							200			
							203			
							202			
							204			
							205			
							206, 207			
							208			
							209			
Hall et al., 1991							ICD-8, 201	57	45.6	Other chemicals and infectious agents
							200–209	2	5.6	
							204–207	31	23.0	
Andjelkovich et al., 1995							ICD-8, 200–209	7	12.0	Silica, PAHs, nickel, and chromium
							200	1	1.8	
							201	1	1.4	
							204–207	2	4.6	
Dell and Teta, 1995			<5		0		ICD-7, 200–205	23	13.63	Asbestos, carbon black, epichlorohydrin, formaldehyde, polyvinyl chloride (PVC), acrylonitrile, styrene, and numerous chemical additives, such, as plasticizers, emulsifiers, and antioxidants
			5–9		10		200	3	2.39	
			10–19		15		204	11	5.56	
			≥20				204.4	8	4.30	
Hansen and Olsen, 1995							ICD-7, 200, 202	32	27.2	Wood dust, other chemicals
							201	12	12.2	
							204	39	47.0	
Chiazze et al., 1997							ICD-7, 200–205	51	10.8	Respirable glass fibers, total particulate, asbestos, refractory ceramic fibers, respirable silica, total chrome, and arsenic
							204		54.11	

Table 1. continued on next page

Table 1. continued.

Reference	Average intensity (ppm)	Cumulative exposure (ppm)	Duration of exposure or length of employment (years)	Time since first exposure (years)	Latency period (years)	Hire date or year of first exposure (year)	ICD code	Observed (total)	Expected (total)	Possible co-exposures discussed
Stellman et al., 1998			<10 10–19 ≥20				ICD-9, 200–208 200, 202 203 204–208	28 11 4 12	NR	Wood dust and asbestos
Marsh et al., 2001							ICD-8, 200–209	199	NR	Fiberglass fibers, arsenic, asbestos, asphalt, epoxy, phenolics, silica, styrene, and urea
Coggon et al., 2003			<1 1–14 ≥15				ICD-9, 201 200, 202, 202.1, 202.8 203 204–208	6 31 15 31	8.5 31.7 17.5 34.1	Asbestos, styrene, ethylene oxide, epichlorhydrin, solvents, chromium, and cadmium
Pinkerton et al., 2004			<3 3–9 ≥10	<1010–19≥20		<19631963–1970≥1971	ICD-9, 200–208 200 201 204–208 202–203	59 5 2 24 28	60.8 5.9 3.6 22.0 28.9	
Ambroise et al., 2005		Four quartiles of exposure					ICD-9, 204–208	1	0.23	Ethylene oxide, insecticides, and rodenticides (over 60 chemicals)
Beane Freeman et al., 2009 (update of Hauptmann et al., 2003)	0 >0–<0.5 0.5–<1.0 ≥1.0	0 >0–<1.5 1.5–<5.5 ≥5.5	Data not shown	0 >0–15 >15–25 >25–35 >35	0 >0–25 >25–42 >42		ICD-8, 200–209 200, 202 201 203 204–207 204 205	286 94 25 48 116 36 44	304.3 110.6 17.6 51.1 113.7 31.3 48.9	Antioxidants, asbestos, benzene, carbon black, dyesand pigments, hexamethylenetetramine, melamine, phenol,plasticizers, urea, and wood dust

Note: NR = not reported. See Table 3 for ICD codes.

lymphohematopoietic cancer (ICD 200–209), cancer of lymphoid origin (ICD 200–204), leukemia (ICD 204–207), hematopoietic cancer of non-lymphoid origin (ICD 205, 206, 208, 209), lymphatic leukemia (ICD 204), myeloid leukemia (ICD 205), other unspecified leukemia (ICD 207), Hodgkin's lymphoma (ICD 201), non-Hodgkin's lymphoma (ICD 200, 202), and multiple myeloma (ICD 203). The majority of studies were subject to confounding by several co-exposures, many of which were not accounted for in statistical analyses.

Several individuals and/or cohorts were analyzed in more than one study. Beane Freeman et al. (2009) conducted the most recent study of the NCI industrial worker cohort, with follow-up through 2004. This cohort was first studied by Blair et al. (1986), who followed workers employed in 10 formaldehyde-producing or -using facilities through 1979. Hauptmann et al. (2003) conducted a follow-up through 1994, although it was noted by Beane Freeman et al. (2009) that 1006 deaths were omitted unintentionally from this analyses (all results presented here are from a reanalysis by Beane Freeman et al. [2009], which included these deaths). To avoid counting information on this cohort more than once, only data from the most recent publication by Beane Freeman et al. (2009) are shown in the tables, but results from the previous studies of this cohort are discussed in the text if they are not consistent with the latest analysis.

Coggon et al. (2003) evaluated a cohort of 14,014 UK workers at factories where formaldehyde was used or produced that had been evaluated previously by Acheson et al. (1984) and Gardener et al. (1993). Acheson et al. (1984) evaluated mortality in 7680 men first employed before 1965 in one of six factories, with follow-up through 1981. Gardener et al. (1993) extended the follow-up of

Table 2. Formaldehyde case control studies.

Reference	Study Population	Job/Exposure Category	Period of Employment	Period of Follow-up	Total Follow-up (person-years)	Minimum Employment (years)	Mean Time-Weighted Average Exposure (ppm)	Peak Exposure (ppm)	Average Intensity (ppm)
Gerin et al., 1989	Canadian population in Montreal	Lifetime job histories obtained by interview and translated into level of exposureto formaldehyde	1979–1985	1979–1985					Low Medium High
Ott et al., 1989	US Union Carbide chemical manufacturing facilities	111 work areas, 21 specific chemicals and 52 chemical-activity groups	1940–1978	1940–1978		≥1 day			
Linos et al., 1990	Iowa and Minnesota Funeral home workers	Funeral service and crematoria workers	NR	NR					
Partanen et al., 1993	Finland Wood production workers	Wood workers (formaldehyde, solvents, wood dust)	1957–1982	1957–1982		≥1			
Tatham et al., 1997	Atlanta, Connecticut, Iowa, Kansas, Miami, San Francisco, Detroit, and Seattle workers	Exposed to formaldehyde or other chemicals	1984–1988	1984–1988		≥1			
Blair et al., 2001	Iowa and Minnesota Industrial workers	15 different industrial and occupational job categories (non-farming)	1980–1983	1980–1983		≥1			
Wang et al., 200b	Connecticut women	Exposure to organic solvents and formaldehyde	1996–2000	1996–2000					
Hauptmann et al., 2009	US Embalmers	Never Embalming Ever Embalming	1932–1986	1960–1986	19,104		0 >0–0.10 >0.10–0.18 >0.18	0 >0–7.0 >7.0–9.3 >9.3	0 >0–1.4 >1.4–1.9 >1.9

Table 2. continued on next page

© 2011 Informa Healthcare USA, Inc.

Table 2. continued.

Reference	Cumulative Exposure (ppm)	Duration of Exposure or Length of Employment (years)	Number of Embalmings	ICD Code	Cases (total)	Controls (total)	Possible Co-Exposures Discussed
Gerin et al., 1989		<10 ≥10		ICD-8, 200, 202 201	206 53	533	
Ott et al., 1989		0 <5 ≥5		Non-Hodgkin's Multiple Myelpoma Leukema	2 1 3	NR	52 chemical groups (e.g., epoxides, halogenated compounds, fused cyclics, nitriles, vinylics)
Linos et al., 1990				Leukemia Non-Hodgkin's	578 622	1245	NR
Partanen et al., 1993				ICD-7, 200–202 201 200, 202 204	5 1 4 2	152	Wood dust, pesticides, chlorophenols, phenol, terpenes, solvents (stains, lacquers, toluene, xylenes, benzene, styrene, butyl acetate, ethyl acetate, butanol, isopropanol, ethanol), aliphatic hydrocarbons (solvent naphtha, white spirits), ketones, glycol ethers, and engine exhaust
Tatham et al., 1997		<10 ≥10		ICD-8, 200, 202	1511	1659	Pesticides, herbicides, wood/saw dust, solvents, shoe/leather dust, meat packaging or processing, metal plating, cutting oils, chlorophenols, heterocyclic nitrogens, carbamates, organophosphates, phenoxy herbicides, chlorinated hydrocarbons, and pyrethroids
Blair et al., 2001	Low High	<10 ≥10		ICD-8, 204–207 205.0 205.1 204.0 204.1	64 14 8 0 30	137	Solvents, paints, metals, solder
Wang et al., 200b	Never Low Medium–High			ICD-9, 200–202	601	717	Organic solvents (benzene, chloroform, carbon tetrachloride, dichloromethane, methyl chloride, trichloroethylene)
Hauptmann et al., 2009	0 >0–4058 >4058–9253 >9253(ppm-h)	0 >0–20 >20–34 >34	0 >0–1422 >1422–3068 >3068	ICD-8, 200–209 200–204 205, 206, 208, 209 205	168 99 48 34	265	Isopropanol, ethylene glycol, methanol, phenol, glutaraldehyde, ionizing radiation, benzene, and cigarette smoking

Note: NR = not reported; AML = acute myeloid leukemia; CML = chronic myeloid leukemia; ALL = acute lymphoid leukemia; CLL = chronic lymphoid leukemia. See Table 3 for ICD codes.

7660 of these workers through 1989, and began following 6357 additional workers who began work after 1964. Coggon et al. (2003) then followed the majority of these workers through 2000. Because results are consistent among the three analyses, only results from Coggon et al. (2003) are discussed here.

Hauptmann et al. (2009) conducted a case-control study based on over 6000 embalmers (NCI embalmers cohort) who died between 1960 and 1985 and were included in proportionate mortality ratio (PMR) studies by Hayes et al. (1990) and Walrath and Fraumeni (1983, 1984). Walrath and Fraumeni (1983) studied embalmers licensed in California, Walrath and Fraumeni (1984) studied those licensed in New York, and Hayes et al. (1990) assembled data on US embalmers and funeral directors who died between 1975 and 1985. In the tables, we present data from both Hauptmann et al. (2009) and Hayes et al. (1990) because they use different methodologies. Data from Walrath and Fraumeni (1983, 1984) are discussed in the text but not the tables, because study subjects are included in the Hayes et al. (1990) analysis and were analyzed in a similar fashion.

Table 3. International disease classification (ICD) codes.

ICD Code	Revision 7	Revision 8	Revision 9
	(200–207) Neoplasms of lymphatic and hematopoietic tissues	(200–209) Neoplasms of lymphatic and hematopoietic tissue	(200–208) Malignant neoplasms of lymphatic and hematopoietic tissue
200	Lymphosarcoma and reticulosarcoma	Lymphosarcoma and reticulum-cell sarcoma	Lymphosarcoma and reticulosarcoma and other specified malignant tumors of lymphatic tissue
201	Hodgkin's disease	Hodgkin's disease	Hodgkin's disease
202	Other forms of lymphoma (reticulosis)	Other neoplasms of lymphoid tissue	Other malignant neoplasms of lymphoid and histiocytic tissue
203	Multiple myeloma	Multiple myeloma	Multiple myeloma and immunoproliferative neoplasms
204	Leukemia & aleukemia	Lymphatic leukemia	Lymphoid leukemia
204.0	Lymphatic leukemia	Acute lymphocytic leukemia	Acute lymphoid leukemia
204.1	Myeloid leukemia	Chronic lymphocytic leukemia	Chronic lymphoid leukemia
204.3	Acute leukemia	—	—
204.4	Other & unspecified leukemia	—	—
205	Mycosis fungoides	Myeloid leukemia	Myeloid leukemia
205.0	—	Acute myeloid leukemia	Acute myeloid leukemia
205.1	—	Chronic myeloid leukemia	Chronic myeloid leukemia
206	Lymphatic system	Monocytic leukemia	Monocytic leukemia
207	Hematopoietic system	Other and unspecified leukemia	Other specified leukemia
208	—	Polycythemia vera	Leukemia of unspecified cell type
209	—	Myelofibrosis	—
238.4	—	—	Polycythemia vera
289.83	—	—	Myelofibrosis
294	Polycythemia	—	—

4.2. Endpoint-by-endpoint analysis

In this section, we discuss each of the individual lymphohematopoietic cancer endpoints analyzed in the epidemiology studies described above. Lymphohematopoietic cancers include a group of hematopoietic and lymphoid cell disorders that have distinct classifications based on morphologic, cytogenic, immunophenotypic, and molecular characteristics (see Vardiman, 2010, for a review of the classifications). We consider various groupings of cancer types as analyzed by study authors, although results from these analyses must be considered carefully because each specific lymphohematopoietic cancer is a different disease. Although some cancer types may have some common mechanisms (e.g., pharmacokinetics), in general, lymphohematopoietic cancers each have a distinct etiology, so an association with one type is not necessarily indicative of risk of another (Schottenfeld and Fraumeni, 2006). That is, if one study reports a statistically significant finding for one cancer type (A) but not another (B), and another study reports a statistically significant finding for cancer type B but not A, this is not consistent evidence of an association. In the same vein, an association between formaldehyde and a group of cancers does not necessarily provide evidence for all cancers in that group, as it may be driven by one cancer type with a distinct mode of action. Thus, it is crucial in a weight-of-evidence analysis to consider each individual cancer type and the implications of analyses of cancer groups.

For each cancer or group of cancers, we evaluated the weight of each study based on several factors, including the study objectives and hypothesis; the study subjects; the exposure and health outcome assessments; the follow-up period; the consideration of bias, confounders, and effect modifiers; the statistical methods; the documentation and interpretation of results; and the external validity (i.e., the bearing on the larger question at hand, formaldehyde as a potential cause of human lymphohematopoietic neoplasms). For each cancer or group of cancers, we also assessed the consistency of findings (which included consideration of the type of exposure metric, e.g., peak vs. cumulative) and whether any exposure-response relationships were evident.

4.2.1. All lymphohematopoietic cancers

The association between formaldehyde exposure and all lymphohematopoietic cancers combined has been investigated in 12 studies (Table 4). Eleven cohort and one case-control study assessed whether study subjects had an increased risk over the general population. Of these, only one reported associations (Hayes et al., 1990). Hayes et al. (1990) found an increased proportion of deaths attributable to lymphohematopoietic cancers among embalmers in the NCI embalmers cohort (PMR = 1.39, 95% confidence interval [CI]: 1.15–1.67).

Lymphohematopoietic cancer risks were also evaluated based on one or more exposure metrics in iron foundry workers, embalmers, and industrial workers. Risks were not increased in formaldehyde-exposed and unexposed US iron foundry workers (Andjelkovich et al., 1995), and risks reported in embalmers and industrial workers were not consistent across exposure metrics (Hauptmann et al., 2009; Beane Freeman et al., 2009).

© 2011 Informa Healthcare USA, Inc.

Table 4a. Association between formaldehyde and all lymphohematopoietic cancers (ICD 200–209).

	Hauptmann et al., 2009					Beane Freeman et al., 2009					Andjelkovich et al., 1995				
	Embalmers case-control					NCI cohort (1934–2004)					Iron foundry workers				
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI
Unexposed/Exposed	Never embalming	24	OR	1.00	—	Unexposed	33	SMR	0.86	0.61–1.21	Unexposed	8	SMR	0.89	0.38–1.76
	Ever embalming	144	OR	1.40	0.80-2.60	Exposed	286	SMR	0.94	0.84–1.06	Exposed	7	SMR	0.59	0.23–1.21
Peak Exposure	0 ppm	24	OR	1.00	—	0 ppm	33	RR	1.07	0.70–1.62					
	>0–<2.0 ppm	48	OR	1.60	0.80–3.20	>0–<2.0 ppm	103	RR	1.00	—					
	2.0–<4.0 ppm	55	OR	1.60	0.90–3.10	2.0–<4.0 ppm	75	RR	1.17	0.86–1.59					
	≥4.0 ppm	41	OR	1.20	0.60–2.30	≥4.0 ppm	108	RR	1.37	1.03–1.81					
	p_{trend} = .302 (exposed)					p_{trend} = .02 (exposed)									
	p_{trend} = .555 (exposed and unexposed)					p_{trend} = .04 (exposed and unexposed)									
Average Intensity	0 ppm	24	OR	1.00	—	0 ppm	33	RR	0.99	0.66–1.48					
	0.1–0.4 ppm	53	OR	1.60	0.90–3.20	>0–<0.5 ppm	164	RR	1.00	—					
	0.5–0.9 ppm	47	OR	1.40	0.70–2.70	0.5–<1.0 ppm	67	RR	1.29	0.97–1.73					
	≥1.0 ppm	44	OR	1.30	0.70–2.50	≥1.0 ppm	55	RR	1.07	0.78–1.47					
	p_{trend} = .443 (exposed)					p_{trend} > .5 (exposed)									
	p_{trend} = .591 (exposed and unexposed)					p_{trend} > .5 (exposed and unexposed)									
Cumulative Exposure	0 ppm-yr	24	OR	1.00	—	0 ppm-yr	33	RR	0.89	0.59–1.34					
	>0–<1.5 ppm-yr	40	OR	1.30	0.60–2.50	>0–<1.5 ppm-yr	168	RR	1.00	—					
	>1.5–<5.5 ppm-yr	49	OR	1.40	0.80–2.80	>1.5–<5.5 ppm-yr	49	RR	0.77	0.56–1.07					
	≥5.5 ppm-yr	55	OR	1.60	0.80–3.00	≥5.5 ppm-yr	69	RR	1.07	0.8–1.42					
	p_{trend} = .753 (exposed)					p_{trend} = .25 (exposed)									
	p_{trend} = .422 (exposed and unexposed)					p_{trend} = .25 (exposed and unexposed)									
Cumulative number of peaks ≥4.0 ppm						No association. Results not shown.									
Exposure/Employment Duration	0 yrs	24	OR	1.00	—	No association. Results not shown.									
	>0–20 yrs	28	OR	0.80	0.40–1.80										
	>20–34 yrs	50	OR	1.50	0.80–2.80										
	>34 yrs	66	OR	1.80	1.00–3.40										
	p_{trend} = .131 (exposed)														
	p_{trend} = .058 (exposed and unexposed)														
Number of Embalmings	0	24	OR	1.00	—										
	>0–1422	29	OR	0.90	0.60–1.80										
	>1422–3068	62	OR	1.90	1.00–3.60										
	>3068	53	OR	1.50	0.80–2.90										
	p_{trend} = .477 (exposed)														
	p_{trend} = .844 (exposed and unexposed)														

Table 4a. continued on next page

© 2011 Informa Healthcare USA, Inc.

Table 4a. continued.

	Hauptmann et al., 2009					Beane Freeman et al., 2009				Andjelkovich et al., 1995			
	Embalmers case-control					NCI cohort (1934–2004)				Iron foundry workers			
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate	95% CI	Category	Obs	Estimate	95% CI
8-Hour Time-Weighted Average Intensity	0	24	OR	1.00	—								
	>0–0.10	47	OR	1.30	0.70–2.60								
	>0.10–0.18	52	OR	1.60	0.80–3.10								
	>0.18	45	OR	1.40	0.70–2.80								
	p_{trend} = .635 (exposed)												
	p_{trend} = .855 (exposed and unexposed)												
Time Since First Exposure						0 yrs	30	RR	0.67	0.31-1.46			
						>0-15 yrs	21	RR	1.00	-			
						>15-25 yrs	46	RR	1.30	0.68-2.49			
						>25-35 yrs	59	RR	0.82	0.40-1.70			
						>35 yrs	163	RR	0.67	0.32-1.41			
Time Since First Exposure ≥4 ppm						0 yrs	211	RR	0.57	0.36-0.88			
						>0-25 yrs	28	RR	1.00	-			
						>25-42 yrs	45	RR	0.69	0.41-1.17			
						>42 yrs	35	RR	0.61	0.34-1.09			

Table 4b. Other cohorts.

Reference	Obs	Estimate		95% CI
Wong et al., 1983	6	SMR	1.36	0.50–2.95
Levine et al., 1984	8	SMR	1.24	—
Hayes et al., 1990	115	PMR	1.39	1.15–1.67
Hall et al., 1991	9 (M)	SMR	1.42	0.65–2.69
Hall et al., 1991	1 (F)	SMR	1.75	0.04–9.77
Matanoski et al., 1991	57	SMR	1.25	0.95–1.62
Bertazzi et al., 1986, 1989	3	SMR	1.73	0.36–5.06
Stellman et al., 1998*	28	RR	1.22	0.84–1.77
Marsh et al., 2001	199	SMR	0.90	0.78–1.04
Pinkerton et al., 2004*	59	SMR	0.97	0.74–1.26

*ICD-8 200–208.

Hauptmann et al. (2009) conducted a case-control study of 168 embalmers (21 with leukemia) from the NCI embalmers cohort (evaluated by Hayes et al., 1990) and examined lymphohematopoietic cancer risks based on seven exposure metrics: exposed (ever/never embalmed), peak exposure, average intensity of exposure when embalming, 8-hour time-weighted average (TWA) exposure, cumulative exposure, exposure duration (years embalming), and number of embalmings. Exposure estimates were developed from a previous exposure-assessment experiment by Stewart et al. (1992). The investigators conducted trend tests for each exposure metric including and excluding unexposed individuals. There were no statistically significant associations between formaldehyde exposure and lymphohematopoietic cancer based on any exposure metric.

Beane Freeman et al. (2009) conducted the most recent study of the NCI industrial worker cohort, with follow-up through 2004. They examined lymphohematopoietic risks based on exposure metrics including exposed (yes/no), peak exposure, number of peak exposures ≥4.0 ppm, duration of exposure, average intensity of exposure, cumulative exposure, years since first exposure, and years since first exposure ≥4 ppm. Beane Freeman et al. (2009) stated that there was no evidence that risks increased with cumulative number of peaks ≥4.0 ppm or for duration of exposure for any lymphohematopoietic cancer evaluated, but they did not present results. An association was observed with the presence of at least one career peak exposure ≥4.0 ppm (risk ratio [RR] = 1.37, 95% CI: 1.03–1.81, $p_{trend} = .02$ based on exposed subjects only and $p_{trend} = .04$ based on all study subjects), but not number of peak exposures ≥4.0 ppm. Risks were also increased with increasing peak intensity with follow-up to 1981 ($p_{trend} = 0.00987$ based on exposed subjects only and $p_{trend} = 0.0485$ based on all study subjects), but not with follow-up from 1981-1994 or 1995-2004. Risks were lower in those with no exposure vs. those with their first exposure to ≥ 4 ppm formaldehyde 0-25 years earlier (RR = 0.57, 95% CI: 0.36-0.88). This was consistent with results of Hauptmann et al. (2003), who followed this cohort through 1994. In their reanalysis of this cohort through 1994, Beane Freeman et al. (2009) found that, of the six exposure metrics, associations were only observed for peak exposure ≥0.04 ppm (RR = 1.48, 95% CI: 1.04–2.12, $p_{trend} = .02$ including or excluding unexposed subjects).

4.2.2. Cancer of lymphoid origin

Risks from cancers of lymphoid origin were examined in four cohorts (Table 5). Both Dell and Teta (1995) and Chiazze et al. (1997) defined cancers of lymphoid origin as those in ICD-7 200–205 categories. Whereas Chiazze et al. (1997) did not report increased risks, Dell and Teta (1995) reported increased risks among plastics manufacturers (standardized mortality rate [SMR] = 1.69, 95% CI: 1.07–2.53). No significant associations were found in the NCI embalmers cohort based on any of the seven exposure metrics evaluated (Hauptmann et al., 2009). Analyses of peak exposure, average intensity, cumulative exposure, cumulative number of peaks ≥4.0 ppm, or duration of employment also did not indicate any associations in the NCI industrial cohort (Beane Freeman et al., 2009).

4.2.3. Leukemia

A large number of investigations have focused on the association between formaldehyde and leukemia (Tables 6, 7, 8, and 9). The types of leukemia investigated vary among studies, and this section focuses on analyses of all leukemia and aleukemias (leukemias in which the circulating white blood cells are normal or decreased in number) combined (ICD-7 204) and lymphatic, myeloid, monocytic, other, and unspecified leukemias combined (ICD-8 204–207 and ICD-9 204–208), whereas later sections discuss assessments of specific types of leukemia. Risk estimates for leukemia among 28 analyses that did not assess exposure-response were generally null (Table 6, table 6C). Only two cohort studies, conducted by Walrath and Fraumeni (1984) and Dell and Teta (1995), reported increased proportions or risks (PMR = 1.5, $p < .05$ and SMR = 2.65, 95% CI: 1.15–5.24, respectively).

There were no increased risks of leukemia in any formaldehyde exposure group among the three studies that assessed exposure-response and, with one exception, no exposure-response associations were reported. Stern et al. (1987) found no association with duration of employment as a leather tannery worker and Pinkerton et al. (2004) found risks in garment workers were not related with duration of exposure, time since first exposure, or year of first exposure (Table 6). Beane Freeman et al. (2009) examined associations with formaldehyde

Table 5a. Association between formaldehyde and cancers of lymphoid origin (ICD 200–204).

	Hauptmann et al., 2009					Beane Freeman et al., 2009				
	Embalmers case-control					NCI cohort (1934-2004)				
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI
Unexposed/Exposed	Never embalming	18	OR	1.00	—					
	Ever embalming	81	OR	1.10	0.50-2.10					
Peak Exposure	0 ppm	18	OR	1.00	—	0 ppm	26	RR	1.17	0.72–1.89
	>0–7.0 ppm	29	OR	1.20	0.60–2.70	>0–<2.0 ppm	73	RR	1.00	—
	>7.0–9.3 ppm	37	OR	1.50	0.70–3.20	2.0–<4.0 ppm	56	RR	1.27	0.89–1.82
	>9.3 ppm	15	OR	0.60	0.20–1.30	≥4.0 ppm	74	RR	1.35	0.97–1.89
	p_{trend} = .111 (exposed)					p_{trend} = .06 (exposed)				
	p_{trend} = .523 (exposed and unexposed)					p_{trend} = .10 (exposed and unexposed)				
Average Intensity	0 ppm	18	OR	1.00	—	0 ppm	26	RR	1.08	0.68–1.71
	>0–1.4 ppm	34	OR	1.40	0.60–2.90	>0–<0.5 ppm	116	RR	1.00	—
	>1.4–1.9 ppm	26	OR	1.00	0.50–2.20	0.5–<1.0 ppm	49	RR	1.36	0.97–1.9
	>1.9 ppm	21	OR	0.90	0.40–1.90	≥1.0 ppm	38	RR	1.05	0.72–1.53
	p_{trend} = .287 (exposed)					p_{trend} > .5 (exposed)				
	p_{trend} = .598 (exposed and unexposed)					p_{trend} > .5 (exposed and unexposed)				
Cumulative Exposure	0 ppm-h	18	OR	1.00	—	0 ppm-yr	26	RR	0.94	0.59–1.49
	>0–4058 ppm-h	23	OR	0.90	0.40–2.00	>0–<1.5 ppm-yr	123	RR	1.00	—
	>4058–9253 ppm-h	33	OR	1.30	0.60–2.80	>1.5–<5.5 ppm-yr	30	RR	0.65	0.44–0.98
	>9253 ppm-h	25	OR	1.00	0.40–2.00	≥5.5 ppm-yr	50	RR	1.06	0.75–1.49
	p_{trend} = .912 (exposed)					p_{trend} > .5 (exposed)				
	p_{trend} = .965 (exposed and unexposed)					p_{trend} > .5 (exposed and unexposed)				
Cumulative number of peaks ≥4.0 ppm						No association. Results not shown.				
Duration of Exposure/ Employment	0 yrs	18	OR	1.00	—	No association. Results not shown.				
	>0–20 yrs	16	OR	0.70	0.30–1.60					
	>20–34 yrs	32	OR	1.20	0.60–2.60					
	>34 yrs	33	OR	1.20	0.60–2.50					
	p_{trend} = .360 (exposed)									
	p_{trend} = .449 (exposed and unexposed)									
Number of Embalmings	0	18	OR	1.00	—					
	>0–1422	17	OR	0.70	0.30–1.60					
	>1422–3068	37	OR	1.50	0.70–3.00					
	>3068	27	OR	1.00	0.50–2.20					
	p_{trend} = .963 (exposed)									
	p_{trend} = .865 (exposed and unexposed)									
8-Hour Time-Weighted Average Intensity	0	18	OR	1.00	—					
	>0–0.10	32	OR	1.20	0.60–2.60					
	>0.10–0.18	25	OR	1.00	0.50–2.10					
	>0.18	24	OR	1.00	0.50–2.10					
	p_{trend} = .766 (exposed)									
	p_{trend} = .605 (exposed and unexposed)									

Table 5b. Other cohorts.

Reference	Obs	Estimate		95% CI
Dell and Teta, 1995*	23	SMR	1.69	1.07–2.53
Chiazze et al., 1997*	5	SMR	0.46	0.15–1.08

Note: *ICD-7 200–205.

in the NCI industrial worker cohort by peak exposure, average intensity, cumulative exposure, cumulative number of peaks ≥4.0 ppm (data not reported), and duration of exposure (data not reported), years since first exposure, and years since first exposure ≥4 ppm, including and excluding a referent group with no exposure. They found no trends except for peak exposure when all exposure groups were included (p_{trend} = .02) but not when the referent group was excluded (p_{trend} = .12). In this cohort, risks were lower in those with no exposure vs. those with their first exposure to ≥4 ppm formaldehyde 0-25 years earlier (RR = 0.34, 95% CI: 0.18-0.67) and also in those whose first exposure to ≥4 ppm formaldehyde was 25-42 years earlier vs. 0-25 years earlier (RR = 0.37, 95% CI: 0.16-0.83). The RR estimates in the NCI industrial worker cohort are similar to those reported in the previous follow-up of this cohort to 1994 (e.g., for peak exposure ≥4.0 ppm, $RR_{through\ 1994}$ = 1.60, 95% CI: 0.90–2.82 vs. $RR_{through\ 2004}$ = 1.42, 95% CI: 0.92–2.18) (Beane Freeman et al., 2009; Hauptman et al., 2003). In this cohort, risks

© 2011 Informa Healthcare USA, Inc.

Table 6a. Association between formaldehyde and leukemia (ICD 204–207).

	Beane Freeman et al., 2009					Pinkerton et al., 2004					Stern et al., 1987				
	NCI cohort (1934–2004)					US garment workers					Leather tannery workers				
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI
Unexposed/exposed	Unexposed	7	SMR	0.48	0.23–1.01	Exposed	24	SMR	1.09	0.70–1.62					
	Exposed	116	SMR	1.02	0.85–1.22										
Peak exposure	0 ppm	7	RR	0.59	0.25–1.36										
	>0–<2.0 ppm	41	RR	1.00	—										
	2.0–<4.0 ppm	27	RR	0.98	0.60–1.62										
	≥4.0 ppm	48	RR	1.42	0.92–2.18										
	$p_{trend} = .12$ (exposed)														
	$p_{trend} = .02$ (exposed and unexposed)														
Average intensity	0 ppm	7	RR	0.54	0.24–1.22										
	>0–<0.5 ppm	67	RR	1.00	—										
	0.5–<1.0 ppm	25	RR	1.13	0.71–1.79										
	≥1.0 ppm	24	RR	1.10	0.68–1.78										
	$p_{trend} > .5$ (exposed)														
	$p_{trend} = .5$ (exposed and unexposed)														
Cumulative exposure	0 ppm-yr	7	RR	0.53	0.23–1.21										
	>0–<1.5 ppm-yr	63	RR	1.00	—										
	>1.5–<5.5 ppm-yr	24	RR	0.96	0.60–1.56										
	≥5.5 ppm-yr	29	RR	1.11	0.70–1.74										
	$p_{trend} = .12$ (exposed)														
	$p_{trend} = .08$ (exposed and unexposed)														
Cumulative number of peaks ≥4.0 ppm	No association. Results not shown.														
Duration of exposure/ employment	No association. Results not shown.					<3 yrs	7	SMR	0.96	—	<1 yr	2	SMR	0.45	0.05–1.68
						3–9 yrs	5	SMR	0.72	—	1–9 yrs	2	SMR	1.00	0.11–3.61
						10+ yrs	12	SMR	1.53	—	10+ yrs	6	SMR	1.70	0.63–3.73
						$p_{trend} > .05$									
Time since first exposure	0 yrs	5	RR	0.28	0.06-1.32	<10 yrs	2	SMR	0.68	—					
	>0-15 yrs	6	RR	1.00	-	<10–19 yrs	3	SMR	0.65	—					
	>15-25 yrs	22	RR	2.13	0.64-7.15	20+ yrs	19	SMR	1.31	—					
	>25-35 yrs	26	RR	0.94	0.25-3.51			$p_{trend} > .05$							
	>35 yrs	64	RR	0.53	0.14-2.09										

Table 6a. continued on next page

Table 6a. continued.

	Beane Freeman et al., 2009					Pinkerton et al., 2004					Stern et al., 1987			
	NCI cohort (1934–2004)					US garment workers					Leather tannery workers			
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate	95% CI
Year of first exposure						<1963	19	SMR	1.23	—				
						1963–1970	4	SMR	0.81	—				
						≥1971	1	SMR	0.56	—				
						$p_{trend} > .05$								
Time Since First Exposure ≥4 ppm	0 yrs	75	RR	0.34	0.18-0.67									
	>0-25 yrs	14	RR	1.00	-									
	>25-42 yrs	16	RR	0.37	0.16-0.83									
	>42 yrs	18	RR	0.51	0.22-1.19									

© 2011 Informa Healthcare USA, Inc.

Table 6b. Other cohorts.

Reference	ICD code	Obs	Estimate		95% CI
Harrington and Shannon, 1975 (pathologists)	204–207	1	SMR	0.63	0.02–3.48
Harrington and Shannon, 1975 (technicians)	204–207	1	SMR	0.45	0.01–2.53
Walrath and Fraumeni, 1983	204–207	12	PMR†	1.40	na
Wong et al., 1983	204–207	2	SMR†	1.18	0.13–4.26
Levine et al., 1984	204–207	4	SMR‡	1.60	0.44–4.10
Walrath and Fraumeni, 1984	204–207	12	PMR†	1.50	$p<.05$
Logue et al., 1986 (radiologists)	204¶	na	SMR	1.55	na
Logue et al., 1986 (pathologists)	204¶	na	SMR	1.06	na
Stroup et al., 1986	204–207	10	SMR	1.50	0.70–2.70
Robinson et al., 1987	204¶	1	SMR¶	0.59	0.01–3.28
Stern et al., 1987 (Plant A)	204¶	4	SMR†	0.70	0.19–1.80
Stern et al., 1987 (Plant B)	204¶	6	SMR†	0.75	0.28–1.64
Stern et al., 1987 (Finishing Department)	204¶	7	SMR†	1.25	0.50–2.58
Ott et al., 1989*	Non-lymphocytic	2	OR	2.6	na
Ott et al., 1989*	Lymphocytic	1	OR	2.60	na
Linos et al., 1990*	na	4	OR	2.10	0.40–10.00
Hall et al., 1991	204–207	4	SMR	2.63	0.41–3.89
Matanoski et al., 1991	na	31	SMR	1.35	0.92–1.92
Partanen et al., 1993	204¶	2	OR	1.40	0.25–7.91
Andjelkovich et al., 1995	204–207	2	SMR†	0.43	0.05–1.57
Dell and Teta, 1995	204¶	8	SMR†	2.65	1.15–5.24
Hansen and Olsen, 1995	204¶	39	SPIR	0.80	0.60–1.60
Chiazze et al., 1997	204¶	1	SMR	0.24	0.006–1.36
Stellman et al., 1998	na	12	RR	0.96	0.54–1.71
Blair et al., 2001*	na	64	OR‡	0.98	0.70–1.36
Coggon et al., 2003	204–208	31	SMR	0.91	0.62–1.29
Ambroise et al., 2005	204–208	1	SMR	4.42	0.11–24.64

Note:
na = not available.
*Case-control study.
†SMR or PMR values divided by 100.
‡Risk estimate not provided in original citation, value calculated by Bachand et al. (2010).
¶ICD-7.

were lower in those with no exposure vs. those with their first exposure to ≥4 ppm formaldehyde 0-25 years earlier (RR = 0.34, 95% CI: 0.18-0.67) and also in those whose first exposure to ≥4 ppm formaldehyde was 25-42 years earlier vs. 0-25 years earlier (RR = 0.37, 95% CI: 0.16-0.83).

4.2.4. Lymphatic leukemia

Results from analyses of lymphatic leukemia (ICD 204) are similar to those reported for all leukemias combined (Tables 6 and 7). Among the four studies that assessed lymphatic leukemia, all risk estimates are null (Table 7, table 7A). Blair et al. (2001) and Pinkerton et al. (2004) reported no association between formaldehyde exposure and mortality from lymphatic leukemia. There were no exposure-response relationships for any of the six exposure metrics evaluated by Beane Freeman et al. (2009) in the NCI industrial cohort. This result was also observed in this cohort with follow-up only through 1994 (Beane Freeman et al., 2009; Hauptmann et al., 2003).

4.2.5. Hematopoietic cancer of non-lymphoid origin

Associations between formaldehyde exposure and hematopoietic cancers of non-lymphoid origin were investigated in several studies (Table 8). In an early analysis of the NCI embalmers cohort, Hayes et al. (1990) found the PMR from polycythaemia vera or myelofibrosis was not higher than expected, but it was from monocytic leukemia, other (i.e., not lymphatic, myeloid, or monocytic), and unspecified leukemias combined (PMR = 2.28, 95% CI: 1.29–3.52). Pinkerton et al. (2004) found no association with monocytic leukemia or leukemia of other or unspecified type among garment workers (SMR = 0.92, 95% CI: 0.34–2.00).

Risks of myeloid leukemia (ICD 205), monocytic leukemia (ICD 206), ploycthaemia vera (ICD 208), and myelofibrosis (ICD 209) combined were examined in recent studies of the NCI industrial worker and embalmer cohorts (Hauptmann et al., 2009; Beane Freeman et al., 2009; Table 8). Beane Freeman et al. (2009) did not report any excess risks in the industrial worker cohort

© 2011 Informa Healthcare USA, Inc.

Table 7a. Association between formaldehyde and lymphatic leukemia (ICD 204).

	Beane Freeman et al., 2009					Blair et al., 2001				
	NCI cohort (1934–2004)					US industrial workers				
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI
						Acute				
Unexposed/Exposed	Unexposed	1	SMR	0.26	0.04–1.82	Low	0	OR	—	—
	Exposed	36	SMR	1.15	0.83–1.59	High	0	OR	—	—
						Chronic				
						Low	29	OR	1.20	0.70–1.80
						High	1	OR	0.60	0.10–5.30
Peak Exposure	0 ppm	1	RR	0.27	0.03–2.13					
	>0–<2.0 ppm	14	RR	1.00	—					
	2.0–<4.0 ppm	8	RR	0.81	0.33–1.96					
	≥4.0 ppm	14	RR	1.15	0.54–2.47					
	$p_{trend} > 0.5$ (exposed)									
	$p_{trend} = 0.3$ (exposed and unexposed)									
Average Intensity	0 ppm	1	RR	0.26	0.03–2.01					
	>0–<0.5 ppm	22	RR	1.00	—					
	0.5–<1.0 ppm	7	RR	0.92	0.39–2.16					
	≥1.0 ppm	6	RR	1.61	0.76–3.39					
	$p_{trend} > 0.5$ (exposed)									
	$p_{trend} > 0.5$ (exposed and unexposed)									
Cumulative Exposure	0 ppm-yr	1	RR	0.24	0.03–1.88					
	>0–<1.5 ppm-yr	21	RR	1.00	—					
	>1.5–<5.5 ppm-yr	5	RR	0.57	0.21–1.54					
	≥5.5 ppm-yr	10	RR	1.02	0.47–2.21					
	$p_{trend} = 0.46$ (exposed)									
	$p_{trend} = 0.41$ (exposed and unexposed)									
Cumulative number of peaks ≥4.0 ppm	No association. Results not shown.									
Duration of Employment	No association. Results not shown.									

Table 7b. Other cohorts.

Reference	ICD Code	Obs	Estimate		95% CI
Hayes et al., 1990	204	7	PMR *	0.7	0.29–1.53
Pinkerton et al., 2004	204	3	SMR	0.60	0.12–1.75

Note: na = not available, *PMR divided by 100.

Table 8a. Association between formaldehyde and cancers of non-lymphoid origin (ICD 205, 206, 208, 209).

	Hauptmann et al., 2009*					Beane Freeman et al., 2009				
	Embalmers case-control					NCI cohort (1934–2004)				
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI
Unexposed/Exposed	Never embalming	4	OR	1.00	—					
	Ever embalming	44	OR	3.00	1.00-9.50					
Peak Exposure	<500 embalmings	9	OR	1.00	—	0 ppm	5	RR	1.01	0.34-2.98
	≤7.0 ppm	10	OR	1.60	0.60–4.50	>0–<2.0 ppm	15	RR	1.00	—
	>7.0–9.3 ppm	12	OR	1.40	0.50–3.70	2.0–<4.0 ppm	11	RR	1.19	0.54–2.62
	>9.3 ppm	17	OR	2.30	0.90–5.60	≥4.0 ppm	21	RR	1.80	0.91–3.57
						$p_{trend} = .09$ (exposed)				
						$p_{trend} = .09$ (exposed and unexposed)				
Average Intensity	<500 embalmings	9	OR	1.00	—	0 ppm	5	RR	0.89	0.32–2.5
	≤1.4 ppm	13	OR	1.70	0.70–4.50	>0–<0.5 ppm	25	RR	1.00	—
	>1.4–1.9 ppm	12	OR	1.70	0.70–4.60	0.5–<1.0 ppm	11	RR	1.40	0.68–2.86
	>1.9 ppm	14	OR	1.80	0.70–4.70	≥1.0 ppm	11	RR	1.51	0.72–3.16
						$p_{trend} > .5$ (exposed)				
						$p_{trend} > .5$ (exposed and unexposed)				
Cumulative Exposure	<500 embalmings	9	OR	1.00	—	0 ppm-yr	5	RR	0.69	0.25–1.95
	≤4058 ppm-h	5	OR	1.10	0.30–3.80	>0–<1.5 ppm-yr	30	RR	1.00	—
	>4058–9253 ppm-h	12	OR	1.40	0.50–3.70	>1.5–<5.5 ppm-yr	7	RR	0.61	0.26–1.41
	>9253 ppm-h	22	OR	2.40	1.00–5.80	≥5.5 ppm-yr	10	RR	0.86	0.41–1.81
						$p_{trend} > .5$ (exposed)				
						$p_{trend} > .5$ (exposed and unexposed)				
Cumulative number of peaks ≥4.0 ppm						No association. Results not shown.				
Duration of Exposue/Employment	<500 embalmings	9	OR	1.00	—	No association. Results not shown.				
	≤20 yrs	2	OR	0.30	0.10–1.70					
	>20–34 yrs	16	OR	2.00	0.80–5.00					
	>34 yrs	21	OR	2.60	1.00–6.40					
Number of Embalmings	<500 embalmings	9	OR	1.00	—					
	≥500–1422	3	OR	0.60	0.20–2.60					
	>1422–3068	15	OR	1.80	0.70–4.60					
	>3068	21	OR	2.30	1.00–5.70					
8-Hour Time-Weighted Average Intensity	<500 embalmings	9	OR	1.00	—					
	≤0.10 ppm	9	OR	1.30	0.50–3.60					
	>0.10–0.18 ppm	16	OR	2.10	0.80–5.30					
	>0.18 ppm	14	OR	1.90	0.70–4.80					

Note: *Results from analyses using those who never embalmed as a referent group (with one myeloid leukemia case) were highly unstable. Results presented here are from analyses using individuals with <500 embalmings as the referent group (see Table 4 in Hauptmann et al., 2009).

Table 8b. Other cohorts.

Reference	Code	Obs	Estimate		95% CI
Pinkerton et al., 2004	206–208	6	SMR	0.92	0.34–2.00
Hayes et al., 1990	206, 207	20	PMR	2.28	1.39–3.52
Hayes et al., 1990	208	3	PMR	3.90	0.80–11.38
Hayes et al., 1990	209	4	PMR	2.62	0.42–3.91

based on analyses by peak exposure, average intensity, cumulative exposure, cumulative number of peaks ≥4.0 ppm, or duration of exposure (Table 8). They also found no exposure-response associations among analyses including or excluding the unexposed population. This is consistent with previous analyses of this cohort (Hauptmann et al., 2003; Blair et al., 1986).

Hautpmann et al. (2009) found that risk estimates from analyses using subjects who never embalmed as a referent category were highly unstable because of the small number of cases in this category ($n = 4$, odds ratio [OR] = 3.0, 95% CI: 1.0–9.5 for ever vs. never embalmed). Still, among six exposure metrics, there were no exposure-response associations reported when unexposed referents (i.e., 0 embalmings) were included or excluded with one exception—there was a trend reported with duration of exposure when the unexposed group was excluded ($p_{trend} = .046$) but not when it was included ($p_{trend} = .348$). Because of the issues with the aforementioned analyses, Hauptmann et al. (2009) also conducted analyses using those who performed <500 embalmings as a referent category. Results from these analyses, which they suggest are more reliable, are presented in Table 8. The majority of risk estimates were null, except for the highest exposure group for cumulative exposure (>34 years, OR = 2.60, 95% CI: 1.0–6.4) and number of embalmings (>3068 embalmings, OR = 2.3, 95% CI: 1.00–5.70). Hauptmann et al. (2009) also reported that among those who embalmed for more than 20 years, a significant increased risk of non-lymphoid cancers was observed (OR = 3.5, 95% CI: 1.1–10.9). The p values reported for the trend tests by Hauptman et al. (2009) are incorrect, as they are the same as those reported for the tests which used 0 embalmers (vs. <500) as the referent category; therefore, they are not reported here.

4.2.6. Myeloid leukemia

Myeloid leukemia was assessed in three case-control studies and four cohort studies, some of which also analyzed acute and/or chronic subtypes (Table 9). Results varied among the four studies that compared risks in exposed vs. unexposed individuals. Stroup et al. (1986) reported an excess in myeloid leukemia in US anatomists (SMR = 8.8, 95% CI: not reported). Similarly, Linos et al. (1990) reported an excess of acute myeloid leukemia in funeral home workers, although this was based on three exposed cases (OR = 6.7, 95% CI: 1.2–36.2). Hayes et al. (1990) reported a significant excess proportion of myeloid leukemia deaths overall in the NCI embalmers cohort (PMR = 1.57, 95% CI: 1.01–2.34), but found no associations in analyses by subtype ($PMR_{acute} = 1.52$, 95% CI: 0.85–2.52; $PMR_{chronic} = 1.84$, 95% CI: 0.79–3.62). Blair et al. (2001) conducted a case-control study of several industrial and occupational job categories in US workers and found no associations between intensity of formaldehyde exposure and acute or chronic myeloid leukemia.

Pinkerton et al. (2004) assessed myeloid leukemia in a cohort of US garment workers and found no association with formaldehyde exposure overall (SMR 1.44, 95% CI: 0.80–2.37) or when examined by subtype ($SMR_{acute} = 1.34$, 95% CI: 0.61–2.54; $SMR_{chronic} = 1.39$, 95% CI: 0.38–3.56). There were also no trends with duration of exposure or time since first exposure ($p > .05$), although risks were increased in workers with 20 or more years since first exposure (SMR = 1.91, 95% CI: not reported). In contrast, there were no increased risks in workers exposed for 10 or more years with 20 or more years since first exposure overall (SMR = 2.43, 95% CI: 0.98–5.01) or in analyses limited to acute myeloid leukemia (SMR = 2.51, 95% CI: 0.81–5.85).

In an analysis of the NCI industrial worker cohort with follow-up through 2004, Beane Freeman et al. (2009) assessed whether myeloid leukemia risk was associated with formaldehyde estimated as peak exposure, average intensity, cumulative exposure, cumulative number of peaks ≥4.0 ppm (data not reported), duration of exposure (data not reported), years since first exposure, and years since first exposure ≥4 ppm. These investigators reported no associations between any exposure metric and myeloid leukemia, including peak exposure (RR = 1.78, 95% CI: 0.87-3.64, $p_{trend} = 0.13$ for exposed groups), except for lower risks in those with no exposure vs. those with their first exposure to ≥ 4 ppm formaldehyde 0-25 years earlier (RR = 0.30, 95% CI: 0.11-0.81) and higher risks with increasing peak intensity with follow-up from 1981-1994 ($p_{trend} = 0.0353$ based on exposed subjects only and $p_{trend} = 0.210$ based on all study subjects), but not with follow-up to 1981 or 1995-2004 (Table 9). These null results were consistent with analyses of this cohort through 1994 based on every exposure metric except peak exposure, for which risks were increased (RR = 2.79, 95% CI: 1.08–7.21, $p_{trend} = .02$ for exposed groups, $p_{trend} = .0087$ for all groups) (Beane Freeman et al., 2009; Hauptmann et al., 2003). There were no associations based on any other metric in analyses.

Hauptmann et al. (2009) conducted a case-control study of professional embalmers, including cases from previous studies (Walrath and Fraumeni, 1983, 1984; Hayes et al., 1990), and assessed myeloid leukemia risk based on seven formaldehyde exposure metrics (Table 9). Having ever embalmed was associated with myeloid leukemia (OR = 11.2, 95% CI: 1.3–95.6, $p_{trend} = .027$), but there was only one case who never embalmed, making this risk estimate highly unreliable. Because of this, Hauptmann et al. (2009) combined unexposed individuals and those with <500 embalmings as a referent group to provide

© 2011 Informa Healthcare USA, Inc.

more conservative and reliable risk estimates; these are discussed here and shown in Table 9. An increased risk for myeloid leukemia was not reported for any exposed group except those with more than 34 years of employment (OR = 3.9, 95% CI: 1.2–12.5), more than 3068 embalmings (OR = 3.0, 95% CI: 1.0–9.2), or more than 9253 ppm-hours of cumulative formaldehyde exposure (OR = 3.1, 95% CI: 1.0–9.6). Hauptmann et al. (2009) conducted similar analyses for acute myeloid leukemia and found no associations in any dose group. Reported p values for trend tests for total and acute myeloid leukemia appear to be those based on analyses using 0 embalmings (vs. <500 embalmings) as a referent category and are not presented here.

Although there are some isolated findings of statistically significant associations between formaldehyde exposure and myeloid leukemia, these have not been found consistently either within or among studies and are far outnumbered by null findings in the more robust studies.

4.2.7. Other unspecified leukemia

Most cohort and case-control studies examined other (i.e., not lymphatic, myeloid, or monocytic) or unspecified leukemias (ICD 207) grouped with other lymphohematopoietic cancer types. The ICD 207 category alone was only examined in the NCI industrial worker cohort (Table 10). Beane Freeman et al. (2009) reported no associations between formaldehyde exposure and other or unspecified leukemia based on peak exposure, average intensity, cumulative exposure, cumulative number of peaks ≥4.0 ppm (data not reported), or duration of exposure (data not reported). These results are consistent with previous evaluations of the NCI industrial worker cohort (Hauptmann et al., 2003; Blair et al., 1986). Hayes et al. (1990) examined monocytic (ICD 206) and other unspecified leukemia (ICD 207) combined in embalmers and reported an increased proportion of deaths (PMR = 2.28, 95% CI: 1.29–3.52). This disease category was not evaluated in the follow-up by Hauptman et al. (2009).

4.2.8. Hodgkin's lymphoma, non-Hodgkin's lymphoma, and multiple myeloma

Cohort and case-control study results for Hodgkin's and non-Hodgkin's lymphoma and multiple myeloma are presented in Tables 11 to 13. Eleven assessments of formaldehyde-exposed vs. unexposed workers did not show associations between exposure and Hodgkin's lymphoma (Table 10). In the one study that evaluated iron foundry workers vs. the general population, risks were also not increased (Andjelkovich et al., 1995). When exposure-response relationships were evaluated in the NCI industrial worker cohort, associations were reported for peak exposure ($p_{trend} = .01$ for exposed groups) and average intensity ($p_{trend} = .05$ for exposed groups) but not for cumulative exposure, cumulative number of peaks ≥4.0 ppm (data not reported), or duration of exposure (data not reported) (Beane Freeman et al., 2009). This is consistent with the earlier examination of the NCI industrial worker cohort, for which exposure-response relationships for peak exposure ($p_{trend} = .04$) and average intensity ($p_{trend} = .03$), but not other exposure metrics, were reported (Beane Freeman et al., 2009; Hauptmann et al., 2003).

None of the 13 epidemiology investigations reported associations between formaldehyde exposure and non-Hodgkin's lymphoma for any exposure metric evaluated (Table 12). There were also no exposure-response associations observed (Table 12).

Multiple myeloma was not associated with formaldehyde exposure in any of the eight groups studied (Table 13). In the NCI industrial worker cohort, multiple myeloma risk was higher in individuals with no exposure based on all measures evaluated (Table 13). The only association with formaldehyde reported was for peak exposure ≥4.0 ppm (RR = 2.04, 95% CI: 1.01–4.12); however, the trend was not significant ($p > .05$) and there was no association with the number of peak exposures ≥4.0 ppm. This finding was consistent with results from earlier follow-ups of this cohort (Beane Freeman et al., 2009; Hauptmann et al., 2003).

4.3. HBWoE evaluation of epidemiology studies

We conducted an HBWoE evaluation of the epidemiology data with regard to an association between formaldehyde exposure and leukemia. Based on review of the data discussed in the previous section, we address the following questions:

1. What are the implications of studies of individual lymphohematopoietic cancers and several groupings of these cancer types (e.g., all cancers of lymphoid origin, all of non-lymphoid origin) regarding leukemia risks from formaldehyde exposure?
2. Were results from the epidemiology data consistent for different types of exposure metrics (e.g., peak exposure, number of peak exposures ≥4.0 ppm, cumulative exposure)? Were results dependent on the robustness of exposure measurements, particularly for the NCI industrial worker and embalmer cohorts?
3. Were co-exposures considered in the interpretation of the study results?
4. Were there consistent exposure-response associations within and across studies?
5. Were there potential statistical limitations among the epidemiology studies?
6. How should latency be considered when interpreting study results? Is it possible that risks decline over time owing to a relatively short induction-incubation period (as proposed by Beane Freeman et al., 2009)?

As a whole, considering these questions allows for an assessment of the extent to which the epidemiology data support either a causal association between formaldehyde exposure and leukemia or an alternative hypothesis. Importantly, one needs to consider the epidemiology

© 2011 Informa Healthcare USA, Inc.

Table 9a. Association between formaldehyde and myeloid leukemia (ICD 205).

	Hauptmann et al., 2009* (ICD 205)					Hauptmann et al., 2009* (ICD 205.0)					Beane Freeman et al., 2009					Pinkerton et al., 2004					Blair et al., 2001				
	Embalmers case-control					Embalmers case-control					NCI Cohort (1934–2004)					US garment workers					US industrial workers				
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI
Unexposed/ Exposed	Never embalming	1	OR	1.00	—						Unexposed	4	SMR	0.65	0.25–1.74						Acute				
	Ever embalming	33	OR	11.20	1.3-95.6						Exposed	44	SMR	0.90	0.67–1.21	Exposed	15	SMR	1.44	0.80–2.37	Low	14	OR	0.9	0.50–1.60
																					High	0	OR	—	—
																Acute	9	SMR	1.34	0.61–2.54					
																Chronic	4	SMR	1.39	0.38–3.56	Chronic				
																Other/ unspecified	1	SMR	2.15	0.05–11.94	Low	7	OR	1.3	0.60–3.10
																					High	1	OR	2.9	0.30–24.50
Peak Exposure	<500 embalmings	5	OR	1.00	—	<500 embalmings	3	OR	1.00	—	0 ppm	4	RR	0.82	0.25–2.67										
	>0–7.0 ppm	9	OR	2.90	0.9–9.8	>0–7.0 ppm	4	OR	1.80	0.4–9.3	>0–<2.0 ppm	14	RR	1.00	—										
	>7.0–9.3 ppm	9	OR	2.00	0.6–6.6	>7.0–9.3 ppm	5	OR	2.10	0.5–9.2	2.0–<4.0 ppm	11	RR	1.30	0.58–2.92										
	>9.3 ppm	11	OR	2.90	0.9–9.5	>9.3 ppm	7	OR	2.90	0.7–12.5	≥ 4.0 ppm	19	RR	1.78	0.87–3.64										
											p_{trend} = .13 (exposed)														
											p_{trend} = .07 (exposed and unexposed)														
Average Intensity	<500 embalmings	5	OR	1.00	—	<500 embalmings	3	OR	1.00	—	0 ppm	4	RR	0.70	0.23–2.16										
	>0–1.4 ppm	10	OR	2.60	0.8–8.7	>0–1.4 ppm	6	OR	2.50	0.6–10.9	>0–<0.5 ppm	24	RR	1.00	—										
	>1.4–1.9 ppm	10	OR	2.80	0.8–9.1	>1.4–1.9 ppm	5	OR	2.00	0.4–9.4	0.5–<1.0 ppm	9	RR	1.21	0.56–2.62										
	>1.9 ppm	9	OR	2.30	0.7–7.5	>1.9 ppm	6	OR	2.30	0.5–10.3	≥ 1.0 ppm	11	RR	1.61	0.76–3.39										
											p_{trend} = .43 (exposed)														
											p_{trend} = .40 (exposed and unexposed)														
Cumulative exposure	<500 embalmings	5	OR	1.00	—	<500 embalmings	3	OR	1.00	—	0 ppm-yr	4	RR	0.61	0.2–1.91										
	>0–4058 ppm-h	5	OR	2.10	0.5–8.1	>0–4058 ppm-h	2	OR	1.30	0.2–9.4	>0–<1.5 ppm-yr	26	RR	1.00	—										

Table 9a. continued on next page

Table 9a. continued.

	Hauptmann et al., 2009* (ICD 205)					Hauptmann et al., 2009* (ICD 205.0)					Beane Freeman et al., 2009					Pinkerton et al., 2004				Blair et al., 2001			
	Embalmers case-control					Embalmers case-control					NCI Cohort (1934–2004)					US garment workers				US industrial workers			
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate	95% CI	Category	Obs	Estimate	95% CI
	>4058–9253 ppm-h	10	OR	2.20	0.7–7.1	>4058–9253 ppm-h	6	OR	1.90	0.4–8.2	>1.5–<5.5 ppm-yr	8	RR	0.82	0.36–1.83								
	>9253 ppm-h	14	OR	3.10	1.0–9.6	>9253 ppm-h	9	OR	3.20	0.8–13.1	≥ 5.5 ppm-yr	10	RR	1.02	0.48–2.16								
											$p_{trend} > .5$ (exposed)												
											$p_{trend} = .44$ (exposed and unexposed)												
Cumulative number of peaks ≥4.0 ppm											No association. Results not shown.												
Duration of exposure/ employment	<500 embalmings	5	OR	1.00	—	<500 embalmings	3	OR	1.00	—						<3 yrs	3	SMR0.83	—				
	>0–20 yrs	2	OR	0.50	0.1–2.9	>0–20 yrs	1	OR	0.40	0.04–4.9	No association. Results not shown.					3–9 yrs	4	SMR1.26	—				
	>20–34 yrs	13	OR	3.20	1.0–10.1	>20–34 yrs	8	OR	2.90	0.7–12.2						10+ yrs	8	SMR2.19	—				
	>34 yrs	14	OR	3.90	1.2–12.5	>34 yrs	8	OR	3.10	0.7–13.7						$p_{trend} > .05$							
Time since first exposure																<10 yrs	1	SMR0.90	—				
																<10–19 yrs	1	SMR0.40	—				
																20+ yrs	13	SMR1.91	†				
																$p_{trend} > .05$							
Number of embalmings	<500 embalmings	5	OR	1.00	—	0	3	OR	1.00	—													
	>0–1422	3	OR	1.20	0.30–5.50	>0–1422	0	OR	0.00	0.0–1.8													
	>1422–3068	12	OR	2.90	0.90–9.10	>1422–3068	8	OR	2.90	0.7–12.0													
	>3068	14	OR	3.00	1.00–9.20	>3068	9	OR	2.90	0.7–11.6													
8-Hour time-weighted average intensity	<500 embalmings	5	OR	1.00	—	0	3	OR	1.00	—													
	>0–0.10	8	OR	2.40	0.70–8.20	>0–0.10	3	OR	1.40	0.3–7.8													
	>0.10–0.18	10	OR	2.60	0.80–8.70	>0.10–0.18	7	OR	2.60	0.6–11.4													
	>0.18	11	OR	2.60	0.80–8.30	>0.18	7	OR	2.60	0.6–11.3													

Table 9b. Other cohorts.

Reference	Obs	Estimate		95% CI
Stroup et al., 1986	5	SMR	8.8	—
Hayes et al., 1990	24	PMR[‡]	1.57	1.01–2.34
Linos et al., 1990 (acute)	3	OR	6.70	1.20–36.20

*Results from analyses using those who never embalmed as the referent group (with one myeloid leukemia case) were highly unstable. Results presented here are from analyses using individuals with <500 embalmings as the referent group. (See Table 4 in Hauptman et al., 2009).
[†]95% CI does not include 1.0.
[‡]PMR divided by 100.

data in the context of the hematotoxicity and mode-of-action data (discussed later), as each of the three lines of evidence inform interpretation of the other; specifically, a claim of causation cannot be solely based on one or the other but has to be reflected consistently across the epidemiology, mode-of-action, and hematoxicity data.

4.3.1. Cancer outcome assessments likely led to disease misclassification

There are several ways in which cancer outcomes were defined and assessed in the formaldehyde epidemiology studies, several of which may have led to disease misclassification and/or misleading results.

With few exceptions, most studies assessed cancer mortality, and several of the larger studies relied on death certificates to determine cause of death (e.g., Beane Freeman et al., 2009; Hauptmann et al., 2003, 2009; Hayes et al., 1990; Pinkerton et al., 2004; Stroup et al., 1986; Walrath and Fraumeni, 1983, 1984). Death certificates do not always identify leukemia subtype, and leukemia diagnosis was considered unreliable prior to 1992 (Collins and Lineker, 2004; Bachand et al., 2010; Miller et al., 1992; Percy et al., 1981, 1990, both as cited by Collins and Lineker, 2004). Thus, relying on death certificates may have led to disease misclassification.

In addition, diagnoses of lymphohematopoietic cancers has evolved in recent decades, and historic records may be inaccurate (Bachand et al., 2010; Collins and Lineker, 2004; Miller et al., 1992; Percy et al., 1990, as cited by Collins and Lineker, 2004; Scott and Chiu, 2006). For example, past classifications of lymphomas do not make distinctions between different cell types (Scott and Chiu, 2006). This means that, within studies that investigated subjects over many decades, individuals assigned the same cancer actually may not have had the same cancer.

There is also an issue with assessing cancers in categories. Each different kind of lymphohematopoietic cancer is a distinct disease with a unique etiology, set of risk factors, and, presumably, mechanism of action. Consequently, grouping cancer types together is not informative regarding risks for a particular cancer type. Any observed increased risks could be driven by risks for one cancer type (e.g., if the majority of cancers in a group were the same type, or one cancer type had very large risks associated with it); a lack of risks could be indicative of no risks among all lymphohematopoietic cancers or that combining cancer types masks true associations with one particular cancer type. For example, in the NCI industrial worker cohort, the two cancers that contribute to the association between peak formaldehyde exposure and all lymphohematopoietic cancers are multiple myeloma and Hodgkin's disease (Beane Freeman et al., 2009). These cancers are not associated with formaldehyde exposure in other studies.

In sum, disease misclassification likely led to uncertain risk estimates. In addition, studies that purport to show associations with a group of cancers that include leukemia do not provide sufficient evidence that risk, if it exists, is for leukemia and not another white cell cancer.

4.3.2. Exposure assessments likely affected by exposure measurement error or misclassification

Because of the difficulty in obtaining exposure data for individuals in cohort and case-control studies, investigators typically estimated exposure from few, if any, measurements of formaldehyde concentrations. For example, Andjelkovich et al. (1995) assigned formaldehyde exposures to each iron foundry worker by job category based on midpoints of ranges from actual sampling data. Pinkerton et al. (2004) conducted analyses based on 1 year of measured data from the 1980s and applied it to the entire follow-up period in garment workers (1955–1998). Exposure estimates in formaldehyde workers in the NCI cohort were developed by assigning job categories from work histories abstracted in 1980 and an expert assessment of job and tasks using current and past measurement data (Stewart et al., 1986; Blair et al., 1986). Although this was considered to be a well-conducted exposure assessment for the time, validation of the exposure matrix was not possible, and exposures to formaldehyde and other potential confounders after 1980 were assumed to be minimal (Beane Freeman et al., 2009; Blair et al., 1986, 1990; Stewart et al., 1986). Peak exposure categories (none, >0 to <0.5 ppm, 0.5 to <2.0 ppm, 2.0 to <4.0 ppm, or ≥4.0 ppm) were estimated and defined as short-term exposures (generally less than 15 minutes) exceeding the 8-hour time-weighted average (TWA8) category (Blair et al., 1986; Beane Freeman et al., 2009). In the NCI embalmers cohort (Hauptmann et al., 2009), questionnaire data were linked to data from an exposure experiment (Stewart et al., 1992). No measurements of peak exposure were available, and average formaldehyde intensity, peak, time-weighted average, and cumulative exposure were estimated using a predictive model. Comparison of modeled average intensity to measurements from independent embalmings suggested the model overestimated exposure by 35%, and peak exposures could not be validated (Hauptmann et al., 2009).

Despite the paucity of exposure information, in two of the largest cohorts evaluated (the NCI industrial worker and embalmer cohorts), several exposure metrics were estimated (e.g., peak exposure, average

© 2011 Informa Healthcare USA, Inc.

Table 10a. Association between formaldehyde and other unspecified leukemia (ICD-8 207).

	Beane Freeman et al., 2009				
	NCI Cohort (1934-2004)				
Measures	Category	Obs	Estimate		95% CI
Peak Exposure	0 ppm	2	RR	0.61	0.13–2.85
	>0–<2.0 ppm	13	RR	1.00	—
	2.0–<4.0 ppm	8	RR	0.86	0.35–2.12
	≥4.0 ppm	13	RR	1.15	0.53–2.53
	$p_{trend} > .5$ (exposed)				
	$p_{trend} = .5$ (exposed and unexposed)				
Average Intensity	0 ppm	2	RR	0.58	0.13–2.62
	>0–<0.5 ppm	21	RR	1.00	—
	0.5–<1.0 ppm	7	RR	0.98	0.42–2.33
	≥1.0 ppm	6	RR	0.84	0.33–2.12
	$p_{trend} > .5$ (exposed)				
	$p_{trend} > .5$ (exposed and unexposed)				
Cumulative Exposure	0 ppm-yr	2	RR	0.77	0.16–3.59
	>0–<1.5 ppm-yr	15	RR	1.00	—
	>1.5–<5.5 ppm-yr	10	RR	1.65	0.73–3.73
	≥5.5 ppm-yr	9	RR	1.44	0.61–3.36
	$p_{trend} = .15$ (exposed)				
	$p_{trend} = .13$ (exposed and unexposed)				
Cumulative number of peaks ≥4.0 ppm	No association. Results not shown.				
Duration of Employment	No association. Results not shown.				

exposure, cumulative exposure, exposure duration). Owing to the importance that peak exposures play in the interpretation of the NCI studies, it is important to note that peaks were not actually measured, but only inferred from job descriptions. Detailed analyses of exposure metrics used for the NCI industrial worker cohort were conducted by Blair et al. (1990) and Stewart et al. (1986). They reported that measures of duration (employment and exposure) and average exposure and level of exposure were highly correlated ($r = .8$). Peak exposures had low to moderate correlations with employment duration ($r = .2$), exposure duration ($r = .3$), cumulative exposure ($r = .3$), average exposure ($r = .5$), and level of exposure ($r = .7$). Average exposure showed little correlation with duration of employment ($r = -.1$) and duration of exposure ($r = .0$). Based on these correlations, it is unclear why lymphohematopoietic and leukemia mortality rates were associated with peak exposure but not with the number of peak exposures ≥4.0 ppm, cumulative exposure, or exposure duration (Beane Freeman et al., 2009). Even if higher exposure intensities are of more consequence as a result of formaldehyde's mode of action, those experiencing higher air concentrations over time with any repeatability would have a higher number of peak exposures ≥4.0 ppm and higher cumulative and average exposures, so these measures ought to show an association as well. This is not the case, indicating the association with peak exposure is not likely to be causal (other issues with this statistic are discussed below).

Table 10b. Other cohorts.

Reference	Code	Obs	Estimate		95% CI
Hayes et al., 1990	206, 207	20	PMR	2.28	1.39–3.52

The lack of precise exposure data likely led to exposure measurement error and/or exposure misclassification in these epidemiology studies. This could have biased results either towards or away from the null (Jurek et al., 2005). Based on the null associations with other exposure metrics, in the case of peak exposure, it appears to be the latter.

4.3.3. Exposures to other chemicals in the workplace may have confounded results

None of the studies adequately addressed co-exposures to other agents. For example, embalmers were exposed to infectious agents and other chemicals in embalming fluid, such as methanol, propylene glycol, industrial methylated spirit, phenol, and glycerol (Coleman and Kogan, 1998; Bachand et al., 2010; Bosetti et al., 2008; Collins et al., 2004). Industrial workers were likely exposed to other chemicals as well (e.g., antioxidants, asbestos, benzene, carbon black, dyes and pigments, hexamethylenetetramine, melamine, phenol, plasticizers, urea, and wood dust) (Beane Freeman et al., 2009). Although benzene is the only known leukemogen among these agents, it is possible that any observed risks, if found to be real, may have been attributable to exposures to other agents.

4.3.4. Exposure-response associations within and among studies are not consistent

If formaldehyde is in fact a causal factor for leukemia, one would expect leukemia risk to increase with formaldehyde exposure both within and among studies. As described below, few studies actually assessed exposure-response (Pinkerton et al., 2004; Stern et al., 1987; Beane Freeman et al., 2009; Hauptmann et al., 2009); among those, consistent associations were not reported. Among studies, leukemia risks appeared to be higher in professionals with lower average formaldehyde exposures (mean TWA8 concentrations <0.5 to 1 ppm in professional settings [e.g., workplaces of histopathologists, embalmers, anatomists]; IARC, 2006), yet more highly exposed industrial workers (mean TWA8 concentrations <1 to >10 ppm in industrial settings [e.g., formaldehyde manufacturing]; IARC, 2006) showed lesser effects, adding to the weight of evidence suggesting formaldehyde is not a causal factor.

In analyses of formaldehyde and risks of all leukemias combined, Pinkerton et al. (2004) found no exposure-response associations with duration of exposure, time since first exposure, or year of first exposure in garment workers. Stern et al. (1987) also found no trend with duration of exposure. There were some statistically significant trends reported in the NCI industrial cohort but,

Table 11a. Association between formaldehyde and Hodgkin's lymphoma (ICD 201).

	Beane Freeman et al., 2009					Andjelkovich et al., 1995				
	NCI cohort (1934–2004)					Iron foundry workers				
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI
Unexposed/Exposed	Unexposed	2	SMR	0.70	0.17–2.80	Unexposed	0	SMR	0.00	0.00–4.12
	Exposed	25	SMR	1.42	0.96–2.10	Exposed	1	SMR	0.72	0.01–4.00
Peak Exposure	0 ppm	2	RR	0.67	0.12–3.60					
	>0–<2.0 ppm	6	RR	1.00	—					
	2.0–<4.0 ppm	8	RR	3.30	1.04–10.50					
	≥4.0 ppm	11	RR	3.96	1.31–12.02					
	p_{trend} = .01 (exposed)									
	p_{trend} = .004 (exposed and unexposed)									
Average Intensity	0 ppm	2	RR	0.53	0.11–2.66					
	>0–<0.5 ppm	10	RR	1.00	—-					
	0.5–<1.0 ppm	9	RR	3.62	1.41–9.31					
	≥1.0 ppm	6	RR	2.48	0.84–7.32					
	p_{trend} = .05 (exposed)									
	p_{trend} = .03 (exposed and unexposed)									
Cumulative Exposure	0 ppm-yr	2	RR	0.42	0.09–2.05					
	>0–<1.5 ppm-yr	14	RR	1.00	—					
	>1.5–<5.5 ppm-yr	7	RR	1.71	0.66–4.38					
	≥5.5 ppm-yr	4	RR	1.30	0.4–4.19					
	p_{trend} = .08 (exposed)									
	p_{trend} = .06 (exposed and unexposed)									
Cumulative number of peaks ≥4.0 ppm	No association. Results not shown.									
Duration of Employment	No association. Results not shown.									

Table 11b. Other cohorts.

Reference	Obs	Estimate		95% CI
Wong et al., 1983	2	SMR	2.40	0.27–8.66
Robinson et al., 1987	2	SMR	3.33	0.59–10.49
Gerin et al., 1989	8	OR	0.50	0.20–1.40
Hayes et al., 1990	3	PMR	0.72	0.15–2.10
Hall et al., 1991	1	SMR	1.31	0.03–7.33
Matanoski et al., 1991	2	SMR	0.36	0.04–1.31
Hansen and Olsen, 1995	12	SPIR	1.00	0.50–1.70
Coggon et al., 2003	6	SMR	0.70	0.26–1.53
Pinkerton et al., 2004	2	SMR	0.55	0.07–1.98
Hauptmann et al., 2009	8	OR	0.50	0.10–2.60

as described below, these findings were not robust or indicative of causation.

In the NCI industrial worker cohort, study subjects were divided into "low" (>0 to <2.0 ppm), "medium" (2.0 to <4.0), and "high" (≥4.0) exposure categories for the inferred lifetime peak level, and analyses were conducted by comparing risks in the medium- and high-exposure categories to those in the low-exposure category. As shown in Figure 1, these analyses showed no statistically significant associations and no exposure-response relationship with leukemia. If, however, a "zero" category was added, comprised of workers from the facilities that were presumably unexposed, the exposure-response trend for leukemia vs. "peak" became statistically significant, as was the contrast between the high vs. the zero (but not vs. the low) category.

Even though the "low" group included people down to zero as the lifetime "peak" exposure, the leukemia risk for the "zero" group was markedly lower. People classified as "zero" must have had systematically different job descriptions than those in the "low" category (for which peak exposure could be as low as zero and still admit them into the "low" group), so the comparability of these groups is in question. Moreover, the "zero" group has leukemia risks that are notably smaller than the general population. Indeed, when analyses were done on an SMR basis, risks, although not statistically significant, were much lower than those for the US population ($SMR_{all\ leukemia}$ = 0.48, 95% CI: 0.23–1.01; $SMR_{lymphatic\ leukemia}$ = 0.26, 95% CI: 0.04–1.82; $SMR_{myeloid\ leukemia}$ = 0.65, 95% CI: 0.25–1.74) (Beane

© 2011 Informa Healthcare USA, Inc.

Table 12a. Association between formaldehyde and non-Hodgkin's lymphoma (ICD 200, 202).

	Beane Freeman et al., 2009					Wang et al., 2009b					Gerin et al., 1989				
	NCI cohort (1934–2004)					Connecticut women (1996–2000)					Montreal workers (1979–1985)				
Measures	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI	Category	Obs	Estimate		95% CI
Unexposed/ Exposed	Unexposed	12	SMR	0.86	0.49–1.52	Never	398	OR	1.00	—					
	Exposed	94	SMR	0.85	0.70–1.05	Ever	203	OR	1.30	1.00–1.70					
Peak Exposure	0 ppm	12	RR	1.06	0.53–2.14										
	>0–<2.0 ppm	39	RR	1.00	—										
	2.0–<4.0 ppm	27	RR	1.08	0.65–1.78										
	≥4.0 ppm	28	RR	0.91	0.55–1.49										
						$p_{trend} > .5$ (exposed)									
						$p_{trend} > .5$ (exposed and unexposed)									
Average Intensity	0 ppm	12	RR	1.08	0.55–2.12	Never	398	OR	1.00	—					
	>0–<0.5 ppm	59	RR	1.00	—	Low	129	OR	1.40	1.00–1.80					
	0.5–<1.0 ppm	22	RR	1.20	0.73–1.96	Medium–High	74	OR	1.20	0.80–1.70					
	≥1.0 ppm	13	RR	0.71	0.39–1.32	$p_{trend} = .21$									
						$p_{trend} > .5$ (exposed)									
						$p_{trend} = .45$ (exposed and unexposed)									
Cumulative Exposure	0 ppm-yr	12	RR	0.94	0.46–1.86						Short	13	OR	0.70	0.30–1.60
	>0–<1.5 ppm-yr	60	RR	1.00	-						Long—low	15	OR	1.10	0.50–2.20
	>1.5–<5.5 ppm-yr	13	RR	0.58	0.31–1.06						Long—medium	14	OR	1.00	0.50–2.10
	≥5.5 ppm-yr	21	RR	0.91	0.54–1.52						Long—high	5	OR	0.50	0.10–1.70
						$p_{trend} > .5$ (exposed)									
						$p_{trend} = .42$ (exposed and unexposed)									
Cumulative number of peaks ≥4.0 ppm						No association. Results not shown.									
Duration of Employment						No association. Results not shown.									

Table 12b. Other cohorts.

Reference	Obs	Estimate		95% CI
Edling et al., 1987	2	SPIR *	2.00	0.50–7.20
Ott et al., 1989†	2	OR	2.00	—
Hayes et al., 1990	34	PMR	1.26	0.87–1.76
Linos et al., 1990†	6	OR	3.2	0.80–13.40
Partanen et al., 1993†	4	OR	4.24	0.68–26.60
Hansen and Olsen, 1995	32	SPIR	0.90	0.60–1.20
Tatham et al., 1997†	93	OR	1.20	0.86–1.50
Stellman et al., 1998	11	RR	0.92	0.50–1.68
Coggon et al., 2003	31	SMR	0.98	0.67–1.39
Hauptmann et al., 2009	NR	OR	0.90	0.40–2.10

Notes *Standardized proportionate incidence ratio.
†Case-control study.

Freeman et al., 2009). In contrast, among the all of the exposed groups, the SMR estimates are more consistent with US expected levels ($SMR_{\text{all leukemia}} = 1.02$, 95% CI: 0.85–1.22; $SMR_{\text{lymphatic leukemia}} = 1.15$, 95% CI: 0.83–1.59; $SMR_{\text{myeloid leukemia}} = 0.90$; 95% CI: 0.67–1.21) (Beane Freeman et al., 2009).

In short, it appears that the reported significant relation of "peak" formaldehyde exposure and leukemia risk depended entirely on a lower-than-usual leukemia rate in the "zero" group rather than any effects among exposed people. "Peaks" were inferred possibilities rather than actual exposures, and they did not account for the duration of time spent in the highest peak-exposure category or the relevant latent period between the date of first highest peak exposure and death (Marsh and Youk, 2004). Individuals with high peak exposures early but not later in their career and those with low peak exposures for the majority of their career but high peaks near the end were likely in the same exposure category. Similarly, those with one peak exposure were likely classified in the same category as those with several peak exposures. In both of these scenarios, individuals with very different exposures were grouped in similar categories. In contrast, when grouping workers by the inferred number of peaks ≥4.0, or by cumulative or average exposure, individuals with similar exposures were more likely to be grouped together. Based on these latter metrics, formaldehyde exposure was not associated with leukemia. Thus, the finding of a significant effect for leukemia in the industrial cohort with peak exposure (Beane Freeman et al., 2009), which relies on picking apparently positive results among several similar analyses demonstrating inconsistent results, is not a substantive or compelling finding.

Table 13a. Association between formaldehyde and multiple myeloma (ICD 203).

	Beane Freeman et al., 2009				
	NCI cohort (1934-2004)				
Measures	Category	Obs	Estimate		95% CI
Unexposed/ eEposed	Unexposed	11	SMR	1.78	0.99–3.22
	Exposed	48	SMR	0.94	0.71–1.25
Peak Exposure	0 ppm	11	RR	2.74	1.18–6.37
	>0–<2.0 ppm	14	RR	1.00	—
	2.0–<4.0 ppm	13	RR	1.65	0.76–3.61
	≥4.0 ppm	21	RR	2.04	1.01–4.12
	$p_{\text{trend}} = .08$ (exposed)				
	$p_{\text{trend}} > .5$ (exposed and unexposed)				
Average Intensity	0 ppm	11	RR	2.18	1.01–4.70
	>0–<0.5 ppm	25	RR	1.00	—
	0.5–<1.0 ppm	11	RR	1.40	0.68–2.86
	≥1.0 ppm	12	RR	1.49	0.73–3.04
	$p_{\text{trend}} > .5$ (exposed)				
	$p_{\text{trend}} > .5$ (exposed and unexposed)				
Cumulative Exposure	0 ppm-yr	11	RR	1.79	0.83–3.89
	>0–<1.5 ppm-yr	28	RR	1.00	—
	>1.5–<5.5 ppm-yr	5	RR	0.46	0.18–1.20
	≥5.5 ppm-yr	15	RR	1.28	0.67–2.44
	$p_{\text{trend}} > .5$ (exposed)				
	$p_{\text{trend}} > .5$ (exposed and unexposed)				
Cumulative number of peaks ≥4.0 ppm	No association. Results not shown.				
Duration of Employment	No association. Results not shown.				

Table 13b. Other cohorts.

Reference	Obs	Estimate		95% CI
Edling et al., 1987	2	SPIR *	4.00	0.50–14.40
Ott et al., 1989†	1	OR	1.00	—
Hayes et al., 1990	20	PMR	1.37	0.84–2.12
Dell and Teta, 1995	5	SMR	2.62	0.85–6.11
Stellman et al., 1998	4	RR	0.74	0.27–2.02
Coggon et al., 2003	15	SMR	0.86	0.48–1.41
Hauptmann et al., 2009	NR	OR	1.40	0.50–5.60

Notes: *Standardized proportionate incidence ratio.
†Case-control study.

It is notable that in analyses limited to myeloid leukemia, a different pattern emerges. Neither Beane Freeman et al. (2009) nor Pinkerton et al. (2004) found any consistent exposure-response associations among the NCI industrial worker cohort and garment workers, respectively, regardless of the exposure metric, whereas Hauptmann et al. (2009) reported trends in embalmers based on peak exposure and duration of exposure, but only when the referent categories (<500 embalmings)

© 2011 Informa Healthcare USA, Inc.

were included in the analyses. No trends were observed when the referent category was excluded or when exposure was defined as average intensity, cumulative exposure, number of embalmings, or 8-hour TWA intensity. Again, this lack of consistency suggests that observed trends are not likely indicative of causation.

Regarding other lymphohematopoietic cancer types or groups, Beane Freeman et al. (2009) observed a trend only with peak exposure and all lymphohematopoietic cancers combined in the NCI industrial cohort. No other trends with all lymphohematopoietic cancers combined were observed in this cohort or the NCI embalmers cohort (Hauptmann et al., 2009). No trends were observed in either cohort for cancers of lymphoid origin and, for cancers of non-lymphoid origin, trends were only observed for duration of exposure in the NCI embalmers cohort when unexposed individuals were excluded (which left two individuals in the lowest category). There were no trends in the NCI industrial workers cohort with lymphatic leukemia, and no trends at all for non-Hodgkin's lymphoma or multiple myeloma. Beane Freeman et al. (2009) observed a trend with peak exposure and average intensity for Hodgkin's lymphoma but not with cumulative exposure, cumulative number of peaks ≥4.0 ppm, or duration of employment.

Regarding exposure-response among studies, professionals—such as embalmers, pathologists, and anatomists—have much lower formaldehyde exposures than industrial workers. Yet Blair et al. (1990) found small excess leukemia risks among professionals but not industrial workers; several meta-analyses have reported similar findings (Bosetti et al., 2008; Collins and Lineker, 2004). Some possible explanations have been put forth to explain these findings, including infectious agents, other chemicals in embalming fluid, occupational and lifestyle factors, observer bias, and a higher degree of scrutiny and medical attention owing to perceived risks (Coleman and Kogan, 1998; Bachand et al., 2010; Bosetti et al., 2008; Collins et al., 2004). Zhang et al. (2010a) suggest that effects in industrial workers may not be observed because analyses were conducted based on combined exposure categories, but this does not explain why effects are observed in professionals and not in industrial workers (who have higher exposures), and results from the few large studies that examined exposure-response associations do not support the hypothesis that formaldehyde is causally associated with leukemia.

Overall, although some statistically significant trends have been noted, these trends were not found consistently within or among studies. The lack of consistent exposure-response associations within or among studies indicates that the few associations noted between formaldehyde and leukemia are not causal.

4.3.5. Statistical limitations may have led to spurious associations

When the same set of data is analyzed in multiple parallel ways using different models, groupings, or summary measures, the meaning of statistical tests becomes distorted by the multiple-comparisons problem. That is, if enough alternatives are tried, some might be "significant" by chance alone (since, at a criterion of $p = .05$, even when there is no effect, 5% of comparisons are ruled "significant").

Data from several studies, including those of the NCI industrial worker and embalmer cohorts, were analyzed many different parallel ways (e.g., average or cumulative or peak exposure; pairwise comparisons or trends; using internal or external controls; with or without a "unexposed" group in the trend test; individual tumor types or various measures of combined tumors). Unless a correction for multiple comparisons is made, finding marginal significance in one or a few such comparisons is not surprising even when there is no true effect. For example, in the NCI industrial worker cohort, associations were reported with peak exposures, but there was no a priori reason to focus on peak exposures. Furthermore, associations were not found for other, more accurate indicators of exposure, such as the number of peak exposures ≥4.0 ppm, cumulative exposure, and average exposure. If there is no a priori reason to choose a superior exposure metric, one should not select a model based solely on statistical performance because choosing the metric with the strongest association with outcome could lead to bias (Kriebel et al., 2007). Instead, one should choose a model based on which is most consistent with the hypothesized mechanism of action. In the case of formaldehyde, peak exposure is clearly an inferior metric (discussed above), and this provides an even stronger argument for not choosing a model based on the strongest statistical association. The result for peaks can at most be a hypothesis-generating observation to be tested on future data. Otherwise, it is post hoc and arbitrary.

4.3.6. The latency argument proposed by Beane Freeman et al. (2009) appears to be a post hoc explanation for the observed effects

Epidemiology studies are often limited in that they are not conducted over long enough periods of time for cancer from particular exposures to develop, in which case causal associations cannot be detected. In the case of leukemia, Beane Freeman et al. (2009) suggest just the opposite, stating that risks for myeloid leukemia may decline over time owing to a relatively short induction-incubation period. In other words, they suggest that, after a latent period during which risks are not increased, increased risks for leukemia will be observed within a certain time period and plateau afterwards. Because of this, they suggest that risks may diminish or not be observed if a study has too long a follow-up period. This appears to be a post hoc explanation for the diminished risks associated with peak exposure observed in the NCI industrial worker cohort with follow-up through 2004 vs. 1994 (in workers who died 16–25 years after the first exposure). It does not explain how risks were only observed with peak exposure, and not with other exposure metrics (particularly cumulative number of peaks ≥4.0 ppm), or how this trend was not

observed in the NCI embalmers cohort (Hauptmann et al., 2003; Walrath and Fraumeni, 1984) or garment workers cohort (Pinkerton et al., 2004), in which risks were only observed with exposures ≥20 years. Also, Beane Freeman et al. (2009) did not test this hypothesis even though they had the data to do so. For all of these reasons, a shorter latency is not a scientifically valid explanation for the lack of observed risks by Beane Freeman et al. (2009).

4.3.7. Recent formaldehyde meta-analyses do not support an association between formaldehyde exposure and leukemia

A number of recent meta-analyses have been conducted on the body of epidemiology studies concerning formaldehyde and leukemia (Bachand et al., 2009; Zhang et al., 2009; Bosetti et al., 2008; Collins and Lineker, 2004; Schwilk et al., 2010). Only the most recent of these analyses (Bachand et al., 2009; Schwilk et al., 2010) include the most recent update to the NCI industrial worker cohort; the others rely on the Hauptmann et al. (2003) analysis, in which 1006 deaths were omitted unintentionally. Of the five meta-analyses, three reported no overall association between formaldehyde and leukemia (Bachand et al., 2009; Bosetti et al., 2008; Collins and Lineker, 2004), and methodological limitations of the other two meta-analyses (Zhang et al., 2009; Schwilk et al., 2010), which are almost the same except for the addition of two studies in the latter, make it challenging to interpret their summary risk estimates (REs).

Bachand et al. (2009) found that, among cohort studies, REs for exposed vs. unexposed ranged from 0.43 to 1.60 for leukemia—none were statistically significant (i.e., no 95% CI excluded 1.0)—and the summary risk estimate indicated no association (RE = 1.05, 95% CI: 0.91–1.20). Data from only two case-control studies were analyzed, neither of which reported increased risks (Blair et al. [2001]: RE = 0.98, 95% CI: 0.70–1.36; Partanen et al.

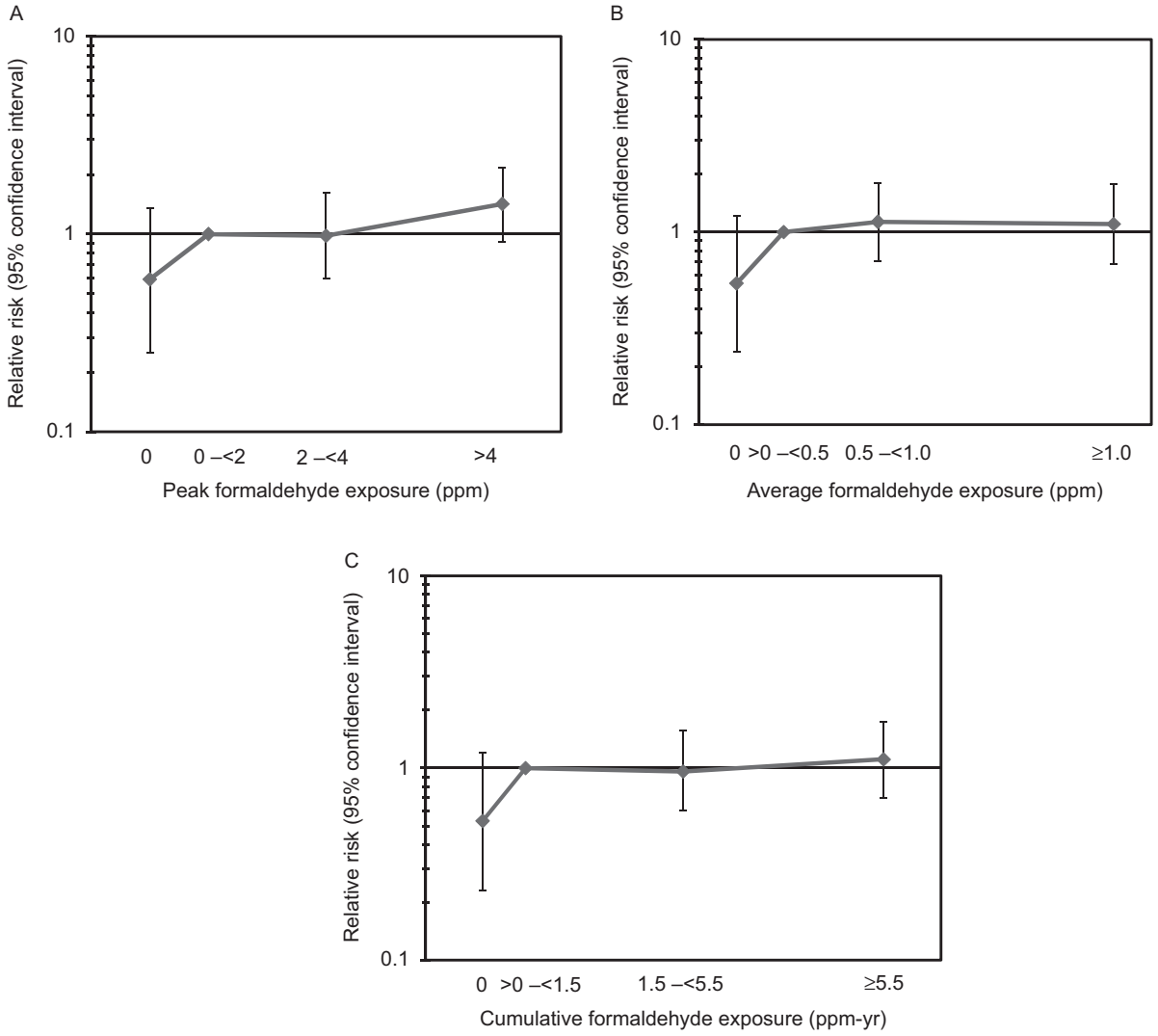

Figure 1. Relative risks of leukemia in the NCI industrial worker cohort compared to study subjects in the low-exposure category for (a) peak, (b) average, and (c) cumulative formaldehyde exposure. The "no exposure" category is comprised of workers from facilities that were presumably unexposed. The only statistically significant trends were for peak exposure when the "no exposure" workers were included in the analyses; all other trend tests including or excluding the "no exposure" workers were null. (See colour version of this figure online at www.informahealthcare.com/txc)

© 2011 Informa Healthcare USA, Inc.

[1993]: RE = 1.40, 95% CI: 0.25–7.91). No associations were observed when analyses were stratified by leukemia type (myeloid, lymphatic/lymphocytic, or other/unspecified), job type (professional/technical or industrial), or region (USA/Canada or Europe).

Bosetti et al. (2008) calculated REs for lymphohematopoietic cancers and leukemia among professionals and industrial workers evaluated in cohort studies published through February 2007. Risks of lymphohematopoietic cancers among professionals were increased (RE = 1.31, 95% CI: 1.16–1.48), but they were decreased among industrial workers (RE = 0.85, 95% CI: 0.74–0.96). Similarly, leukemia risks were elevated among professionals (RE = 1.39, 95% CI: 1.15–1.68), but not industrial workers (RE = 0.90, 95% CI: 0.75–1.07). Based on analyses of 18 studies published through December 1, 2003, Collins and Lineker (2004) found similar results. They reported that leukemia risks were not increased among industrial workers (RE = 0.9, 95% CI: 0.8–1.0) or pathologists and anatomists (RE = 1.4, 95% CI: 1.0–1.9), but were increased among embalmers (RE = 1.6, 95% CI: 1.2–2.0), who had among the lowest exposures.

In contrast to the three meta-analyses discussed above, Zhang et al. (2009) found a significant effect across industries (RE = 1.54, 95% CI: 1.18–2.00). This can be explained by the unusual means of selecting and combining studies: they used different measures of exposure, selecting only one from each study even if several were examined, resulting in selection of peak exposure for studies where available, then average exposure and cumulative exposure for others, and, finally, exposure duration if none of the other metrics were analyzed. Zhang et al. (2009) claim this is because average and cumulative exposure may be less accurate measures of true exposure if workers with very high exposure also have long intervening periods with little or no exposure, but they have not considered whether these metrics are relevant for assessing risk. Moreover, if several categories or levels of exposure were examined, Zhang et al. (2009) took data from only the highest among them. What constituted a "high" category also varied considerably among studies, depending on how each study established gradations of exposure. As a consequence, the comparisons across studies are very heterogeneous, and it is not clear whether a comparable question was being examined in each case. Furthermore, by not using the entire range of exposure estimates (i.e., by examining risks in the high-exposure group vs. the low-exposure group only), exposure-response could not be assessed, which likely generated misleading results, since a lack of exposure-response can indicate a lack of a causal association. Finally, Zhang et al. (2009) did not use the most recent NCI industrial worker cohort data, instead relying on data from the Hauptmann et al. (2003) study, which didn't account for over 1000 deaths in the cohort and only reported increased risks of myeloid leukemia based on internal comparisons that depended on the reference category and category cut points.

Schwilk et al. (2010) updated the Zhang et al. (2009) analysis by including the most recent NCI industrial worker and embalmer cohort studies (Hauptmann et al., 2009; Beane Freeman et al., 2009) and reported increased risks of leukemia (RR = 1.53, 95% CI: 1.11–2.21) and myeloid leukemia (RR = 2.47, 95% CI: 1.42–2.47). Because Schwilk et al. (2010) use similar methods as Zhang et al. (2009), their study suffers from the same limitations. In addition, Schwilk et al. (2010) use one-sided p values, which increased the likelihood of false-positive results. They also reported several exposure-response relationships in six studies and concluded an exposure-response association exists, but they did not discuss these associations for all exposure metrics from each study. As we have shown, had they done this, it would be evident that there are no consistent exposure-response associations between formaldehyde exposure and leukemia. Because of all these limitations, the results of the Zhang et al. (2009) and Schwilk et al. (2010) analyses should be interpreted with caution, especially in view of the substantial heterogeneity and their lack of concordance with other meta-analyses.

Overall, results from the meta-analyses of formaldehyde and leukemia are consistent with a lack of association and the results of our weight-of-evidence evaluation.

4.4. Summary

As a whole, the available formaldehyde epidemiology studies do not support a causal association between formaldehyde exposure and leukemia. As demonstrated in the endpoint-by-endpoint analysis and Tables 4 to 13, there is no lymphohematopoietic cancer or group of lymphohematopoietic cancers for which associations with formaldehyde were found consistently within or across studies. Although some statistically significant associations were reported, these were outnumbered by null findings in the more robust studies using related exposure metrics, and there were no consistent exposure-response relationships observed. Limitations in exposure and cancer outcome assessments as well as statistical analyses also likely affected calculations of risk.

If formaldehyde were truly a causal factor for leukemia, consistent observations of effect should have been observed, with increased risks found with increased exposures. Because this is not the case, it is most likely that any observed effects were a result of confounders, limitations in statistical methods (e.g., multiple comparisons), disease misclassification, and/or exposure misclassification/measurement error.

This question can be further explored by considering information on toxicology and mode-of-action studies. It is a precept of the HBWoE approach that one considers the cross-discipline integration of hypothesized effects. To the degree that consideration of animal data and dosimetry casts doubt on the ability

of inhaled formaldehyde to interact with and perturb hematopoiesis, this increases the relative plausibility of a conclusion that those associations seen in purely epidemiologic investigations are not in fact causal but are the result of chance, co-exposures, or confounding, compared to an analysis that relies solely on the observed patterns seen among the epidemiology studies themselves. Moreover, if the dependence of effect on peak formaldehyde exposures that has been suggested in some epidemiology studies is indeed important in understanding the patterns among human studies, this dependence ought to be reflected in information about dosimetry and hypothesized modes of action. Conversely, if hypothesized modes of action, if operating, would not be expected to produce a dependence on peak exposures, then the role of peaks in explaining positive and null results among human studies is weakened. These issues will be discussed further below.

5. Weight of evidence regarding hematotoxicity from formaldehyde exposure

In the following analysis, we examine a potential association between formaldehyde exposure and leukemia in animals and hematotoxicity as reflected by changes in peripheral blood hematology in both humans and animals. We conducted a literature search, using PubMed, for all human studies measuring or estimating formaldehyde exposure and the changes in peripheral blood hematology, in addition to all short- and long-term animal studies that investigated either potential formaldehyde-associated leukemogenicity or hematology changes in peripheral blood. Search terms included "hematology," "hematotoxicity," "leukemia," "lymphoma," "lymphohematopoietic," "formaldehyde," "rat," "mouse," "rodent," "human," and "occupational." We also relied on references within the papers that we found in the PubMed search and on non-peer-reviewed analyses of animal studies, which are part of the current debate on potential formaldehyde leukemogenicity.

Hematotoxicity may be defined as an insult that can be identified in blood and blood components. This toxicity is reflected in the production or loss of blood components, including red blood cells (RBCs, erythrocytes), white blood cells (WBCs, leukocytes), platelets, and hemoglobin (Hb) (found in RBCs), responsible for carrying oxygen. Hematopoietic progenitor cells (HPCs) in bone marrow give rise to RBCs and WBC subtypes—neutrophils, lymphocytes (B and T types), monocytes, eosinophils, basophils, and megakaryocytes, from which platelets are derived (Cotran et al., 1999). There are three main types of myeloid progenitor cells that undergo several stages of differentiation to give rise to the blood cells (e.g., granulocyte-macrophage colony-forming unit [CFU-GM] gives rise to granulocytes and macrophages) (Cotran et al., 1999). Blood-forming cells normally leave the bone marrow only when fully differentiated, but a small number of progenitor cells can leave the bone marrow and circulate in blood (Aster and Kumar, 1999).

A decline in peripheral blood of one or more WBC type counts can result in leukopenia, that of RBCs, anemia, and when all cell types in peripheral blood decline, pancytopenia. When the oxygen-carrying capacity of RBCs is compromised, new RBCs can be manufactured at a faster-than-usual rate, which may result in larger mean RBC size (mean corpuscular volume, MCV). Numerous factors can influence changes in blood components. These include, but are not limited to, certain infections, nutrient imbalance, xenobiotic insults to either blood components directly or to the bone marrow progenitor cells, alcohol intake, smoking, excessive bleeding, menstruation, and certain medications (Mayo Clinic, 2008). Leukemogenesis may be viewed as a multistage process that involves interruption of the normal cellular differentiation process in the bone marrow and accumulation of the undifferentiated cells in bone marrow, a condition that crowds and suppresses the remaining normal hematopoietic progenitor cells. This suppression of normal hematopoiesis can result in anemia, leukopenia, and pancytopenia. Eventually, the undifferentiated, and abnormally functional, cells in bone marrow spill into peripheral blood and become the predominant cells there (Irons and Stillman, 1996; Aster and Kumar, 1999). As discussed in Section 5.3.3 below, most known leukemogens can cause pancytopenia (a decline in all cell types in peripheral blood) that is secondary to bone marrow toxicity. In addition to pancytopenia, bone marrow toxicity has been associated with decreased counts or viability of circulating blood cells, including progenitor cells (Dempster and Snyder, 1991; Toft et al., 1982, both as cited in ATSDR 2007).

5.1. Formaldehyde hematotoxicity in animals

5.1.1. Hematology

Several animal studies have assessed the hematotoxic potential of formaldehyde via both oral and inhalation routes. As shown in Table 14, these studies ranged from subacute to chronic in duration (4 weeks to 24 months) and used a wide range of exposure concentrations; the highest was 5000 ppm in drinking water (Tobe et al., 1989) and 20 ppm in air (Woutersen et al., 1987).

As shown in Table 14, the results from the inhalation studies generally show that formaldehyde does not induce changes in standard hematology parameters in peripheral blood. One study (Dean et al., 1984) showed a significant ($p < .05$) decrease in monocytes, but not other leukocytes, in the blood of B6C3F1 mice after 3 weeks of exposure to 15 ppm formaldehyde, but no exposure-related changes in either bone marrow cellularity or CFU progenitor cell counts. In contrast, a *longer-term* study by Kerns et al. (1983) in the same mouse strain found no changes in hematology.

The results of the ingestion exposure studies are generally not indicative of a hematotoxic effect of

© 2011 Informa Healthcare USA, Inc.

formaldehyde. For example, the studies by Appelman et al. (1988), Johannsen et al. (1986), and Til et al. (1988, 1989) did not find any changes in hematology with exposure. Tobe et al. (1989) reported statistically significant lowered RBC counts and Hb concentrations, but these changes were not exposure-concentration dependent. The results by Tobe et al. (1989) were contradicted by Vargova et al. (1993), who reported increased hematocrit (Hct) and Hb concentrations and RBC counts in blood. Furthermore, statistically significant changes in WBC counts following oral exposure to formaldehyde were found by Vargova et al. (1993) as increased monocyte counts and decreased lymphocyte counts only following exposure to very high doses (as high as 80 mg/kg for 4 weeks, equivalent to ~800 ppm in drinking water).

5.1.2. Leukemia

We reviewed eight animal studies investigating the tumorigenic potential of formaldehyde by inhalation (Kerns et al., 1983; Kamata et al., 1997; Albert et al., 1982; Feron et al., 1988) and ingestion (Tobe et al., 1989; Til et al., 1989; Takahashi et al., 1986; Soffritti et al., 1989, 2002). We also considered two unpublished and non-peer-reviewed analyses (DeVoney et al., 2010, poster abstract only; Woutersen, 2007) of data from the Battelle Columbus Laboratories (1981) study that was later published by Kerns et al. (1983). Animal inhalation studies generally showed significantly increased rates of nasal tumors (Kerns et al., 1983; Kamata et al., 1997; Sellakumar et al., 1985) but, as shown in Table 15, not of leukemias or lymphomas, when these endpoints were investigated. Ingestion studies showed neither increased rates of nasal nor hematopoietic malignancies in rats. We have limited our review to analyzing potential formaldehyde-associated changes in leukemia and lymphoma rates in the exposed animals. Table 15 describes the animal species, exposure characteristics, and hematopoietic malignancy outcomes of the studies discussed in this section.

Battelle Columbus Laboratories (1981) exposed Fischer 344 rats and B6C3F1 mice via inhalation to formaldehyde for 2 years. As shown in Table 15, tumor incidence data from Battelle Columbus Laboratories (1981) were analyzed by Kerns et al. (1983), DeVoney et al. (2010 poster abstract), and Woutersen (2007) but with different outcomes. Whereas DeVoney et al. (2010 poster abstract) reported elevated lymphoma incidence in female B6C3F1 mice and elevated leukemia incidence in female Fischer 344 rats, Woutersen's analysis found increased lymphoma incidence only in female mice, immediately after exposure, although this trend showed no statistically significant association for formaldehyde when the 3-month period following exposure was considered (Woutersen, 2007). Kerns et al. (1983) reported no formaldehyde-associated elevated rates of leukemia or lymphoma in this study in either rats or mice. It is noteworthy that the leukemias found in Fischer 344 rats by Battelle Columbus Laboratories (1981) likely included mononuclear cell leukemias (MCLs), which are usually observed in ~50% and ~28% of unexposed male and female Fischer 344 rats, respectively (Haseman et al., 1998). This high background incidence of MCLs in rats brings into question the outcomes and resulting conclusions of these non-peer-reviewed results by DeVoney et al. (2006 poster, 2010 poster abstract). Furthermore, as discussed by Ishmael and Dugard (2006), any MCL counterpart in humans is rare, MCLs are more likely elevated by chemical exposure in Fischer rats but not in Osborne Mendel or Sprague Dawley rats, and MCL incidence can be reduced by type of chemical delivery vehicle, such as corn oil, all which suggest that any positive findings involving MCL incidence in animals may not be relevant in humans.

Four studies, all of which used rats, assessed the tumorigenicity of formaldehyde from drinking water. The exposure concentrations were as high as 5000 ppm in water. Exposure durations were either 2 years (Soffritti et al., 1989, 2002; Til et al., 1989; Tobe et al., 1989) or 32 weeks preceded by 8 weeks of treatment with a tumor initiator, *N*-methyl-*N'*-nitrosoguanidine (Takahashi et al., 1986). Of these studies, only Soffritti et al. (1989, 2002) reported statistically significantly increased hematopoietic malignancies (i.e., lymphomas and leukemias).

Soffritti et al. (1989) performed two experiments. In the first experiment, the authors exposed male and female Sprague Dawley rats to 0, 10, 50, 100, 500, 1000, or 1500 ppm of formaldehyde in drinking water for 2 years. The authors reported an increase in leukemia incidence at concentrations above 50 ppm (specifically, lymphoblastic leukemias and lymphosarcomas, immunoblastic lymphosarcomas, and "other leukemias," although the anatomic location of these neoplasms was not indicated). The increase, particularly in immunoblastic lymphosarcomas, was not exposure related, however. Moreover, the lack of statistical analysis of the data in this report does not allow a full assessment of cause and effect. In the second experiment, the authors exposed male and female Sprague-Dawley breeder rats and their male and female offspring to regular drinking water and drinking water containing 2500 ppm formaldehyde. The authors reported increased leukemia rates (specifically immunoblastic lymphosarcomas and "other leukemias") for each of the male and female breeder groups and the male offspring group, but there was no incidence in the female offspring group. The lack of statistical analysis for this experiment also precludes proper data assessment. In their subsequent report of this same study, Soffritti et al. (2002) presented the results from only the first experiment mentioned above, but these results differed from the earlier report by Soffritti et al. (1989). The authors neither explained why they included only one experiment in this report nor addressed the differences in reported outcomes between reports.

5.2. Formaldehyde hematotoxicity in humans

There are limited, mostly occupational, studies in humans of the hematotoxic effects of exposure to formaldehyde. Tang et al. (2009) recently abstracted data from eight studies conducted in China that assessed WBC and platelet counts and Hb concentration in subjects occupationally exposed to formaldehyde (Yang et al., 2007; Kuo et al., 1997; Cheng et al., 2004; Xu et al., 2007; Qian et al., 1988; Feng et al., 1996; Tang et al., 2003; Tong et al., 2007, all as cited by Tang et al., 2009). The findings by Kuo et al. (1997), the only study of hematological effects cited by Tang et al. (2009) available in English, are associated with several uncertainties that weaken the conclusions drawn by Tang et al. (2009) about these effects. Many questions arise about the outcomes and exposure-related uncertainties in the findings of the other, untranslated, studies cited by Tang et al. (2009) (Section 5.3.2 provides further discussion of this point). We found four other studies that assessed hematological parameters associated with formaldehyde exposure (Ye et al., 2005; Lyapina et al., 2004; Srivastava et al., 1992; Zhang et al., 2010b). Table 16 describes the exposure characteristics and hematology outcomes, as available, of the studies discussed in this section.

In the China-based hematotoxicity studies reported in Tang et al. (2009), the leukocyte counts in the exposed subjects were generally lower than in the control subjects, but differences were statistically significant in only four (Yang et al., 2007; Cheng et al., 2004; Qian et al., 1988; Tong et al., 2007, all as cited by Tang et al., 2009) of the eight studies. Feng et al. (1996, as cited by Tang et al., 2009), Xu et al. (2007, as cited by Tang et al., 2009), and Tang et al. (2003, as cited by Tang et al., 2009) did not find statistically significant differences in leukocyte counts related to formaldehyde exposure. Several studies found an inverse correlation between duration of exposure to formaldehyde and leukocyte counts (Tong et al., 2007; Yang et al., 2007; Tang et al., 2003; Kuo et al., 1997, all as cited by Tang et al., 2009), but the relationship was only reported as statistically significant in the study by Kuo et al. (1997, as cited by Tang et al., 2009).

Ye et al. (2005) assessed lymphocyte subset counts in peripheral blood in student controls (living in dorms), factory workers (8.6 years mean duration of exposure), and ballroom waiters (12 week exposure duration), all non-smokers. The formaldehyde concentration was 0.8 ppm in the factory, 0.09 ppm in the ballroom, and 0.009 ppm in the dorms. As shown in Table 16, differences in percentage of lymphocyte subset counts among groups were limited to statistically significantly increased B lymphocytes, and decreased CD3 (total T cells) and CD8 (T-cytotoxic), but not CD4 (T-helper-inducer), T lymphocytes, in the workers as compared with the students. The change in CD3 cells (~5% decrease) appeared to be driven by CD8 cells (~25% decrease). It is not apparent why CD3 and CD8 cells decrease, but not CD4. Since total T cells were decreased, however, it is possible that there was no need for the helper T cells (CD4) and therefore no change was observed relative to controls. Nevertheless, this is an interesting finding by Ye et al. (2005) and warrants further investigation of lymphocyte subset dynamics. Further, since this study shows statistically significant changes in both B- and T-cell populations, it is likely that any effect attributed to formaldehyde exposure is immune and acquired and did not originate from an insult to the bone marrow. It would be interesting to know the counts in peripheral blood of RBCs and WBCs, other than lymphocytes, to shed light on potential bone marrow involvement, but these data were not available for this study.

Lyapina et al. (2004) found no significant differences in standard hematology tests of workers applying formaldehyde-carbamide glue when compared with those from subjects with no known appreciable formaldehyde exposure (formaldehyde levels not reported for either group). The authors reported a statistically significant inverse relationship, however, between the duration of occupational exposure to formaldehyde and RBC count, but there was no relationship with WBC count. Similarly, Srivastava et al. (1992) found decreased Hb concentrations in the blood of three of six workers who were involved in producing and preparing melamine-formaldehyde resin. However, it is not readily apparent whether the differences in RBC count and hemoglobin are related to formaldehyde exposure since this study included a limited number of subjects. In addition, this study found increased total lymphocytes (>3200 per mm^3 of blood) in three of six workers. The authors of this study indicated that the subjects were exposed to relatively high levels of formaldehyde at different times during the day.

Recently, Zhang et al. (2010b) investigated the associations between formaldehyde exposure and various hematology parameters in subjects working with formaldehyde-melamine resin in two factories in China (median, 1.28 ppm formaldehyde) as compared with volunteer subjects from three other factories with lower formaldehyde levels (median, 0.026 ppm). Formaldehyde can dissociate from the melamine resin, become airborne, and be inhaled by the factory workers. The workers were possibly also exposed to formaldehyde dermally, and subject to potential formaldehyde-induced skin reactions, if they touched the resin with their bare skin. The authors assessed personal workplace exposures to formaldehyde in air on 3 days for a full shift (>6 hour/shift) and to other volatile organic compounds (VOCs), including benzene, on two or three occasions within a 3-week period. The authors matched exposed and control subjects by age and sex; however, there were considerably different rates of current alcohol drinkers (yes/no answer) and recent respiratory infections (yes/no answer) in the exposed as compared with the control subjects (26% vs. 41% and 40% vs. 29%, respectively). Alcohol intake and recent respiratory infections can influence WBC counts. As shown in Table 16, Zhang et al. (2010b) report statistically significant lower counts of total WBCs, lymphocytes, granulocytes, platelets, and RBCs in addition to a higher MCV,

© 2011 Informa Healthcare USA, Inc.

Table 14. Formaldehyde animal hematotoxicity studies.

			Exposure		Outcomes	
Study	Species	Sex	Concentration	Duration	RBC, Hct, Hb	WBC
Inhalation studies						
Monticello et al., 1989	Monkey, rhesus	Male	0, 6 ppm	6 h/d, 5 d/wk, 6 wk	NS	NS
Appelman et al., 1988	Rat, Wistar	Male	0, 0.1, 1, 10 ppm	6 h/d, 5 d/wk, 13 or 52 wk	NS	NS
Holmstrom et al., 1989	Rat, Sprague-Dawley	Female	0, 12.6 ppm	6 h/d, 5 d/wk, 22 mo	NS	NS
Kerns et al., 1983	Rat, Fischer 344	Female, Male	0, 2.0, 5.6, 14.3 ppm	6 h/d, 5 d/wk, 24 mo with follow-up till 30 mo	NS	NS
As above	Mouse, B6C3F1	Female, Male	As above	As above	NS	NS
Dean et al., 1984	Mouse, B6C3F1	Female	0, 15 ppm	6 h/d, 5 d/wk, 3 wk	NS	↓ monocytes Other cell types NS
Kamata et al., 1997	Rat, Fischer 344	Male	0, 0.3, 2, 15 ppm	6 h/d, 5 d/wk, 28 mo	NS	NS
Woutersen et al., 1987	Rat, Wistar	Female, Male	0, 1, 10, 20 ppm	6 h/d, 5 d/wk, 13 wk	NS	NS
Ingestion Studies						
Vargova et al., 1993	Rat, Wistar	Male	0, 20, 40, and 80 mg/kg bw/d	5 d/wk, 4 wk; gastric intubation	↑ at 40, 80 mg/kg/d	↑ monocytes ↓ lymphocytes
Til et al., 1989	Rat, Wistar	Female	0, 1.8, 21, 109 mg/kg bw/d	Daily, 24 mo, drinking water	NS	NS
As above	Rat, Wistar	Male	0, 1.2, 15 or 82 mg/kg bw/d	Daily, 24 mo, drinking water	NS	NS
Til et al., 1988	Rat, Wistar	Female, Male	5, 25 and 125 mg/kg bw/d	Daily, 4 wk	NS	NS
Johannsen et al., 1986	Rat, Sprague-Dawley	Female, Male	0, 50, 100, 150 mg/kg/d	Daily, 3 mo	NS	NS
As above	Dog, Beagle	Female, Male	0, 50, 75, 100 mg/kg/d	Daily, 3 mo	NS	NS
Tobe et al., 1989	Rat, Wistar	Female, Male	0, 200, 1000, 5000 ppm (0, 10, 50, 300 mg/kg bw/d)	Daily, 24 mo, drinking water	↓, not concentration dependent	Assessed, NR

Note. WBC = white blood cell count in peripheral blood; RBC = red blood cell count in peripheral blood; Hct = hematocrit; Hb = hemoglobin concentration in blood; NS = not statistically significant; ↓ = statistically significant decrease; ↑ = statistically significant increase; mg/kg bw/d = milligram per kilogram body weight per day; ppm = parts per million; wk = week(s); mo = month(s).

but not monocytes or Hb concentration in exposed subjects relative to controls. In addition, the authors found no statistically significant difference in the growth of circulating CFU-GM hematopoietic progenitor cells ex vivo between exposed and control subjects.

In the following section, we weigh the evidence from human and animal studies discussed in Sections 5.1 and 5.2 to assess the likelihood of a hematotoxic role for formaldehyde.

5.3. HBWoE evaluation of formaldehyde hematotoxicity studies

Based on the available data summarized above, some have hypothesized that formaldehyde may cause hematotoxicity and leukemia in humans. We ask the following questions with regard to this hypothesis:

1. Do animal studies suggest formaldehyde exposure is causally associated with hematotoxicity and leukemia?
2. What do the human studies tell us about potential formaldehyde hematotoxicity in humans? Are the results of human studies consistent with those of animal studies?
3. What is known about the hematotoxicity of known leukemogens (e.g., benzene) from animal and human studies and how does that compare to formaldehyde?
4. Are there alternative explanations for decreased WBC and RBC counts?

As a whole, considering these questions allows for an assessment of the extent to which the hematotoxicity and animal leukemia data support either a causal association between formaldehyde exposure and leukemia or an alternative hypothesis. Importantly, one needs to consider the hematotoxicity and animal leukemia data in the context of the epidemiology and mode-of-action data, as each of the three lines of evidence inform interpretation of each other.

© 2011 Informa Healthcare USA, Inc.

Table 15. Formaldehyde animal carcinogenicity studies.

Study	Species	Sex	Exposure		Tumor rate vs. control (%)	Background tumor rate
			Concentration	Duration	Hematopoietic malignancies	Mean (range)
Inhalation studies						
Albert et al., 1982	Rat, Sprague-Dawley	Male	0, 14.7 ppm	6 h/d, 5 d/wk, 19.4 mo	NR (Authors performed complete necropsy and histological sections taken from organs with gross pathological alterations)	
As above	As above	Male	As above	As above	NS	
Kerns et al., 1983	Rat, Fischer 344	Female	0, 2.0, 5.6, 14.3 ppm	6 h/d, 5 d/wk, 24 mo with follow-up till 30 mo	NS	
As above	As above	Male	As above	As above	NS	
As above	Mouse, B6C3F1	Female	As above	As above	NS	
As above	As above	Male	As above	As above	NS	
DeVoney et al., 2006 poster, 2010 poster abstract (analysis of Battelle Columbus Laboratories [1981])*	Rat, Fischer 344	Female	0, 2.0, 5.6, 14.3 ppm	6 h/d, 5 d/wk, 24 mo with follow-up till 30 mo	Leukemia; 24% vs. 15% (14.3 ppm) (no statistical test identified)	Leukemia; 37.3% (24%–54%)†
As above	As above	Male	As above	As above	NS	
As above	Mouse, B6C3F1	Female	As above	As above	Lymphoma; 28% vs. 18% (14.3 ppm) (no statistical test identified)	Lymphoma; 19.9% (6%–44%)†
As above	As above	Male	As above	As above	NS	
Woutersen, 2007* (analysis of Battelle Columbus Laboratories [1981])	Rat, Fischer 344	Female	0, 2.0, 5.6, 14.3 ppm	6 h/d, 5 d/wk, 24 mo with follow-up till 30 mo	NS	
As above	As above	Male	As above	As above	NS	
As above	Mouse, B6C3F1	Female	As above	6 h/d, 5 d/wk, 24 mo with follow-up till 27 mo	Lymphoma; considered data immediately after 24-month exposure: 17%, 16%, 9%, 29% (0, 2.0, 5.6, 14.3 ppm, respectively);considered data 3 months after 24-month exposure: 50%, 20%, 15%, 45% (0, 2.0, 5.6, 14.3 ppm, respectively)	Lymphoma; 19.9% (6%–44%)†
As above	As above	Male	As above	As above	NS	
Sellakumar et al., 1985 (extended analysis of the study by Albert et al. [1982]).	Rat, Sprague-Dawley	Male	0, 14.8 ppm (with or without ~10 ppm HCl)	6 h/d, 5 d/wk, >28 mo	NS	
Kamata et al., 1997	Rat, Fischer 344	Male	0, 0.3, 2, and 15 ppm	6 h/d, 5 d/wk, 28 mo	NR (No hematological changes were found. Also, the authors examined femur, mesenteric lymph nodes, many other organs, and "any other gross lesions"	

Table 15. continued on next page

Table 15. continued..

Study	Species	Sex	Exposure		Tumor rate vs. control (%)	Background tumor rate
			Concentration	Duration	Hematopoietic malignancies	Mean (range)
Feron et al., 1988	Rat, Wistar	Male	0, 10, 20 ppm	6 h/d, 5 d/wk, (4, 8 or 13 wk); follow-up >28 mo	NR (Animals were autopsied and examined for gross pathological changes. The authors found no gross pathological changes)	
Ingestion studies						
Til et al., 1989	Rat, Wistar	Female	0, 1.8, 21, 109 mg/kg bw/d (0, 20, 260, 1900 ppm)	Daily, 24 mo, drinking water	NS	
As above	Rat, Wistar	Male	0, 1.2, 15, 82 mg/kg bw/d (0, 20, 260, 1900 ppm)	As above	NS	
Soffritti et al., 1989	Rat, Sprague-Dawley	Female	Experiment 1: 0, 10, 50, 100, 500, 1000, 1500 (rats were 7 weeks old at start) Experiment 2: 0, 2500 ppm (rats were 25 week old breeder and their offspring)	Daily, 24 mo, drinking water with lifetime follow-up >36 mo	All hematopoietic malignancies Experiment 1: 14% vs. 3% Experiment 2: Breeders: 11.1% vs. 5% Offspring: 0% vs. 6.1%	Up to 19% in males and 14% in males and females combined‡
As above	Rat, Sprague-Dawley	Male	As above	As above	All hematopoietic malignancies Experiment 1: 22% vs. 4%) Experiment 2: Breeders: 11.1% vs. 0% Offspring: 11.1% vs. 5.1%	Up to 19% in males and 14% in males and females combined‡
Tobe et al., 1989	Rat, Wistar	Female	0, 200, 1000, 5000 ppm	Daily, 24 mo, drinking water	NR (No hematological changes were found. Also, authors examined lymph nodes and several other organs and "tumorous tissues")	
As above	As above	Male	As above	As above	As above	
Takahashi et al., 1986	Rat, Wistar	Male	0, 5000 ppm	Daily, 32 wk after 8 wk exposure to *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG), drinking water	NR (Animals were necropsied and "no malignant tumors found outside the gastroduodenal tract")	

Note. NR = not reported; NS = not statistically significant; NA = not applicable/available; mg/kg bw/d = milligram per kilogram body weight per day; ppm = parts per million; wk = week(s); mo = month(s).

*To adjust for early deaths, Woutersen (2007) used the Peto mortality prevalence trend test, and DeVoney (2006 poster, 2010 poster abstract) "adjusted for early deaths and time to tumor observation." It should be noted that these two references are a conference presentation and a conference poster, respectively, and are not peer-reviewed publications.

†Data from Haseman et al. (1998); background tumor rates from NTP studies data based on spontaneous tumor rates in approximately 1000 animals.

‡Data from the review by Feron et al. (1990) of background leukemia incidence in rats from the same colony used in the study by Soffritti et al. (1989).

5.3.1. Key animal studies do not provide strong evidence of an association between formaldehyde exposure and hematotoxicity and leukemia

5.3.1.1. Hematology Known leukemogens, such as benzene, can cause bone marrow toxicity, which affects the ability of bone marrow cells to produce blood-forming cells (ATSDR, 2007). As discussed in the introduction to this section, this toxicity can be manifested in bone marrow suppression and pancytopenia, a generalized decrease of blood cellular components. This insult to bone marrow may progress to a malignancy that shows a predominance in production of one or more cell types in bone marrow that spill into peripheral blood.

We examined the available formaldehyde animal studies for signs of hematotoxicity as reflected in peripheral blood. Of 12 studies we reviewed, 9 reported no change in hematology parameters (see Table 14). These studies, which ranged from exposures lasting a few weeks to longer than 2 years, spanned a range of concentrations and durations that would be sufficient to show any changes in hematology.

Three studies reported a change in one or more hematology parameter: Dean et al. (1984) by inhalation, and Tobe et al. (1989) and Vargova et al. (1993) by oral exposure. However, these outcomes were mixed. For example, Dean et al. (1984) reported a decrease in monocytes but no other hematology parameter, whereas Vargova et al. (1993) reported an *increase* in monocytes but a decrease in lymphocyte counts. Vargova et al. (1993) also found increased Hct, Hb concentration, and RBC counts in blood, in contrast to results from Tobe et al. (1989), which indicated *decreased* Hb concentrations and RBC counts. It is noteworthy that the changes in the aforementioned studies resulted from very high exposures, particularly in oral exposure studies, ranging from 40 to 300 mg/kg body weight/day (approximately 800–5000 ppm in drinking water), as noted in Table 14.

Overall, the hematological ingestion and inhalation studies of formaldehyde we reviewed are inconsistent and eclipsed by overwhelming evidence from the same and other species of animals that show no change in hematology parameters. When found, the statistically significant changes are likely not related to formaldehyde exposure, particularly because they arise among many other statistically insignificant associations. Further, if bone marrow toxicity had occurred, it is likely that declines in more than one blood cell type would have been observed, such as is established for benzene (discussed in Section 5.3.3), and that was not reported by the authors of any of the studies we reviewed.

5.3.1.2. Leukemia We also analyzed the outcomes of animal studies that examined the carcinogenicity of formaldehyde by inhalation or ingestion. The majority of these studies, listed in Table 15 and discussed in Section 5.1, found no excess hematopoietic malignancies associated with formaldehyde exposure (Albert et al., 1982; Kerns et al., 1983; Sellakumar et al., 1985; Kamata et al., 1997; Feron et al., 1988; Til et al., 1989; Tobe et al., 1989; Takahashi et al., 1986). However, the studies by Soffritti et al. (1989, 2002), the unpublished data from the DeVoney et al. (2010 poster abstract), and Woutersen's (2007 presentation) analyses of data from Battelle Columbus Laboratories (1981) found increased incidence of hematopoietic malignances from formaldehyde ingestion and inhalation, respectively.

Soffritti et al. (1989) performed two carcinogenicity experiments by ingestion and reported statistically significantly increased hematopoietic malignancies in two reports (Soffritti et al., 1989, 2002) described in Section 5.1 and Table 15, the results of which are inconsistent and have been criticized by both ATSDR (1999) and the European Food Safety Authority (EFSA, 2006) as unreliable. The lack of confidence in the results by Soffritti et al. (1989, 2002) stems, in part, from concerns about the rodent colony where the experiments occurred. Feron et al. (1990) suggested that the elevated leukemia incidence might have been "unrelated to formaldehyde ingestion," because of the wide range of incidence rates of hematopoietic malignancies in control animals from the same colony—as high as 19%. Moreover, a possible infection of the rat colony by *Mycoplasma pulmonis* (an organism that preferentially colonizes the respiratory tract in rats and secretes substances that can promote mitogenesis in lymphocytes) also has been presented as a potential confounder for hematolymphopoietic malignancies reported by Soffritti et al. (1989, 2002).

Other studies examining formaldehyde carcinogenicity via the oral route did not indicate an increased incidence of hematopoietic malignancies relative to control exposures. The longer-term carcinogenicity studies by both Til et al. (1989) and Tobe et al. (1989) showed no increase in these malignancies after 2 years of exposure and follow-up, consistent with Takahashi et al. (1986) who only followed animals for 40 weeks.

As reported in conference posters, DeVoney et al. (2006 poster, 2010 poster abstract) reevaluated data from Battelle Columbus Laboratories (1981) and reported increased lymphoma incidence rates for female B6C3F1 mice exposed to ~15 ppm by inhalation (28% vs. 18% in exposed vs. control mice, respectively). According to Haseman et al. (1998), the background rate for lymphoma in these female mice, based on 1092 control mice used in National Toxicology Program (NTP) studies, is 19.9%, with a range of 6% to 44% in all studies examined. Therefore, the rates reported by DeVoney et al. (2006 poster, 2010 poster abstract) fall within background tumor rates for the rodent species use in Battelle Columbus Laboratories (1981) and do not provide sufficient evidence for formaldehyde leukemogenicity. These non-peer-reviewed results by DeVoney et al. (2006 poster, 2010 poster abstract) are contrasted by another evaluation of Battelle Columbus Laboratories (1981) data by Woutersen (2007) (as presented at the Formaldehyde International Science Conference), who found a statistically significant increased trend in lymphoma for

© 2011 Informa Healthcare USA, Inc.

Table 16. Human hematotoxicity studies.

Study	Exposure (ppm)	Number of persons (exposed, control)	% change in WBC count between groups or % of subjects with decreased hematology parameters	Peripheral blood changes			Other findings
				Total WBC	Platelets	Hb	
Yang et al., 2007*	0.018–0.036	239, 200	WBCs: 14% (E) vs. 4% (C) Platelets: 11% (E) vs. 1% (C) RBCs: 32% (E) vs. 21.5% (C)	↓↓	↓↓	↓↓	
Kuo et al., 1997	ND–0.054 (personal samples) 0.006–0.237(area samples)	50, 71	NA	↓	NS	NS	Statistically significant (−0.33, $p < .05$) inverse relationship between FA and WBC counts, but not 11 other hematology parameters
Cheng et al., 2004*	0.2–0.76	72, 150	WBCs: 14% in E vs. 5% in C	↓	NA	NA	
Lyapina et al., 2004	0.52–1.049 (mean, 0.71)	29, 21	NA	NS	NS	NR	Statistically significant inverse relationship between duration of exposure to formaldehyde and RBC counts and Hct
Xu et al., 2007*	0.36–5.56	10, 10	−11.4	NS	NS	NS	
Ye et al., 2005	0.8 (8-h TWA), 1.38 (max) in workers; vs. 0.009 (mean), 0.012 (max) in controls	36, 6	NA	NA	NA	NA	Workers vs. students in dorms. Statistically significantly increased B lymphocytes. Statistically significantly decreased CD3 and CD8 but not CD4 T lymphocytes in peripheral blood
As above	0.09 (5-h TWA), 0.24 (max); vs. 0.009 (mean), 0.012 (max) in controls	18, 6	NA	NA	NA	NA	Waiters vs. students in dorms. No change in B or T lymphocytes counts in peripheral blood
Zhang et al., 2010b	0.63–2.51 (mean, 1.28)	43, 51	−13.5	↓↓	↓	NS	Decreased RBC; increased MCV (statistically significant)
Qian et al., 1988*	2.44 (estimated)	55, 41	−13.3	↓↓↓	NA	NA	Increase in immunoglobulins (Ig) IgM and IgA, and eosinophils (no statistical significance reported)
Feng et al., 1996*	0.57–15.61	104, 68	NA	NS	NA	NS	
Srivastava et al., 1992	NR	6, 0	Increased blood lymphocyte counts in 3 of 6 subjects	NA	NA	Decrease	Decreased Hb in 4 of 6 subjects
Tang et al., 2003*	NR	110, 120	−17.1	NS	NA	NA	Decreased WBC count with increasing work years (no statistical significance reported)
Tong et al., 2007*	NR	65, 70	−18	↓↓↓	↓↓↓	NS	WBC and platelet counts decreased with increasing work years (no statistical significance reported)

Note. E = exposed group; C = control group; NR = not reported; NS = not statistically significant; NA = not applicable/available; WBC = white blood cell count in peripheral blood; RBC = red blood cell count in peripheral blood; Hct = hematocrit; Hb = hemoglobin content of RBCs; ↓ = statistically significant decrease ($p < .05$); ↓↓ = statistically significant decrease ($p < .01$); ↓↓↓ = statistically significant decrease ($p < .001$).
* As cited in Tang et al. (2009). These studies are in Chinese and are not available on PubMed.

only female mice among the rodent species and sexes examined. Moreover, there was no statistically significant association with formaldehyde when post-exposure follow-up data (3-month period) was considered. When considering the three distinct analyses of the Battelle Columbus Laboratories (1981) carcinogenicity data, in addition to the high background tumor rates in the rodent species examined, it becomes less likely that the

reported inconsistent increased lymphoma incidence is related to formaldehyde exposure.

The rates presented by DeVoney et al. (2006 poster, 2010 poster abstract) for hematopoietic malignancies in female Fischer 344 rats demonstrate no dose-response for leukemia; the authors found 25%, 23%, and 24% leukemia incidence in rats exposed to 2, 6, and 15 ppm formaldehyde vs. 15% in control rats. Similarly, the leukemia rates in both exposed and control rats were either below or within the range of leukemia incidence in the controls used in NTP studies, which Haseman et al. (1998) lists as 37% (range, 24% to 54%) for leukemia in female Fischer 344 rats. Noteworthy is that the leukemias found in Fischer 344 rats likely included MCLs, which are usually observed in ~50% and ~28% of unexposed male and female Fischer 344 rats, respectively (Haseman et al., 1998). This high background incidence of MCLs in rats brings into question the validity of the unpublished results by DeVoney et al. (2006 poster; 2010 poster abstract). Further, as discussed in Section 5.1.2, above, MCLs are not observed in humans, and findings involving MCL incidence in animals may not be relevant in humans.

The results by Soffritti et al. (1989, 2002) and the unpublished reanalysis by DeVoney et al. (2006 poster; 2010 poster abstract) of the Battelle Columbus Laboratories (1981) data do not provide appreciable support for formaldehyde-induced leukemia in rodents. These results are unlikely to indicate formaldehyde leukemogenicity, particularly when weighed against the relatively high background rates of hematopoietic malignancies in the mouse and rat species used in these studies and the overwhelmingly negative results from nine other carcinogenicity studies in rats and mice. Finally, it is not surprising that most studies show no change in leukemia incidence with formaldehyde exposure since most studies, short- and long-term, showed no change in hematology parameters, which are important precursors in the chain of events for chemically induced leukemia.

5.3.2. Key human studies do not provide strong evidence of an association between formaldehyde exposure and hematotoxicity

The studies that associate hematology parameters in humans with formaldehyde exposure are generally cross-sectional in nature. Cross-sectional environmental toxicology studies frequently involve concurrent observation of a biological endpoint and exposure to an environmental agent at a single point in time or over a short time duration. Except for a few studies, no information was available on either the methods of exposure assessment of formaldehyde or the assessment of potential confounding effects from known hematotoxicants such as benzene. For some studies, formaldehyde exposure information was absent altogether (see Table 16).

We investigated whether a possible exposure-response pattern existed between the reported airborne formaldehyde concentrations and the reported hematology responses in the studies that we reviewed. The results of studies based in China and cited by Tang et al. (2009) do not show an exposure-response relationship between formaldehyde concentrations and hematology parameters. For example, the study by Feng et al. (1996, as cited by Tang et al., 2009) reportedly showed no association between very high formaldehyde exposures (range, 0.57–15.61 ppm) and either changes in WBC counts or Hb concentrations in peripheral blood. Xu et al. (2007, as cited by Tang et al., 2009) also found no significant differences in Hb concentrations or WBC and platelet counts in association with relatively elevated formaldehyde exposures ranging from 0.36 ppm to 5.56 ppm. In contrast, Qian et al. (1998, as cited by Tang et al., 2009) estimated formaldehyde exposure to be 2.44 ppm, and found a statistically significant association between this concentration and a lower WBC count. In addition, some studies with *lower* formaldehyde exposures were statistically significantly associated with *decreased* WBC counts and other hematology parameters. For example, Cheng et al. (2004, as cited by Tang et al., 2009) found decreased WBC counts in individuals exposed to formaldehyde concentrations ranging from 0.2 ppm to 0.76 ppm, whereas Yang et al. (2007, as cited by Tang et al., 2009) found similar associations at lower concentrations not exceeding 0.036 ppm. When examining the hematology outcomes of the studies we reviewed (shown in Table 16), we found no consistent exposure-dependent pattern in either qualitative or quantitative changes.

Moreover, using the data presented by Tang et al. (2009), we determined the percent change or difference in total WBC counts between exposure groups when these counts were available; as shown in Table 16, WBC counts were between 11% and 19% lower in exposed vs. control subjects. However, when examining the studies altogether, we did not find dose dependency in the WBC, platelet, or Hb associations with formaldehyde exposure measurements.

Some of the Chinese occupational studies that reported significantly lower WBC concentrations in exposed subjects also reported formaldehyde concentrations in air that were lower than those expected outside of work. For example, the air concentration ranges reported by Zhang et al. (2010b) (median, 0.026 ppm), by Yang et al. (2007, as cited by Tang et al., 2009) (0.018–0.036 ppm), and by Kuo et al. (1997) (ND–0.054 ppm) overlap with concentrations reported in indoor public places in several Chinese cities (0.12 ppm [range, 0.02–0.31 ppm] as reported by Tang et al. [2009] from the Chinese Ministry of Health). These data suggest that results based on workplace exposure to formaldehyde may be confounded by non-occupational exposures, which can be as high or even higher than the occupational exposures. Appreciable non-occupational sources of formaldehyde exposure exist, particularly in China, where most of the human formaldehyde hematology studies have been conducted. Formaldehyde concentrations in indoor air of homes have been measured

© 2011 Informa Healthcare USA, Inc.

up to 0.5 ppm in China (mean, 0.19 ppm) and certain dietary items have been found to contain several hundreds or even thousands of milligrams of formaldehyde per kilogram (as reviewed by Tang et al., 2009). The authors' lack of accounting for non-occupational sources of formaldehyde exposure adds to the uncertainties in the study outcomes and diminishes their credibility for a hematotoxic role for formaldehyde.

There are many uncertainties about the Chinese hematology studies as reported by Tang et al. (2009). For example, Tang et al. (2009) suggest that Kuo et al. (1997) shows a statistically significant inverse relationship between formaldehyde concentrations and WBC counts. Although Kuo et al. (1997), the only study among those cited by Tang et al. (2009) available in English, indeed shows such an association (−.33, $p < .05$), it does not show a significant relationship between exposure measurements and 11 other hematology parameters, including RBC and individual WBC-type counts. Further, the authors collected peripheral blood on two occasions 1 year apart and found associations only with the second but not the first blood sample. Subjects in the Kuo et al. (1997) study were employed for an average of 3 years at the study locations; if the association between formaldehyde exposure and this lone hematology parameter is real, then it should have been consistent in both blood samples collected. When considering the uncertainties associated with this readily obtainable study (Kuo et al., 1997), many questions arise about the full outcomes and exposure-related uncertainties in the findings of the other studies cited by Tang et al. (2009) (which are unavailable in English).

The most recent study to assess hematotoxicity in humans exposed to formaldehyde is Zhang et al. (2010b). The study provides some associations between formaldehyde exposure and changes in hematology, but also demonstrates serious weaknesses in the study design. The hematology findings by Zhang et al. (2010b) do not consistently support a hematotoxic role for formaldehyde. For example, the authors found statistically significant lower WBC, RBC, and platelet counts in exposed vs. control factory workers, yet they found no statistically significant difference between exposed and control subjects in relation to colony formation of myeloid progenitor cells (CFU-GM), which give rise to granulocytes and macrophages, cultured from blood. If it were possible for formaldehyde to cause direct or indirect toxicity to bone marrow, a decreased ability of CFU-GM to grow in culture would likely be observed, but this did not occur.

It is not clear from the Zhang et al. (2010b) data whether all subjects with decreased WBC counts also had decreased RBC counts and vice versa. If bone marrow toxicity was indeed in progress in the subjects exposed to higher levels of formaldehyde, then both WBC and RBC counts would be lower in the same individuals. In addition, because the WBC and RBC counts were pooled, it is impossible to determine if outliers in either group might have influenced the results, since subject-specific hematology parameters are not reported. Therefore, one cannot make definitive conclusions concerning these data.

We investigated the consistency between the animal and human study outcomes. As we discuss, animal studies generally show no evidence of formaldehyde-induced hematotoxicity. Human studies, on the other hand, show inconsistent associations between formaldehyde exposure measurements and hematologic parameters. Even when hematology changes, such as depressed WBC counts in the blood, are associated with formaldehyde, these associations are not exposure related, and do not agree with the findings from animal studies. This lack of concordance between human and animal studies does not provide evidence to support an argument for formaldehyde-induced hematotoxicity, unlike benzene and other leukemogens that show concordance between animal and human data (as discussed in Section 5.3.3). However, the limited number of human studies available, and the inconsistencies among them, warrants the need for well-controlled human studies with respect to exposure assessment and subject-matching between the exposed and control groups.

The available human hematotoxicity studies are cross-sectional in nature. In cross-sectional studies, both exposure and outcome are evaluated at the same time. A major weakness of cross-sectional environmental toxicology studies is that a chemical measurement at one point of time may not be indicative of earlier exposures that may have caused the biological outcome. Also, inappropriate subject-group matching (e.g., for smoking, drinking, age, sex) may result in findings of differences in biological outcomes associated with the agent in question when there are in fact none. In particular, Zhang *et al.* (2010b) only matched subjects by age and sex; however, there was a considerably higher rate of recent respiratory infections (yes/no answer) in the exposed vs. control subjects (40% vs. 29%, respectively). Zhang et al. (2010b) report that subjects were screened by physicians and trained questionnaire administrators. However, no listing of medications or medical conditions is available for the subjects in this study. We discuss possible confounders in the subsequent sections, particularly from dermal exposure to formaldehyde and respiratory infections that can possibly modulate the associations in the studies we reviewed for human formaldehyde hematotoxicity.

5.3.3. If formaldehyde causes leukemia in humans, it is likely due to a mechanism that is different from that observed with known leukemogens

Hematotoxicity has been demonstrated in both animals and humans exposed to leukemogens (i.e., benzene, chemotherapeutic alkylating agents, and x-ray and gamma radiation). This hematotoxicity can be illustrated with benzene. Benzene has been frequently found to cause pancytopenia in animals (e.g., Aksoy et al., 1972; Farris et al., 1997, both as cited in ATSDR, 2007) and humans (e.g., Kipen et al., 1989; Schnatter et al., 2010;

ATSDR, 2007). Moreover, numerous studies have shown that exposure to benzene can cause leukemia and other hematopoietic malignancies in animals by inhalation and oral routes in several rodent species of both sexes (Snyder et al., 1984; Cronkite et al., 1984, 1985, 1989; all as cited in ATSDR, 2007). Many epidemiology studies have also shown robust associations between exposure to benzene and increased risk of leukemia (e.g., Rinsky et al., 1987, 2002; Yin et al., 1996; Infante et al., 1977; ATSDR, 2007). The mechanism for benzene hematotoxicity and leukemogenicity is dependent on its metabolism to reactive intermediates and is well established as having the ability to affect bone marrow cells directly (ATSDR, 2007).

Other leukemogens have cytotoxic and genotoxic properties similar to benzene. These agents can affect all cells, particularly rapidly dividing cells such as bone marrow. For example, cyclophosphamide, a chemotherapeutic alkylating agent, has produced hematopoietic malignancies in exposed animals (Schmahl and Habs, 1979) as well as leukopenia in humans (Bower et al., 2004; Tjan-Heijnen et al., 2001) and animals (Wang et al., 2002; Nohynek et al., 1997). X-ray and gamma-radiation also have been repeatedly shown to cause leukemia and bone marrow toxicity in animals and humans (IARC, 2000).

As discussed in the preceding section, the available human studies lack the appropriate exposure and subject information and the consistent outcomes to make a convincing case for formaldehyde leukemogenicity. Upon reviewing the available studies of formaldehyde hematology effects, we found no consistent evidence of hematotoxicity in humans. Moreover, the animal studies using mice, rats, dogs, and monkeys, often of both sexes, overwhelmingly reported no evidence of changes in hematology parameters, as shown in Table 14. Therefore, if formaldehyde causes leukemia in humans, it must be by a mechanism that is different from that observed with known leukemogens and is likely specific to humans and not common to rodents. Table 17 illustrates the divergence of formaldehyde from known leukemogens in terms of hematotoxicity indicators.

5.3.4. There are alternative explanations for the pancytopenia reported by Zhang et al. (2010b) and the leukopenia reported by other studies

As reviewed above, animal studies generally do not show a hematotoxic effect of formaldehyde. However, several human studies report that formaldehyde is associated with lowered WBC and RBC counts in peripheral blood (see Table 16). Our review of available human studies of formaldehyde hematotoxicity finds that these studies do not sufficiently explain some of the formaldehyde-associated depression in RBC and WBC counts in "exposed" vs. "control" subjects. When we consider the uncertainties in these associations and the absence of a clear dose-response in the available studies, we find that there are many potential confounders to a possible formaldehyde-associated decline in WBC and RBC counts in blood. Some of these confounders include inappropriate matching by exposure due to consideration of only airborne measurements of formaldehyde (and not dermal or oral), no reported assessment of non-occupational exposures to formaldehyde, differences among groups in alcohol intake and respiratory infections, and the possible effect of formaldehyde on hematology parameters in peripheral blood (via dermal irritation and sensitization, as discussed in Section 5.3.4.1) in addition to other issues such as nutrient imbalance and certain medications, all of which may have a significant effect on hematology parameters. Further, unlike established leukemogens such as benzene (as discussed in Section 5.3.3), formaldehyde has not been associated with bone marrow toxicity or aplastic anemia in occupationally exposed subjects.

5.3.4.1. Subjects exposed to formaldehyde share common immunology markers with subjects having dermatitis or other inflammatory conditions As reviewed by Deane and Hickey (2009), epidermal inflammation in atopic dermatitis, psoriasis, and allergic contact dermatitis involves the movement of leukocytes from peripheral blood to skin. Singbartl and Ley (2004) also describe the process of leukocyte recruitment to inflamed tissues in the case of acute renal failure as occurring in a cascade-like fashion that encompasses capture, rolling, activation, firm adhesion, and tissue translocation of leukocytes. These mechanisms may contribute to a decrease in blood cells from peripheral blood in subjects with certain inflammatory conditions in the skin or other organs.

The ability of liquid formaldehyde to cause dermatitis, skin irritation, and immune modulation in occupational and non-occupational settings is well documented. For example, Nethercott and Holness (1988, as cited in ATSDR 1999) showed an 11% prevalence of contact skin dermatitis (3% positive formaldehyde skin-patch tests) in embalmers working at funeral homes vs. 0% in controls. Similar results have been reported for nurses exposed to formaldehyde disinfectant (Rudzki et al., 1989, as cited in ATSDR 1999). Further, in a review of formaldehyde in cosmetic products, de Groot and Maibach (2010) find that formaldehyde applied to skin has been shown to induce dermatitis from short-term use. Finally, in a review of skin sensitivity to formaldehyde in various populations, de Groot et al. (2009) found a 4.1% prevalence in one study of Chinese subjects and up to 9.2% in studies from the United States. If occupational dermal exposure to formaldehyde results in skin irritation and dermatitis, which influence changes in concentrations in peripheral blood of leukocytes and other hematology parameters, these changes may explain the heterogeneity in response with different exposure concentrations in the human studies summarized in Table 16. It is noteworthy that bone marrow is dynamic in that cell loss is compensated by cell production, and this characteristic should be the subject of further study in the case of dermal reactions.

© 2011 Informa Healthcare USA, Inc.

There are similarities between the observed effects in peripheral blood of some formaldehyde-exposed subjects (via inhalation and possibly dermally, as discussed above) and in subjects with dermatitis conditions. For example, Yoshino et al. (2000) found evidence that the degree of clinical dermatitis was associated (positively or negatively) with peripheral mononuclear WBC counts and that the proliferation of peripheral mononuclear cells may be suppressed in severe atopic dermatitis cases; the authors suggest that this is related to the high rate of T-cell apoptosis in severe atopic dermatitis. Similarly, Forte et al. (2009) found that a reduction in chemotactic response and phagocytic activity by neutrophilic and/or mononuclear phagocytes in the majority of patients with atopic dermatitis ranged from moderate to severe. Further, Lebre et al. (2008) found that myeloid dendritic cells and plasmacytoid dendritic cells from patients with atopic dermatitis showed defective interleukin (IL)-12, tumor necrosis factor (TNF)-α, and interferon (IFN)-α production; the authors suggest that these immune indicators may contribute to the maintenance of an allergic state in these patients. Dermatitis conditions have been associated with increased eosinophil counts (290 vs. 153.3 cells/mm^3, $p<.05$) in the blood of patients with atopic dermatitis vs. healthy subjects, respectively (Jenerowicz et al., 2007). From the peripheral blood eosinophil count and eosinophil percent of blood cells, we calculated a lower mean WBC count in subjects having dermatitis vs. healthy subjects (4581 vs. 4746 WBCs/mm^3, respectively; 30 subjects per group), but it was not possible to assess whether these counts were statistically significantly different from each other, since the raw data were not provided for the individual subjects in this study. Since hematology changes can be associated with allergic reactions in general, and dermatitis in particular, and because formaldehyde can cause dermatitis and dermal sensitivity, an assessment of skin reactions to formaldehyde is necessary when investigating formaldehyde-induced hematology effects.

If skin reactions are indeed present in the study subjects, they may confound the hematology findings reported by many of the aforementioned studies. Dermal exposure to formaldehyde and its effects on clinical and subclinical skin sensitivity reactions are not reported in the human studies we reviewed. However, it is possible that the subjects in the studies by Zhang et al. (2010b), Lyapina et al. (2004), Srivastava et al. (1992), and in the studies cited by Tang et al. (2009) were exposed dermally to formaldehyde. In addition, the status of sensitization or inflammation in exposed vs. control individuals in these studies is largely unknown. Further, Farage (2008) reported that skin reactions may not be easily diagnosed by visual inspection and may require more sophisticated technology that is not widely available. Therefore, even if subjects with higher air exposures to formaldehyde had some form of skin reaction to formaldehyde, the possibility exists that this condition would not be detected by a clinician.

5.3.4.2. A recent respiratory infection can result in hematological changes—Subjects with exposure to formaldehyde in the study by Zhang et al. (2010b) were more likely than control subjects to have had recent respiratory tract infections Several studies suggest that respiratory infections can be associated with leukopenia, or decreased WBC counts in peripheral blood, in humans. Cummins et al. (1998) found a 5% decrease ($p=.02$) in total leukocyte counts in blood and a 9% decline ($p=.001$) in lymphocyte counts in 70 elderly subjects 4 weeks after they received an influenza vaccine. These results are supported by those from three cases of pediatric influenza infections that were associated with declines in peripheral blood WBC counts (Rice and Resar, 1998). Further, a study by Shen et al. (2008) showed that not only do infections modify hematology parameters, but also that the type of infection could be important. For example, children with influenza B infection had a significantly lower total WBC count than those with influenza A infection. Influenza infection has also been shown to cause or exacerbate bone marrow suppression in mice (Lavrov and Semenkov, 1991; Hyland et al., 2005). Shen et al. (2008) found that leukopenia is not an uncommon occurrence in influenza infections and that this decline in WBC counts is possibly related to B-lymphocyte apoptosis in bone marrow.

The higher rate of recent respiratory infections in the exposed vs. control groups of the study by Zhang et al. (2010b) could have resulted in confounding of hematology parameters. Alternatively, it may be argued that the higher rate of recent respiratory infections in the exposed workers is due to lower WBC counts or that respiratory tract infections could either increase or decrease WBC counts in peripheral blood (Mayo Clinic, 2008) and may, therefore, be unrelated to the findings in this study. However, all reported WBC counts for formaldehyde-exposed subjects and their controls are above 4900 cells/mm^3 of blood (e.g., Tang et al., 2009; Zhang et al., 2010b), which are higher than the Mayo Clinic's benchmark of 3500 cells/mm^3 for leukopenia (Mayo Clinic, 2008). Therefore, it is less likely that the lower WBC count is the cause of the recent infections in the exposed subjects. Better-matched exposure and control groups in future studies may eliminate this potential confounder.

5.3.4.3. Other unmeasured potential confounders As discussed in the preceding subsections, several conditions may be associated with decreased WBC and RBC counts in peripheral blood. In a study of adult Japanese male office workers, Nakanishi et al. (2003) found that WBC counts increased with increasing body mass index and smoking, but decreased with alcohol intake, nutritional balance, and hours worked per day. Here we focus on two possible confounders of hematology parameters that may be associated with oral or inhalation exposure to formaldehyde: (1) the effects of formaldehyde on the hypothalamic/pituitary/adrenal (HPA) axis involvement in WBC count modulation; (2) the effect, on the kidneys,

Table 17. Comparison of formaldehyde hematotoxicity to known leukemogens.

	Benzene	Cyclophosphamide	Radiation	Formaldehyde
Pancytopenia in animals	Yes	Yes	Yes	No
Hematopoietic malignancies in animals	Yes	Yes	Yes	No
Pancytopenia in humans	Yes	Yes	Yes	More research needed

of possible exposure to melamine from formaldehyde-melamine resins (which were the source of formaldehyde exposures in the study by Zhang et al., 2010b); kidneys are important in producing erythropoietin, the hormone responsible for inducing RBC production.

There is evidence to suggest that formaldehyde exposure can modulate WBC counts in peripheral blood via an endocrine pathway. Brondeau et al. (1990) found that exposure to airborne irritants (including formaldehyde) caused leukopenia at irritant levels (≥43 ppm for formaldehyde, in an exposure-dependent manner) over a 4-hour exposure, and that this effect was prevented by removal of the adrenal gland, suggesting a possible role for this gland in apparent hematological effects. Sari et al. (2004) also found that relatively low concentrations (0.08, 0.4, and 2.0 ppm) of formaldehyde increased hypothalamus/pituitary/adrenal (HPA) axis activity by increasing the numbers of both hypothalamus corticotropin releasing hormone-immunoreactive neurons and pituitary adrenocorticotropin hormone (ACTH)-immunoreactive cells in mice. These neuronal changes were paralleled by increased mRNA expression of pituitary ACTH, which functions in regulating adrenal gland function. It is also well established that the adrenal glands produce glucocorticoids, mainly cortisol, a steroid that can suppress the immune response (Cotran et al., 1999). Although more research is needed to investigate the potential effect of formaldehyde on WBC counts via an endocrine pathway, particularly at occupationally relevant concentrations, there is suggestive evidence that it occurs. This HPA pathway should be considered when evaluating hematology data associated with formaldehyde exposure.

The recent discovery that melamine can cause or contribute to renal toxicity may have implications for the consumption of melamine when it is either present as an adulterant in food or inhaled during its manufacture or processing. In patients who have kidney disease it is likely that production of erythropoietin, the hormone responsible for inducing RBC production, is depressed. As a result, the bone marrow makes fewer RBCs and therefore patients with kidney disease often have to take erythropoietin supplements (NIH, 2008). This notion of melamine-induced changes in RBC counts and hemoglobin concentration finds support from Srivastava et al. (1992) who reported declines in hemoglobin concentration and *elevations* in lymphocyte counts, but not other hematology indicators, in workers with considerable exposure to formaldehyde-melamine resin. Moreover, Dobson et al. (2008) found that melamine or melamine cyanuric acid ingestion in rats caused renal toxicity. Further, acute renal failure has also been reported in human infants in Beijing, China, who were exposed to melamine via a popular Chinese brand of milk formula, "Sanlu" (Sun et al., 2010). The role for melamine in renal toxicity and how it might be related to changes in RBC counts and Hb concentration in blood is important, particularly when analyzing the results of the study by Zhang et al. (2010b). Consideration of melamine exposure in the Zhang et al. (2010b) study participants is important, since the study participants were potentially exposed by inhalation and ingestion to formaldehyde-melamine resins at work, and they were possibly exposed to melamine from food items made or contaminated with melamine-adulterated milk powder outside work.

Finally, there are several other potential confounders of hematology parameters in occupational studies. For example, alcohol consumption, which is a potential confounder in the study by Zhang et al. (2010b), can modulate immune function (Szabo and Mandrekar, 2009). Also, nutritional deficiencies in folic acid and cyanocobalamin (vitamin B12) have been associated with megaloblastic anemia, which manifests with faulty RBCs that are larger than normal (increased MCV) (Morris et al., 2007). Further, certain herbal supplements, such as Echinacea, have been associated with depressed WBC counts after chronic ingestion (Kemp and Franco, 2002). It is thus possible that WBC and RBC counts in peripheral blood are modulated by formaldehyde exposure in mechanisms involving extramedullary systems (i.e., outside the bone marrow); this would contribute to confounding in epidemiology studies that result in observed differences between exposed and controls.

Future studies investigating a possible association between exposure to formaldehyde and hematotoxicity should consider a number of confounders, including, but not limited to, the ones discussed here.

5.4. Summary

As a whole, the available studies of formaldehyde hematotoxicity in both animals and humans provide little evidence to support the account that formaldehyde exposure is causally associated with leukemia. The animal studies generally reported neither hematotoxicity nor leukemia associated with formaldehyde inhalation or ingestion. The two studies, one of which is not peer-reviewed, that reported some evidence of formaldehyde-induced leukemia are not convincing of such an association due to (1) inconsistent and potentially flawed data that has been dismissed by both EFSA and ATSDR (as discussed in Section 5.3.1.2) (Soffritti et al., 1989, 2002); (2) the high background tumor rate in the animal models used (DeVoney et al., 2006 poster, 2010 poster abstract); (3) the

© 2011 Informa Healthcare USA, Inc.

lack of corroboration from numerous other studies that examined the same endpoints in animals.

A few human studies, as cited by secondary sources, may be consistent with hematotoxicity, but they are inconsistent with other study findings and plagued by possible confounding. As discussed in Section 5.3.2, the studies suggestive of hematotoxicity are reported by Tang et al. (2009) and Zhang et al. (2010b). However, because the only study of hematological effects cited by Tang et al. (2009) available in English (Kuo et al., 1997) is associated with several uncertainties, the conclusions drawn by Tang et al. (2009) are weakened. Until the other studies cited by Tang et al. (2009) are translated, many questions exist about the outcomes and exposure-related uncertainties in the findings. Many medical and lifestyle factors can contribute to changes in hematology, particularly declines in WBC and RBC counts. The study by Zhang et al. (2010b) provides some evidence to support an association between formaldehyde and hematotoxicity; however, as discussed in Sections 5.3.2 and 5.3.4, this study's outcomes are mixed and may suffer from potential confounding of results by recent respiratory tract infections and leukopenia resulting from possible dermatitis. When considering the many possible known and unknown confounders in the studies we reviewed, such as dermatitis, respiratory infection, alcohol consumption, non-occupational sources of formaldehyde, etc., it is impossible to rule out confounding. In addition, many of the human studies are cross-sectional and therefore cannot adequately show cause and effect. Moreover, the available data from human studies do not provide sufficient proof for formaldehyde-induced hematotoxicity particularly when animal studies provide strong evidence *against* it. If formaldehyde is hematotoxic in humans, this toxicity would likely be via a mechanism not feasible in rodents, rhesus monkeys, or beagle dogs, since formaldehyde exposure does not cause hematotoxicity in these animals, therefore bringing into the question of biological plausibility of formaldehyde-induced hematotoxicity in humans.

Finally, the question of potential formaldehyde-induced hematoxicity can be explored by considering information on epidemiology and mode-of-action studies. As part of the HBWoE approach, one considers the cross-discipline integration of hypothesized effects. As discussed herein, the epidemiology and mode-of-action data cast doubt on the ability of inhaled formaldehyde to interact with and perturb hematopoiesis, which complicates further the plausibility of a conclusion of causal association based on the observations in the hematotoxicity and animal leukemia studies.

6. Weight of evidence regarding a plausible mode of action for formaldehyde leukemogenesis

In the following analysis, we examine the data relevant to the modes of action that have been proposed for formaldehyde leukemogenesis. We focused on studies that examined formaldehyde metabolism and distribution, and genotoxicity in animals, humans, and in vitro. We conducted literature searches, using PubMed and several search terms in combination with "formaldehyde": "genom*," "chromosom*," "micronuclei," "cytogenetic," "DNA damage," "genotox*," "mutagen*," "metabol*," "toxicokinetic," and "pharmacokinetic." We also relied on recent key review articles and agency reports (IARC, 2006; US EPA, 2010; Heck and Casanova, 2004; Pyatt et al., 2008; Golden et al., 2006), as well as references within those reports and papers found in the PubMed search.

As discussed, the epidemiology data do not support a causal association between formaldehyde exposure and leukemia. In addition, the available studies of formaldehyde hematotoxicity in both animals and humans provide little evidence for formaldehyde-associated leukemia. The animal studies generally reported neither hematotoxicity nor leukemia associated with formaldehyde exposure, and although a few human study findings are consistent with hematotoxicity, they are inconsistent with other study findings and plagued by possible confounding.

Despite these findings, three modes of action for formaldehyde leukemogenesis have been hypothesized by Zhang (2009, 2010a) and are also discussed in US EPA's recent draft toxicological profile for formaldehyde (US EPA, 2010). The proposed modes of action are as follows:

1. Formaldehyde targeting bone marrow hematopoietic stem cells—formaldehyde complexes as a hydrate [$CH_2(OH)_2$] that could potentially reach the bone marrow, where it could directly induce DNA damage and chromosomal aberrations in hematopoietic stem or progenitor cells, leading to leukemia.
2. Formaldehyde targeting nasal stem cells (nasal-associated lymphoid tissue, or NALT)—nasal stem cells are damaged by formaldehyde, released from the nasal passage, circulate in the blood, and are eventually incorporated into bone marrow leading to leukemia.
3. Formaldehyde targeting circulating hematopoietic stem cells—stem cells circulate from marrow to nasal tissue where they are transformed by formaldehyde (pre-mutagenic lesions), and then migrate back to bone marrow, eventually leading to leukemia.

Here we first describe what is known about formaldehyde metabolism, biological distribution, and genotoxicity. We then provide a weight-of-evidence analysis of the formaldehyde data with regard to the three proposed modes of action.

6.1. Formaldehyde toxicokinetics

The toxicokinetics of formaldehyde has been extensively studied and is summarized in recent reviews and agency toxicological profiles (ATSDR, 1999; ATSDR, 2010;

Heck and Casanova, 2004; IARC, 2006; US EPA, 2010). Formaldehyde is a normal by-product of several metabolic pathways in mammals, and is naturally present in tissues, cells, and biological fluids. Under physiological conditions, it exists in equilibrium, predominantly in its hydrated form methanediol [$CH_2(OH)_2$], with less than 0.1% as free formaldehyde. Formaldehyde is water soluble and highly reactive; therefore, it is readily absorbed and metabolized in biological systems. It is primarily metabolized by glutathione-dependent formaldehyde dehydrogenase (FALDH) and aldehyde dehydrogenases (ALDHs). Formaldehyde enters the "one-carbon" pool and is readily incorporated into macromolecules in the body. In rats exposed to [^{14}C]formaldehyde via inhalation (0.63 or 13 ppm), the exhaled fraction was independent of exposure concentration, with 40% of the ^{14}C incorporated into macromolecules and 40% exhaled as $^{14}CO_2$, and the remainder was excreted in the feces and urine, and incorporation into macromolecules in the blood was via the one-carbon pool and not through DNA or protein adducts (Heck et al., 1983, as cited in Heck and Casanova, 2004). Salthammer et al. (2010) discusses a median concentration of 4.3 ppb formaldehyde in human breath that is likely due to endogenous sources.

The concentration of endogenous formaldehyde in human blood is approximately 0.1 mM and, as discussed in Heck and Casanova (2004), this concentration is not increased in humans who inhale 2 ppm formaldehyde for 40 minutes or in monkeys inhaling 6 ppm for 4 weeks. The inability of exogenous formaldehyde to increase blood concentrations of formaldehyde was confirmed in an analysis by Franks (2005) using a sophisticated mathematical model. These data strongly suggest that, at concentrations to which humans might be exposed, formaldehyde does not move beyond the nasal mucosa to cause effects at distant sites. Recent dosimetry, cytotoxicity, and genomics studies conducted by Andersen et al. (2010) suggest that exposure to formaldehyde concentrations of 1 to 2 ppm would not affect formaldehyde homeostasis or increase genotoxic and cytotoxic effects in the nose or in any other tissue. Andersen et al. (2010) developed a pharmacokinetic model to estimate various forms of formaldehyde and glutathione (GSH) tissue concentrations, accounting for enogenous levels of formaldehyde, and applied the model to compare tissue concentrations with histopathology and gene expression changes in the nasal epithelium of rats. The study found that at high exposure concentrations (6 to 15 ppm), gene expression changes reflected pathways involved in cell cycle control, DNA repair, and apoptosis, with tissue responses including cell proliferation, erosion, necrosis, and increased severity of squamous metaplasia—cellular responses potentially associated with carcinogenesis. At lower exposure concentrations (less than 1 to 2 ppm), the gene expression changes likely represented extracellular responses (such as responses to irritancy and to export GSH to extracellular spaces), with tissue responses at 2 ppm reflecting mild squamous metaplasia.

6.2. Formaldehyde genotoxicity

Formaldehyde induces a variety of genotoxic and mutagenic effects, including DNA protein cross-links (DPX), DNA adducts, point mutations, DNA strand breaks, chromosomal aberrations (CA), deletions, sister-chromatid exchange (SCE), and micronucleus (MN) formation (ATSDR, 1999; ATSDR, 2010; Heck and Casanova, 2004; IARC, 2006; US EPA, 2010).

6.2.1. DNA adducts and protein cross-links

At high exposure concentrations, formaldehyde causes DNA-protein cross-links (DPX) in the nasal mucosa of rats, upper respiratory tract of monkeys, and in vitro in human cells (Heck and Casanova, 2004; ATSDR, 1999; ATSDR, 2010; IARC, 2006; US EPA, 2010). Pharmacokinetic models have been used to study the disposition of inhaled [^{14}C]formaldehyde in the respiratory tract. At very low concentrations of formaldehyde, nearly 100% is eliminated through metabolism or through non-saturable pathways other than DPX (such as protein adducts), with very little (7×10^{-6}%) bound as DPX. At higher concentrations (6 ppm, 6 hours) in rats and Rhesus monkeys, 91% and 96% of the [^{14}C]formaldehyde in the DNA was due to metabolic incorporation, and approximately 9% and 4% of the [^{14}C]formaldehyde in the DNA was bound as DPX in the nasal respiratory mucosa, respectively (Heck and Casanova, 2004). Studies suggest that formaldehyde-induced DPX are rapidly removed (24 hours) from human blood cultures treated in vitro (Schmid and Speit, 2007), and from the nasal respiratory mucosa of rats exposed via inhalation to formaldehyde (6, 10 ppm) (Heck and Casanova, 2004).

There is no strong evidence to suggest that formaldehyde causes DPX in bone marrow or WBCs (discussed in more detail in the next section). A recent study by Wang et al. (2009a) found higher levels of the formaldehyde-DNA adduct N^6-hydroxymethyldeoxyadenosine (N^6-HOMe-dA) in leukocytes of smokers vs. non-smokers. The authors suggest that N^6-HOMe-dA adducts in leukocyte DNA may be potentially important as a cause of cancer from smoking. A recent study by Lu et al. (2010), in which rats were exposed via inhalation to 10 ppm deuterium-labeled formaldehyde (i.e., [$^{13}CD_2$]formaldehyde) to trace the disposition of exogenous vs. endogenous formaldehyde in DNA, found exogenous formaldehyde-DNA adducts in the nasal respiratory mucosa but not at distant sites (including WBCs and bone marrow). In addition, Lu et al. (2010) found that exogenous formaldehyde caused only N^2-HOMe-dG adducts in the nasal mucosa and no N^6-HOMe-dA adducts; however, both adducts were found in distant sites but only from endogenous formaldehyde. Another study by Lu et al. (2011) examined molecular dosimetry (0.7, 2, 5.8, 9.1, and 15.2 ppm [$^{13}CD_2$]formaldehyde for 6 hours) of endogenous and exogenous N^2-HOMe-dG adducts in the nasal mucosa of rats. The authors found that endogenous adducts dominated at low exposure concentrations (more than 99% and 97% endogenous at 0.7 and 2 ppm,

© 2011 Informa Healthcare USA, Inc.

respectively). Further, the authors examined the levels of endogenous and exogenous N^2-HOMe-dG adducts in bone marrow from exposure to 15.2 ppm formaldehyde and found that exogenous adducts were not detectable. A similar study conducted by the same group (Moeller et al., 2011) examined the levels of endogenous and exogenous N^2-HOMe-dG adducts in the nasal mucosa and bone marrow of cynomolgus macaques exposed to 1.9 and 6.1 ppm [$^{13}CD_2$]formaldehyde for 6 hours a day for 2 consecutive days. The authors observed readily detectable levels of exogenous and endogenous adducts in the nasal mucosa at both exposures; however, only endogenous adducts were detectable in the bone marrow. These data strongly suggest that the results observed by Wang et al. (2009a) may be specific to effects from cigarette smoke (i.e., the generation of formaldehyde from metabolism of *N*-nitrosodimethylamine [NDMA] and 4-(methylnitorosamino)-1-(3-pyridyl)-1-butanone [NNK]) and not from exogenous formaldehyde. In addition, Lu et al. (2010, 2011) and Moeller et al. (2011) provide strong evidence to support the biological implausibility of distant site carcinogenicity, such as leukemia, from inhaled formaldehyde, while providing evidence that formaldehyde inhalation can lead to DNA adducts in respiratory nasal epithelium. In addition, Neuss et al. (2010) show that human nasal epithelial cells pre-exposed in vitro to high concentrations of formaldehyde do not cause DNA damage (DPX) in co-cultivated isolated human lymphocytes, lending further support that formaldehyde that has entered the nasal epithelial cells does not move beyond these cells to damage DNA in other cells in close proximity (discussed in more detail below with respect to the NALT hypothesis).

6.2.2. Clastogenic and cytogenetic effects

In vivo mammalian formaldehyde genotoxicity assays have examined clastogenic and cytogenetic effects (CA, SCE, and MN formation) in rodents and humans, and the results have been summarized (ATSDR, 1999; ATSDR, 2010; Heck and Casanova, 2004; IARC, 2006; US EPA, 2010). As presented in these reviews, the cytogenetic results in humans and animals are conflicting, showing both positive and negative effects. In humans, the majority of these studies have been carried out in nasal or oral mucosa (to examine site of direct contact) and in peripheral blood lymphocytes (PBLs) (to examine distant-site toxicity). As reviewed by Speit and Schmid (2006) and agency reviews (IARC, 2006; US EPA, 2010), the published studies suggest that inhalation of formaldehyde leads to increased MN frequencies in nasal and/or buccal mucosa cells. There are a number of issues with these studies, however, including incomplete information on study design, exposure, and potential confounding factors (Speit and Schmid, 2006). Speit and Schmid (2006) suggest that because of this, it is not yet possible to make meaningful conclusions regarding local genotoxic effects of formaldehyde.

From our review of the current literature, and from studies summarized in recent agency reviews (IARC, 2006; US EPA, 2010; Jakab et al., 2010; Jiang et al., 2010; Pala et al., 2008), to date, approximately 20 studies have examined the cytogenetic effects of formaldehyde in human PBLs, as a means for examining distant-site toxicity. These data are insufficient and conflicting, with both positive and negative results. As discussed in several recent reviews (Heck and Casanova, 2004; Pyatt et al., 2008; Golden et al., 2006), and in more detail in the next section, interpretation of the positive findings in humans, particularly in the context of leukemia, is problematic given (1) potential confounding in the studies, including diet and life style differences, or the lack of good exposure information; (2) the lack of evidence to suggest that DNA damage in human PBLs is a model for DNA damage in stem cells, since these effects have not been shown to occur in stem cells that can transition to leukemia; and (3) similar results have not been found in controlled animal studies. For example, Kligerman et al. (1984) found no statistically significant increase in SCE or chromosome breakage in PBLs of rats exposed to formaldehyde (0.5, 6, or 15 ppm). A similar study carried out recently by Speit et al. (2009) found that formaldehyde (0.5, 1, 2, 6, 10, and 15 ppm) did not induce any significant genotoxic effects (DPX, SCE, or MN) in PBLs of rats.

6.3. HBWoE evaluation of the proposed modes of action for formaldehyde as a leukemogen

The plausibility of the three proposed modes of action has been extensively reviewed by others (Pyatt et al., 2008; Golden et al., 2006). We have considered these reviews, in addition to the primary formaldehyde inhalation toxicity literature, and have come to the following questions with regard to the proposed modes of action:

1. What is the evidence that formaldehyde exposure induces carcinogenic (or genotoxic) transformation directly in bone marrow?
2. What is the evidence that formaldehyde can induce carcinogenic (or genotoxic) transformation in nasal-associated lymphoid tissue (NALT), or peripheral hematopoietic stem cells (HSCs)?
3. Is the DNA damage observed in the formaldehyde genotoxicity studies consistent with DNA damage associated with leukemia?
4. If formaldehyde could induce systemic DNA damage, what concentrations in the nose would it take to reach levels higher than endogenous formaldehyde DNA adduct levels in the NALT or circulating HSCs to cause a sufficient level of DNA damage that would induce cell proliferation in the bone marrow? Would these concentrations be relevant to typical human formaldehyde exposures? How do these concentrations compare to levels that would also cause irritation?
5. If formaldehyde could induce DNA adducts above endogenous levels in NALT or circulating HSCs, what is the likelihood that these cells would home back to healthy bone marrow to cause leukemia?

As a whole, considering these questions allows for an assessment of the extent to which the genotoxicity and mode-of-action data support either a causal association between formaldehyde exposure and leukemia or an alternative hypothesis. Importantly, one needs to consider the mode-of-action data in the context of the epidemiology and hematotoxicity data, as each of the three lines of evidence inform interpretation of the other.

6.3.1. There is no consistent evidence that inhaled formaldehyde induces genotoxicity in bone marrow, NALT, or peripheral HSCs that might lead to leukemia

Although the evidence clearly indicates that formaldehyde induces DPX in nasal mucosa of rats and the upper respiratory tract of monkeys (Heck and Casanova, 2004), a large body of evidence suggests that formaldehyde does not move beyond the respiratory mucosa to induce systemic genotoxic effects and cellular transformation (Heck and Casanova, 2004; Pyatt et al., 2008; Golden et al., 2006, Andersen, et al., 2010). These data are discussed in more detail below in the context of the distant sites (bone marrow, NALT, or peripheral HSCs) relevant to the proposed formaldehyde leukemogenic modes of action.

6.3.1.1. Bone marrow Zhang et al. (2009) hypothesize that formaldehyde may potentially reach the bone marrow in its hydrated methanediol form where some level of free formaldehyde may exist in equilibrium with methandiol so that it could react with bone marrow stem cells to cause leukemia. This is very unlikely, however, given that, as discussed above, the levels of formaldehyde in the blood do not increase even with reasonably high exposure levels in humans (2 ppm). As discussed below, there are studies to support the implausibility of this mechanism.

As discussed in Heck and Casanova (2004), studies using radiolabeled formaldehyde have shown that there is a lack of detectable DPX in the bone marrow of rats exposed to 15 ppm formaldehyde (Casanova-Schmitz et al., 1984), in bone marrow of GSH-depleted rats exposed to 10 ppm formaldehyde (Casanova and Heck, 1987), and in Rhesus monkeys exposed to formaldehyde at concentrations as high as 6 ppm (Heck and Casanova, 2004). Further, as discussed above, recent studies (Lu et al., 2010, 2011; Moeller et al., 2011), using [$^{13}CD_2$]formaldehyde, clearly indicate that exogenous formaldehyde does not induce DNA damage beyond the nasal tissue (i.e., bone marrow).

In addition, cytogenetic assays in bone marrow of Sprague-Dawley rats (Dallas et al., 1992) exposed to 15 ppm formaldehyde, and mice exposed to formaldehyde via intraperitoneal injection (Natarajan et al., 1983 as cited in US EPA, 2010; Gocke et al., 1981), observed no significant increase in CA or MN in bone marrow cells relative to controls. In contrast, one study by Kitaeva et al. (1990, abstract only) of Wistar rats exposed to very low concentrations of formaldehyde (0.4 to 1.2 ppm) observed an increased incidence of CA in bone marrow cells relative to controls. This one study is not supported by results from the other three studies discussed. In addition, as discussed in Heck and Casanova (2004) and in Golden et al. (2006), this study is hampered by a lack of critical experimental details (i.e., dose levels and other experimental procedures are not clear, and statistical methods were not described properly) that precludes its use in drawing any meaningful conclusions.

Overall, the weight of evidence does not support the proposed mode of action that inhaled formaldehyde moves beyond the nasal respiratory mucosa to cause genotoxicity in the bone marrow.

6.3.1.2. Stem cells in the NALT Zhang et al. (2009) hypothesize another potential mode of action involving direct induction of mutations in the pluripotent stem cells of the nasal passage (or the NALT), and that these stem cells could then be released into the circulation where they could eventually make their way to the bone marrow. There are several lines of evidence, discussed below, that suggest the implausibility of this proposed mechanism.

First, if precursor cells in nasal tissue were acted upon in this way, there should also be generation of chloromas in the nasal tissue, since isolated accumulations of myeloid tumor cells would be expected to originate from the same proposed precursor cells in nasal tissue. There is no sign of chloromas, however, among formaldehyde-exposed workers in the current literature. Further, as discussed in Pyatt et al. (2008), all lymphoid tumors arising from the NALT have been classifiable as non-Hodgkin's lymphoma (NHL), which is not elevated in the formaldehyde occupational epidemiology studies. The lack of chloromas and NHL arising in the NALT (nasal lymphomas) in the epidemiology data provide strong evidence against this mode of action.

Recent experimental evidence directly examining this proposed mechanism suggests its implausibility. Kuper et al. (2009) examined the proliferative effect of formaldehyde on the NALT and local lymph nodes in F344 rats and B6C3F1 mice exposed to 0, 0.5, 1, 2, 6, 10, and 15 ppm formaldehyde for 28 days. The authors found an increased proliferation rate in the nasal epithelial cells and a slight to moderate simple hyperplasia of the NALT in rats exposed to 15 ppm but not at lower concentrations, and no increases were observed at any concentration in mice, suggesting that at concentrations of less than 15 ppm formaldehyde, sufficient levels of formaldehyde do not move beyond the nasal mucosa to the NALT to induce cell proliferation. Given these observations, it is worth considering whether it is biologically plausible to incur enough damage in the NALT tissue, from typical human formaldehyde exposures, that would be sufficient to have other manifestations. Although levels lower than 15 ppm formaldehyde do not induce proliferation in the NALT, one might argue that DNA damage may still occur at low levels of exposure; if this damage is in a pluripotent stem cell that is released into the circulation and the DNA is sufficiently

© 2011 Informa Healthcare USA, Inc.

damaged such that carcinogenic initiation could occur, this cell might home back to bone marrow to cause leukemia. But, one must ask whether this is quantitatively plausible, particularly since mucosa-associated lymphoid tissue (such as the NALT) represent small concentrations of tissue. Stochastic models of carcinogenesis have been developed that suggest human cancers are the result of a multistage process requiring at least two genetic alterations for carcinogenic transformation (Moolgavkar et al., 1999). With the understanding that malignant tumors arise from a single malignant progenitor cell, we must ask whether there is a strong enough stochastic argument to support the hypothesis that formaldehyde exposure (at typical human exposure concentrations of 2 ppm or less) would hit enough stem cells in the NALT such that there is a reasonable likelihood that the critical genes, in at least one of the stem cells that is released into the circulation, would be sufficiently damaged to cause carcinogenic initiation, and further that there is a reasonable likelihood that the initiated stem cell will home back to healthy bone marrow to cause leukemogenesis. Given the stochastic nature of carcinogenesis, the relatively small amount of NALT tissue, and the gene expression results of Andersen et al. (2010) that suggest 2 ppm formaldehyde exposure is not likely to increase genotoxic and cytotoxic effects in the nose or in any other tissue, the probability that there is enough damage in the NALT to lead to further carcinogenic manifestations beyond the nose is likely very small at typical human exposure concentrations. Further, the level of damage required to reach quantitative plausibility would likely result in other manifestations in the nose, such as chloromas, which are rarely observed.

In another study, Neuss et al. (2010) show that human nasal epithelial cells pre-exposed in vitro to high concentrations of formaldehyde do not cause DNA damage (DPX) in co-cultivated isolated human lymphocytes, lending further support that formaldehyde that has entered the nasal epithelial cells does not move beyond these cells to damage other cells in close proximity, such as progenitor stem cells in the nasal mucosa.

Zhang et al. (2009) cite a study by Murell et al. (2005) in support of the NALT mode of action, since this study provides some support for the ability of rat olfactory epithelial cells to repopulate hematopoietic tissue in bone marrow of irradiated rats. The olfactory mucosa stem cells used in the Murell et al. (2005) study, however, were tested for their ability to repopulate ablated irradiated rat bone marrow. As discussed in more detail below, a number of studies (McKinney-Freeman and Goodell, 2004; Abkowitz et al., 2003; Edgren et al., 2007) suggest that the majority of circulating stem cells do not efficiently home back to bone marrow under homeostatic conditions.

Overall, the weight of evidence does not support the proposed mode of action that formaldehyde exposure, at reasonably expected concentrations in humans, targets stem cells in the NALT, such that these cells would then be released into the circulation to home back to the bone marrow to cause leukemia.

6.3.1.3. Circulating peripheral HSCs Zhang et al. (2009) propose another mode of action for formaldehyde-induced leukemia, suggesting that formaldehyde could move beyond the nasal tissue into the circulation where it may transform circulating HSCs that could travel back to the bone marrow.

As discussed already, many studies have examined the cytogenetic effects of formaldehyde in human PBLs as a means for examining distant-site toxicity, but these data are conflicting, with both positive and negative results. In addition, controlled animal studies did not find any significant genotoxic effects (SCE, MN, or CA) in PBLs of rats exposed to high levels of formaldehyde (15 ppm) (Kligerman et al., 1984; Speit et al., 2009). Furthermore, although it is not an unreasonable assumption, observations from studies of circulating blood lymphocytes should not necessarily be taken to mean that the same effects will occur in circulating stem cells that then could transition to leukemia. Only one study to date has examined whether cytogenetic effects in cultured hematopoietic progenitor cells from peripheral blood were increased in workers exposed to formaldehyde (Zhang et al., 2010b), and as discussed in more detail below, there are several problems with interpretation of this study. Therefore, interpretation of the positive cytogenetic findings (beyond the nasal mucosa) in humans, particularly in the context of leukemia, is problematic.

First, as discussed earlier, there is a large body of evidence suggesting that inhaled formaldehyde does not move beyond the nasal respiratory mucosa to cause genotoxicity at distant sites, including lymphocytes (Heck and Casanova, 2004; Pyatt et al., 2008; Golden et al., 2006; Schmid and Speit, 2007; Speit et al., 2009; Lu et al., 2010, 2011; Moeller et al., 2011; Neuss et al., 2010). Although Shaham et al. (1996, 1997, 2003) reported increased levels of protein-associated DNA (presumed to be DPX) in the lymphocytes of hospital workers (laboratory assistants and technicians, physicians, orderlies, and pathologists), as discussed by Heck and Casanova (2004) and Pyatt et al. (2008), there are many problems with these studies. For example, the authors claimed that DPX could be detected down to 0.001 mM; however, their data do not provide any evidence of a concentration-response relationship for DPX below 0.3 mM. Further, Shaham et al. (1996, 1997) indicate that DPX are persistent and can accumulate in lymphocytes. Their data, however, do not support their assertion and are contradictory to studies showing the rapid removal of DPX from formaldehyde-exposed human blood in culture (Schmid and Speit, 2007), and from the nasal respiratory mucosa of rats exposed to formaldehyde via inhalation (Heck and Casanova, 2004). Further, with regard to chromosomal aberrations observed in PBLs, as shown by Schmid and Speit (2007), SCEs are formed from DNA synthesis through DPX

during S-phase in human blood cultures. These results suggest that, given the rapid removal of DPX, it is unlikely that a sufficient amount of formaldehyde-induced DPX would persist through DNA replication in occupationally exposed workers. This further suggests that reported SCE frequencies in PBLs of workers exposed to formaldehyde are unrelated to any formaldehyde exposure. The authors extend this argument for other cytogenetic events as well (MN and CA).

Second, interpretation of many of the human PBL studies of formaldehyde-exposed workers is limited due to the lack of reliable exposure information and potential confounding by exposures to other chemicals in the workplace or other factors that may impact background levels of CA and MN. Several studies (Battershill et al., 2008; Iarmarcovai et al., 2008, 2007) suggest that many factors, including age, gender, smoking status, alcohol consumption, disease conditions and infections, physical exercise, and vitamin B12 and folate status impact background levels of CA and MN in PBLs (albeit some factors stronger than others). Battershill et al. (2008) suggest that the evaluation of PBLs as genotoxicity biomarkers is complex, requiring good exposure data, appropriate stratification of exposed groups, and appropriate statistical power. Given the general limitations in the human PBL studies, it is not surprising that the results with respect to formaldehyde are inconsistent.

Third, observations from studies of circulating blood lymphocytes should not necessarily be taken to mean that the same effects will occur in circulating stem cells that then could transition to leukemia. In fact, CA and MN in PBLs are associated with many types of cancers, and they appear to be a general marker for increased cancer risk, not specific to leukemia (Bonassi et al., 2008; Murgia et al., 2008). In these studies, it is noteworthy that increased CA in PBLs are not associated with occupational exposures to genotoxic agents. Further, as discussed in Pyatt et al. (2008), there are many commonly used drugs with clastogenic properties in vitro and in vivo (methotrexate), and in human lymphocytes in vitro (including antibiotics metronidazole, trimethoprin, and hydrochlorothiazide). Therefore, there is limited value in using clastogenic effects in human lymphocytes as being predictive of leukemic potential.

Only one study (Zhang et al., 2010b) reports increased cytogenetic effects (aneuploidy) in cultured myeloid progenitor cells from 10 workers exposed to formaldehyde (mean of 2 ppm). Zhang et al. (2010b) report an increased loss of chromosome 7 (monosomy 7) and gain of chromosome 8 (trisomy 8) in exposed relative to the unexposed control group. There are several problems, however, with this study.

- First, the study group was very small (10 exposed vs. 12 control) and the results were pooled. Individual results for monosomy 7 and trisomy 8 should have been provided so that the exact nature of aneuploidy could have been assessed on an individual basis, and so it would be clear whether there was a consistent increase for all subjects, or if some were much higher than others, or if some had just one change or both, etc.
- Second, were other chromosome changes looked for and not found? Or did the authors only look for these particular changes? It is not clear, as there is no discussion beyond monosomy 7 and trisomy 8. This is particularly relevant because, although aneuploidy of chromosomes 7 and 8 have been shown to be associated with leukemia (Johnson and Cotter, 1997; Rowley, 2000; Paulsson and Johansson, 2007), they are not the only chromosome changes that are associated with the disease. In fact, as discussed in Johnson and Cotter (1997) and Paulsson and Johansson (2007), monosomy 7 and trisomy 8 are not likely to be initiating events in leukemogenesis, and trisomy 8 alone is not sufficient for leukemogenesis. Trisomy 8 has been shown to occur as a secondary change to primary inversions of other chromosomes (i.e., chromosomes 9 and 11) (Paulsson and Johansson, 2007).
- Third, Zhang et al. (2010b) note a high monosomy 7 incidence in the controls and indicate that this could be due to artifactual chromosome loss during metaphase spread preparation; therefore, there is inherent bias in the sampling technique that could bias the results.
- And finally, myeloid associated monosomy 7 and trisomy 8 have been shown to be correlated with other exposures. Smoking has been shown to cause trisomy 8 (Paulsson and Johansson, 2007; Moorman, 2002), and other occupational exposures (e.g., pesticides, organic solvents, and petroleum compounds) have been shown to cause monosomy 7 (Johnson and Cotter, 1997). A recent formaldehyde occupational exposure study (Iarmarcovai et al., 2007), where increased MN were observed in exposed vs. controls, found that alcohol consumption had a potential confounding effect on chromosome loss. Approximately 40% of the control and exposed subjects in the Zhang et al. (2010b) study were smokers, and about 20% in each group consumed alcohol. Although the percent smokers and alcohol consumers was roughly the same in the exposed and control groups, there was no attempt to determine the degree of smoking or alcohol consumption among the subjects. Therefore, potential confounding from these exposures could have biased the results, particularly given the small sample size. Individual data could provide more insight into potential confounding associations.
- Overall, given the small study group, lack of a thorough examination of chromosomal effects, potential confounding of observed effects (i.e., other potential exposures, smoking, alcohol consumption), and the possibility of artifactual chromosome loss during sample preparation, it is possible to attribute the chromosomal changes reported by Zhang et al. (2010b) to chance.

© 2011 Informa Healthcare USA, Inc.

Given the strong evidence that inhalation exposure to formaldehyde (at reasonably expected concentrations for humans) does not increase the level of formaldehyde in the blood and does not cause DNA damage and cellular transformation beyond the nasal respiratory mucosa, in combination with the inconsistent effects observed in PBLs of humans occupationally exposed to formaldehyde (likely due to confounding and lack of good formaldehyde exposure information), and the fact that there is little support for the use of PBLs as a marker for effects in HSCs and leukemia, the PBL data from formaldehyde occupation studies, taken as a whole, provide little (if any) support for the proposed modes of action for formaldehyde as a leukemogen. Finally, the recent study by Zhang et al. (2010b) is hampered by potential confounding, a small study group, sampling artifacts, and lacks reporting of critical information, such that the reported chromosome changes in this study are impossible to interpret.

Therefore, the weight of available evidence does not support the proposed mode of action that formaldehyde might target circulating HSCs that might then home back to the bone marrow to cause leukemia.

6.3.2. Formaldehyde exposure would have to be very high to induce DNA damage above endogenous levels in the bone marrow, NALT, or circulating HSCs, and would likely be associated with a high degree of irritation

As discussed already, there is a large body of supportive evidence that inhalation exposure to formaldehyde at reasonably expected concentrations for humans (less than 2 ppm) does not result in increased blood levels of formaldehyde (Heck and Casanova, 2004; Franks, 2005; Andersen et al., 2010), likely due to normal metabolic processes that prevent formaldehyde from readily entering the circulation. Further, there is evidence to suggest that DNA damage does not occur in the blood or bone marrow of animals even at concentrations as high as 6–15 ppm. Schmid and Speit (2007) propose that, due to the rapid removal of DPX, very high concentrations of formaldehyde would be required (higher than what would be expected for humans occupationally exposed to formaldehyde) to produce enough DPX that would persist until DNA replication could lead to a permanent genotoxic effect (i.e., SCE, CA, or MN).

It is important to consider these concentrations in the context of what concentrations of formaldehyde are known to cause sensory irritation. Arts et al. (2006) conducted a review of the formaldehyde respiratory irritation and carcinogenicity data and found that overall, formaldehyde sensory irritation is first observed at 1 ppm in animals and humans, with eye and nasal irritation occurring at concentrations ≥1 and ≥2 ppm, and throat irritation occurring at ≥3 ppm, and more severe irritation occurring at concentrations ≥6 ppm. Therefore, sensory irritation occurs at concentrations well below those that would likely be necessary to cause sufficient DNA damage in blood, NALT, or bone marrow, and therefore the formaldehyde exposure concentrations necessary to cause such DNA damage would likely not be tolerated by humans.

6.3.3. Circulating HSCs may not readily home back to healthy bone marrow to cause leukemia

A critical assumption in the proposed modes of action that formaldehyde either targets stem cells in the NALT or circulating in the blood is that these damaged cells will travel back to and become incorporated into the bone marrow where they could then cause leukemia. Although much of the evidence suggests that these proposed modes of action are not biologically plausible, there is still a general assumption that if the exposure conditions were such that even one cell was transformed, either directly in circulating HSCs or in the NALT and then released into the circulation, that this cell would then readily home back to the bone marrow. The current evidence is not clear, however, with regard to this assumption for people with healthy bone marrow (McKinney-Freeman and Goodell, 2004; Abkowitz et al., 2003; Abrams et al., 1980; Wright et al., 2001; Schulz et al., 2009), which would be the majority of the population for which the regulatory outcome of these studies and proposed mechanisms would seek to protect. And, in fact, a number of studies suggest that the majority of circulating HSCs may not efficiently home to bone marrow.

For example, using genetically marked parabiosed CD45 congenic mice (surgically joined and sharing a common circulation), McKinney-Freeman and Goodell (2004) found that although there was a small percent of partner-derived stem cells present in the bone marrow, the majority of animals were not stably engrafted with partner HSCs when tested for functional HSC activity, suggesting that although a small percent of circulating HSC can reenter the bone marrow during homeostasis (i.e., in the absence of cytokine mobilization), this reentrance is transient and unstable, and functional HSCs do not persist in the bone marrow after returning from the circulation. The results of this study are supported by Abkowitz et al. (2003), who also used genetically marked parabiosed mice in a similar experiment and found similar results. These results suggest that HSC homeostasis is primarily maintained by endogenous stem cells in the bone marrow, and not from the return of stem cells from the circulation. The authors propose that "[b]ecause the HSC replication rate is high [in the bone marrow], the new HSCs outnumber the few HSCs entering the marrow from the peripheral blood. Once HSCs exit bone marrow, their lifespan in the circulation is extremely short, contributing to the competitive advantage of endogenously generated cells."

There may be additional support for the idea that circulating HSCs do not readily home back to bone marrow in that that there is no evidence that blood transfusions from precancerous (leukemia) blood donors are associated with increased risk of leukemia in recipients (Edgren et al., 2007). It is not unreasonable

to assume that blood donors who were later diagnosed with leukemia had circulating progenitor cells that had genetic damage or were transformed. Although it would need to be confirmed that preleukemogenic individuals have precancerous circulating HSCs, if preleukemogenic cells did exist in a blood donation, and these cells readily home back to bone marrow, Edgren et al. should have seen an increased risk of leukemia in the blood recipients, but they did not. The authors cite other studies that were inconclusive with regard to this question.

There are conflicting studies that appear to suggest that HSCs do efficiently home to bone marrow under homeostatic conditions (Wright et al., 2001). A recent review by Schulz et al. (2009), however, indicates that the mechanisms involved in the control of hematopoietic stem or progenitor cell function remain largely unknown. The authors indicate that, in addition to recirculation to the bone marrow, HSCs migrate to peripheral tissue during inflammation to respond to tissue damage. Therefore, it appears that there is much to learn with regard to mechanisms involved in homing of HSCs to bone marrow under homeostatic conditions. Consequently, the assumption that damaged HSCs or NALT stem cells would readily return to bone marrow where they could then cause leukemia should be questioned, and further studies are necessary to assess the extent to which this might occur under homeostatic conditions.

Therefore, aside from the questions put forth with regard to the implausibility that exogenous formaldehyde could sufficiently damage NALT stem cells or circulating HSCs, there are clearly also questions regarding the extent to which these stem cells would then migrate back to the bone marrow. Consequently, these studies add further to the questions regarding the plausibility of the proposed modes of action. Moreover, it is critical that we try to better understand HSC trafficking in and out of bone marrow under normal physiological conditions before accepting any mode of action that relies so heavily on this mechanism.

6.4. Summary

As a whole, the available formaldehyde toxicokinetic, mode-of-action, and genotoxicity studies provide little evidence for support of the account that formaldehyde exposure is causally associated with leukemia. The ad hoc assumptions that have been put forth in support of the three proposed modes of action are not consistent with the full body of evidence. To support the proposed modes of action, one must assume,

1. with regard to targeting circulating hematopoietic stem cells, that formaldehyde can move beyond the nasal respiratory mucosa to increase levels in the blood to a sufficient degree that would result in carcinogenic initiation of progenitor cells, and the weight of evidence does not suggest this, at least for levels to which humans are likely to be exposed and that could be tolerated (due to irritation at higher levels of exposure);
2. with regard to targeting bone marrow, that formaldehyde can travel in its hydrated methanediol form to the bone marrow where it will be in equilibrium with free formaldehyde that can cause DNA damage and cellular transformation, even though this is biologically implausible and the weight of evidence strongly suggests that exogenous formaldehyde does not cause DNA damage in any tissue other than the nasal respiratory mucosa;
3. with regard to targeting stem cells in the NALT, that formaldehyde somehow moves beyond the nasal respiratory mucosa and causes sufficient damage to nasal stem cells, such that further carcinogenic manifestations could occur (leukemia), without causing any nasal lymphomas or chloromas in the nasal tissue, even though it is biologically and quantitatively implausible that the level of damage likely required in the NALT to cause further carcinogenic manifestations would not also lead to chloromas and nasal lymphomas;
4. with regard to targeting circulating HSCs, that formaldehyde somehow moves beyond the nasal respiratory mucosa and causes DNA damage or transformation of circulating stem cells, even though the majority of evidence provided as support for this mechanism is from a large number of inconsistent PBL cytogenetic studies of formaldehyde-exposed workers and likely confounded by exposures to other chemicals in the workplace or by effects from smoking or alcohol consumption (in addition to the assumption that chromosomal effects in PBLs are good biomarkers for effects in HSCs and leukemia, and there is little support for this in the literature); or
5. the chromosome aneuploidy in cultured myeloid progenitor cells of 10 formaldehyde exposed workers reported in the Zhang et al. (2010b) study somehow suggests that these workers may be at a higher risk for leukemia, even though this study is hampered by potential confounding, a small study group, sampling artifacts (e.g., possible artifactual chromosomal loss during metaphase spread preparation), and lacks reporting of critical information, such that the reported chromosome changes in this study are impossible to interpret; and
6. even if one accepts, or it is somehow shown, that formaldehyde is capable of transforming stem cells in the NALT or circulating HSCs, that these cells will then readily home back to the bone marrow, even though currently there is evidence to suggest that these cells infrequently home back to healthy bone marrow.

Moreover, beyond the lack of support provided by the current mechanistic weight of evidence, as discussed, the epidemiology data, human and animal hematotoxicity data, and animal leukemia studies do not provide any

© 2011 Informa Healthcare USA, Inc.

support for the proposed modes of action for formaldehyde leukemogenesis.

It is worth pointing out an inconsistency with respect to data that have been put forth in the context of the three proposed modes of action for formaldehyde leukemogenesis. That is, reported observations of formaldehyde-induced hematotoxicity have been generally discussed as indicating a causal association with leukemia, and the proposed association has been discussed in the context of three possible modes of action. Bone marrow toxicity, however, can only occur if the chemical interacts directly with the bone marrow, which would only happen in the proposed mode of action that targets bone marrow. If the alternative modes of action are plausible (targeting circulating hematopoietic stem cells or NALT stem cells), formaldehyde would not be expected to cause hematotoxicity because it would not be directly acting on bone marrow. Instead, it likely would not be until tumor formation in bone marrow that one would expect a change in blood cell counts (likely increase in WBCs). That is, hematotoxicity would not be expected to occur in the exposed workers in the Zhang et al. (2010b) study if the mode of action was through formaldehyde damage to circulating progenitor cells or NALT stem cells. Interestingly, there are no other leukemogens that do not also show hematotoxicity, and therefore these leukemogens likely act by directly damaging the bone marrow. So, acceptance of one of these two modes of action (targeting circulating HSCs or NALT stem cells) suggests formaldehyde acts via a completely different mechanism from other leukemogens (i.e., in the absence of hematotoxicity), further suggesting biological implausibility.

Finally, it is informative to consider the phenomenon of apparent dependence of increased leukemia risk in certain epidemiology studies on peak exposure rather than on average or cumulative exposure. As we described in Section 4, in the NCI industrial worker cohort, Beane Freeman et al. (2009) found that the presence in a worker's career of peak exposures >4 ppm was associated with increased leukemia risk. In that section, we questioned whether this dependence on peaks was merely a matter of choosing among several dose metrics considered based on its outcome. But if the dependence on peaks is a real effect—if it is a discovery of the epidemiology investigations—there should be some correspondent peak-dependent aspects evident when proposed modes of action are investigated. It is not clear from consideration of the modes of action that have been proposed how such a peak dependence could work. If formaldehyde has to leave the respiratory tract and be redistributed to distant tissues such as marrow, the sharpness of a peak of exposure would be greatly attenuated as the absorbed formaldehyde mixed into the general circulation. Similarly, if susceptible cells are to migrate from marrow to the respiratory tract and back to the marrow, or from NALT in the respiratory tract to the marrow—processes that are hypothesized to be occurring at a low but ongoing level—it is not clear how peak inhalation exposures could have special effect. One would expect associations with other measures of exposure besides peak if this were the case. If genotoxic modes of action are proposed (so as to form the basis for concern regarding potential cancer risks to people experiencing low environmental exposures), the accumulation of risk of transforming mutations similarly must be an ongoing process that does not readily explain the apparent dependence on peak exposure as noted by Beane Freeman et al. (2009). In our view, such considerations illustrate the importance of integrating weight of evidence across disciplines, not just in combining conclusions from different disciplines, and in using a hypothesis-based framework to assess the consistency of analyses and their interpretations with mutual illumination across disciplines.

7. Discussion

The most current draft of the US EPA assessment of formaldehyde's human health risks (US EPA, 2010) states, "[h]uman epidemiological evidence is sufficient to conclude a causal association between formaldehyde exposure and … all leukemias, myeloid leukemia and lymphohematopoietic cancers as a group," but it also notes that "[l]imited evidence from animal bioassays is available to support the conclusion from human epidemiologic data that formaldehyde causes some types of lymphohematopoietic cancers." As is clear from the US EPA statement, these conclusions are backed by evaluations based initially on a judgment about the human data alone, conducted according to the approaches that epidemiologists use to evaluate whether the patterns observed among human studies of apparent associations between inhaled formaldehyde and lymphohematopoietic cancers are, in the judges' view, sufficiently indicative of a causative process. It is only afterward that the compatibility of this conclusion with information from animal studies or mode-of-action data is considered, and, if the human-data-only conclusion is one of causation, the presence or (as in the case of formaldehyde) lack of additional support is noted.

The most recent update of the IARC monograph (IARC, 2009) states that, with regard to formaldehyde, "the epidemiological evidence on leukaemia has become stronger, and new mechanistic studies support a conclusion of *sufficient evidence* in humans. This highlights the value of mechanistic studies, which in only 5 years have replaced previous assertions of biological implausibility with new evidence that formaldehyde can cause blood-cell abnormalities that are characteristic of leukaemia development." IARC further states that "[t]he Working Group was almost evenly split on the evaluation of formaldehyde causing leukaemias in humans, with the majority viewing the evidence as *sufficient* for carcinogenicity and the minority viewing the evidence as *limited*. Particularly relevant to the discussions regarding sufficient evidence was a recent study accepted for publication which, for the first time, reported aneuploidy in blood of exposed workers

characteristic of myeloid leukaemia and myelodysplastic syndromes with supporting information suggesting a decrease in the major circulating blood cell types and in circulating haematological precursor cells." Although the IARC monograph highlights mechanistic studies, it appears that "viewing the evidence as sufficient" stems predominantly from one human occupational study (likely Zhang et al., 2010b, although not cited by IARC, 2009).

Our concern with such a process is that it fails to appreciate the role that animal, toxicokinetic, and mode-of-action data can and should have, not just in the overall conclusion, but in the interpretation of the meaning of the epidemiological data themselves. If one concludes that the epidemiological data show causation, then there is an implicit conclusion that some mechanism for this causal process is not merely conceivable or not yet disproven, but it must actually exist. If it is firmly concluded that something is causal, it must also be firmly concluded that a means for that causation exists, even if it is not named. If animal studies or other mode-of-action studies are not in concordance with the human-data-only conclusion (epidemiology and the key mechanistic occupational study referenced by IARC), acceptance of the apparent causation in humans necessarily includes a further conclusion that the discordance is explicable—that the causes invoked for the human data either do not operate in animals or, for some scientifically plausible reason, are not manifested in observable consequences.

Our HBWoE approach calls attention to this and recognizes that the weight-of-evidence evaluation should evaluate these subsidiary conclusions about the plausibility of human mechanisms and their concordance or lack of concordance with mechanisms in animals. It is important to evaluate these subsidiary conclusions explicitly rather than leave them implicit. It is particularly important when, as is the case for formaldehyde, our understanding of these other aspects is not merely non-supporting of the human-data-only conclusion but actually conflicts with it. If inhaled formaldehyde is indeed a human leukemogen, then something about what is commonly understood, related to possible mechanisms and their potential operation in humans and rodents, is in error. Conversely, if it is indeed right to doubt the scientific plausibility of suggested mechanisms, their operation in human exposures, and their lack of operation in animal studies, then it is wrong to interpret the patterns among human studies as indicative of causality. Because the epidemologists' evaluation of causality from the human data entails judging how well a common causal explanation is supported by the array of observations compared to alternative explanations that attribute the apparent patterns to other, non-causal influences (such as chance and confounding), the scientific plausibility of the causal interpretation of the human-data patterns in view of other, non-epidemiologic data is an important part of a sound evaluation.

We have attempted to carry out a more complete evaluation across scientific disciplines for the case of inhaled formaldehyde and hematopoietic cancers in humans. In our reading of the weight of evidence, the conclusion that formaldehyde can cause such effects is not well supported.

In summary, the HBWoE evaluation for formaldehyde considers two alternative accounts. One account consists of acceptance of the epidemiology evidence as sufficiently compelling that, even in the face of weak hematological and carcinogenic evidence in animals and weak and inconsistent hematological evidence in humans, one of the proposed modes of action for formaldehyde leukemogenesis must be right, since its manifestations as increased leukemia risks are seen in the human studies. Moreover, the arguments against the biological plausibility of these modes of action must in some way be incorrect. Acceptance of this account is associated with many unanswered questions and post hoc explanations for how the current data should be interpreted as supporting it. This account requires that one accepts the reported exposure and disease information in the epidemiology studies as true, even though the lack of precise exposure data likely led to exposure measurement error and/or exposure misclassification that could have biased results, and disease misclassification in these studies likely led to unreliable risk estimates. It requires that all the many human studies that failed to show increased leukemia risk did so for plausible reasons, such that the lack of effects does not contradict formaldehyde's asserted general property of leukemogenicity. This account requires that one accepts an existence of an exposure-response relationship, despite the lack of consistently observed exposure-response associations within or among the epidemiology studies. It requires that one accepts the post hoc explanation of short latency for the increased risks associated with peak exposure observed in the NCI industrial worker cohort with follow-up through 1994, but not when follow-up was continued through 2004, even though this does not explain how this trend was not observed in the NCI embalmers cohort (Hauptmann et al., 2003; Walrath and Fraumeni, 1983, 1984) or garment workers cohort (Pinkerton et al., 2004), in whom risks were only observed with exposures over 20 years. This account requires that, although the epidemiology data were statistically analyzed in many different parallel ways with many finding no significant association, one chooses to focus only on the few marginally significant findings while ignoring the others as part of the evidence as a whole. For example, in the NCI industrial worker cohort, associations were reported with peak exposures, but there was no a priori reason to focus on peak exposures. These results should at most be treated as hypothesis-generating observation to be tested empirically. Otherwise, it is post hoc and arbitrary.

Moreover, this account (that formaldehyde is causally associated with leukemia) requires inclusion of an explanation as to why controlled animal experiments fail

© 2011 Informa Healthcare USA, Inc.

to show hematological or leukemogenic effects at high formaldehyde exposure concentrations (6 to 15 ppm). That is, what is being argued to be happening in humans (to allow the leukemogenic effect) must for some reason not be happening in the experimental animals, or else they would have been seen to have parallel hematotoxic and leukemogenic effects, as well as evidence of other consequences of operation of the proposed modes of action. It is not beyond reason that a leukemogenic effect of formaldehyde might be confined to humans, but there has been no explanation offered for why this might be so. Further, the proposed modes of action that would enable an effect in humans do not have any evident basis to be absent in rodents—indeed, some of the elements (migrating stem cells, effects on NALT), both consistent with or contrary to this account, are based on rat data. As it stands, the reasons for rodents not being subject to the proposed causative processes in humans constitutes an unstated corollary—one without empirical support or plausible basis—to the theories of human leukemogenesis of formaldehyde.

One needs to account for the inconsistencies among studies regarding the hematological evidence in humans; if there is an effect of formaldehyde inhalation, then what reasons are proposed for why it is not seen in many of the studies (and not seen at all in animals)? Only some of the hypothesized modes of action entail hematotoxicity, and so a proposal of its role in human leukemia depends on the particular variety of proposed mode of action being considered, with observations in favor of one mode tending to contradict other modes and hence in need of explanation for why such conflicts are not refuting. Finally, because of the weak and inconsistent epidemiological and toxicological evidence for a causal association, this account requires that one rely heavily on the truth of toxicokinetic and mechanistic hypotheses that permit a plausible biological mode of action. To accept this account as true, one must accept that somehow formaldehyde can move beyond the nasal respiratory mucosa to ultimately cause DNA damage and cellular transformation in bone marrow, circulating hematopoietic stem cells, or the NALT, even though there is a large body of evidence to suggest that inhaled formaldehyde (at reasonably high exposure concentrations for humans, 2 ppm) does not increase levels in the blood and does not cause DNA damage in cells and tissues beyond the nasal respiratory mucosa to a sufficient degree that would manifest as leukemia. If one is to conclude that formaldehyde is a "known" human leukemogen, one must assert not only that these hypothesized modes of action are conceivably true but that it is indeed known that one of them is true, for otherwise an essential and utterly necessary element of the causal conclusion is missing.

For this account (formaldehyde is causally associated with leukemia), there is a very large degree of ad hoc argument. That is, the elements of this account are chosen so as to fit the hypothesis already put forth, not based purely on an evaluation of the weight of the evidence as a whole and how it may (or may not) support the proposed hypothesis. Consequently, alternative accounts need to be considered.

An alternative, and contrasting, account is that it is not possible for formaldehyde to move beyond the nasal respiratory mucosa to cause systemic DNA damage and cellular transformation (in the bone marrow, circulating hematopoietic stem cells, or the NALT), and therefore there is no biologically plausible mechanism for formaldehyde leukemogenesis. This is supported by a large body of toxicokinetic and mechanistic data in animals and in vitro, and by inconsistent cytogenetic peripheral blood lymphocytes data in humans that are likely confounded by other exposures and a lack of reliable formaldehyde exposure information, in addition to the fact that there is little evidence to support the use of peripheral blood lymphocytes data as a biomarker for effects in hematopoietic stem cells or for leukemia. Further, the lack of toxicokinetic and mechanistic biological plausibility is supported by the largely negative toxicological evidence and a significant number of null epidemiology findings, which are considered under this account to be the true results, whereas those relatively isolated and unrepeated positive results are considered as false positives attributable to confounding by other exposures or to chance. If this account is true, an association between inhalation of formaldehyde and leukemia would be understood as not plausible for humans, and the few positive associations that have been observed would be attributed to alternative explanations (i.e., to other chemical exposures in the workplace, or lifestyle-related exposures such as smoking or alcohol consumption, or simply to chance).

In comparing these two accounts, neither is proven or disproven, but when assessing the weight of the available evidence in support of either account, it is clear that the first account requires far more ad hoc assumptions and post hoc explanations. In the first account, the inferences regarding potential human risk are not coming from the data themselves, but from assumptions invoked after the fact to fit the hypotheses put forth and without the evidence that would tie the weak epidemiological, toxicological, and mode-of-action data causally to formaldehyde inhalation exposure. Therefore, the weight of evidence for this account (i.e., exposure to formaldehyde in air is causally associated with leukemia in humans) is weak in comparison to the more substantial weight of evidence supporting the lack of a causal association.

Acknowledgments

The authors would like to acknowledge Robert Golden of ToxLogic LLC and Peter de la Cruz of Keller and Heckman LLP for providing critical review of the draft manuscript.

Declaration of interest

All of the authors are employees of Gradient, a private consulting firm that provides services to both private

and public organizations on toxicological and human health risk assessment issues. Individually, the authors have offered scientific testimony on formaldehyde risks to public organizations. The preparation of this review was sponsored by the Formaldehyde Council. The Council was provided the opportunity to review a draft of the paper and offer comments for consideration by the authors. The authors have sole responsibility for the content and the writing of the paper. The interpretations and views expressed in the paper are not necessarily those of the Formaldehyde Council.

References

Abkowitz JL, Robinson AE, Kale S, Long MW, Chen J. (2003). Mobilization of hematopoietic stem cells during homeostasis and after cytokine exposure. Blood 102(4):1249–1253.

Abrams RA, Glaubiger D, Appelbaum FR, Deisseroth AB. (1980). Result of attempted hematopoietic reconstitution using isologous, peripheral blood mononuclear cells: A case report. Blood 56(3):516–520.

Acheson ED, Barnes HR, Gardner MJ, Osmond C, Pannett B, Taylor CP. (1984). Formaldehyde in the British chemical industry. An occupational cohort study. Lancet 1(8377):611–616.

Agency for Toxic Substances and Disease Registry (ATSDR). (1999). Toxicological Profile for Formaldehyde (Update). Atlanta, GA: ATSDR.

Agency for Toxic Substances and Disease Registry (ATSDR). (2010). Addendum to the Toxicological Profile for Formaldehyde. Atlanta, GA: ATSDR.

Agency for Toxic Substances and Disease Registry (ATSDR). (2007). Toxicological Profile for Benzene. [Online] Available at: http://www.atsdr.cdc.gov/toxprofiles/tp3.pdf. Accessed on 7 April 2008.

Albert RE, Sellakumar AR, Laskin S, Kuschner M, Nelson N, Snyder CA. (1982). Gaseous formaldehyde and hydrogen chloride induction of nasal cancer in the rat. J Natl Cancer Inst 68(4):597–603.

Ambroise D, Moulin JJ, Squinazi F, Protois JC, Fontana JM, Wild P. (2005). Cancer mortality among municipal pest-control workers. Int Arch Occup Environ Health 78(5):387–393.

Andersen ME, Clewell HJ III, Bermudez E, Dodd DE, Wilson GA, Campbell JL, Thomas RS. (2010). Formaldehyde: Integrating dosimetry, cytotoxicity and genomics to understand dose-dependent transitions for an endogenous compound. Toxicol Sci 118 (2):716-731.

Andjelkovich DA, Janszen DB, Brown MH, Richardson RB, Miller FJ. (1995). Mortality of iron foundry workers: IV. Analysis of a subcohort exposed to formaldehyde. J Occup Environ Med 37(7):826–837.

Appelman LM, Woutersen RA, Zwart A, Falke HE, Feron VJ. (1988). One-year inhalation toxicity study of formaldehyde in male rats with a damaged or undamaged nasal mucosa. J Appl Toxicol 8(2):85–90.

Arts JH, Rennen MA, de Heer C. (2006). Inhaled formaldehyde: Evaluation of sensory irritation in relation to carcinogenicity. Regul Toxicol Pharmacol 44(2):144–160.

Aster J, Kumar, V. (1999). White cells, lymph nodes, spleen, and thymus. In: Cotran RS, Kumar V, Collins T, eds. Robbins Pathologic Basis of Disease. (Sixth edition). Philadelphia: W.B. Saunders Co., 644–695.

Bachand A, Mundt KA, Mundt DJ, Montgomery RR. (2010). Epidemiological studies of formaldehyde exposure and risk of leukemia and nasopharyngeal cancer: A meta-analysis. Crit Rev Toxicol 40(2):85–100.

Batelle Columbus Laboratories. (1981). Final Report on a Chronic Inhalation Toxicology Study in Rats and Mice Exposed to Formaldehyde. Volume I (Excerpts). Report to Chemical Industry Institute of Toxicology. Columbus, OH: Batelle Columbus Laboratories.

Battershill JM, Burnett K, Bull S. (2008). Factors affecting the incidence of genotoxicity biomarkers in peripheral blood lymphocytes: Impact on design of biomonitoring studies. Mutagenesis 23(6):423–437.

Beane Freeman LE, Blair A, Lubin JH, Stewart PA, Hayes RB, Hoover RN, Hauptmann M. (2009). Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries: The National Cancer Institute Cohort. J Natl Cancer Inst 101(10):751–761.

Bertazzi PA, Pesatori AC, Radice L, Zocchetti C, Vai T. (1986). Exposure to formaldehyde and cancer mortality in a cohort of workers producing resins. Scand J Work Environ Health 12(5):461–468.

Bertazzi PA, Pesatori AC, Guercilena S, Consonni D, Zocchetti C. (1989). Cancer risk among workers producing formaldehyde-based resins: Extension of follow-up. Med Lav 80:111–122.

Blair A, Stewart P, O'Berg M, Gaffey W, Walrath J, Ward J, Bales R, Kaplan S, Cubit D. (1986). Mortality among industrial workers exposed to formaldehyde. J Natl Cancer Inst 76:1071–1084.

Blair A, Saracci R, Stewart PA, Hayes RB, Shy C. (1990). Epidemiologic evidence on the relationship between formaldehyde exposure and cancer. Scand J Work Environ Health 16(6):381–393.

Blair A, Zheng T, Linos A, Stewart PA, Zhang YW, Cantor KP. (2001). Occupation and leukemia: A population-based case-control study in Iowa and Minnesota. Am J Ind Med 40:3–14.

Bonassi S, Norppa H, Ceppi M, Stromberg U, Vermeulen R, Znaor A, Cebulska-Wasilewska A, Fabianova E, Fucic A, Gundy S, Hansteen IL, Knudsen LE, Lazutka J, Rossner P, Sram RJ, Boffetta P. (2008). Chromosomal aberration frequency in lymphocytes predicts the risk of cancer: Results from a pooled cohort study of 22,358 subjects in 11 countries. Carcinogenesis 29(6):1178–1183.

Bosetti C, McLaughlin JK, Tarone RE, Pira E, La Vecchia C. (2008). Formaldehyde and cancer risk: A quantitative review of cohort studies through 2006. Ann Oncol 19(1):29–43.

Bower M, McCall-Peat N, Ryan N, Davies L, Young AM, Gupta S, Nelson M, Gazzard B, Stebbing J. (2004). Protease inhibitors potentiate chemotherapy-induced neutropenia. Blood 104(9): 2943–2946.

Brondeau MT, Bonnet P, Guenier JP, Simon P, de Ceaurriz J. (1990). Adrenal-dependent leucopenia after short-term exposure to various airborne irritants in rats. J Appl Toxicol 10(2):83–86.

Casanova M, Heck Hd'A. (1987). Further studies of the metabolic incorporation and covalent binding of inhaled [3H]- and [14C] formaldehyde in Fischer-344 rats: Effects of glutathione depletion. Toxicol Appl Pharmacol 89(1):105–121.

Casanova-Schmitz M, Starr TB, Heck HD. (1984). Differentiation between metabolic incorporation and covalent binding in the labeling of macromolecules in the rat nasal mucosa and bone marrow by inhaled [14C]- and [3H]formaldehyde. Toxicol Appl Pharmacol 76(1):26–44.

Chiazze L Jr, Watkins DK, Fryar C. (1997). Historical cohort mortality study of a continuous filament fiberglass manufacturing plant. I. White men. J Occup Environ Med 39(5):432–441.

Coggon D, Harris EC, Poole J, Palmer KT. (2003). Extended follow-up of a cohort of British chemical workers exposed to formaldehyde. J Natl Cancer Inst 95(21):1608–1615.

Coleman R, Kogan I. (1998). An improved low-formaldehyde embalming fluid to preserve cadavers for anatomy teaching. J Anat 192(Pt. 3):443–446.

Collins JJ, Lineker GA. (2004). A review and meta-analysis of formaldehyde exposure and leukemia. Regul Toxicol Pharmacol 40(2):81–91.

Cotran RS, Kumar V, Collins T. (1999). Robbins Pathologic Basis of Disease. (Sixth edition). Philadelphia, PA: W.B. Saunders Co.

Cummins D, Wilson ME, Foulger KJ, Dawson D, Hogarth AM. (1998). Haematological changes associated with influenza vaccination in people aged over 65: Case report and prospective study. Clin Lab Haematol 20(5):285–287.

Dallas CE, Scott MJ, Ward JB Jr, Theiss JC. (1992). Cytogenetic analysis of pulmonary lavage and bone marrow cells of rats after repeated formaldehyde inhalation. J Appl Toxicol 12(3):199–203.

de Groot AC, Flyvholm MA, Lensen G, Menne T, Coenraads PJ. (2009). Formaldehyde-releasers: Relationship to formaldehyde

© 2011 Informa Healthcare USA, Inc.

contact allergy. Contact allergy to formaldehyde and inventory of formaldehyde-releasers. Contact Dermatitis 61(2): 63–85.

de Groot AC, Maibach HI. (2010). Does allergic contact dermatitis from formaldehyde in clothes treated with durable-press chemical finishes exist in the USA? Contact Dermatitis 62(3): 127–136.

Dean JH, Lauer LD, House RV, Murray MJ, Stillman WS, Irons RD, Steinhagen WH, Phelps MC, Adams DO. (1984). Studies of immune function and host resistance in B6C3F1 mice exposed to formaldehyde. Toxicol Appl Pharmacol 72(3):519–529.

Deane JA, Hickey MJ. (2009). Molecular mechanisms of leukocyte trafficking in T-cell-mediated skin inflammation: Insights from intravital imaging. Expert Rev Mol Med 11:e25.

Dell L, Teta MJ. (1995). Mortality among workers at a plastics manufacturing and research and development facility: 1946–1988. Am J Ind Med 28(3):373–384.

DeVoney D, Vandenberg J, Whalan J. (2006). Hypothesized mode of action for formaldehyde-induced lymphoid malignancies. Poster presented at EPA Science Forum. Washington, DC: US EPA National Center for Environmental Assessment (NCEA). 1p. (poster at EPA Science Forum)

DeVoney D, Whalan J, Sonawane B. (2010). Evidence of formaldehyde-induced leukemia and lymphoma in long-term rodent bioassays. Toxicologist 114(1):113. [Abstract No. 520].

Dobson RL, Motlagh S, Quijano M, Cambron RT, Baker TR, Pullen AM, Regg BT, Bigalow-Kern AS, Vennard T, Fix A, Reimschuessel R, Overman G, Shan Y, Daston GP. (2008). Identification and characterization of toxicity of contaminants in pet food leading to an outbreak of renal toxicity in cats and dogs. Toxicol Sci 106(1):251–262.

Edgren G, Hjalgrim H, Reilly M, Tran TN, Rostgaard K, Shanwell A, Titlestad K, Adami J, Wikman A, Jersild C, Gridley G, Wideroff L, Nyren O, Melbye M. (2007). Risk of cancer after blood transfusion from donors with subclinical cancer: A retrospective cohort study. Lancet 369(9574):1724–1730.

Edling C, Järvholm B, Andersson L, Axelson O. (1987). Mortality and cancer incidence among workers in an abrasive manufacturing industry. Br J Ind Med 44:57–59.

European Food Safety Authority (EFSA). (2006). Opinion of the Scientific Panel on Food Additives, Flavourings, Processing Aids and Materials in Contact with Food (AFC) on a request from the Commission related to use of formaldehyde as a preservative during the manufacture and preparation of food additives: Question number EFSA-Q-2005-032. EFSA J. 415: 1–10.

Farage MA. (2008). Enhancement of visual scoring of skin irritant reactions using cross-polarized light and parallel-polarized light. Contact Dermatitis 58(3):147–155.

Feron VJ, Bruyntjes JP, Woutersen RA, Immel HR, Appelman LM. (1988). Nasal tumours in rats after short-term exposure to a cytotoxic concentration of formaldehyde. Cancer Lett 39(1):101–111.

Feron VJ, Til HP, Woutersen RA. (1990). Response to Soffritti et al. [re: Formaldehyde: An experimental multipotential carcinogen.] ([letter).]. Toxicol Ind Health 6(6):637–639.

Forte WC, Guardian VC, Mantovani PA, Dionigi PC, Menezes MC. (2009). Evaluation of phagocytes in atopic dermatitis. Allergol Immunopathol (Madr) 37(6):302–308.

Franks SJ. (2005). A mathematical model for the absorption and metabolism of formaldehyde vapour by humans. Toxicol Appl Pharmacol 206(3):309–320.

Gardner MJ, Pannett B, Winter PD, Cruddas AM. (1993). A cohort study of workers exposed to formaldehyde in the British chemical industry: An update. Br J Ind Med 50(9): 827–834.

Gerin M, Siemiatycki J, Nadon L, Dewar R, Krewski D. (1989). Cancer risks due to occupational exposure to formaldehyde: Results of a multi-site case-control study in Montreal. Int J Cancer 44(1):53–58.

Gocke E, King MT, Eckhardt K, Wild D. (1981). Mutagenicity of cosmetics ingredients licensed by the European Communities. Mutat Res 90(2):91–109.

Golden R, Pyatt D, Shields PG. (2006). Formaldehyde as a potential human leukemogen: An assessment of biological plausibility. Crit Rev Toxicol 36(2):135–153.

Hall A, Harrington JM, Aw TC. (1991). Mortality study of British pathologists. Am J Ind Med 20(1):83–89.

Hansen J, Olsen JH. (1995). Formaldehyde and cancer morbidity among male employees in Denmark. Cancer Causes Control 6(4):354–360.

Harrington JM, Shannon HS. (1975). Mortality study of pathologists and medical laboratory technicians. Br Med J 4(5992):329–332.

Haseman JK, Hailey JR, Morris RW. (1998). Spontaneous neoplasm incidences in Fischer 344 rats and B6C3F1 mice in two-year carcinogenicity studies: A National Toxicology Program update. Toxicol. Pathol 26(3):428–441.

Hauptmann M, Lubin JH, Stewart PA, Hayes RB, Blair A. (2003). Mortality from lymphohematopoietic malignancies among workers in formaldehyde industries. J Natl Cancer Inst 95(21):1615–1623.

Hauptmann M, Stewart PA, Lubin JH, Beane Freeman LE, Hornung RW, Herrick RF, Hoover RN, Fraumeni JF, Blair A, Hayes RB. (2009). Mortality from lymphohematopoietic malignancies and brain cancer among embalmers exposed to formaldehyde. J Natl Cancer Inst 101(24):1696–1708.

Hayes RB, Blair A, Stewart PA, Herrick RF, Mahar H. (1990). Mortality of U.S. embalmers and funeral directors. Am J Ind Med 18(6):641–652.

Heck H, Casanova M. (2004). The implausibility of leukemia induction by formaldehyde: A critical review of the biological evidence on distant-site toxicity. Regul Toxicol Pharmacol 40(2):92–106.

Hill AB. (1965). The environment and disease: Association or causation? Proc R Soc Med 58:295–300.

Holmstrom M, Wilhelmsson B, Hellquist H. (1989). Histological changes in the nasal mucosa in rats after long-term exposure to formaldehyde and wood dust. Acta Otolaryngol 108(3-4):274–283.

Hyland L, Villarreal-Ramos B, Clarke B, Baaten B, Hou S. (2005). Bone marrow immunosuppression in *Salmonella*-infected mice is prolonged following influenza virus infection. Exp Hematol 33(12):1477–1485.

Iarmarcovai G, Bonassi S, Sari-Minodier I, Baciuchka-Palmaro M, Botta A, Orsiere T. (2007). Exposure to genotoxic agents, host factors, and lifestyle influence the number of centromeric signals in micronuclei: A pooled re-analysis. Mutat Res 615:18–27.

Iarmarcovai G, Bonassi S, Botta A, Baan RA, Orsiere T. (2008). Genetic polymorphisms and micronucleus formation: A review of the literature. Mutat Res 658(3):215–233.

Infante PF, Wagoner JK, Rinsky RA, Young RJ. (1977). Leukemia in benzene workers. Lancet 76(3):76–78.

International Agency for Research on Cancer (IARC). (2000). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans:. Volume 75: Ionizing Radiation, Part 1: X- and Gamma (γγ)-Radiation, and Neutrons. Lyon, France: International Agency for Research on Cancer (IARC), World Health Organization (WHO). [Online] Available at: http://monographs.iarc.fr/ENG/Monographs/vol75/mono75.pdf. Accessed on 9 March 2009.

International Agency for Research on Cancer (IARC). (2006). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 88: Formaldehyde, 2-Butoxyethanol and 1-tert-Butoxypropan-2-ol. Geneva, Switzerland)Lyon, France: International Agency for Research on Cancer (Lyon, France);, World Health Organization (WHO) (IARC Monograph No. 88. [Online]. Available at: http://monographs.iarc.fr/ENG/Monographs/vol88/mono88.pdf. Accessed on 3 May 2010.

International Agency for Research on Cancer (IARC). (2009). A Review of Human Carcinogens—Part F: Chemical Agents and Related Occupations. [Online]Available at: http://monographs.iarc.fr/pdfnews/WG-100F.pdf. Accessed on 11 January 2011.

Irons RD, Stillman WS. (1996). The process of leukemogenesis. Environ Health Perspect 104(Suppl. 6):1239–1246.

Ishmael J, Dugard PH. (2006). A review of perchloroethylene and rat mononuclear cell leukemia Regul Toxicol Pharmacol 45(2):178–184.

Jakab MG, Klupp T, Besenyei K, Biro A, Major J, Tompa A. (2010). Formaldehyde-induced chromosomal aberrations and apoptosis

in peripheral blood lymphocytes of personnel working in pathology departments. Mutat. Res. 698(1-2):11–17.

Jenerowicz D, Czarnecka-Operacz M, Silny W. (2007). Peripheral blood eosinophilia in atopic dermatitis. Acta Dermatovenerol. Alp. Panonica Adriat. 37(6): 302–308.

Jiang S, Yu L, Cheng J, Leng S, Dai Y, Zhang Y, Niu Y, Yan H, Qu W, Zhang C, Zhang K, Yang R, Zhou L, Zheng Y. (2010). Genomic damages in peripheral blood lymphocytes and association with polymorphisms of three glutathione S-transferases in workers exposed to formaldehyde. Mutat. Res. 695(1-2):9–15.

Johannsen FR, Levinskas GJ, Tegeris AS. (1986). Effects of formaldehyde in the rat and dog following oral exposure. Toxicol. Lett. 30(1):1–6.

Johnson E, Cotter FE. (1997). Monosomy 7 and 7q—Associated with myeloid malignancy. Blood Rev. 11(1):46–55.

Jurek AM, Greenland S, Maldonado G, Church TR. (2005). Proper interpretation of non-differential misclassification effects: Expectations vs observations. Int J Epidemiol 34(3):680–687.

Kamata E, Nakadate M, Uchida O, Ogawa Y, Suzuki S, Kaneko T, Saito M, Kurokawa Y. (1997). Results of a 28-month chronic inhalation toxicity study of formaldehyde in male Fisher-344 rats. J Toxicol Sci 22(3):239–254.

Kemp DE, Franco KN. (2002). Possible leukopenia associated with long-term use of echinacea. J Am Board Fam Pract 15(5):417–419.

Kerns WD, Pavkov KL, Donofrio DJ, Gralla EJ, Swenberg JA. (1983). Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. Cancer Res 43(9):4382–4392.

Kipen HM, Cody RP, Goldstein BD. (1989). Use of longitudinal analysis of peripheral blood counts to validate historical reconstructions of benzene exposure. Environ Health Perspect 82:199–206.

Kitaeva LV, Kitaev EM, Pimenova MN. (1990). The cytopathic and cytogenetic sequelae of chronic inhalational exposure to formaldehyde on female germ cells and bone marrow cells in rats. Tsitologiia 32(12):1212–1216.

Kligerman AD, Phelps MC, Erexson GL. (1984). Cytogenetic analysis of lymphocytes from rats following formaldehyde inhalation. Toxicol Lett 21(3):241–246.

Kriebel D, Checkoway H, Pearce N. (2007). Exposure and dose modelling in occupational epidemiology. Occup Environ Med 64(7):492–498.

Krimsky, S. (2005). The weight of scientific evidence in policy and law. Am J Public Health 95:S129–S136.

Kuo H, Jian G, Chen C, Liu C, Lai J. (1997). White blood cell count as an indicator of formaldehyde exposure. Bull Environ Contam Toxicol 59(2):261–267.

Kuper FK, van Oostrum L, Ma-Hock L, Durrer S, Woutersen RA. (2009). Hyperplasia of the lymphoepithelium of NALT in rats but not in mice upon 28-day exposure to 15ppm formaldehyde vapor. Exp Toxicol Pathol 63(1-2):25-32.

Lavrov VF, Semenkov VF. (1991). The infection of mice with the influenza virus suppresses bone marrow precursor differentiation in the erythroid and granulocyte directions. Vopr Virusol 36(4):284–286.

Lebre MC, van Capel TM, Bos JD, Knol EF, Kapsenberg ML, de Jong EC. (2008). Aberrant function of peripheral blood myeloid and plasmacytoid dendritic cells in atopic dermatitis patients. J Allergy Clin Immunol 122(5):969–976.

Levine RJ, Andjelkovich DA, Shaw LK. (1984). The mortality of Ontario undertakers and a review of formaldehyde-related mortality studies. J Occup Med 26(10):740–746.

Linkov I, Loney D, Cormier S, Satterstrom FK, Bridges T. (2009). Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. Sci Total Environ 407(19):5199–5205.

Linos A, Blair A Cantor KP, Burmeister L, VanLier S, Gibson RW, Schuman L, Everett G. (1990). Leukemia and non-Hodgkin's Hodgkin's lymphoma among embalmers and funeral directors ([letter).]. J Natl Cancer Inst 82(1):66.

Logue JN, Barrick, MK, Jessup GL Jr. (1986). Mortality of radiologists and pathologists in the Radiation Registry of Physicians. J Occup Med 28(2):91–99.

Lu K, Collins LB, Ru H, Bermudez E, Swenberg JA. (2010). Distribution of DNA adducts caused by inhaled formaldehyde is consistent with induction of nasal carcinoma but not leukemia. Toxicol Sci 116(2):441–451.

Lu K, Moeller B, Doyle-Eisele M, McDonald J, Swenberg JA. (2011). Molecular dosimetry of *N*(2)-hydroxymethyl-dG DNA adducts in rats exposed to formaldehyde. Chem Res Toxicol 24(2):159-61.

Lyapina M, Zhelezova G, Petrova E, Boev M. (2004). Flow cytometric determination of neutrophil respiratory burst activity in workers exposed to formaldehyde. Int Arch Occup Environ Health 77(5):335–340.

Marsh GM, Youk AO, Stone RA, Buchanich JM, Gula MJ, Smith TJ, Quinn MM. (2001). Historical cohort study of US man-made vitreous fiber production workers: I. 1992 fiberglass cohort follow-up: Initial findings. J Occup Environ Med 43(9):741–756.

Marsh GM, Youk AO. (2004). Reevaluation of mortality risks from leukemia in the formaldehyde cohort study of the National Cancer Institute. Regul Toxicol Pharmacol 40(2):113–124.

Marsh GM, Youk AO. (2005). Reevaluation of mortality risks from nasopharyngeal cancer in the formaldehyde cohort study of the National Cancer Institute. Regul Toxicol Pharmacol 42:275–283.

Marsh GM, Youk AO, Morfeld P. (2007a). Mis-specified and non-robust mortality risk models for nasopharyngeal cancer in the National Cancer Institute formaldehyde worker cohort study. Regul Toxicol Pharmacol 47(1):59–67.

Marsh GM, Youk AO, Buchanich JM, Erdal S, Esmen NA. (2007b). Work in the metal industry and nasopharyngeal cancer mortality among formaldehyde-exposed workers. Regul Toxicol Pharmacol 48(3):308–319.

Matanoski GM. (1989.). Risks of Pathologists Exposed to Formaldehyde (Final). Washington, DC: National Technical Information Service (NTIS). NTIS PB91–173682.

Mayo Clinic Staff. (2008.). Low white blood cell count. [Online] Available at: http://www.mayoclinic.com/health/low-white-blood-cell-count/my00162. Accessed on 30 June 2010.

McKinney-Freeman S, Goodell MA. (2004). Circulating hematopoietic stem cells do not efficiently home to bone marrow during homeostasis. Exp Hematol 32(9):868–876.

Miller BA, Ries LA, Hankey BF, Kosary CL, Edwards BK. (1992). Cancer Statistics Review. 1973–1989. Frederick, MD: National Cancer Institute. NIH Pub. lication No. 92–2789.

Moeller BC, Lu K, Doyle-Eisele M, McDonald J, Gigliotti A, Swenberg JA. (2011). Determination of *N*(2)-hydroxymethyl-dG adducts in the nasal epithelium and bone marrow of nonhuman primates following (13)CD(2)-formaldehyde inhalation exposure. Chem Res Toxicol 24(2):162-4.

Monticello TM, Morgan KT, Everitt JI, Popp JA. (1989). Effects of formaldehyde gas on the respiratory tract of rhesus monkeys. Pathology and cell proliferation. Am J Pathol 134(3):515–527.

Moolgavkar S, Krewski D, Schwartz M. (1999). Mechanisms of carcinogenesis and biologically based models for estimation and prediction of risk. In: Moolgavkar, S, Krewski, D, Zeise, L, Cardis, L, and Moller, H, eds. Quantitative Estimation and Prediction of Human Cancer Risks. IARC Scientific Publications No. 31. Lyon, France: International Agency for Research on Cancer.

Moorman AV, Roman E, Cartwright RA, Morgan GJ. (2002). Smoking and the risk of acute myeloid leukaemia in cytogenetic subgroups. Br J Cancer 86(1):60–62.

Morris MS, Jacques PF, Rosenberg IH, Selhub J. (2007). Folate and vitamin B-12 status in relation to anemia, macrocytosis, and cognitive impairment in older Americans in the age of folic acid fortification. Am J Clin Nutr 85(1):193–200.

Murgia E, Ballardin M, Bonassi S, Rossi AM, Barale R. (2008). Validation of micronuclei frequency in peripheral blood lymphocytes as early cancer risk biomarker in a nested case-control study. Mutat Res 639(1-2):27–34.

Murrell W, Feron F, Wetzig A, Cameron N, Splatt K, Bellette B, Bianco J, Perry C, Lee G, Mackay-Sim A. (2005). Multipotent stem cells from adult olfactory mucosa. Dev Dyn 233(2):496–515.

© 2011 Informa Healthcare USA, Inc.

Nakanishi N, Suzuki K, Tatara K. (2003). Association between lifestyle and white blood cell count: A study of Japanese male office workers. Occup Med (Lond) 53(2):135–137.

National Institutes of Health. (2008). Anemia in Kidney Disease and Dialysis. Bethesda, MD: National Institutes of Health. NIH Publication No. 09-4619. [Online] Available at: http://kidney.niddk.nih.gov/kudiseases/pubs/anemia/. Accessed on 4 June 4 2010.

Neuss S, Moepps B, Speit G. (2010). Exposure of human nasal epithelial cells to formaldehyde does not lead to DNA damage in lymphocytes after co-cultivation. Mutagenesis 25(4):359-64.

Nohynek GJ, Plard JP, Wells MY, Zerial A, Roquet F. (1997). Comparison of the potency of glycosylated and nonglycosylated recombinant human granulocyte colony-stimulating factors in neutropenic and nonneutropenic CD rats. Cancer Chemother Rep 39(3):259–266.

Ott MG, Teta MJ, Greenberg HL. (1989). Lymphatic and hematopoietic tissue cancer in a chemical manufacturing environment. Am J Ind Med 16(6):631–643.

Pala M, Ugolini D, Ceppi M, Rizzo F, Maiorana L, Bolognesi C, Schiliro T, Gilli G, Bigatti P, Bono R, Vecchio D. (2008). Occupational exposure to formaldehyde and biological monitoring of Research Institute workers. Cancer Detect Prev 32(2):121–126.

Partanen T, Kauppinen T, Luukkonen R, Hakulinen T, Pukkala E. (1993). Malignant lymphomas and leukemias, and exposures in the wood industry: An industry-based case-referent study. Int Arch Occup Environ Health 64(8):593–596.

Paulsson K, Johansson B. (2007). Trisomy 8 as the sole chromosomal aberration in acute myeloid leukemia and myelodysplastic syndromes. Pathol Biol (Paris) 55(1):37–48.

Pinkerton LE, Hein MJ, Stayner LT. (2004). Mortality among a cohort of garment workers exposed to formaldehyde: An update. Occup Environ Med 61:193–200. [Online] Available at: http://oem.bmj.com/. Accessed on 25 July 25 2005.

Pyatt D, Natelson E, Golden R. (2008). Is inhalation exposure to formaldehyde a biologically plausible cause of lymphohematopoietic malignancies? Regul Toxicol Pharmacol 51(1):119–133.

Rhomberg LR Bailey LA Goodman JE. (2010). Hypothesis-based weight of evidence: A tool for evaluating and communicating uncertainties and inconsistencies in the large body of evidence in proposing a carcinogenic mode of action — Naphthalene as an example. Critical Reviews in Toxicology. Crit Rev Toxicol 40(8):671–696.

Rice J, Resar LM. (1998). Hematologic abnormalities associated with influenza A infection: A report of 3 cases. Am J Med Sci 316(6):401–403.

Rinsky RA, Hornung RW, Silver SR, Tseng CY. (2002). Benzene exposure and hematopoietic mortality: A long-term epidemiologic risk assessment. Am J Ind Med 42(6):474–480.

Rinsky RA, Smith AB, Hornung R, Filloon TG, Young RJ, Okun AH, Landrigan PJ. (1987). Benzene and leukemia: An epidemiologic risk assessment. N Engl J Med 316:1044–1050.

Robinson C, Fowler D, Brown DP, Lemen RA. (1987). Plywood Mill Workers' Workers' Mortality Patterns 1945–1977 (Revised). Springfield, VA: National Technical Information Service (NTIS). NTIS PB90–147075.

Rowley JD. (2000). Molecular genetics in acute leukemia. Leukemia 14(3):513–517.

Salthammer T, Mentese S, Marutzky R. (2010). Formaldehyde in the indoor environment. Chem Rev 110(4):2536–2572

Sari DK, Kuwahara S, Tsukamoto Y, Hori H, Kunugita N, Arashidani K, Fujimaki H, Sasaki F. (2004). Effect of prolonged exposure to low concentrations of formaldehyde on the corticotropin releasing hormone neurons in the hypothalamus and adrenocorticotropic hormone cells in the pituitary gland in female mice. Brain Res 1013(1):107–116.

Schmahl D, Habs M. (1979). Carcinogenic action of low-dose cyclophosphamide given orally to Sprague-Dawley rats in a lifetime experiment. Int J Cancer 23(5):706–712.

Schmid O, Speit G. (2007). Genotoxic effects induced by formaldehyde in human blood and implications for the interpretation of biomonitoring studies. Mutagenesis 22(1):69–74.

Schnatter AR, Kerzic PJ, Zhou Y, Chen M, Nicolich MJ, Lavelle K, Armstrong TW, Bird MG, Lin L, Fu H, Irons RD. (2010). Peripheral blood effects in benzene-exposed workers. Chem Biol Interact 184(1-2):174–181.

Schottenfeld D, Fraumeni JF. (2006). Cancer Epidemiology and Prevention. (Third edition). New York, NY: Oxford University Press.

Schulz C, von Andrian UH, Massberg S. (2009). Hematopoietic stem and progenitor cells: Their mobilization and homing to bone marrow and peripheral tissue. Immunol Res 44(1-3):160–168.

Schwilk E, Zhang L, Smith MT, Smith AH, Steinmaus C. (2010). Formaldehyde and leukemia: An updated meta-analysis and evaluation of bias. J Occup Environ Med 52(9):878–886.

Scott CS, Chiu WA. (2006). Trichloroethylene cancer epidemiology: A consideration of select issues. Environ Health Perspect 114(9):1471–1478.

Sellakumar AR, Snyder CA, Solomon JJ, Albert RE. (1985). Carcinogenicity of formaldehyde and hydrogen chloride in rats. Toxicol. Appl. Pharmacol. 81(3 Pt 1):401–406.

Shaham J, Bornstein Y, Meltzer A, Kaufman Z, Palma E, Ribak J. (1996). DNA-protein crosslinks, a biomarker of exposure to formaldehyde—In vitro and in vivo studies. Carcinogenesis 17:121–125.

Shaham J, Bomstein Y, Melzer A, Ribak J. (1997). DNA-Protein crosslinks and sister chromatid exchanges as biomarkers of exposure to formaldehyde. Int J Occup Environ Health 3:95–104.

Shaham J, Bomstein Y, Gurvich R, Rashkovsky M, Kaufman Z. (2003). DNA-protein crosslinks and p53 protein expression in relation to occupational exposure to formaldehyde. Occup Environ Med 60(6):403–409.

Shen CF, Huang SC, Wang SM, Wang JR, Liu CC. (2008). Decreased leukocytes and other characteristics of laboratory findings of influenza virus infections in children. J Microbiol Immunol Infect 41(4):294–300.

Singbartl K, Ley K. (2004). Leukocyte recruitment and acute renal failure. J Mol Med 82(2):91–101.

Soffritti M, Maltoni C, Maffei F, Biagi R. (1989). Formaldehyde: An experimental multipotential carcinogen. Toxicol Ind Health 5:699–730.

Soffritti M, Belpoggi F, Lambertini L, Lauriola M, Padovani M, Maltoni C. (2002). Results of long-term experimental studies on the carcinogenicity of formaldehyde and acetaldehyde in rats. Ann N Y Acad Sci 982:87–105.

Speit G, Schmid O. (2006). Local genotoxic effects of formaldehyde in humans measured by the micronucleus test with exfoliated epithelial cells. Mutat Res 613:1–9.

Speit G, Zeller J, Schmid O, Elhajouji A, Ma-Hock L, Neuss S. (2009). Inhalation of formaldehyde does not induce systemic genotoxic effects in rats. Mutat. Res. 677(1-2):76–85.

Srivastava AK, Gupta BN, Gaur JS, Bihari V. (1992). Clinical evaluation of workers handling melamine formaldehyde resin ([letter).]. J Toxicol Clin Toxicol 30(4):677–681.

Stellman SD, Demers PA, Colin D, Boffetta P. (1998). Cancer mortality and wood dust exposure among participants in the American Cancer Society Cancer Prevention Study-II (CPS-II). Am J Ind Med 34(3):229–237.

Stern FB, Beaumont JJ, Halperin WE, Murthy LI, Hills BW, Fajen JM. (1987). Mortality of chrome leather tannery workers and chemical exposures in tanneries. Scand. J. Work Environ. Health 13(2):108–117.

Stewart PA, Blair A, Cubit DA, Bales RE, Kaplan SA, Ward J, Gaffey W, O"Berg MT, Walrath J. (1986). Estimating historical exposures to formaldehyde in a retrospective mortality study. Appl Ind Hyg 1(1):34–41.

Stewart PA, Herrick RF, Feigley CE, Utterback DF, Hornung R, Mahar H, Hayes R, Douthit DE, Blair A. (1992). Study design for assessing

exposures of embalmers for a case-control study. Part I. Monitoring results. Appl Occup Environ Hyg 7(8):532–540.

Stroup NE, Blair A, Erikson GE. (1986). Brain cancer and other causes of death in anatomists. J Natl Cancer Inst 77(6):1217–1224.

Sun Q, Shen Y, Sun N, Zhang GJ, Chen Z, Fan JF, Jia LQ, Xiao HZ, Li XR, Puschner B. (2010). Diagnosis, treatment and follow-up of 25 patients with melamine-induced kidney stones complicated by acute obstructive renal failure in Beijing Children's Children's Hospital. Eur J Pediatr 169(4):483–489.

Szabo G, Mandrekar P. (2009). A recent perspective on alcohol, immunity, and host defense. Alcohol Clin Exp Res 33(2):220–232.

Takahashi M, Hasegawa R, Furukawa F, Toyoda K, Sato H, Hayashi Y. (1986). Effects of ethanol, potassium metabisulfite, formaldehyde and hydrogen peroxide on gastric carcinogenesis in rats after initiation with *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. Jpn J Cancer Res 77(2):118–124.

Tang X, Bai Y, Duong A, Smith MT, Li L, Zhang L. (2009). Formaldehyde in China: Production, consumption, exposure levels, and health effects. Environ Int 35(8):1210–1224.

Tatham L, Tolbert P, Kjeldsberg C. (1997). Occupational risk factors for subgroups of non-Hodgkin's Hodgkin's lymphoma. Epidemiology 8(5):551–558.

Til HP, Woutersen RA, Feron VJ, Clary JJ. (1988). Evaluation of the oral toxicity of acetaldehyde and formaldehyde in a 4-week drinking-water study in rats. Food Chem Toxicol 26(5):447–452.

Til HP, Woutersen RA, Feron VJ, Hollanders VH, Falke HE, Clary JJ. (1989). Two-year drinking-water study of formaldehyde in rats. Food Chem Toxicol 27(2):77–87.

Tjan-Heijnen VC, Postmus PE, Ardizzoni A, Manegold CH, Burghouts J, van Meerbeeck J, Gans S, Mollers M, Buchholz E, Biesma B, Legrand C, Debruyne C, Giacoone G. (2001). Reduction of chemotherapy-induced febrile leucopenia by prophylactic use of ciprofloxacin and roxithromycin in small-cell lung cancer patients: An EORTC double-blind placebo-controlled phase III study. Ann Oncol 12(10):1359–1368.

Tobe M, Naito K, Kurokawa Y. (1989). Chronic toxicity study on formaldehyde administered orally to rats. Toxicology 56(1):79–86.

US EPA. (2010). Toxicological Review of Formaldehyde — Inhalation Assessment (CAS No. 50-00-0) in Support of Summary Information on the Integrated Risk Information System (IRIS). Volumes I-IV (Draft). Washington, DC: US EPA. EPA/635/R-10/002A.

Vardiman JW. (2010). The World Health Organization (WHO) classification of tumors of the hematopoietic and lymphoid tissues: An overview with emphasis on the myeloid neoplasms. Chem Biol Interact 184(1-2):16–20.

Vargova M, Wagnerova J, Liskova A, Jakubovsky J, Gajdova M, Stolcova E, Kubova J, Tulinska J, Stenclova R. (1993). Subacute immunotoxicity study of formaldehyde in male rats. Drug Chem Toxicol 16(3):255–275.

Walrath J, Fraumeni JF Jr. (1983). Mortality patterns among embalmers. Int J Cancer 31(4):407–411.

Walrath J, Fraumeni JR Jr. (1984). Cancer and other causes of death among embalmers. Cancer Res 44(10):4638–4641.

Wang E, Simard M, Ouellet N, Bergeron Y, Beauchamp D, Bergeron MG. (2002). Pathogenesis of pneumococcal pneumonia in cyclophosphamide-induced leukopenia in mice. Infect Immun 70(8):4226–4238.

Wang M, Cheng G, Balbo S, Carmella SG, Villalta PW, Hecht SS. (2009a). Clear differences in levels of a formaldehyde-DNA adduct in leukocytes of smokers and nonsmokers. Cancer Res 69(18):7170–7174.

Wang R, Zhang Y, Lan Q, Holford TR, Leaderer B, Zahm SH, Boyle P, Dosemeci M, Rothman N, Zhu Y, Qin Q, Zheng T. (2009b). Occupational exposure to solvents and risk of non-Hodgkin lymphoma in Connecticut women. Am J Epidemiol 169:176–185.

Weed DL. (2005). Weight of evidence: A review of concepts and methods. Risk Anal 25:1545–1557.

Wong O. (1983). An epidemiologic mortality study of a cohort of chemical workers potentially exposed to formaldehyde, with a discussion on SMR and PMR. In: Gibson, JE, ed. Formaldehyde Toxicity. Washington, DC: Hemisphere Publishing Corp., 256–272.

Woutersen R. (2007). Indications for leukemia and lymphoma in former animal studies. Presented atIn: Formaldehyde International Science Conference; held September 20–21, 2007 in Barcelona, Spain.

Woutersen RA, Appelman LM, Wilmer JW, Falke HE, Feron VJ. (1987). Subchronic (13-week) inhalation toxicity study of formaldehyde in rats. J Appl Toxicol 7(1):43–49.

Wright DE, Wagers AJ, Gulati AP, Johnson FL, Weissman IL. (2001). Physiological migration of hematopoietic stem and progenitor cells. Science 294(5548):1933–1936.

Ye X, Yan W, Xie H, Zhao M, Ying C. (2005). Cytogenetic analysis of nasal mucosa cells and lymphocytes from high-level long-term formaldehyde exposed workers and low-level short-term exposed waiters. Mutat Res 588(1):22–27.

Yin SN, Hayes RB, Linet MS, Dosemeci M, Travis LB, Li CY, Zhang ZN, Li DG, Chow WH, Wacholder S, Wang YZ, Jiang ZL, Dai TR, Zhang WY, Chao XJ, Ye PZ, Kou QR, Zhang XC, Lin XF, Meng JF, Ding CY, Zho JS, Blot WJ. (1996). A cohort study of cancer among benzene-exposed workers in China: Overall results. Am J Ind Med 29(3):227–235.

Yoshino T, Asada H, Sano S, Nakamura T, Itami S, Tamura M, Yoshikawa K. (2000). Impaired responses of peripheral blood mononuclear cells to staphylococcal superantigen in patients with severe atopic dermatitis: A role of T cell apoptosis. J Invest Dermatol 114(2):281–288.

Zhang L, Steinmaus C, Eastmond DA, Xin XK, Smith MT. (2009). Formaldehyde exposure and leukemia: A new meta-analysis and potential mechanisms. Mutat Res 681(2-3):150–168.

Zhang L, Freeman LE, Nakamura J, Hecht SS, Vandenberg JJ, Smith MT, Sonawane BR. (2010a). Formaldehyde and leukemia: Epidemiology, potential mechanisms, and implications for risk assessment. Environ Mol Mutagen 51:181–191.

Zhang L, Tang X, Rothman N, Vermeulen R, Ji Z, Shen M, Qui C, Guo W, Liu S, Reiss B, Freeman LB, Ge Y, Hubbard AAE, Hua M, Blair A, Galvan N, Ruan X, Alter BP, Xin KX, Li S, Moore LE, Kim S, Xie Y, Hayes RB, Azuma M, Hauptmann M, Xiong J, Stewart P, Li L, Rappaport SM, Huang H, Fraumeni JF ?Smith MT, Lan Q. (2010b). Occupational exposure to formaldehyde, hematotoxicity, and leukemia-specific chromosome changes in cultured myeloid progenitor cells. Cancer Epidemiol Biomarkers Prev 19(1):80–88.

© 2011 Informa Healthcare USA, Inc.

Prueitt *et al*., 2012

Critical Reviews in Toxicology, 2011; 41(10): 822–903
© 2011 Informa Healthcare USA, Inc.
ISSN 1040-8444 print/ISSN 1547-6898 online
DOI: 10.3109/10408444.2011.616877

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

REVIEW ARTICLE

Hypothesis-based weight-of-evidence evaluation of the neurodevelopmental effects of chlorpyrifos

Robyn L. Prueitt, Julie E. Goodman, Lisa A. Bailey, and Lorenz R. Rhomberg

Gradient, Cambridge, Massachusetts, USA

Abstract

We used a hypothesis-based weight-of-evidence (HBWoE) approach to analyze the evidence regarding the hypothesis that chlorpyrifos can cause neurodevelopmental effects below the threshold for inhibition of acetylcholinesterase activity in the nervous system, which is an established mode of action for chlorpyrifos neurotoxicity. The epidemiology data do not consistently demonstrate associations between chlorpyrifos exposure and neurodevelopmental toxicity, and the animal toxicity data do not provide clear evidence that neurodevelopmental effects occur at doses below the threshold for acetylcholinesterase inhibition. The alternative mechanisms proposed to underlie potential neurodevelopmental effects in humans have been observed in the absence of acetylcholinesterase inhibition in a few *in vitro* studies but not in the developing brain *in vivo*. We provide perspective on the HBWoE approach compared with frameworks developed by the United States Environmental Protection Agency and the European Center for Ecotoxicology and Toxicology of Chemicals. We suggest that our HBWoE approach offers advantages over these frameworks in providing a better perspective on how to integrate all of the relevant data and how to use each line of evidence to inform the integration of other kinds of data or compare alternative hypotheses. Based on an HBWoE analysis, we conclude that a causal association between chlorpyrifos exposure and neurodevelopmental effects in the absence of acetylcholinesterase inhibition in the brain is not plausible in humans, and the few positive associations observed in epidemiology studies are most likely attributable to alternative explanations.

Keywords: Risk assessment, epidemiology, neurotoxicity, developmental toxicity, mechanism of action, acetylcholinesterase inhibition, pesticides, neurobehavior, cognitive and motor development, child behavior, ECETOC framework, US EPA framework

Copyright © 2011 Informa Healthcare
Not for Sale or Commercial Distribution
Unauthorized use prohibited. Authorized users can download, display, view and print a single copy for personal use

Contents

Abstract 822
1. Introduction 824
2. Hypothesis-based weight of evidence 824
3. Chlorpyrifos case study 827
  3.1. Chlorpyrifos background 827
  3.2. Epidemiology studies of neurodevelopmental effects 829
    3.2.1. Overview of epidemiology studies 829
      3.2.1.1. Cohort studies of chlorpyrifos 829
      3.2.1.2. Chlorpyrifos exposure in the cohort studies 833
      3.2.1.3. Other human studies of chlorpyrifos 834

Address for Correspondence: Julie E. Goodman, Gradient, 20 University Rd, Cambridge, MA 02138, USA. Tel: 617-395-5000. E-mail: jgoodman@gradientcorp.com

(Received 13 April 2011; revised 18 July 2011; accepted 18 August 2011)

RIGHTSLINK

3.2.2. Endpoint-by-endpoint analysis of neurodevelopmental effects ... 834
3.2.2.1. Newborn head circumference ... 834
3.2.2.2. Infant neurobehavior ... 835
3.2.2.3. Cognitive and motor development ... 835
3.2.2.4. Child behavioral outcomes ... 839
3.2.3. Analysis of human data ... 840
3.2.3.1. Exposure assessment ... 840
3.2.3.2. Outcome assessment ... 846
3.2.3.3. Clinical significance ... 847
3.2.3.4. Confounding and bias ... 847
3.2.3.5. Exposure-response ... 849
3.2.3.6. Statistical analyses ... 849
3.2.4. Conclusions for human data ... 850
3.3. Neurodevelopmental toxicity studies in animals ... 851
3.3.1. Overview of animal studies ... 851
3.3.2. Endpoint-by-endpoint analysis of neurodevelopmental effects in animals ... 851
3.3.2.1. Social and maternal behavior ... 855
3.3.2.2. Emotion and anxiety ... 859
3.3.2.3. Motor function ... 860
3.3.2.4. Cognitive function ... 870
3.3.3. Analysis of animal data ... 876
3.3.3.1. Adequacy of study design ... 876
3.3.3.2. Consistency of outcomes within and across studies ... 877
3.3.3.3. Exposure-response ... 879
3.3.3.4. Biological significance of responses ... 880
3.3.4. Conclusions for animal data ... 880
3.4. Evaluation of mechanistic data ... 880
3.4.1. Introduction ... 880
3.4.2. Analysis of mechanistic data ... 881
3.4.2.1. Neuronal differentiation ... 881
3.4.2.2. Oxidative stress ... 881
3.4.2.3. cAMP-related cell signaling ... 882
3.4.2.4. Serotonergic dysfunction ... 883
3.4.3. Conclusions for mechanistic data ... 884
3.5. HBWoE evaluation of the potential neurodevelopmental toxicity of chlorpyrifos ... 884
3.5.1. Hypotheses under consideration ... 885
3.5.2. Evaluation of hypotheses for each line of evidence ... 886
3.5.3. Evaluation of hypotheses for all lines of evidence together ... 888
3.5.4. Evaluation of alternative accounts ... 890
4. US EPA, ECETOC, and HBWoE frameworks ... 891
4.1. US EPA framework for incorporating human epidemiology and incident data in health risk assessment ... 891
4.2. ECETOC framework for integration of human and animal data in chemical risk assessment ... 892
4.2.1. Human data quality in the ECETOC framework ... 893
4.2.2. Animal data quality in the ECETOC framework ... 893
4.2.3. Relevance of animal data in human risk assessment ... 894
4.2.4. Integration of animal and human data ... 894
4.2.5. Practical application of WoE in the ECETOC framework ... 895
4.3. Comparison of HBWoE, US EPA, and ECETOC frameworks ... 896
4.4. Comparison of frameworks in the context of chlorpyrifos ... 897
5. Conclusions ... 898
Declaration of interest ... 899
References ... 899

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

1. Introduction

Regulatory agencies are moving toward making greater use of human data in risk assessments, especially assessments of pesticides. The United States Environmental Protection Agency (US EPA) and the European Center for Ecotoxicology and Toxicology of Chemicals (ECETOC) have proposed frameworks for incorporating human data into chemical risk assessments (US EPA, 2010; ECETOC, 2009). We have recently developed a hypothesis-based weight-of-evidence (HBWoE) approach that not only parallels several key aspects of the US EPA and ECETOC frameworks but also provides guidance regarding how to integrate all of the data (positive, null, or of varying quality) that are relevant to determining human disease causation (Rhomberg *et al.*, 2010, 2011).

Human studies will likely be a large focus of US EPA's review of the pesticide chlorpyrifos, as there are now several epidemiology studies examining chlorpyrifos exposure and neurodevelopmental effects. The results of studies indicating neurodevelopmental effects in humans at low exposure levels are not consistent with the well-established animal and mode-of-action (MoA) data, which indicate neurological effects only at high chlorpyrifos exposures, and it will be a challenge to determine how to assess all of these data together.

In the present paper, we have two aims: first, to apply the HBWoE framework to a case study of chlorpyrifos, and second, to provide perspective on our approach compared with those put forth by others, by describing and evaluating the US EPA and ECETOC frameworks and contrasting their rationales with that of our own approach. The aim of our HBWoE analysis is to provide a critical review and synthesis of all the evidence regarding the hypothesized ability of chlorpyrifos to cause neurodevelopmental effects in humans in a transparent manner to determine whether they support a causal association at low chlorpyrifos exposures. This analysis not only provides insights about chlorpyrifos toxicity, it also addresses the larger issue of combining human, animal, and mechanistic data in risk assessments for pesticides and identifies ways to improve the US EPA and ECETOC frameworks.

2. Hypothesis-based weight of evidence

Incorporating human data into risk assessment is a critical aspect of evaluating the causes of human disease. It is important, however, to evaluate the question of human disease causation in the context of all relevant data, including epidemiology, animal toxicology, MoA (*e.g.*, *in vitro* and *in silico* approaches), and pharmacokinetics. Organizing data, evaluating data quality, and summarizing results of all the relevant studies are critical steps in evaluating all of the data relevant to the causal question.

An important further step in evaluating the relevant data is weighing all of the evidence in a clear, logical, and non-biased way so that judgments can be made based solely on the data at hand, rather than simply noting selected instances of consistency with (or contradiction of) pre-conceived ideas. Although "weight of evidence" (WoE) is often discussed as a necessary part of evaluating a causal association between a given disease and chemical exposures, there is little explicit guidance on how to weigh all of the evidence in a manner that can be documented and that for which the outcome can be used in risk management decisions. Weed (2005) points out that the term "weight of evidence" is often used loosely; he calls on practitioners to articulate what they mean by the phrase and to specify their approach. Clearly, professional judgment is involved, but it is not enough simply to name the evidence at hand and then announce one's conclusion. Flexibility in how evidence should be weighed is also necessary. Weighing all of the pertinent data, in all its diversity of study designs and complexity of bearing on the questions at hand, can be an overwhelming task when faced with a question of human disease causation. It is not a straightforward task to strike the proper balance between rigidly prescriptive guidelines (which tend to dictate scientific interpretations) and flexible, less structured guidelines that nonetheless provide some useful perspective on how, in practice, one should actually proceed and that provide adequate documentation of the basis for scientific professional judgments. There will likely be many proposed approaches to WoE; we believe, however, that there are several key aspects that should be central to a scientifically based WoE evaluation, and we have based our approach on them. These key aspects are:

1. Systematically review individual studies potentially relevant to causal question at hand (*e.g.*, epidemiology, MoA, pharmacokinetic, toxicology), with focus on evaluation of the quality of all individual studies (both negative and positive, of varying qualities).
2. Within a realm of investigation (*e.g.*, epidemiology, animal toxicology, or MoA studies), systematically examine the data for particular endpoints across studies, evaluating consistency, specificity, and reproducibility of outcomes.
3. Identify and articulate lines of argument (or "hypotheses"), newly proposed or those already put forth (if available), that bear on the available data. Discuss how available studies are used for each hypothesis to infer the existence, nature, or magnitude of human risk.
4. Evaluate the logic of the proposed hypotheses with respect to each line of evidence to determine how well the hypotheses are supported by the available data.
5. Evaluate the logic of the proposed hypotheses with respect to all lines of evidence holistically so that all of the data are considered and integrated and allowed to inform interpretation of one another.

RIGHTSLINK

6. Describe and compare (if more than one hypothesis has been put forth) the various alternate accounts of the observations at hand. That is, describe how well each overarching hypothesis is supported by all of the available data, discussing the uncertainties and inconsistencies in the data set and *ad hoc* assumptions required to support each hypothesis. This step involves presenting the lines of reasoning, based on the science and integration of the lines of evidence, so that the data will speak for themselves in supporting (or not supporting) the overarching hypotheses that have been put forth.
7. Formulate discussion and conclusion regarding the WoE, and proposed next steps.

These steps are intended to provide general guidance on how to weigh all of the evidence in a systematic way, but are also intended to be flexible. That is, every causal question has a different data set that will require a somewhat different specific approach for presentation and systematic review of the data at hand, but should generally follow these seven steps.

Analyses of various technical approaches to WoE have been offered by Krimsky (2005) and Linkov *et al.* (2009). Several additional frameworks have been put forth specifically as guidance for weighing evidence in the context of evaluating potential human disease causation. For example, US EPA's Office of Pesticide Programs (OPP) and the European Center for Ecotoxicology and Toxicology of Chemicals (ECETOC) have proposed frameworks for incorporating human data into chemical risk assessments (US EPA, 2010; ECETOC, 2009). Other human relevance frameworks have been put forth that provide guidance on incorporating MoA data into human risk assessment (Sonich-Mullin *et al.*, 2001; Meek *et al.*, 2003; Seed *et al.*, 2005; Boobis *et al.*, 2006; Boobis *et al.*, 2008), and US EPA (2010) and ECETOC (2009) have incorporated aspects of these guidelines into their frameworks.

We have developed the HBWoE framework, which has been described recently and applied to other chemical causation questions (Rhomberg *et al.*, 2010, 2011). It is hypothesis-based in the sense that it emphasizes articulation of the proposed bases for the relevance of the data to the causal question at hand, specifying the logic and reasoning. The approach weighs all of the data (*e.g.*, epidemiology, animal toxicology, MoA), both positive and negative, in terms of quality and relevance to humans in a way that allows each data set to inform the other, and further synthesizes all of the data to determine overall plausibility for causality in humans, considering uncertainties and inconsistencies in the data sets and *ad hoc* assumptions that may be required for some of the hypotheses put forth.

The HBWoE framework emphasizes articulation of the logic and reasoning that form the bases of various lines of argument (or "overarching hypotheses") that are either newly proposed or have already been put forth for a given question regarding human disease causation. A key aspect of the HBWoE framework is the importance of analysis of these lines of argument, or consideration of alternate "accounts" (or interpretations) of the available data and how each is supported by the available data. Hill (1965) makes explicit the importance of considering alternative "accounts" of the observations at hand in stating:

> None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non. What they can do with greater or less strength, is to help us to make up our minds on **the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?** (Hill, 1965) [emphasis added]

The key outcome of the HBWoE framework is the evaluation and comparison of alternative and contrasting accounts. In the end, each account (that is, each tentative "story" as to why the facts are as they are) can be compared to other accounts. In this way, various competing overarching hypotheses can be weighed by comparing their relative success at explaining phenomena seen in the data, the relative reasonableness of *ad hoc* assumptions needed for each, and the relative naturalness and plausibility of the means whereby potentially refuting observations are reconciled with the account's central hypothesis. Although it is hard to reduce this evaluative process into checklists, scores, or enumerations, the hope is that, by not simply conducting such evaluations of alternative accounts but also by writing them down to be scrutinized and debated, the relative explanatory success of each account, and the relative "epistemological baggage" associated with defending each alternative interpretation, will be evident. This can then serve as the basis for assigning the relative degree of credence that should be given to an account that asserts the existence of a causal role of the exposures of interest in the disease *versus* accounts that ascribe any apparent patterns of association of exposure and disease that appear among the data to other, noncausal factors. In addition, from this assessment, one can more clearly define hypotheses and propose areas of research needed to fill data gaps for each account or to put their hypotheses to the test.

As part of the comparison of accounts, the HBWoE approach considers all data relevant to the causal question at hand, even negative data and (particularly when they are the bases for a particular line of argument) data of questionable quality or from studies with significant design shortcomings. In this last case, it is important to demonstrate the analysis and logic of how poor quality data have been interpreted within an account, how critical they are to the account's assertions, and the *ad hoc*

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

assumptions required to fit these data to the proposed hypothesis. In the HBWoE framework, such questionable data are automatically downweighted by their poor ability to discriminate between accounts, as their face-value interpretation is not markedly more compelling than alternative explanations that ascribe the outcomes to those extraneous factors or alternative possible causes that better-designed studies would have eliminated. That is, the results are relatively easily and credibly explained away as artifacts.

As described by Rhomberg *et al.* (2010, 2011), the goal of the HBWoE approach is broad in that the relative degrees of credence that should be placed in alternative possible interpretations of hypotheses are expressed in a way that shows how such credence is tied to specific scientific interpretations, considering consistencies, inconsistencies, and contradictions within and across the various data sets. The explanations in each account need not be proven—what is important is that one set out the following questions to be considered throughout the evaluation:

- What is being proposed as causal and generalizable phenomena (*i.e.*, what constitutes the basis for applying observations of biological perturbations or realized risks in other contexts to project potential risks to humans as they are exposed)?
- What is being proposed as the basis for deviations that lead to observations that do not fit the hypothesized causal model (*i.e.*, that would otherwise be counterexamples or refutations)?
- What assumptions are made that are *ad hoc* (*i.e.*, to explain particulars, but for which the evidence consists of their plausibility and the observations they are adduced to explain)?
- What further auxiliary assumptions have to be made, and how reasonable are they in view of our wider knowledge and understanding?
- What is relegated to error, happenstance, or other causes not relevant to the question at hand?
- For those events or processes proposed as critical for a given account, what other observable manifestations should they have? Are these other manifestations indeed found?
- If either the operation or necessity of the proposed critical events for a given account were disproven, how else would one explain the array of outcomes?

The HBWoE framework generally consists of the seven key aspects of WoE evaluations outlined above and in Table 1. First, the framework evaluates the intrinsic quality of the individual studies, and evaluates the data for consistency, specificity and reproducibility across various lines of evidence (*e.g.*, epidemiology, animal toxicology, and MoA studies), including both positive and negative studies and studies of varying quality. The next step involves articulation of various lines of argument that have been put forth within the scientific community to explain the observations at hand. The proffered explanations are based on the notion that true causal effects should be repeatable with some specificity and should be generally operating in all relevant test systems—or at least there should be reasoning as to why exceptions to this exist. In weighing the evidence, the framework focuses on critical evaluation of these various lines of argument, specifying the data on which each are based, and the reasoning for why these data are (or are not) informative about the human risk question at hand. These lines of argument are the "hypotheses" of the HBWoE framework, and they are articulated so that one can evaluate, throughout the process of weighing all of the evidence, how well they are in agreement with all of the data, how well they would explain patterns in the data if they were true, and what other consequences should have been observed if they were true and whether in fact these consequences are observed.

The HBWoE framework then traces the logic and reasoning within each line of evidence, in the context of the various hypotheses. The aim is to establish how well the hypotheses being examined comport with and help explain common patterns in the data, what data seem to constitute exceptions or contrary outcomes to the hypothesized causal principles, and what reasons for such exceptions might be proposed. The framework then traces through the logic regarding each line of evidence and how the animal tests, human studies, and MoA data inform interpretation of one another within the context of the various proposed hypotheses. The question is whether explanations or hypothesized causal factors proposed in one realm (*e.g.*, epidemiology) have aspects that should be observable in others (*e.g.*, MoA studies), enabling evaluation of whether signs of those causal processes do or do not appear where expected.

The final, and key, step to the HBWoE framework, as discussed above and in more detail by Rhomberg *et al.* (2010, 2011), is formulating alternate accounts of the observations at hand, and comparing these accounts. Clearly, there may be many accounts, but the major contending accounts will be those that require the fewest *ad hoc* explanations for why certain observations do not fit with the data at hand. As an explicit process to the HBWoE framework, the scientific judgment (or logical rationale) required for each account needs to be illustrated and discussed in narrative text to describe how the data are being weighed, and what *ad hoc* assumptions are required to account for some of the problematic facts within the observations at hand. Different methods can be applied (*e.g.*, organizational tables or figures), depending on the nature of the data, to organize and illustrate the consistencies and inconsistencies of the data as applied to various lines of evidence and various accounts. The point is to illustrate how one is tracing the logic through various competing accounts, and this will vary depending on the

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Table 1. Comparison of frameworks for integrating human data into risk assessment.

Key aspects of a weight-of-evidence evaluation	US EPA framework	ECETOC framework	HBWoE framework
1. Systematically review individual studies potentially relevant to causal question at hand (*e.g.*, epidemiology, mode of action, pharmacokinetic, toxicology), both negative and positive, and of varying quality.	Yes (focus on epidemiology)	Yes (focus on epidemiology and animal toxicity studies)	Yes
2. Within a realm of investigation (*e.g.*, epidemiology, animal toxicology, or mode of action studies), systematically examine the data for particular endpoints across studies, evaluating consistency, specificity, and reproducibility of outcomes.	Yes (focus on epidemiology)	Yes (focus on epidemiology and animal toxicity studies)	Yes
3. Identify and articulate lines of argument (or "overarching hypotheses"), newly proposed or those already put forth (if available), that bear on the available data.	No explicit guidance	No explicit guidance	Yes
4. Evaluate the logic of the proposed hypotheses with respect to each line of evidence or realm of investigation (*e.g.*, separate evaluation of epidemiology, animal toxicology, and mode of action data).	No explicit guidance	No explicit guidance	Yes
5. Evaluate the logic of the proposed hypotheses with respect to all lines of evidence holistically so that all of the data are considered and integrated and allowed to inform interpretation of one another.	No explicit guidance (although importance of integrating is discussed)	No explicit guidance (although importance of integrating is discussed)	Yes
6. Describe and compare (if more than one hypothesis has been put forth) the various alternate accounts of the observations at hand. That is, describe how well each overarching hypothesis is supported by all of the available data, discussing the uncertainties and inconsistencies in the data set and *ad hoc* assumptions required to support each hypothesis.	No explicit guidance	No explicit guidance	Yes
7. Propose next steps (*e.g.*, sharpening of proposed hypothesis already put forth, propose additional testing to clarify data gaps).	Yes	Yes	Yes

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

data set, likely requiring illustration as well as narrative text. Therefore, the HBWoE framework is intended to be flexible so that each analysis can be constructed in a way that optimizes transparency and logic for the particular set of relevant data.

Below, we apply the HBWoE framework to evaluate the WoE regarding a causal association between exposure to chlorpyrifos and adverse effects on neurodevelopment in humans.

3. Chlorpyrifos case study

In this section, we apply the HBWoE framework in a case study of chlorpyrifos, to evaluate the evidence regarding the hypothesized ability of chlorpyrifos to cause neurodevelopmental effects in humans. The results of several epidemiology studies are not consistent with the well-established animal and MoA data, which indicate neurological effects only at high chlorpyrifos exposures. This case study will provide a critical review and synthesis of all of the relevant evidence in a transparent manner to determine whether they support a causal association between low chlorpyrifos exposures and neurodevelopmental effects. We begin with a discussion on the general background of chlorpyrifos, then we present the results of our evaluation of the epidemiology, animal toxicity, and mechanistic data.

3.1. Chlorpyrifos background

Chlorpyrifos is an organophosphorus (OP) insecticide that is the active component in pesticide formulations such as Dursban and Lorsban. Chlorpyrifos was widely used for agricultural and residential pest control until 2001, when restriction of its non-agricultural use began in the United States (US EPA, 2002). Current use of chlorpyrifos is mainly limited to controlling insect damage in agricultural settings worldwide. Human exposure to chlorpyrifos can occur through oral, dermal, and inhalation pathways. Inhalation and dermal exposures have likely been the predominant pathways for occupational exposure, and ingestion from residues in the diet is likely the predominant pathway for non-occupational exposures today (Eaton *et al.*, 2008). Human exposures to chlorpyrifos are estimated based on several biomarkers,

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

including various metabolites of chlorpyrifos that are described below.

Chlorpyrifos is well-absorbed after oral and inhalation exposures (Nolan *et al.*, 1984; Bakke *et al.*, 1976; Smith *et al.*, 1967; Ahdaya *et al.*, 1981), but dermal absorption is relatively low unless skin integrity is compromised (Aprea *et al.*, 1994; Shah *et al.*, 1987). Once absorbed into the body, chlorpyrifos is readily distributed to all organs and undergoes rapid metabolism. Oxidative desulfuration of chlorpyrifos *via* cytochrome P-450 (CYP450) enzymes to chlorpyrifos-oxon, the principal toxic metabolite, occurs predominantly in the liver, but extrahepatic metabolism has been reported, including in the brain (Chambers and Chambers, 1989). Chlorpyrifos-oxon is rapidly hydrolyzed by A-esterases, including paraoxonases such as PON1, to form diethylphosphate (DEP) and 3,5,6-trichloro-2-pyridinol (TCPy) (Sultatos and Murphy, 1983a,b). Because of this rapid metabolism, chlorpyrifos-oxon does not escape the liver once steady-state conditions are reached (Sultatos and Murphy, 1983a,b) and has not been detected in human blood or urine after oral administration (Timchalk *et al.*, 2002). Chlorpyrifos-oxon is detectable in rat blood, at concentrations close to the analytical limits of quantitation, but only after exposure to high doses (Timchalk *et al.*, 2002).

Chlorpyrifos itself undergoes oxidative dearylation *via* CYP450 enzymes to an unstable intermediate that is hydrolyzed to diethylthiophosphate (DETP) and TCPy (Timchalk *et al.*, 2002). TCPy is the major chlorpyrifos metabolite identified in the urine of both humans and animals (Bakke *et al.*, 1976; Nolan *et al.*, 1984), and its glucuronide and sulfate conjugates, as well as DEP and DETP, are also excreted in the urine. Detoxification of chlorpyrifos to DETP and TCPy occurs predominantly in the liver and plasma, and is also rapid and extensive. In humans, Nolan *et al.* (1984) estimated an elimination half-life of 27 hours for chlorpyrifos following oral or dermal exposure, and more than 90% of chlorpyrifos was eliminated within 48 hours in rats after a single-dose oral exposure (Bakke *et al.*, 1976; Smith *et al.*, 1967). The distribution and elimination of chlorpyrifos follow a two-compartment model, however, with the portion of chlorpyrifos that is partitioned into body fat or tightly bound to plasma proteins having much slower elimination. For example, Smith *et al.* (1967) reported an elimination half-life of 10–16 hours for chlorpyrifos in various rat organs except body fat, which had an estimated half-life of 62 hours.

Chlorpyrifos-oxon binds to and irreversibly inhibits cholinesterases, such as acetylcholinesterase (AChE), and inhibition of AChE in the nervous system is the mechanism through which chlorpyrifos toxicity is hypothesized to occur (Richardson, 1995). Chlorpyrifos itself can also inhibit cholinesterases, but has been reported to be two to five orders of magnitude less potent at inhibition of AChE than chlorpyrifos-oxon (Huff *et al.*, 1994; Das and Barone, 1999). The chlorpyrifos metabolites TCPy, DEP, and DETP are not considered to make a significant contribution to AChE inhibition and, thus, are not considered to be toxic. AChE is also associated with erythrocytes (red blood cells), and other esterases occur in several tissues at much higher concentrations than AChE, including plasma butyrylcholinesterase (BuChE) and various carboxylesterases in the liver and other organs (ATSDR, 1997; Eaton *et al.*, 2008). Chlorpyrifos-oxon also binds to and inhibits carboxylesterases and these, as well as erythrocyte and plasma cholinesterases, act as a metabolite "sink" to reduce the amount of chlorpyrifos-oxon that can reach the nervous system (Misulis *et al.*, 1993; Chanda *et al.*, 1997). Overall, chlorpyrifos toxicity through AChE inhibition results from a balance of activation and detoxification of chlorpyrifos-oxon through multiple pathways, including various CYP450 enzymes, PON1, circulating cholinesterases, and carboxylesterases.

AChE is an enzyme that terminates the action of the neurotransmitter acetylcholine at cholinergic synapses in the central and peripheral nervous system and at neuromuscular junctions (Palmer, 1980). Inhibition of AChE leads to an accumulation of acetylcholine at cholinergic synapses and overstimulation of nicotinic and muscarinic receptors throughout the body (Richardson, 1995; ATSDR, 1997; Eaton *et al.*, 2008). Acute cholinergic toxicity occurs when cholinesterase inhibition exceeds 70% (Clegg and van Gemert, 1999), and includes effects such as increased salivation and sweating, changes in blood pressure and heart rate, confusion, headache, nausea, diarrhea, muscle tremors, and, with very high doses, convulsions, respiratory failure, and death (ATSDR, 1997; Eaton *et al.*, 2008). These effects usually appear within a few minutes to 24 hours after exposure and are transient for non-fatal exposures, with clinical signs lasting for weeks following exposure in some cases (Lotti *et al.*, 1986).

Treatment of acute cholinergic toxicity is through the cholinergic muscarinic antagonist, atropine, which blocks the accumulation of acetylcholine on muscarinic receptors to relieve receptor hyperstimulation (Aiuto *et al.*, 1993; Namba *et al.*, 1971). Oximes, such as pralidoxime, can also be used for treatment if given shortly after exposure. Pralidoxime can displace chlorpyrifos-oxon from AChE and restore its activity, but only if the covalent bond between them has not undergone the process of "aging," in which the stability of the bond is enhanced (Eyer, 2003). In the absence of oximes, recovery of enzyme activity depends on synthesis of new enzyme, a process that may take days. In 20–50% of cases, an intermediate syndrome develops during or just after recovery from acute cholinergic toxicity (ATSDR, 1997; Eaton *et al.*, 2008). This syndrome involves weakness of the neck, limb, and respiratory muscles, and the underlying mechanism is not known. A delayed peripheral neuropathy can develop several weeks after cholinergic toxicity and the intermediate syndrome (Richardson, 1995: Moretto and Lotti, 1998; Albers *et al.*, 1999). The

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

clinical manifestations of this include motor weakness, with some involvement of peripheral sensory and autonomic function. These symptoms eventually stabilize, and recovery of strength and sensory function occurs, although residual sensory and autonomic dysfunction may persist for years after exposure cessation (ATSDR, 1997; Eaton *et al.*, 2008).

Several investigators have shown that young animals are more susceptible than adults to the acute toxicity of chlorpyrifos, with neonate animals having over an order of magnitude lower LD_{50} values than adults (Pope and Chakraborti, 1992; Pope *et al.*, 1991; Whitney *et al.*, 1995). This age-dependent susceptibility is likely attributable to different detoxication abilities between young animals and adults. In rodents, carboxylesterase activity is much lower in weanling animals than in adults (Karanth and Pope, 2000), and in both rodents and humans, PON1 activity is very low at birth and increases over time, reaching a plateau around postnatal day 21 in rodents and between 6 and 15 months of age in humans (Mueller *et al.*, 1983; Cole *et al.*, 2003; Li *et al.*, 1997). Developing organisms recover more quickly from cholinesterase inhibition than comparably-exposed adults, however, largely because of rapid synthesis of new cholinesterase molecules (Pope and Chakraborti, 1992; Pope *et al.*, 1991).

Human exposures to chlorpyrifos have been measured using several different metrics. Chlorpyrifos can be measured directly in blood, although usually in trace concentrations because of its rapid metabolism. The chlorpyrifos metabolites TCPy, DEP, and DETP can be measured in urine, but they have limitations as biomarkers of exposure to chlorpyrifos. TCPy in urine originates from exposure to not only chlorpyrifos, but from exposure to the pesticide chlorpyrifos-methyl and the herbicide triclopyr, as well as to pre-formed TCPy in the environment (MacIntosh *et al.*, 1999; Needham, 2005; Morgan *et al.*, 2005). Urinary DEP and DETP can also originate from exposure to other pesticides, such as diazinon and disulfoton, and to pre-formed, environmental DEP and DETP (Needham, 2005; Wessels *et al.*, 2003). Thus, these metabolites are not specific to chlorpyrifos and their use as an exposure metric, especially from environmental exposures, can overestimate chlorpyrifos exposure. Activities of erythrocyte AChE or plasma BuChE have also been used as biomarkers of chlorpyrifos exposure, but these activities are also not specific to chlorpyrifos, as other chemicals, including other OPs and N-methyl carbamate pesticides, inhibit cholinesterases (ATSDR, 1997; Barr and Angerer, 2006).

Controlled human exposure studies (Coulston *et al.*, 1972; Kisicki *et al.*, 1999; Nolan *et al.*, 1984) and occupational epidemiology studies (*e.g.*, Albers *et al.*, 2004a, 2004b, 2004c; 2007) have not reported clinical signs of cholinergic toxicity associated with low chlorpyrifos exposures in adults (*e.g.*, controlled human exposures used 0.01–2 mg/kg exposures). Because chlorpyrifos readily passes through the placenta, however, it has been hypothesized that exposure to chlorpyrifos may be associated with neurodevelopmental effects at doses below the threshold for AChE inhibition, through a non-enzymatic role of AChE in brain development or by other non-cholinergic mechanisms in the developing nervous system. Next, we describe the epidemiology studies that have been conducted to address the potential association of chlorpyrifos exposure with neurodevelopmental effects.

3.2. Epidemiology studies of neurodevelopmental effects

We evaluated the available epidemiology data that are relevant to determining whether there is sufficient evidence to support an association between chlorpyrifos exposure and adverse neurodevelopmental effects. Below, we provide a brief overview of the epidemiology literature followed by an endpoint-by-endpoint analysis of each neurodevelopmental outcome that has been investigated. Then, we critically evaluated the epidemiology data as a whole, considering many factors such as the weight of the exposure metric used, outcome assessed, clinical significance of reported effects, control of confounding factors, exposure-response relationships, and statistical limitations. For our analysis, we first conducted a literature search, using PubMed and Toxline, for all human studies measuring or estimating chlorpyrifos exposure and neurodevelopmental outcomes. Search terms included: "chlorpyrifos," "neurological," "neurobehavioral," "neurotoxicity," "behavior*," "birth outcomes," "cognitive," "intelligence," and epidemiol*." We also relied on the reference lists of several review articles (*e.g.*, Eaton *et al.*, 2008; Needham, 2005).

3.2.1. Overview of epidemiology studies

3.2.1.1. Cohort studies of chlorpyrifos Several cohort studies examining the association between chlorpyrifos exposure and neurodevelopmental outcomes in newborns and young children have been conducted, with multiple studies stemming from each cohort. Participants in these cohort studies were likely exposed to many classes of pesticides and other environmental chemicals, but residential or agricultural uses of chlorpyrifos were large sources of their pesticide exposure (Needham, 2005). Among the cohort studies, several different exposure metrics were used to assess chlorpyrifos exposure. Some studies measured chlorpyrifos directly in maternal prenatal and postnatal blood and in cord blood (Perera *et al.*, 2003; Whyatt *et al.*, 2004; Rauh *et al.*, 2006, 2011; Barr *et al.*, 2010). Other studies measured maternal urinary concentrations of TCPy, or both maternal and child urinary concentrations of total diethyl phosphate metabolites (DEPs), which include DEP and DETP (Eskenazi *et al.*, 2004, 2007; Berkowitz *et al.*, 2004; Young *et al.*, 2005; Engel *et al.*, 2007, 2011; Wolff *et al.*, 2007). Enzymatic activities of cholinesterases in whole blood and BuChE

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

in plasma were used in two studies as a general marker of OP exposure (Eskenazi *et al.*, 2004; Wolff *et al.*, 2007). Finally, one study used measurements of chlorpyrifos in ambient air through personal monitoring (Whyatt *et al.*, 2004).

The cohort studies evaluated multiple neurodevelopmental endpoints, which are described in more detail in a later section. Briefly, newborn head circumference was reported in the infants' medical records following delivery. Infant neurobehavioral capacities were measured with the Brazelton Neonatal Behavioral Assessment Scale (BNBAS). The BNBAS scores infant behavior in seven domains: habituation (ability to respond to stimuli while asleep), orientation (attention to visual and auditory stimuli and quality of alertness), motor performance, range of state (arousal and state lability), regulation of state (in the face of increasing levels of stimulation), autonomic stability (signs of stress related to homeostatic adjustments of the CNS), and primitive reflexes (Lester *et al.*, 1982). Cognitive and motor development were assessed using the Bayley Scales of Infant Development II (BSID-II). The BSID-II is a widely used test for identifying young children at risk for developmental delay, and it yields scores for Mental Development Index (MDI) and Psychomotor Development Index (PDI). Cognitive development was also assessed using the Weschler Intelligence Scales for Children (WISC-IV), which yields scores for four areas of mental functioning that are associated with overall IQ. The four indices are verbal comprehension (verbal concept formation), working memory (ability to memorize new information, concentrate, and manipulate information), perceptual reasoning (non-verbal and fluid reasoning), and processing speed (ability to focus attention and quickly order visual information). Behavioral outcomes such as Attention Deficit Hyperactivity Disorder (ADHD), Pervasive Development Disorder (PDD), and attention problems were measured through reporting by the mothers on the Child Behavior Checklist (CBCL). Several studies also evaluated birth weight and birth length, but we did not include these outcomes in our analysis because they are general measures of fetal growth rather than specific neurodevelopmental endpoints.

Each cohort and the studies evaluating neurodevelopmental endpoints are described below and summarized in Table 2. Each study, grouped by cohort, is presented in the rows of Table 2, with separate columns for each of the various exposure metrics and outcomes examined. This provides an overview of the exposure metrics and outcomes that were analyzed across studies and cohorts and is useful for the evaluation of the data regarding the association between chlorpyrifos exposure and neurodevelopmental effects presented in the studies.

Columbia cohort: A cohort of Dominican and African-American mother-newborn pairs living in inner-city neighborhoods in New York City was studied by researchers at Columbia University. Residential pesticide use is widespread among minority populations in New York City, with chlorpyrifos being the most heavily applied pesticide prior to the restriction of its residential use (Landrigan *et al.*, 1999; Whyatt *et al.*, 2002). In the Columbia cohort, 85% of the mothers reported using some form of pest control measure during pregnancy, including sticky traps, gels, can sprays, and pest bombs, and 35% reported using an exterminator (Whyatt *et al.*, 2002). Many of these measures were performed repeatedly (Whyatt, *et al.*, 2002, 2004), increasing the likelihood of repeated inhalation exposure. Deposition of chlorpyrifos on surface areas within the residences most likely led to dermal exposure, and potentially to exposure *via* ingestion beyond that from residues in the diet. The mothers delivered at New York Presbyterian Medical Center, Harlem Hospital, or their satellite clinics between 1998 and 2002. Mothers were eligible for the cohort if they were non-smokers, aged 18–35, were free of diabetes, hypertension, or known HIV, and resided in the area for a minimum of 1 year.

Perera *et al.* (2003) examined the association between chlorpyrifos exposure and newborn head circumference measured at birth in 113 mother-newborn pairs of this cohort. The authors measured chlorpyrifos levels in cord plasma collected at delivery and in maternal plasma collected within one day postpartum. Maternal and cord plasma chlorpyrifos levels were highly correlated ($r = 0.76$), so the authors only used the cord plasma levels in their analysis. Whyatt *et al.* (2004) also considered the association between cord plasma levels of chlorpyrifos with newborn head circumference in an expanded number of mother-newborn pairs (287), although they did not state whether all subjects from the Perera *et al.* (2003) study were included in their analysis. Whyatt *et al.* (2004) also conducted personal air monitoring of chlorpyrifos for 48 hours during the third trimester for 271 women and evaluated the association between these measurements and head circumference.

In an effort to evaluate cognitive and behavioral outcomes, Rauh *et al.* (2006) extended the follow-up through the first 3 years of life for a subset of 254 infants. The authors examined associations between the chlorpyrifos levels in cord plasma, as measured by Perera *et al.* (2003) and Whyatt *et al.* (2004), and cognitive and motor development, *via* scores for MDI and PDI at 12, 24, and 36 months. They also examined the association between chlorpyrifos exposure and behavioral outcomes, including ADHD, PDD, and attention problems, using the CBCL at 36 months of age. In a recent study, Rauh *et al.* (2011) used the WISC-IV to examine associations between cord plasma chlorpyrifos levels and cognitive development at 7 years of age for 265 children in the cohort.

CHAMACOS cohort: Researchers in the Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS) at the University of California at Berkeley studied neurodevelopmental endpoints in a cohort of mother-newborn pairs living in an agricultural community in the Salinas Valley of California. Exposures to chlorpyrifos were primarily from agricultural spraying of

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 2. Overview of cohort studies of chlorpyrifos exposure and neurodevelopmental outcomes.

		Exposure Metric							Outcome assessed					
Study	Subjects	Chlorpyrifos in cord blood	Chlorpyrifos in maternal blood	Urinary TCPy	Urinary DEPs	Cholinesterases in cord blood	Cholinesterases in maternal blood	Personal air monitoring	Age at outcome assessment	Head circumfer-ence	Infant neurobehav-ior (BNBAS)	Cognitive and motor develop-ment (BSID-II)	Cognitive development (WISC-IV)	Behavioral outcomes (ADHD, PDD, and attention problems) (CBCL)
Columbia Cohort (Infants born between 1998–2002)														
Perera *et al.*, 2003	113 African-American and Dominican mother-newborn pairs living in inner-city (NYC)	Plasma levels, collected at delivery	Plasma levels, collected within 1 day postpartum						Birth	X				
Whyatt *et al.*, 2004	Expanded number of mother-newborn pairs ($n = 287$) from Perera *et al.* (2003)	Plasma levels, collected at delivery	Plasma levels, collected within 2 days postpartum					Collected for 48 hours during third trimester	Birth	X				
Rauh *et al.*, 2006	254 children from Whyatt *et al.* (2004)	Plasma levels, collected at delivery	Plasma levels, collected within 1 day postpartum						12 months			X		
									24 months			X		
									36 months			X		X
Rauh *et al.*, 2011	265 children from Whyatt *et al.* (2004)	Plasma levels, collected at delivery	Plasma levels, collected within 2 days postpartum						7 years				X	
CHAMACOS Cohort (Infants born between 2000–2001)														
Eskenazi *et al.*, 2004	488 mother-newborn pairs living in agricultural community (Salinas Valley, CA)			Maternal levels, collected at 13 and 26 weeks gestation	Maternal levels, collected at 13 and 26 weeks gestation	ChE activity in whole blood; BuChE activity in plasma; collected at delivery	ChE activity in whole blood; BuChE activity in plasma; collected at 26 weeks gestation and delivery		Birth	X				
Young *et al.*, 2005	381 mother-newborn pairs from Eskenazi *et al.* (2004)				Maternal levels, collected at 14 and 26 weeks gestation and within one week of delivery				0 to 62 days		X			

(Continued)

© 2011 Informa Healthcare USA, Inc.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 2. (Continued).

		Exposure Metric							Outcome assessed					
Study	Subjects	Chlorpyrifos in cord blood	Chlorpyrifos in maternal blood	Urinary TCPy	Urinary DEPs	Cholinesterases in cord blood	Cholinesterases in maternal blood	Personal air monitoring	Age at outcome assessment	Head circumfer-ence	Infant neuro-behavior (BNBAS)	Cognitive and motor develop-ment (BSID-II)	Cognitive development (WISC-IV)	Behavioral outcomes (ADHD, PDD, and attention problems) (CBCL)
Eskenazi *et al.*, 2007	396 children from Eskenazi *et al.* (2004)			Maternal levels, collected at 14 and 26 weeks gestation	Maternal levels, collected at 14 and 26 weeks gestation; child levels, collected at 6, 12, and 24 months of age				6 months			X		
									12 months			X		
									24 months			X		X
Mount Sinai Cohort (Infants born between 1998–2002)														
Berkowitz *et al.*, 2004	404 mother-newborn pairs living in NYC			Maternal levels, collected during third trimester					Birth	X				
Wolff *et al.*, 2007	404 mother-newborn pairs from Berkowitz *et al.* (2004)				Maternal levels, col-lected dur-ing third trimester		BuChE activity in plasma; collected during third trimester		Birth	X				
Engel *et al.*, 2007	311 mother-newborn pairs from Berkowitz *et al.* (2004)				Maternal levels, collected during third trimester				0 to 5 days		X			
Engel *et al.*, 2011	360 children from Berkowitz *et al.* (2004)				Maternal levels, collected during third trimester				12 months			X		
									24 months			X		
									7–9 years				X	
UMDNJ Cohort (Infants born between 2003–2004)														
Barr *et al.*, 2010	150 mother-newborn pairs scheduled for elective cesarean delivery at Saint Peter's University Hospital, New Brunswick, New Jersey	Serum levels, collected at delivery	Serum levels, collected just prior to delivery						Birth	X				

DEPs = diethyl phosphates; TCPy = 3,5,6-trichloro-2-pyridinol; ChE = cholinesterase; BuChE = butyrylcholinesterase.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

several pesticides, pesticide drift from the spraying, and pesticide residues brought into homes on workers' clothing, with potential exposures from residential pesticide use and dietary exposure (Needham, 2005; Castorina *et al.*, 2003). A total of 488 pregnant women participated in the study. During pregnancy, 28% of the women worked in the fields, 14% worked at other agricultural jobs, and 85% lived in homes with agricultural workers (Eskenazi *et al.*, 2004). The mothers delivered at Natividad Medical Center during 2000 and 2001 and were excluded from the study if they were less than 18 years of age or had gestational or pre-existing diabetes, hypertension, twin births, or stillbirths.

Eskenazi *et al.* (2004) studied all 488 mother-newborn pairs, including 11 infants diagnosed with congenital anomalies at birth because, according to the authors, their exclusion did not materially affect the results. Concentrations of TCPy and DEPs were measured in maternal urine collected at two time periods during pregnancy (first period, mean 13 weeks gestation, range 4–29 weeks; second period, mean 26 weeks gestation, range 18–39 weeks) and these measurements were averaged for each mother. Maternal blood specimens were collected during the second period and cord blood was collected from newborns for measurements of total cholinesterase activity in whole blood and BuChE activity in plasma. The authors evaluated the association between these exposure estimates and head circumference measured at birth.

Young *et al.* (2005) further examined a subset of 381 mother-newborn pairs and evaluated the association between concentrations of DEPs in maternal urine collected at 14 and 26 weeks gestation and at one time point after delivery (usually within one week) and the BNBAS scores of infant behavior assessed at 2 months of age or younger.

Eskenazi *et al.* (2007) evaluated the association between chlorpyrifos exposure and MDI and PDI scores in a subset of 396 infants at 6, 12, and 24 months of age. The authors also examined the association between chlorpyrifos exposure and behavioral outcomes (ADHD, PDD, and attention problems) at 24 months of age. Infants with a medical condition that could affect assessment (Down syndrome, deafness, and hydrocephalus) were excluded from the study. The authors used measurements of TCPy and DEPs in maternal urine collected at 14 and 26 weeks gestation and DEPs in the children's urine collected at 6, 12, and 24 months of age, as surrogates for chlorpyrifos exposure.

Mount Sinai cohort: Researchers at Mount Sinai Medical School studied a multi-ethnic cohort of mother-infant pairs living in New York City (predominantly in East Harlem) who delivered between 1998 and 2002 at Mount Sinai Hospital. Approximately 71% of the mothers reported potential exposure to indoor application of pesticides (Berkowitz *et al.*, 2004). Mothers were excluded if they had serious chronic diseases such as diabetes, hypertension, or thyroid disease or if they developed a serious pregnancy complication that could affect fetal development. Mothers were also excluded if they consumed more than two alcoholic beverages per day or used illegal drugs, or if their child was born with a congenital malformation or severe prematurity (< 1,500 grams birth weight or < 32 weeks gestation).

In an initial study of this cohort, Berkowitz *et al.* (2004) evaluated the association between TCPy concentration in maternal urine collected during the third trimester and newborn head circumference measured at birth in 404 mother-infant pairs. Wolff *et al.* (2007) also studied head circumference in the 404 infants, using maternal urinary concentrations of DEPs and BuChE activity in plasma to estimate chlorpyrifos exposure.

In order to evaluate potential neurobehavioral outcomes, Engel *et al.* (2007) examined the association between maternal urinary concentrations of DEPs and behavior in a subset of 311 newborns, as assessed through administration of the BNBAS within five days of delivery.

Engel *et al.* (2011) also examined associations between maternal urinary DEP levels and cognitive and motor development *via* MDI and PDI scores for 177 children at 12 months of age and 247 children at 24 months of age. Cognitive development was also assessed using the WISC-IV in 114 children between seven and 9 years of age.

UMDNJ cohort: A cohort of 150 mother-newborn pairs was examined in a study led by researchers at the University of Medicine and Dentistry of New Jersey (UMDNJ; Barr *et al.*, 2010). More than half of the mothers reported using some type of pesticide during pregnancy (Barr *et al.*, 2010). The mothers underwent elective cesarean delivery at term (≥ 37 weeks) at Saint Peter's University Hospital in New Jersey between July, 2003 and May, 2004. Mothers were excluded if their hemoglobin levels were ≥ 8 mg/dL, they were taking medications that could potentially interfere with metabolism of environmental chemicals, or their fetus had congenital anomalies. Maternal blood was collected pre-operation and cord blood was collected within 15 minutes of birth for measurement of serum chlorpyrifos levels, and the authors evaluated the association between these levels and newborn head circumference.

3.2.1.2. Chlorpyrifos exposure in the cohort studies

Air concentrations of chlorpyrifos were measured for the Columbia and CHAMACOS cohorts. Whyatt *et al.* (2004) reported a mean and standard deviation of 15.3 ± 31.8 ng/m^3 chlorpyrifos in personal air samples collected for 271 of the women in the Columbia cohort. Bradman *et al.* (2007) measured chlorpyrifos in indoor air in the homes of 20 children from the CHAMACOS cohort, reporting a median concentration of 11 ng/m^3 (range: 4.0–36). Using these measurements, Eaton *et al.* (2008) estimated the daily intake of chlorpyrifos from inhalation at 0.003–0.07 μg/kg-day for mothers in the Columbia cohort and 0.002–0.007 μg/kg-day for mothers in the CHAMACOS cohort. Eaton *et al.* (2008) noted

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

that, although chlorpyrifos exposures in air were not measured for the Mount Sinai cohort, other reports of exposure in this cohort indicate that the estimated daily intake from inhalation should be similar to that of the Columbia cohort. Based on FDA market basket surveys, Eaton *et al.* (2008) also estimated the daily intake of chlorpyrifos from consuming a typical US diet as 0.005 μg/kg-day for an average adult. If this intake is added to the estimated intake for inhalation exposure in the cohorts, with the assumption that inhalation contributes approximately one-third of the total exposure, the average daily exposure for mothers in the Columbia cohort is estimated at 0.008 μg/kg-day and for the CHAMACOS cohort at 0.007 μg/kg-day. These estimates are only slightly higher than the estimates for current exposures in the general population, which Eaton *et al.* (2008) estimate as 0.004–0.006 μg/kg-day for adults and less than 0.01 μg/kg-day for children, but they do not account for dermal exposure or for exposure *via* ingestion beyond typical dietary exposure.

Chlorpyrifos was measured in blood samples from the Columbia and UMDNJ cohorts. For the Columbia cohort, mean chlorpyrifos concentrations in cord plasma were 7.5 ± 7.5 pg/g for 113 subjects in the initial cohort (Perera *et al.*, 2003) and 4.0 ± 6.1 pg/g for 287 subjects in the expanded cohort (Whyatt *et al.*, 2004). These concentrations are similar to those measured in the general population, based on a mean chlorpyrifos level of 9 pg/g measured in serum from blood bank donors in Cincinnati, Ohio (Barr *et al.*, 2002). For the UMDNJ cohort, Yan *et al.* (2009) reported mean chlorpyrifos concentrations of 0.55 pg/g in cord serum ($n = 148$) and 0.09 pg/g in maternal serum ($n = 138$). Barr *et al.* (2010) reported the same values, but with units of ng/g. This is likely an error, given that the values would be two orders of magnitude greater than those for the Columbia cohort if these units were correct, yet they were measured after the 2001 ban of residential chlorpyrifos use.

Urinary metabolites of chlorpyrifos were measured in each cohort. Median urinary concentrations of DEPs in maternal urine were reported as 22 nmol/L for the CHAMACOS cohort (Eskenazi *et al.*, 2004), 12.97 nmol/L for the UMDNJ cohort (Yan *et al.*, 2009), and ranged from 18.8–24.7 nmol/L for the Mount Sinai cohort (Wolff *et al.*, 2007; Engel *et al.*, 2007, 2011). Eskenazi *et al.* (2007) determined urinary levels of DEPs in 6- to 24-month-old children of the CHAMACOS cohort and reported geometric mean levels ranging from 10.5–15.2 nmol/L. Median concentrations of TCPy in maternal urine ranged from less than the detection limit to 0.61 μg/L for samples collected at various time points during the third trimester in the Columbia cohort (Whyatt *et al.*, 2009). The median concentration of TCPy in maternal urine in the CHAMACOS cohort was reported as 3.3 μg/L (Eskenazi *et al.*, 2004) and in the Mount Sinai cohort as 7.6 μg/L (Berkowitz *et al.*, 2004). Yan *et al.* (2009) determined the mean concentration of TCPy in maternal urine for the UMDNJ cohort as 1.515 μg/L.

For most of the cohorts, urinary levels of TCPy are similar to those among the general US population, as determined by levels of these metabolites in the 1999–2000 National Health and Nutrition Examination Survey (NHANES; CDC, 2009). The median TCPy level in urine in the total NHANES population was 1.70 μg/L, which is slightly greater than the median reported for the Columbia and UMDNJ cohorts. The median TCPy level in the CHAMACOS cohort is equivalent to the 75th percentile reported for Mexican-Americans (3.20 μg/L) in the 1999–2000 NHANES data, and the median level in the Mount Sinai cohort is similar to the 90th percentile reported for the total NHANES population (7.30 μg/L).

3.2.1.3. Other human studies of chlorpyrifos There are many other studies of the effects of chlorpyrifos in humans besides the cohort studies described above. Several epidemiology studies have been conducted to examine the effects of chronic exposure in workers involved in the manufacture or application of chlorpyrifos (Albers *et al.*, 2004a, 2000b, 2000c; 2007). Controlled exposure studies of chlorpyrifos have been conducted in healthy adult volunteers to determine the safety and pharmacokinetics of this compound in humans (Coulston *et al.*, 1972; Kisicki *et al.*, 1999; Nolan *et al.*, 1984). There have also been many case studies of acute chlorpyrifos exposure reported after accidental or intentional poisoning incidents (as reviewed by ATSDR, 1997; Eaton *et al.*, 2008). Because these studies do not evaluate potential neurodevelopmental effects of chlorpyrifos, they will not be discussed further.

3.2.2. Endpoint-by-endpoint analysis of neurodevelopmental effects

In this section, we discuss the individual neurodevelopmental outcomes analyzed in the cohort studies described above. For each outcome, we assessed the consistency of findings across studies, including consideration of the type of exposure metric, and whether any exposure-response relationships were evident.

3.2.2.1. Newborn head circumference The association between chlorpyrifos exposure and newborn head circumference has been investigated in six of the cohort studies (Table 3). Three of these studies used blood levels of chlorpyrifos as the exposure metric. In studies of the Columbia cohort, Perera *et al.* (2003) and Whyatt *et al.* (2004) reported no association between cord plasma levels of chlorpyrifos and head circumference using regression models with chlorpyrifos as a continuous variable ($p = 0.28$ and $p = 0.86$, respectively). Similarly, Barr *et al.* (2010) reported no association between increased chlorpyrifos levels in maternal ($p = 0.229$) or cord serum ($p = 0.989$) and head circumference in their study of the UMDNJ cohort, using exposure indicators of chlorpyrifos

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

concentrations > 75th percentile and ≤ 75th percentile in the regression models.

Three studies examined associations between maternal urinary concentrations of the metabolites TCPy and/or DEPs and head circumference. In a study of the CHAMACOS cohort, Eskenazi *et al.* (2004) reported no association between TCPy levels categorized above ($p = 0.85$) or below the median value ($p = 0.78$), compared to the referent group with non-detectable levels, and head circumference. The authors also reported no association between urinary levels of DEPs, which were analyzed as continuous variables on a $\log_{10}$ scale, and head circumference ($p = 0.07$). Berkowitz *et al.* (2004) reported no association between urinary levels of TCPy, above and below the limit of detection (LOD), and head circumference in the Mount Sinai cohort. The authors also examined maternal PON1 activity and reported a trend of decreased head circumference ($p = 0.014$) in mothers whose TCPy levels were above the LOD and whose PON1 activity was lowest, but the test for interaction among these was not statistically significant ($p > 0.05$). Wolff *et al.* (2007) reported no association between urinary levels of DEPs, as a continuous variable on a $\log_{10}$ scale, and head circumference ($p = 0.67$) in their study of the Mount Sinai cohort.

Two studies examined head circumference using cholinesterase activities, analyzed as continuous variables on a $\log_{10}$ scale, as estimates of exposure to OPs, including chlorpyrifos. For the CHAMACOS cohort, Eskenazi *et al.* (2004) reported no association between cholinesterase activity in maternal blood during pregnancy ($p = 0.45$) or at delivery ($p = 0.27$) and head circumference. Cholinesterase activity in cord blood was also not associated with this outcome ($p = 0.65$). The authors also reported no associations between head circumference and any of several measures of exposure, including BuChE activity in maternal plasma during pregnancy ($p = 0.58$), maternal plasma at delivery ($p = 0.73$), or cord plasma ($p = 0.91$). Wolff *et al.* (2007) reported no association between BuChE activity in maternal plasma and head circumference ($p = 0.728$) in the Mount Sinai cohort.

Only one study used air concentrations of chlorpyrifos as the exposure metric. Whyatt *et al.* (2004) reported no association in the Columbia cohort between maternal air concentrations of chlorpyrifos as assessed with personal monitors during the third trimester (as a continuous variable) and head circumference ($p = 0.59$), which is consistent with their results for cord plasma levels of chlorpyrifos.

Overall, studies of associations between chlorpyrifos exposure and newborn head circumference have reported consistently null results across cohorts and exposure measures. One study reported an association between maternal PON1 activity and head circumference when maternal TCPy levels in urine were considered jointly, but the test of interaction among these factors was not statistically significant.

3.2.2.2. Infant neurobehavior Two studies examined the association between maternal urinary concentrations of DEPs and infant neurobehavior as assessed by BNBAS scores (Table 4). In a study of the CHAMACOS cohort, Young *et al.* (2005) reported no association between the mean of urinary levels of DEPs at 14 and 26 weeks gestation as a continuous variable and each of the seven BNBAS cluster scores assessed within two months of birth, with the exception of the score for primitive reflexes. Increasing urinary DEPs were associated with an increased number of abnormal reflexes ($p < 0.05$). The authors stratified the results by age at BNBAS assessment and reported an association between levels of DEPs and abnormal reflexes for infants assessed after the first three days of life ($p < 0.05$) but not for those assessed within the first three days of life ($p > 0.05$). Among infants older than three days at assessment, the proportion with more than three abnormally-rated reflexes was examined by quintiles of average $\log_{10}$ total DEPs during pregnancy. A marginally statistically significant ($p = 0.05$) trend of increasing proportion of more than three abnormal reflexes with increasing DEPs quintiles was reported. For all infants in the study, the odds of having more than three abnormal reflexes (compared to three or fewer) increased with each 10-fold increase in maternal levels of DEPs during pregnancy (Odds Ratio [OR] = 3.4, 95% Confidence Interval [CI]: 1.2–9.9). The authors noted that there were no associations between DEPs measured in the post-delivery sample of maternal urine and performance in any BNBAS cluster, but the results were not shown.

Engel *et al.* (2007) used the BNBAS to assess neurobehavior in infants of the Mount Sinai cohort within five days of delivery. Urinary levels of DEPs during the third trimester were analyzed as a continuous variable on a $\log_{10}$ scale for all BNBAS clusters except for primitive reflexes, which were analyzed with Poisson regression because of the count nature of the data for this cluster. As with the CHAMACOS cohort results reported by Young *et al.* (2005), there were no associations between levels of DEPs and any of the BNBAS cluster scores, except for scores of abnormal primitive reflexes (Relative Risk [RR] = 1.49, 95% CI: 1.12–1.98, per 10-fold increase in DEPs). The authors also examined the association between DEPs and the number of abnormally-rated reflexes, and reported that maternal concentrations of DEPs above the median were associated with two or more abnormal reflexes (RR = 2.3, 95% CI: 1.1–5.0).

Overall, the two studies examining infant neurobehavior reported associations between increasing maternal urinary levels of DEPs and BNBAS scores for abnormal reflexes, but not with less-than-optimal scores for any other BNBAS cluster. No studies are available that assessed infant neurobehavior using measurements of chlorpyrifos itself as the exposure metric.

3.2.2.3. Cognitive and motor development The association between chlorpyrifos exposure and cognitive and motor development, as measured by MDI and PDI

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 3. Results of cohort studies assessing head circumference.

Cohort	Study	Exposure metric/group	Exposure	N	Estimate	95% CI	p-value	Statistical method
Columbia	Perera *et al.*, 2003	Chlorpyrifos in cord plasma	Mean: 7.5 ± 7.5 pg/g	113	$\beta = -0.005$	NR	0.28	Multiple regression with log-transformed chlorpyrifos as a continuous variable
	Whyatt *et al.*, 2004	Chlorpyrifos in cord plasma	Mean: 4.0 ± 6.1 pg/g	287	$\beta = -0.01$	(−0.13, 0.11)	0.86	Multiple regression with log-transformed chlorpyrifos as a continuous variable
		Chlorpyrifos in maternal air	Mean: 15.3 ± 31.8 ng/m^3	271	$\beta = -0.04$	(−0.18, 0.10)	0.59	
CHAMACOS	Eskenazi *et al.*, 2004	Maternal urinary TCPy	Median: 3.3 μg/L					Linear regression with TCPy analyzed as categorical variables
		Detectable levels < median		220	$\beta = 0.06$	(−0.37, 0.49)	0.78	
		Detectable levels > median		221	$\beta = 0.04$	(−0.39, 0.47)	0.85	
		Maternal urinary DEPs	Median: 22 nmol/L	486	$\beta = 0.28$	(−0.02, 0.59)	0.07	Linear regression with log-transformed DEP levels as a continuous variable
		ChE (μmol/min/mL)						Linear regression with log-transformed cholinesterase levels as a continuous variable
		Maternal blood, pregnancy	Median: 5.7 μmol/min/mL	340	$\beta = 0.06$	(−0.09, 0.21)	0.45	
		Maternal blood, delivery	Median: 5.1 μmol/min/mL	357	$\beta = -0.07$	(−0.19, 0.05)	0.27	
		Cord blood	Median: 3.8 μmol/min/mL	292	$\beta = -0.04$	(−0.23, 0.14)	0.65	
		BuChE (μmol/min/mL)						
		Maternal blood, pregnancy	Median: 1.4 μmol/min/mL	340	$\beta = 0.12$	(−0.31, 0.56)	0.58	
		Maternal blood, delivery	Median: 1.4 μmol/min/mL	357	$\beta = -0.07$	(−0.45, 0.31)	0.73	
		Cord blood	Median: 1.2 μmol/min/mL	292	$\beta = -0.03$	(−0.50, 0.45)	0.91	
Mount Sinai	Berkowitz *et al.*, 2004	Maternal urinary TCPy	Median: 7.6 μg/L					Generalized linear models with TCPy analyzed as categorical variables
		TCPy < LOD		216	NR	NR	NS	
		TCPy > LOD		171				
		TCPy > LOD, Low PON1		47	NR	NR	0.014	
		TCPy > LOD, Medium PON1		57	NR	NR	NS	
		TCPy > LOD, High PON1		55	NR	NR	NS	
	Wolff *et al.*, 2007	Maternal urinary DEPs	Median: 18.8 nmol/L	318	$\beta \pm SE = -0.052 \pm 0.12$		0.67	Generalized linear models with log-transformed DEP levels or BuChE activity as continuous variables
		BuChE enzymatic activity	Range: <1.9–13.8 mg/ml * min	NR	$\beta \pm SE = 0.44 \pm 1.27$		0.728	
UMDNJ	Barr *et al.*, 2010	Chlorpyrifos in maternal serum	Mean: 0.09 ng/g					Multivariable linear regression with exposure indicators > *versus* ≤ 75th percentile
		> 75th percentile versus ≤ 75th percentile		138	NR	NR	0.229	
		Chlorpyrifos in cord serum	Mean: 0.55 ng/g					
		> 75th percentile versus ≤ 75th percentile		148	NR	NR	0.989	

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 4. Results of cohort studies assessing infant neurobehavior using the BNBAS.

Cohort	Study	Exposure metric/ group	Exposure concentration	N	Estimate	95% CI	p-value	Statistical method
CHAMACOS	Young *et al.*, 2005	Maternal urinary DEPs	Median: 21 nmol/L	175	Habituation; $\beta = 0.33$	(–0.06, 0.72)		Regression models with log-transformed DEP levels as a continuous variable for all clusters except abnormal reflexes; Poisson regression for abnormal reflexes
				379	Orientation; $\beta = -0.32$	(–0.66, 0.03)		
				381	Motor Performance; $\beta = 0.10$	(–0.06, 0.27)		
				381	Range of State; $\beta = -0.02$	(–0.27, 0.24)		
				381	Regulation of State; $\beta = -0.15$	(–0.47, 0.17)		
				381	Autonomic Stability; $\beta = 0.06$	(–0.15, 0.27)		
				381	Reflexes; $\beta = 0.22$	(0.04, 0.40)	< 0.05	
		Maternal urinary DEPs, ≤ 3 days at assessment		109	Habituation; $\beta = 0.47$	(–0.05, 0.99)		Regression estimates stratified by age at BNBAS assessment (≤ 3 days)
				197	Orientation; $\beta = -0.11$	(–0.65, 0.43)		
				197	Motor Performance; $\beta = 0.08$	(–0.17, 0.33)		
				197	Range of State; $\beta = -0.21$	(–0.54, 0.12)		
				197	Regulation of State; $\beta = -0.08$	(–0.52, 0.37)		
				197	Autonomic Stability; $\beta = 0.31$	(0.01, 0.61)	< 0.05	
				197	Reflexes; $\beta = 0.08$	(–0.16, 0.32)		
		Maternal urinary DEPs, > 3 days at assessment		66	Habituation; $\beta = 0.20$	(–0.43, 0.83)		Regression estimates stratified by age at BNBAS assessment (> 3 days)
				182	Orientation; $\beta = -0.33$	(–0.73, 0.08)		
				184	Motor Performance; $\beta = 0.17$	(–0.05, 0.38)		
				184	Range of State; $\beta = 0.20$	(–0.21, 0.62)		
				184	Regulation of State; $\beta = -0.24$	(–0.72, 0.24)		
				184	Autonomic Stability; $\beta = -0.16$	(–0.47, 0.14)		
				184	Reflexes; $\beta = 0.37$	(0.09, 0.64)	< 0.05	

(Continued)

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 4. (Continued).

Cohort	Study	Exposure metric/ group	Exposure concentration	N	Estimate	95% CI	p-value	Statistical method
		Maternal urinary DEPs, > 3 days at assessment, Quintile 1	0.51 to 0.90 nmol/L	37	χ^2 trend = 4.0		0.05	Regression estimates for proportion of > 3 abnormal reflexes, stratified by age at BNBAS assessment (> 3 days) with log-transformed DEP levels categorized by quintiles
		Quintile 2	0.90 to 1.10 nmol/L	37				
		Quintile 3	1.11 to 1.27 nmol/L	37				
		Quintile 4	1.28 to 1.58 nmol/L	37				
		Quintile 5	1.58 to 2.35 nmol/L	36				
		Maternal urinary DEPs		N = 381	Total DEPs, OR = 3.4	(1.2, 9.9)		Logistic regression for > 3 abnormal reflexes *versus* ≤ 3 abnormal reflexes; OR per 10-fold increase in log-transformed DEP levels
Mount Sinai	Engel *et al.*, 2007	Maternal urinary DEPs	Median: 24.7 nmol/L	144	Habituation; $\beta = 0.168$	(–0.23, 0.57)		Generalized linear models with log-transformed DEP levels as a continuous variable for all clusters except abnormal reflexes; Poisson regression for abnormal reflexes
				233	Orientation; $\beta = -0.106$	(–0.41, 0.20)		
				249	Motor Performance; $\beta = 0.049$	(–0.08, 0.17)		
				253	Range of State; $\beta = 0.035$	(–0.12, 0.19)		
				253	Regulation of State; $\beta = -0.047$	(–0.30, 0.21)		
				253	Autonomic Stability; $\beta = -0.154$	(–0.38, 0.08)		
				239	Reflexes; RR = 1.49	(1.12, 1.98)		
		Maternal urinary DEPs > median versus ≤ median		239	RR = 2.3	(1.1, 5.0)		Multivariable logistic regression with number of abnormal reflexes dichotomized at two or more

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

scores or by the WISC-IV, was assessed in four studies (Table 5). In a study of the Columbia cohort, Rauh *et al.* (2006) dichotomized the cord plasma chlorpyrifos levels measured by Perera *et al.* (2003) and Whyatt *et al.* (2004) to classify subjects into high exposure (> 6.17 pg/g) or lower exposure (≤ 6.17 pg/g) groups, because previous analyses had indicated associations with birth weight in this cohort at exposure levels greater than 6.17 pg/g (Perera *et al.*, 2003; Whyatt *et al.*, 2004). The authors reported no association between increased chlorpyrifos exposure measured at birth and MDI scores at 12, 24, or 36 months of age using adjusted multivariate regression analyses. An association between increased chlorpyrifos exposure and lower PDI scores at 36 months was reported ($p = 0.003$), but no associations were reported at 12 or 24 months. The risks of mental or psychomotor delays, as determined by MDI and PDI scores, were assessed using adjusted logistic regression. The authors reported increased risks for mental (OR = 2.37, 95% CI: 1.08–5.19) and motor (OR = 4.52, 95% CI: 1.61–12.70) delays for children in the high exposure group at 36 months and no increase in risk for either type of delay in the high exposure group at 12 and 24 months of age compared to those in the low exposure group.

Eskenazi *et al.* (2007) evaluated the association between chlorpyrifos exposure and MDI and PDI scores in children of the CHAMACOS cohort at 6, 12, and 24 months of age. The authors measured TCPy and DEPs in maternal urine collected at 14 and 26 weeks gestation and DEPs in child urine at the time of assessment with the BSID-II. The metabolite concentrations were $\log_{10}$-transformed and the maternal concentrations were averaged to create prenatal exposure values. A large proportion of the maternal urine samples had non-detectable levels of TCPy, so levels of this metabolite were categorized into three groups: < LOD for both measurements, and subdivided below and above the median of the average level for those with at least one detectable level. Adjusted multiple regression models revealed no associations between maternal TCPy levels and MDI or PDI scores at any age. The authors also examined the change in MDI and PDI scores associated with a 10-fold increase in DEPs. An overall pattern of negative associations between prenatal DEPs and MDI scores and positive associations between concurrent child DEPs and MDI scores was observed, but these associations were not statistically significant, with the exception of the positive association with concurrent DEPs at the 12-month assessment ($\beta = 1.89$, 95% CI: 0.21–3.58, $p \leq 0.05$). Neither prenatal nor concurrent urinary levels of DEPs were associated with PDI scores at any age.

Engel *et al.* (2011) used generalized linear models to examine associations between DEPs in maternal urine collected during the third trimester and MDI and PDI scores assessed at 12 and 24 months of age in children of the Mount Sinai cohort. The authors reported no associations between levels of DEPs and both MDI and PDI scores at either age.

Cognitive development was also assessed *via* the WISC-IV during the early school-age years in two studies. Rauh *et al.* (2011) examined the association between cord plasma chlorpyrifos, analyzed as a continuous variable, and log (ln)-transformed WISC-IV scores in the Columbia cohort at age seven using linear regression models adjusted for multiple covariates. The authors determined that a 1 pg/g increase in cord blood chlorpyrifos was associated with an estimated decrement in working memory scores ranging from 0.35 to 0.81 points ($p = 0.003$) and in full-scale IQ ranging from 0.20 and 0.40 points. Although the latter deficit was statistically significant ($p = 0.048$), the 95% confidence interval ranged from −0.006 to 0.001. Chlorpyrifos exposure was not associated with changes in scores for the WISC-IV indices of verbal comprehension, perceptual reasoning, or processing speed.

Engel *et al.* (2011) examined the association between maternal urinary levels of DEPs during the third trimester and performance on the WISC-IV in 7- to 9-year-old children from the Mount Sinai cohort. The authors used generalized linear models and reported no associations between DEPs and any of the WISC-IV indices.

Overall, the four studies examining chlorpyrifos exposure and cognitive and motor development are not consistent with respect to the exposure metric and timing of outcome assessment. Rauh *et al.* (2006) reported an association between increasing chlorpyrifos exposure and lower PDI scores at 36 months, as well as increased risks of mental and motor delays in highly-exposed children at this age, but Eskenazi *et al.* (2007) and Engel *et al.* (2011) assessed children only to 24 months of age. No associations between increasing chlorpyrifos exposure and lower MDI scores were reported at any age in each of the studies examining this endpoint, and both Eskenazi *et al.* (2007) and Engel *et al.* (2011) reported no associations with PDI scores. Rauh *et al.* (2011) reported a decrement in working memory scores on the WISC-IV associated with chlorpyrifos exposure in 7-year-old children, whereas Engel *et al.* (2011) reported no association with changes in scores on the WISC-IV in children between seven and 9 years of age.

3.2.2.4. Child behavioral outcomes Associations between chlorpyrifos exposure and behavioral outcomes in children were assessed in two of the studies that examined cognitive and motor development (Table 6). In an analysis of the Columbia cohort, Rauh *et al.* (2006) again used cord plasma chlorpyrifos levels dichotomized into high and low exposure categories and logistic regression to determine the risks of behavior problems (as assessed by the mothers' reporting on the CBCL) in children at 36 months of age. The authors reported associations between high chlorpyrifos exposure and attention problems (OR = 11.26, 95% CI: 1.79–70.99), ADHD (OR = 6.50, 95% CI: 1.09–38.69), and PDD (OR = 5.39, 95% CI: 1.21–24.11). These risk estimates were based on 3.9–4.9% of the subjects, or up to 11 children, who scored in the

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

clinical (abnormal) range for these problems. Because of the small study size and because the fraction affected was so small, these risk estimates have large confidence intervals and are highly unstable.

For the CHAMACOS cohort, Eskenazi *et al.* (2007) examined associations between levels of TCPy and DEPs in maternal urine and CBCL outcomes in the clinical or borderline clinical range in children at 24 months of age. Urinary DEPs were also measured for the children at the time of assessment with the CBCL. The authors noted that there were no associations between maternal TCPy levels and attention problems in the borderline clinical range, ADHD in the borderline clinical range, or PDD in the clinical range (results were not shown). Maternal and concurrent child levels of DEPs were also not associated with attention problems or ADHD, but each 10-fold increase in concurrent child DEPs was associated with an increased risk of PDD scores in the clinical range (OR = 1.72, 95% CI: 1.12–2.64).

Overall, the two studies examining behavioral outcomes in children with chlorpyrifos exposure reported associations with PDD scores based on maternal reporting, although few subjects scored in the clinical range for this outcome in the study by Rauh *et al.* (2006). A low percentage of scores in the clinical range was also noted for the associations with attention problems and ADHD reported by Rauh *et al.* (2006).

3.2.3. Analysis of human data

To evaluate the relationship between chlorpyrifos exposure and neurodevelopmental effects, we first considered the overall consistency of the reported results across studies. Null results were reported for the outcome of head circumference in all of the cohorts, regardless of the exposure metric used. The two studies assessing infant neurobehavior in the CHAMACOS and Mount Sinai cohorts reported associations between increasing levels of DEPs and abnormal reflexes. The three studies of cognitive and motor development assessed *via* the BSID-II in the Columbia, CHAMACOS, and Mount Sinai cohorts were not consistent with respect to the exposure metric and timing of assessment, with associations being reported only in children at 36 months of age in the Columbia cohort, whereas children were only examined up to 24 months of age, with null results, in the CHAMACOS and Mount Sinai cohorts. Two of these same studies reported inconsistent results for child behavior outcomes, with the Columbia cohort study reporting associations with all three behavioral outcomes examined at 36 months, whereas the CHAMACOS cohort study only reported an association with PDD at 24 months of age with DEPs, but not TCPy, as the exposure metric. The two studies of cognitive development assessed *via* the WISC-IV were also inconsistent, with associations reported for decrements in working memory scores in the Columbia cohort but no associations with WISC-IV scores in the Mount Sinai cohort. Although it is presumed that studies of the same cohort represent different analyses of the same subjects, loss to follow-up and other factors led to different sample sizes in these studies. Because of this, results across studies of the same cohort are difficult to compare on an individual-by-individual basis. One cannot assess correlations between outcomes across these studies but can only look for trends within cohorts.

Because of the inconsistencies within and among studies, we critically examined the epidemiology data as a whole to assess whether the weight of the epidemiology evidence supports the hypothesis that chlorpyrifos is associated with neurobehavioral effects. This evaluation considered the reliability of the different types of exposure metrics used and whether results were dependent on the robustness of the exposure measurements. It also considered the validity of each neurodevelopmental outcome assessment and whether the results of these evaluations have clinical significance. We also determined whether potential confounding factors in these cohorts were addressed and whether any observed risks may have been attributable to them. Finally, we assessed whether there were consistent exposure-response associations within and across studies and potential statistical limitations among studies. Together, these analyses allowed for an assessment of which study results are likely to be the most valid, based on the weight of each of the factors examined, and whether they provide sufficient evidence to support the hypothesis that chlorpyrifos causes neurodevelopmental effects.

3.2.3.1. Exposure assessment The epidemiology studies examining neurodevelopmental effects potentially related to chlorpyrifos exposure used the following exposure metrics, most of which were measured at only one point in time: chlorpyrifos levels in maternal and/or cord blood; maternal levels of TCPy in urine; maternal or child levels of DEPs in urine; cholinesterase activity in whole blood or plasma; and chlorpyrifos concentrations in ambient air.

Five of the cohort studies used chlorpyrifos levels in cord plasma or serum as an exposure metric (Perera *et al.*, 2003; Whyatt *et al.*, 2004; Rauh *et al.*, 2006, 2011; Barr *et al.*, 2010). Cord blood levels reflect the amount of chlorpyrifos absorbed by the mother and also the amount of absorbed dose transferred to the developing fetus (Needham, 2005), although, in being measured at birth, they most strongly reflect recent exposure rather than exposure earlier in development, which may differ if exposure levels are not consistent throughout pregnancy. As discussed in Section 3.1, chlorpyrifos is oxidized to chlorpyrifos-oxon, through which toxicity from cholinesterase inhibition is hypothesized to occur in the brain. Chlorpyrifos-oxon is rapidly hydrolyzed to TCPy or DEP, and has not been detected in human blood or urine after oral administration (Timchalk *et al.*, 2002). For exposure assessment of chlorpyrifos, direct measurement of the parent compound more accurately reflects the chlorpyrifos dose in the brain than do measurements of its metabolites in urine, which are not on the hypothesized

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 5. Results of cohort studies assessing cognitive and motor development using the BSID-II or WISC-IV.

Outcome assessed	Cohort	Study	Exposure metric/group	Exposure concentration	Age at assessment	N	Estimate	95% CI	p-value	Statistical method
MDI scores	Columbia	Rauh *et al.*, 2006	Chlorpyrifos in cord blood	Range: ND to 63 pg/g						Multivariate linear regression with log-transformed, dichotomized exposure variable
			High exposure (>6.17 pg/g) compared to low exposure (≤ 6.17 pg/g)		12 months	229	$\beta = -0.344 \pm 1.66$		0.836	
					24 months	225	$\beta = -1.480 \pm 2.03$		0.466	
					36 months	228	$\beta = -3.327 \pm 1.76$		0.6	
			High exposure (>6.17 pg/g) compared to low exposure (≤ 6.17 pg/g)		12 months	228	OR = 1.22	(0.48, 3.06)		Logistic regression with log-transformed, dichotomized exposure variable; odds of mental delay
					24 months	228	OR = 1.75	(0.86, 3.60)		
					36 months	228	OR = 2.37	(1.08, 5.19)		
	CHAMACOS	Eskenazi *et al.*, 2007	Maternal urinary TCPy	Median: 3.54 μg/L						Multiple linear regression with log-transformed metabolite levels as continuous variables
			< Median detected		6 months	395	$\beta = 0.24$	(-2.12, 2.60)		
			≥ Median detected				$\beta = 0.08$	(-2.29, 2.44)		
			< Median detected		12 months	393	$\beta = -0.45$	(-3.67, 2.76)		
			≥ Median detected				$\beta = -0.65$	(-3.88, 2.58)		
			< Median detected		24 months	369	$\beta = -1.02$	(-5.34, 3.31)		
			≥ Median detected				$\beta = -1.94$	(-6.26, 2.37)		
			Maternal urinary DEPs	Geometric mean: 18.1 nmol/L						
					6 months	395	$\beta = -0.16$	(-1.96, 1.65)		
					12 months	393	$\beta = -1.14$	(-3.51, 1.22)		
					24 months	369	$\beta = -0.85$	(-3.98, 2.27)		
			Concurrent child urinary DEPs	Geometric mean: 10.6 nmol/L	6 months	395	$\beta = 0.24$	(-0.78, 1.25)		
				Geometric mean: 15.2 nmol/L	12 months	393	$\beta = 1.89$	(0.21, 3.58)	≤ 0.05	
				Geometric mean: 10.5 nmol/L	24 months	369	$\beta = 1.02$	(-0.52, 2.57)		
	Mount Sinai	Engel *et al.*, 2011	Maternal urinary DEPs	Median: 20.2 nmol/L	12 months	149	$\beta = 0.03$	(-2.23, 2.29)		Generalized linear models with log-transformed DEP levels as a continuous variable
					24 months	208	$\beta = -1.47$	(-3.99, 1.04)		

(Continued)

© 2011 Informa Healthcare USA, Inc.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 5. (Continued).

Outcome assessed	Cohort	Study	Exposure metric/group	Exposure concentration	Age at assessment	N	Estimate	95% CI	p-value	Statistical method
PDI scores	Columbia	Rauh *et al.*, 2006	Chlorpyrifos in cord blood	Range: ND to 63 pg/g						Multivariate linear regression with log-transformed, dichotomized exposure variable
			High exposure (>6.17 pg/g) compared to low exposure (≤ 6.17 pg/g)		12 months	228	$\beta = -3.30 \pm 2.11$		0.12	
					24 months	227	$\beta = 1.17 \pm 1.98$		0.56	
					36 months	228	$\beta = -6.46 \pm 2.18$		0.003	
			High exposure (>6.17 pg/g) compared to low exposure (≤ 6.17 pg/g)		12 months	228	OR = 1.88	(0.78, 4.53)		Logistic regression with log-transformed, dichotomized exposure variable; odds of psychomotor delay
					24 months	228	OR = 1.01	(0.37, 2.76)		
					36 months	228	OR = 4.52	(1.61, 12.70)		
	CHAMACOS	Eskenazi *et al.*, 2007	Maternal urinary TCPy	Median: 3.54 μg/L						Multiple linear regression with log-transformed metabolite levels as continuous variables
			< Median detected		6 months	395	$\beta = -0.56$	(−4.03, 2.91)		
			≥ Median detected				$\beta = -0.21$	(−3.69, 3.27)		
			< Median detected		12 months	393	$\beta = -0.70$	(−5.26, 3.86)		
			≥ Median detected				$\beta = -1.62$	(−6.20, 2.96)		
			< Median detected		24 months	369	$\beta = -2.65$	(−6.50, 1.21)		
			≥ Median detected				$\beta = -2.72$	(−6.57, 1.12)		
			Maternal urinary DEPs	Geometric mean: 18.1 nmol/L						
					6 months	395	$\beta = 0.02$	(−2.63, 2.67)		
					12 months	393	$\beta = 0.30$	(−3.03, 3.63)		
					24 months	369	$\beta = -0.86$	(−3.64, 1.92)		
			Concurrent child urinary DEPs	Geometric mean: 10.6 nmol/L	6 months	395	$\beta = 0.60$	(−0.89, 2.09)		
				Geometric mean: 15.2 nmol/L	12 months	393	$\beta = 1.91$	(−0.46, 4.27)		
				Geometric mean: 10.5 nmol/L	24 months	369	$\beta = 0.30$	(−1.07, 1.67)		

(Continued)

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 5. (Continued).

Outcome assessed	Cohort	Study	Exposure metric/group	Exposure concentration	Age at assessment	N	Estimate	95% CI	p-value	Statistical method
	Mount Sinai	Engel *et al.*, 2011	Maternal urinary DEPs	Median: 20.2 nmol/L	12 months	149	$\beta = -0.20$	(−3.28, 2.87)		Generalized linear models with log-transformed DEP levels as a continuous variable
					24 months	210	$\beta = 0.67$	(−1.72, 3.06)		
WISC-IV scores	Columbia	Rauh *et al.*, 2011	Chlorpyrifos in cord blood	Range: ND to 63 pg/g	7 years	265	Full-Scale IQ; $\beta = -0.003$	(−0.006, 0.001)	0.048	Multivariate linear regression with exposure as a continuous variable and ln-transformed WISC-IV scores
							Working Memory; $\beta = -0.006$	(−0.010, −0.002)	0.003	
							Verbal Comprehension; $\beta = -0.002$	(−0.005, 0.001)	0.208	
							Perceptual Reasoning; $\beta = -0.002$	(−0.006, 0.002)	0.29	
							Processing Speed; $\beta = 0.001$	(−0.004, 0.005)	0.728	
	Mount Sinai	Engel *et al.*, 2011	Maternal urinary DEPs	Median: 20.2 nmol/L	7–9 years	114	Full-Scale IQ; $\beta = -3.15$	(−7.19, 0.89)		Generalized linear models with log-transformed DEP levels as a continuous variable
							Working Memory; $\beta = -3.48$	(−7.29, 0.34)		
							Verbal Comprehension; $\beta = -0.08$	(−3.91, 3.76)		
							Perceptual Reasoning; $\beta = -4.37$	(−9.10, 0.36)		
							Processing Speed; $\beta = -2.11$	(−6.81, 2.59)		

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 6. Results of cohort studies assessing child behavioral outcomes using the CBCL.

Outcome assessed	Cohort	Study	Exposure metric/group	Exposure concentration	Age at assessment	N	Estimate	95% CI	p-value	Statistical method
ADHD	Columbia	Rauh *et al.*, 2006	Chlorpyrifos in cord blood	Range: ND to 63 pg/g						Logistic regression with log-transformed, dichotomized exposure variable
			High exposure (> 6.17 pg/g) compared to low exposure (≤ 6.17 pg/g)		36 months	228	OR = 6.50	(1.09, 38.69)		
	CHAMACOS	Eskenazi *et al.*, 2007	Maternal urinary DEPs	Geometric mean: 18.1 nmol/L	24 months	356	OR = 0.59	(0.21, 1.68)		Multiple regression; OR per 10-fold increase in log-transformed DEP levels
			Concurrent child urinary DEPs	Geometric mean: 10.5 nmol/L	24 months	356	OR = 1.18	(0.72, 1.94)		
Attention problems	Columbia	Rauh *et al.*, 2006	Chlorpyrifos in cord blood	Range: ND to 63 pg/g						Logistic regression with log-transformed, dichotomized exposure variable
			High exposure (> 6.17 pg/g) compared to low exposure (≤ 6.17 pg/g)		36 months	228	OR = 11.26	(1.79, 70.99)		
	CHAMACOS	Eskenazi *et al.*, 2007	Maternal urinary DEPs	Geometric mean: 18.1 nmol/L	24 months	356	OR = 0.78	(0.26, 2.31)		Multiple regression; OR per 10-fold increase in log-transformed DEP levels
			Concurrent child urinary DEPs	Geometric mean: 10.5 nmol/L	24 months	356	OR = 1.02	(0.61, 1.71)		
PDD	Columbia	Rauh *et al.*, 2006	Chlorpyrifos in cord blood	Range: ND to 63 pg/g						Logistic regression with log-transformed, dichotomized exposure variable
			High exposure (> 6.17 pg/g) compared to low exposure (≤ 6.17 pg/g)		36 months	228	OR = 5.39	(1.21, 24.11)		
	CHAMACOS	Eskenazi *et al.*, 2007	Maternal urinary DEPs	Geometric mean: 18.1 nmol/L	24 months	355	OR = 0.88	(0.37, 2.07)		Multiple regression; OR per 10-fold increase in log-transformed DEP levels
			Concurrent child urinary DEPs	Geometric mean: 10.5 nmol/L	24 months	355	OR = 1.72	(1.12, 2.64)	≤ 0.05	

causal pathway, or measurement of cholinesterase activity, which is affected by other compounds in addition to chlorpyrifos.

The study by Barr *et al.* (2010) also used maternal levels of chlorpyrifos in serum as an exposure metric. Maternal blood levels of chlorpyrifos may be a surrogate for fetal levels of chlorpyrifos, because chlorpyrifos is distributed through the body in the lipids of the blood, brain, and other tissues (Needham, 2005). They are not as good a metric as cord blood levels, however, because they do not reflect the absorbed dose transferred to the fetus. If the exposure route is ingestion, much of the absorbed chlorpyrifos may undergo first-pass metabolism to TCPy before it reaches the maternal systemic blood supply and, hence, the placenta (Needham, 2005). In addition, blood chlorpyrifos concentrations depend somewhat on the equilibrium between concentrations in adipose tissue and blood (Needham, 2005), and blood lipid levels can increase up to four-fold during pregnancy, resulting in the distribution of chlorpyrifos in blood being affected by these lipid changes (McMullin *et al.*, 2008). If not accounted for, this can result in overestimation of chlorpyrifos concentrations in maternal and fetal serum or plasma. Concentrations of chlorpyrifos in blood can be evaluated on both a concentration basis and a lipid basis to adjust for this, but this was not done in any of the cohort studies.

Whyatt *et al.* (2004) used personal monitoring of prenatal chlorpyrifos concentrations in ambient air for 48 hours during the third trimester in their study of the Columbia cohort. Exposure to chlorpyrifos can come from multiple sources (diet, residential and workplace use) and *via* multiple routes (ingestion, inhalation, dermal absorption), however. Although inhalation is a likely route of exposure in the Columbia cohort, air concentrations do not reflect the amount of chlorpyrifos absorbed from all of the potential sources and routes. Because of this, they are not as good of an exposure measure as biomarkers that are internal dosimeters.

Three studies used measurements of TCPy in urine as an exposure metric (Eskenazi *et al.*, 2004, 2007; Berkowitz *et al.*, 2004). Urinary TCPy is not a reliable biomarker of exposure to chlorpyrifos because TCPy in urine originates from exposure to not only chlorpyrifos, but to chlorpyrifos-methyl, triclopyr, and TCPy itself (MacIntosh *et al.*, 1999; Needham, 2005; Morgan *et al.*, 2005). Thus, measures of TCPy in urine can overestimate exposure to chlorpyrifos in both the mother and fetus. Exposure to pre-formed TCPy can occur through multiple environmental sources, including the diet (Morgan *et al.*, 2005). TCPy is not cholinesterase-inhibiting (Morgan *et al.*, 2005) and has not been shown to be associated with adverse effects (Eaton *et al.*, 2008).

Six of the studies used urinary levels of DEPs as the exposure metric (Eskenazi *et al.*, 2004, 2007; Engel *et al.*, 2007, 2011; Young *et al.*, 2005; Wolff *et al.*, 2007). Urinary DEPs originate from exposure to diazinon, disulfoton, and DEPs themselves in addition to chlorpyrifos, so their measurement is also not a specific biomarker of chlorpyrifos exposure and can overestimate chlorpyrifos levels in the mother and fetus.

Two studies used cholinesterase activities in whole blood or plasma as an exposure metric (Eskenazi *et al.*, 2004; Wolff *et al.*, 2007). Although the interrelationship of the different exposure metrics is not well studied, Eskenazi *et al.* (2004) found no correlation between concurrent measurements of DEPs in urine and either total cholinesterase activity in blood or BuChE activity in plasma in the CHAMACOS cohort. Measurement of cholinesterase activity is not a chlorpyrifos-specific exposure metric because other chemicals, including other OPs and N-methyl carbamate pesticides, inhibit cholinesterases. Another drawback of monitoring cholinesterase activity is that large doses of chlorpyrifos are required for significant inhibition; therefore, such measures are more appropriately used as indicators of toxicity at high exposure levels and are relatively insensitive at the low exposure levels observed in the cohort studies (Wessels *et al.*, 2003).

Most of the exposure metrics used in the cohort studies were measured at only one point in time. Exposures to chlorpyrifos and other pesticides can be transient and highly variable from day to day, so a single measurement may not represent average exposure over time or exposure at some earlier or later time. For example, large intraindividual variability in maternal TCPy and DEP metabolite levels was reported in the Columbia and CHAMACOS cohorts (Whyatt *et al.*, 2009; Eskenazi *et al.*, 2004, 2007; Young *et al.*, 2005), and Eskenazi *et al.* (2007) reported that urinary DEPs in children measured more than a few days apart are uncorrelated, suggesting considerable intraindividual variability in each study.

Another issue contributing to exposure measurement error with the use of pesticide biomarkers is that OPs and their metabolites have a short residence time in the body. Once absorbed, they are rapidly eliminated with biologic half-lives on the order of hours to days in adults (Barr *et al.*, 2002). Thus, any measure of chlorpyrifos or its metabolites in blood or urine at a single time point reflects exposure during the brief period of time prior to measurement and may not accurately reflect exposure throughout the entire critical period of neurodevelopment unless exposure was continuous and yielded a steady-state concentration. For example, chlorpyrifos measurements at or near the time of delivery, such as in cord blood, would reflect only exposure during late pregnancy. The period of vulnerability to chlorpyrifos may begin earlier in pregnancy and extend through the period of synaptic modeling, which continues well into childhood and adolescence (Slotkin, 2004), but the data do not allow conclusions regarding earlier or later exposures. Even a small difference between the measured levels and the actual levels prior to conception, during early pregnancy, or during early childhood could lead to a relatively high degree of exposure measurement error, biasing the results, especially if the day-to-day variation is

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

substantial compared to the amount of variation among subjects in their single measurements. As is the case for biomarkers, the single prenatal air sample measured in the study by Whyatt *et al.* (2004) may not be representative of average exposure during a particular trimester or the entire pregnancy.

In some studies, many of the measured biomarker concentrations were at or near the LOD (*e.g.*, Barr *et al.*, 2010; Berkowitz *et al.*, 2004; Eskenazi *et al.*, 2004, 2007). The uncertainty associated with such measurements can lead to exposure measurement error, biasing results toward or away from the null.

Direct measurement of chlorpyrifos in cord blood is the most reliable exposure metric, and studies using this method should carry more weight in assessing the hypothesis that chlorpyrifos exposure causes neurodevelopmental effects.

3.2.3.2. Outcome assessment The epidemiology studies considered the following neurodevelopmental outcomes: head circumference; infant neurobehavior as assessed by the BNBAS; cognitive development as assessed by MDI scores on the BSID-II and by the WISC-IV; motor development as assessed by PDI scores on the BSID-II; and behavioral outcomes (attention problems, ADHD, and PDD) as assessed by the CBCL. Measurements of these outcomes can be informative for assessing neurodevelopmental effects, but the cohort study results are limited by the sensitivity and predictive ability of these standardized developmental tests in the first few years of life.

Head circumference: Head circumference was examined in six of the cohort studies (Perera *et al.*, 2003; Whyatt *et al.*, 2004; Eskenazi *et al.*, 2004; Berkowitz *et al.*, 2004; Wolff *et al.*, 2007; Barr *et al.*, 2010). Head circumference correlates with brain weight (Lemons *et al.*, 1981) and some studies have reported that reduction in head circumference correlates with lower IQ and poorer cognitive functioning in childhood (Hack *et al.*, 1991; Lasky *et al.*, 1981; Ounsted *et al.*, 1988; Rushton and Ankney, 2009), whereas others have reported no strong influence of brain volume on overall cognitive performance (Schoenemann *et al.*, 2000). The studies that examined head circumference compared cord blood levels to head circumference only at birth, so the exposures measured do not necessarily precede the outcome measured, and this single measure can lead to misclassification of outcome.

Infant neurobehavior: Infant neurobehavior was examined in two of the cohort studies (Young *et al.*, 2005; Engel *et al.*, 2007). Both cohort studies measured infant behavior using the BNBAS, which consists of 28 behavioral items scored on a nine-point scale and 18 reflex items scored on a four-point scale. Each of these items were reduced to seven clusters based on the scoring method developed by Lester *et al.* (1982): habituation, orientation, motor performance, range of state, regulation of state, autonomic stability, and primitive reflex. The six behavioral cluster scores are calculated such that higher scores represent more optimal functioning, whereas the reflex cluster score is the total number of reflexes coded as abnormal, so that higher scores indicate less optimal functioning. Young *et al.* (2005) stated that the 18 reflex items of the BNBAS are not designed to provide a neurological diagnosis, but can potentially identify gross neurologic abnormalities, as more than three abnormally rated reflexes may be clinically relevant (Brazelton and Nugent, 1995).

Young *et al.* (2005) stated that the BNBAS can be appropriately administered through the first two months of life, and the authors assessed infants between 0 and 62 days of age. Engel *et al.* (2007) conducted assessments within five days of birth and noted that 23% of the initial cohort was not assessed for various reasons, including weekend delivery. This is a potential source of selection bias, as any factors related to weekend delivery, such as fewer induced deliveries, may be underrepresented among the subjects who were tested. Engel *et al.* (2007) stated that even if those factors are related to exposure or disease status, this alone does not impose a bias on their findings, as fewer induced deliveries with pesticide exposure alone would affect precision but would not affect the validity of the estimates.

In both studies, each infant was assessed only once. The prognostic utility of a single assessment of infant reflexes shortly after birth is unclear and the study results may be subject to misclassification of outcome from the use of a single measurement of neurobehavior.

Cognitive and motor development: Three of the cohort studies examined associations between chlorpyrifos exposure and cognitive and motor development, as measured by the BSID-II within the first 3 years of life (Eskenazi *et al.*, 2007; Rauh *et al.*, 2006; Engel *et al.*, 2011). The BSID-II is used to identify young children at risk for mental and motor delay. Rauh *et al.* (2006) stated that when administered at 3 years of age, the BSID-II demonstrates moderate predictive power for subsequent intelligence and school performance but is only clinically useful for children performing in the subnormal range.

Two studies assessed cognitive development between the ages of seven and 9 years through scores on the WISC-IV (Rauh *et al.*, 2011; Engel *et al.*, 2011). The WISC-IV measures four indices of cognitive functioning (verbal comprehension, working memory, perceptual reasoning, and processing speed) that are combined to yield a full-scale IQ score. Rauh *et al.* (2011) noted that WISC-IV scores can be influenced by factors such as socioeconomic status (SES) and child behavior problems.

Behavioral outcomes: Two studies examined behavioral outcomes as measured by scores on the 99-item CBCL for ages 18 months to 5 years (Rauh *et al.*, 2006; Eskenazi *et al.*, 2007). The CBCL is a widely used measure to assess children's emotional and behavioral problems and competencies during the previous two months, and its validity and reliability have been documented

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

(Achenbach *et al.*, 2003). It was administered to mothers in the Columbia cohort when the children were 36 months of age (Rauh *et al.*, 2006) and to mothers in the CHAMACOS cohort when their children were 24 months of age (Eskenazi *et al.*, 2007). Both studies examined the results of three scales of the CBCL: attention problems, ADHD, and PDD. The attention problem scale rates behaviors related to concentration and sitting still. The ADHD scale includes additional attention items such as "gets into everything," and criteria for this scale are derived from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), which is used as a diagnostic tool (APA, 2000, as cited by Eskenazi *et al.*, 2007). The PDD scale criteria are also derived from the DSM-IV, and they rate behaviors that are consistent with Asperger's Disorder and Autistic Disorder, and include items such as avoiding eye contact, rocking of the head and body, and unresponsiveness to affection. Scores on each scale are considered of borderline clinical significance if they are > 93rd percentile of the national norms and of clinical significance if > 97th percentile (Achenbach and Rescorla, 2000).

The single assessment of behavioral outcomes at one age per study may lead to outcome misclassification, particularly when the measure is based on maternal report, which is subject to reporting bias. Rauh *et al.* (2006) noted that the DSM-IV, from which the CBCL criteria for ADHD are derived, has low sensitivity for assessing the inattentiveness of preschool-aged children, limiting the usefulness of this outcome measure.

3.2.3.3. Clinical significance To determine whether any reported outcomes associated with chlorpyrifos exposure are adverse effects, one should assess whether they are clinically significant (Goodman *et al.*, 2010). That is, whether or not a statistically significant association has been observed, one should determine whether an outcome actually constitutes an adverse effect. If the outcomes observed in the cohort studies are clinically significant, they are considered to constitute an adverse effect.

Two studies used the BNBAS to assess seven clusters of infant neurobehavior, including abnormal reflexes (Young *et al.*, 2005; Engel *et al.*, 2007). More than three abnormally rated reflexes may be clinically relevant, often resulting in a more intensive neurologic examination and possible intervention (Brazelton and Nugent, 1995).

Three studies examined associations between chlorpyrifos exposure and cognitive and motor development as assessed by MDI and PDI scores on the BSID-II (Rauh *et al.*, 2006; Eskenazi *et al.*, 2007; Engel *et al.*, 2011). Scores for the BSID-II have a mean (standard deviation [SD]) of 100 (15) for "normal," so, assuming a normal distribution, in 68% of a standard population, the scores range from 85 to 115. As noted above, the BSID-II is only clinically useful for predicting subsequent intelligence and school performance in children scoring in the subnormal range (< 85).

Two studies used the WISC-IV to examine associations between chlorpyrifos exposure and cognitive development (Rauh *et al.*, 2011; Engel *et al.*, 2011). Like the BSID-II, scores for the WISC-IV have a mean and standard deviation of 100 and 15, respectively, so "normal" scores range from 85 to 115.

Two studies assessed the association between chlorpyrifos exposure and the risks of scoring in the clinical range of the CBCL for attention problems, ADHD, and PDD (Rauh *et al.*, 2006; Eskenazi *et al.*, 2007). Scores on each of these scales are considered of borderline clinical significance if they are > 93rd percentile of the national norms and of clinical significance if > 97th percentile. Eskenazi *et al.* (2007) reported a small percentage of children scoring in the clinical range for attention problems and ADHD (2.0–2.8%), so they used the less-conservative borderline cut-points for these outcomes. The true clinical significance of scores for DSM-IV-oriented scales, such as ADHD and PDD, is still unknown because scores in the clinical range of these scales based on maternal report are not directly equivalent to a DSM diagnosis (Achenbach and Rescorla, 2000).

3.2.3.4. Confounding and bias Many genetic and environmental factors are hypothesized to affect neurodevelopmental outcomes, and these factors could be correlated with exposure to chlorpyrifos or other pesticides, generating confounding. Only a small number of these factors were considered as confounders in the cohort studies. Because several studies examined the same cohorts, confounding factors that affect outcomes in a given cohort will do so across all studies of that cohort. Lack of adjustment for these factors in statistical models decreases the likelihood that any observed effects are attributable to chlorpyrifos exposure.

Each cohort was exposed to multiple types of pesticides besides chlorpyrifos. While studies of the Columbia and Mount Sinai cohorts used chlorpyrifos levels as the exposure metric, confounding by other pesticides could have occurred. Other studies relied on non-specific metabolite levels that reflect exposure to other pesticides in addition to chlorpyrifos, so it cannot be known whether any observed associations are attributable to chlorpyrifos, these other pesticides, or some other confounder. Chlorpyrifos exposures were often correlated with other pesticides, and only one of the studies controlled for other pesticide exposures to ensure that the main effects of chlorpyrifos were not attributable to other pesticides. Whyatt *et al.* (2004) controlled for diazinon and the carbamate insecticide propoxur (*via* its metabolite, 2-isopropoxyphenol) in models predicting the associations between chlorpyrifos exposure and birth weight and birth length, but did not control for these pesticides in models for head circumference. For these reasons, any observed associations in the cohort studies may be attributable to pesticides other than chlorpyrifos.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Exposure to lead has been associated with adverse cognitive, motor, and behavioral outcomes in children, and cohort studies of these effects in early childhood considered potential confounding by lead exposure. Rauh *et al.* (2006, 2011) measured lead levels as well as chlorpyrifos levels in cord blood. They reported that lead and chlorpyrifos levels were not correlated, so they did not include lead as a covariate in their regression models for each outcome. Eskenazi *et al.* (2007) also measured lead biomarkers and modeled them simultaneously with pesticide exposures for MDI scores, for which no associations were reported with TCPy and DEPs, but did not include lead exposure in the models for PDD, for which an association was reported with levels of DEPs.

Exposure to environmental tobacco smoke (ETS) has also been associated with adverse cognitive, motor, and behavioral outcomes. In studies of the Columbia cohort, the authors attempted to eliminate active smokers by excluding women with plasma cotinine levels > 25 ng/ml from the analyses (Perera *et al.*, 2003). Approximately 43% of mothers reported a smoker in the home and also had cotinine values that reflected tobacco exposure. Cotinine levels have a half-life during pregnancy of 8.8 hours, and maternal plasma samples were obtained within one (Perera *et al.*, 2003) or two days (Whyatt *et al.*, 2004) after delivery, increasing the likelihood that cotinine status from this measure is considerably underestimated. All mothers also had detectable inhalation levels of one or more PAH, but Perera *et al.* (2003) reported no significant interactions between PAHs and chlorpyrifos.

Cotinine levels were not measured in the CHAMACOS and Mount Sinai cohorts, and information regarding smoking use in the mothers was obtained solely by questionnaires, a practice that is subject to reporting bias in that mothers may have underreported this factor because it is not socially desirable. There was no information collected on smoking in the study of the UMDNJ cohort (Barr *et al.*, 2010).

Maternal alcohol has been shown to be a significant predictor of adverse behaviors in children, and attention problems in particular (Sood *et al.*, 2001). Information regarding alcohol use was obtained solely by questionnaires in each cohort, which, as noted above, is subject to reporting bias. Perera *et al.* (2003) noted that 24% of the mothers in the Columbia cohort reported alcohol use during pregnancy, and in analyses of behavioral outcomes in this cohort, Rauh *et al.* (2006) did not control for alcohol exposure. Mothers in the Mount Sinai cohort were excluded if they drank more than two alcoholic drinks per day, which eliminates only those with heavy consumption and introduces uncertainty regarding alcohol exposure levels across the remainder of the cohort. There was no information collected on alcohol use in the study of the UMDNJ cohort (Barr *et al.*, 2010).

The cohorts in which associations were reported all come from populations with low SES. The effects of low SES on fetal and child neurodevelopment have been demonstrated in a range of populations (Rauh *et al.*, 2004); thus, it is expected that children in these cohorts should have lower scores in the outcome measures examined. Both Rauh *et al.* (2006) and Eskenazi *et al.* (2007) observed a large increase in the percentage of children with deficits in MDI scores at the final time point in their study in which the children were examined (36 months and 24 months, respectively), which could be attributed to a lack of stimulating environments, which is common in low SES populations, or to this measure not being clinically valid in Spanish or Latino immigrant communities, as the BSID-II is confounded by significant language demands (Youngstrom *et al.*, 2010). Engel *et al.* (2007) noted that the exclusions of mothers after study entry because they moved out of the area, were lost to follow-up, or lacked prenatal biological specimens may be a potential source of selection bias because these exclusion factors largely reflect socioeconomic condition. These reasons for exclusion may also be associated with other lifestyle factors that are correlated with chlorpyrifos exposure, providing other possibilities for confounding. Of the studies reporting associations with neurodevelopmental outcomes, all examined at least some confounding factors related to SES (*e.g.*, mother's education; household income; quality of home environment) and those that were associated with outcomes were included as covariates in the final models.

Maternal IQ is also a potential confounder for cognitive outcomes, and it was controlled for only in the studies by Rauh *et al.* (2006, 2011). Rauh *et al.* (2006) used the sample mean to substitute for IQ scores that were missing for 29 of the 254 mothers, however, and such imputation of missing data adds considerable uncertainty to this factor.

All of the studies, with the exception of the UMDNJ cohort study, were conducted during a time period spanning the phase-out of residential uses of chlorpyrifos that began on January 1, 2001. After the ban, it was possible that some families continued to use chlorpyrifos products purchased before the ban, so it is difficult to determine the date when exposure stopped. It is possible that other factors distinguished the pre- and post-ban periods, and that any of the reported associations with neurodevelopmental outcomes may have been attributable to these factors. This possibility was addressed only in studies of the Columbia cohort. Whyatt *et al.* (2004) reported that levels of chlorpyrifos in ambient air and blood samples from this cohort decreased substantially between 1998 and 2002. When their analysis with chlorpyrifos air data was stratified among newborns born before *versus* on or after January 1, 2001, there was still no association with head circumference. The authors did not do this analysis using chlorpyrifos levels in blood as the exposure metric. Rauh *et al.* (2006) examined associations with MDI and PDI scores with chlorpyrifos exposures pre-ban, mid-ban, and post-ban, and observed statistically significant increases in MDI and PDI scores

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

from pre-ban to mid-ban, but slight decreases in these scores post-ban that were not statistically significant, indicating no improvements in mental or motor function after the ban. These results indicate that there were no confounding factors that distinguished the two periods for this cohort.

Although each cohort study adjusted for several different confounders, other factors that could affect the results may not have been accounted for, which increases the likelihood that there are alternative explanations for the observed outcomes other than exposure to chlorpyrifos. Thus, studies that do not adequately account for potential confounders carry less weight in addressing the hypothesis that chlorpyrifos causes neurodevelopmental effects.

3.2.3.5. Exposure-response If chlorpyrifos is a causal factor for neurodevelopmental effects, one would expect the risks for these effects to increase with exposure both within and among studies. Most of the chlorpyrifos cohort studies did not fully assess the exposure-response relationship. While many studies analyzed the exposure metric as a continuous variable (Perera *et al.*, 2003; Whyatt *et al.*, 2004; Wolff *et al.*, 2007; Young *et al.*, 2005; Engel *et al.*, 2007, 2011; Eskenazi *et al.*, 2004, 2007; Rauh *et al.*, 2011), some studies dichotomized exposure levels (Barr *et al.*, 2010; Eskenazi *et al.*, 2004, 2007; Berkowitz *et al.*, 2004; Rauh *et al.*, 2006). This sacrificed exposure information and does not allow full exploration of subtle relationships between exposure and outcome.

Only two studies performed analyses assessing exposure-response relationships. Rauh *et al.* (2006) originally categorized exposure into tertiles, with an additional referent group having undetectable chlorpyrifos levels, and they observed lower mean MDI and PDI scores in the highest tertile and the referents compared with the low and middle tertiles, indicating a lack of an exposure-response relationship. The authors reported these results as "preliminary" and used the results from a *post hoc* dichotomization of exposure levels above and below the cut-off for the highest tertile in the main analyses.

Young *et al.* (2005) examined the exposure-response relationship for the proportion of infants with more than three abnormally rated reflexes across quintiles of maternal levels of DEPs. Although there was an increasing trend of this outcome with increasing quintiles of exposure, the result was only marginally statistically significant ($p = 0.05$).

It is not possible to assess clearly whether an exposure-response relationship exists among the studies because most studies reporting an association with a specific outcome did not use the same exposure metric. For example, studies of the Columbia cohort used blood levels of chlorpyrifos, whereas those of the CHAMACOS cohort used urinary metabolites. Overall, there is a lack of clear exposure-response information in most of the cohort studies, which limits their ability to aid in assessing the hypothesis that chlorpyrifos causes neurodevelopmental effects.

3.2.3.6. Statistical analyses The statistical strength of epidemiology studies is related, in part, to the number of study subjects, or sample size. The sample sizes in the chlorpyrifos cohort studies ranged from 113 to 486, depending on the exposure metric and outcome, as some studies reported missing measurements for particular metrics or the inability to assess outcomes in some of the children for various reasons. Studies with sample sizes near the high end of this range have a higher likelihood of observing an association, if indeed there truly is one, than those with sample sizes near the low end of the range. If studies with smaller sample sizes report associations when studies with larger sample sizes do not, it may be that the reported associations are statistical anomalies. Of the five studies that reported associations between chlorpyrifos exposure and neurodevelopmental outcomes, three had sample sizes in the lower half of this range (Engel *et al.*, 2007; Rauh *et al.*, 2006, 2011), and two were in the upper half (Young *et al.*, 2005; Eskenazi *et al.*, 2007).

When data are analyzed in multiple ways or for multiple outcomes, the meaning of the statistical tests that are used becomes distorted by the multiple comparisons problem, such that if enough tests are run, it becomes more likely that several results will be statistically significant by chance. For example, Young *et al.* (2005) performed several different analyses of data relating to BNBAS scores and reported statistically significant associations for abnormal reflexes in infants greater than three days of age, and also for autonomic stability in infants ≤ 3 days of age. The latter result is contrary to the *a priori* hypotheses of a detrimental impact of chlorpyrifos, however, and the authors noted that this may be the result of multiple testing, as there is no explanation for a biologically protective effect of chlorpyrifos on infant neurobehavior. It is also possible that their reported association for abnormal reflexes could be the result of multiple testing as well, particularly because the magnitudes of both associations were almost the same. Eskenazi *et al.* (2007) and Rauh *et al.* (2006) also conducted many different analyses and did not adjust for multiple comparisons.

The cohort studies that dichotomized exposure values into low and high groups used cut-off values with no biological basis. For example, Berkowitz *et al.* (2004) divided urinary TCPy concentrations by the LOD of their analytical method because of the large number of concentrations below this limit. Thus, this cut-off value was chosen on an analytical, rather than biological, basis and may have no biologic relevance. The authors did not report the use of other cut-offs with which to compare their results. Other studies used cut-off points such as the median (Eskenazi *et al.*, 2004, 2007; Engel *et al.*, 2007), the 75th percentile (Barr *et al.*, 2010), or the cut-off

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

value for the highest tertile (Rauh *et al.*, 2006). None of these cut-offs have a biological basis. Sensitivity analyses can be used to assess whether results are dependent on where the cut-off value is chosen, but this type of analysis was not routinely performed in the chlorpyrifos cohort studies.

Overall, studies with small sample sizes, potential issues with multiple comparisons, and arbitrary cut-offs for exposure with no biological relevance should carry less weight in assessing the hypothesis that chlorpyrifos causes neurodevelopmental effects because they may support the alternative hypothesis that the observed associations are statistical anomalies.

3.2.4. Conclusions for human data

There are many inconsistencies in the results both within and among the cohort studies examining the association between chlorpyrifos exposure and neurodevelopmental outcomes in newborns and young children. We assessed which study results are likely to be the most valid, considering the exposure metric used, outcome assessed, clinical significance of reported effects, control of confounding factors, exposure-response relationships, and statistical limitations. Based on these factors, we determined whether the studies provide sufficient evidence to support the hypothesis that chlorpyrifos causes neurodevelopmental effects, or whether there are alternative hypotheses that are more likely to explain the results.

Newborn head circumference was examined as an outcome in at least one study of each of the four cohorts. All studies examining this outcome reported no association, regardless of the exposure metric used. Each study measured exposure at only one point in time, which could lead to a high degree of exposure measurement error. Three of the studies used the most reliable exposure metric, which is direct measurement of chlorpyrifos in blood (Perera *et al.*, 2003; Whyatt *et al.*, 2004; Barr *et al.*, 2010). None of the studies adequately controlled for exposure to other pesticides or ETS, maternal alcohol use, low SES, or other potential confounding factors. While a few of the studies had very small sample sizes (*e.g.*, < 150; Perera *et al.*, 2003; Barr *et al.*, 2010), others had relatively larger sample sizes (*e.g.*, 486; Eskenazi *et al.*, 2004). Regardless of the weight of each of these factors, null results were reported across all studies of newborn head circumference, increasing the likelihood that the overall findings are robust and that there is no association between chlorpyrifos exposure and decreased newborn head circumference.

The two studies of effects on infant neurobehavior, as assessed by the BNBAS, reported associations between chlorpyrifos exposure and abnormal reflexes in the CHAMACOS (Young *et al.*, 2005) and Mount Sinai (Engel *et al.*, 2007) cohorts. These studies used only urinary metabolite levels, measured at one point in time, as the exposure metric. Outcome was only assessed at one time point as well, and Young *et al.* (2005) reported potential impacts of the wide variation in the age of infants in their assessment on their results. Their report of a biologically protective effect of exposure for one of the seven cluster scores increases the plausibility of the interpretation that their results are attributable to multiple testing, rather than a true effect. The prognostic utility of a single measurement of infant reflexes shortly after delivery is unclear. Although both studies showed effects related to the same cluster of the BNBAS, these studies suffer from potential exposure measurement error from the use of urinary DEPs as the exposure metric, and misclassification of outcome from a single assessment of neurobehavior. Because of these uncertainties, their results may also support the hypothesis that other factors are causal for this outcome or that the observed associations are statistical anomalies; thus, they should carry less weight in the assessment of neurodevelopmental effects from chlorpyrifos exposure.

The three studies that examined cognitive and motor development *via* the BSID-II used different exposure metrics, and reported no associations between chlorpyrifos exposure and MDI or PDI scores up to 24 months of age in the Columbia (Rauh *et al.*, 2006), CHAMACOS (Eskenazi *et al.*, 2007), and Mount Sinai (Engel *et al.*, 2011) cohorts. Only Rauh *et al.* (2006) examined these outcomes in children at 36 months of age, using a single measure of chlorpyrifos in cord blood as the exposure metric. At this age, associations were reported for lower PDI scores and for mental and motor delays. It is unclear whether these reported effects are clinically significant, however. The difference in mean PDI scores was a modest −6.5 points between the low and high exposure groups, and the mean score of the high-exposure group was 95.69, which is well within the normal range (> 85). Similarly, the difference in mean MDI scores between low and high exposure groups was −3.3 points, with a mean score of 87.39 for the high exposure group. Both studies conducted many different analyses without adjustment for multiple comparisons, did not provide evidence of an exposure-response relationship, and did not adequately control for confounding by other exposures and SES. Although Rauh *et al.* (2006) used a more reliable exposure metric, Eskenazi *et al.* (2007) had a sample size that was almost twice as large. Two studies also examined cognitive development through scores on the WISC-IV administered during the early school-age years. Rauh *et al.* (2011) reported an association between chlorpyrifos concentrations in cord blood and decrements in working memory scores at age seven in the Columbia cohort. By contrast, Engel *et al.* (2011) reported no associations between maternal urinary DEPs and scores on any WISC-IV index in 7- to 9-year-olds in the Mount Sinai cohort. Because of the lack of confirmation of the associations at the same age and the methodological issues described, the studies of cognitive and motor development do not carry enough weight to decrease the likelihood that there are alternative explanations for the observed outcomes other than

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

exposure to chlorpyrifos or that the observed associations are statistical anomalies.

Two of the same studies that measured cognitive and motor development using the BSID-II also examined behavior outcomes reported by mothers on the CBCL, although at different ages (36 months in the study by Rauh *et al.*, 2006; 24 months in the study by Eskenazi *et al.*, 2007). The only consistent association between these two studies was with increased risk of PDD. Rauh *et al.* (2006) reported associations between chlorpyrifos levels in cord blood and all three scales examined (attention problems, ADHD, and PDD) at 36 months in the Columbia cohort. Eskenazi *et al.* (2007) reported no associations between maternal TCPy or DEPs and any of these outcomes, but did report an association between concurrent child DEPs and increased risk of PDD scores in the clinical range in the CHAMACOS cohort. In both studies, very few children scored in the clinical ranges of the CBCL, limiting the clinical significance of the results. Although associations between exposure and risk of PDD were reported in both cohorts, different exposure metrics were used, and children were assessed at different ages. The single assessment of behavioral outcomes at one age per study may lead to outcome misclassification, particularly when the measure is based on reporting by mothers, which is subject to reporting bias. The study using the more robust exposure metric at only one time point (Rauh *et al.*, 2006) reported associations with all three scales examined, whereas the study with the larger sample size but using a less-reliable exposure metric at two time points (Eskenazi *et al.*, 2007) only reported an association with PDD. As with their assessment of effects on cognitive and motor development, these studies do not report consistent results and do not carry enough weight to decrease the likelihood that there are alternative explanations for the observed outcomes other than exposure to chlorpyrifos or that the observed associations are statistical anomalies.

In conclusion, the chlorpyrifos cohort studies do not report consistent results. There are very few studies of each specific neurodevelopmental outcome, limiting the ability to look for consistency across studies or cohorts. The studies with more robust factors, such as reliable exposure metrics or larger sample sizes, do not appear to be more likely to report associations with adverse neurodevelopmental effects. Overall, the epidemiology data are not sufficiently robust to support the hypothesis that chlorpyrifos is a causal factor in neurodevelopmental effects. In the following sections, we will describe the toxicology and mechanistic data regarding chlorpyrifos. We then use the HBWoE approach to evaluate the weight of the evidence regarding a causal association between chlorpyrifos exposure and adverse neurodevelopmental effects.

3.3. Neurodevelopmental toxicity studies in animals

We evaluated the available neurodevelopmental toxicity animal data that are relevant to determining whether sufficient evidence is available to support the hypothesis that chlorpyrifos causes adverse neurodevelopmental effects. Below, we provide a brief overview of these data followed by an endpoint-by-endpoint analysis of the different neurodevelopmental outcomes that have been investigated. Then, we critically assess the data as a whole for consistency of the observed outcomes, evaluate the evidence for exposure-response relationships, and discuss the weight of studies based on factors related to study design. To conduct our analysis, we searched PubMed and Toxline for animal studies that investigated neurodevelopmental effects of chlorpyrifos exposure. Search terms included: "chlorpyrifos," "animal" "cognitive," "neurodevelopmental," "behavior*," "motor," "reflex," "learning," and "memory." We also relied on the reference lists within the studies that we found in the literature search. We limited our evaluation to studies that examined neurodevelopmental effects that are similar to those examined in the cohort studies (*i.e.*, infant neurobehavior, motor and cognitive effects, and general behavioral effects). We identified 22 studies that investigated these potential neurodevelopmental effects in rats and mice.

3.3.1. Overview of animal studies

Many studies have examined the potential association between chlorpyrifos and neurodevelopmental responses in rodents. In the studies that we reviewed, the exposure periods generally ranged from the first gestational day (GD) to the post-weanling period, up to postnatal day (PND) 25 (*i.e.*, the 25th day after birth or extraction from the womb) and encompassed either prenatal time points alone, prenatal and postnatal time points, or postnatal time points alone. Across studies, exposure concentrations ranged from 0.03 to 40 mg/kg chlorpyrifos and were delivered mostly *via* subcutaneous injection or oral gavage, although one study exposed rats dermally and another used intraperitoneal injection as the exposure route. Inhalation exposures to chlorpyrifos were not examined in the animal studies. The neurodevelopmental tests of these exposures were conducted at various time points, ranging from immediately following exposure to adulthood, or even after mating and subsequent reproduction. These tests examined potential effects of chlorpyrifos on social behavior (including maternal behavior), emotion and anxiety, motor function (including locomotor activity, neuromuscular and neuromotor function, and sensorimotor reflexes), and cognitive function (*i.e.*, learning and memory). Table 7 summarizes these tests and their psychobiological significance. In addition, many of the studies described below that investigated the aforementioned effects also assessed brain AChE activity.

3.3.2. Endpoint-by-endpoint analysis of neurodevelopmental effects in animals

In this section, we discuss the animal studies that evaluated the individual neurodevelopmental outcomes

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 7. Neurodevelopmental tests used in animal studies of chlorpyrifos.

Test	Description	Psychobiological significance	References
Social behavior			
Sociability	An unfamiliar animal of the same strain, sex, and age is placed in a side chamber to serve as a social stimulus; this allows nose contact through the bars, but prevents fighting. The other side chamber contains an empty wire cage identical to that enclosing the social stimulus, serving as a control for exploring a novel object/environment.	Sociability and social preference	Venerosi *et al.* (2008)
Socioagonistic behavior	The animal undergoes an agonistic encounter with an unfamiliar partner of the same sex and age and the behavior is video-recorded. Nonsocial and social behaviors are scored; these usually include general activity, such as exploring, rearing, and digging, as well as social responses, including attack, often accompanied by biting attempts; aggressive grooming; tail rattling; offensive upright posture; defensive upright posture; and submissive upright posture.	Social aggression	Ricceri *et al.* (2006)
Social novelty preference	The test animal is restricted in a chamber and an unfamiliar partner is introduced. The test mouse is allowed to choose between a familiar partner and the novel, unfamiliar partner. The mouse is observed for choice of partner and time spent with partner.	Sociability and social preference	Venerosi *et al.* (2008)
Maternal behavior			
Maternal aggressive behavior	After delivery, pups are removed and dam is exposed to an intruder male in its home cage. Maternal behaviors are recorded.	Maternal aggression against an intruder	Venerosi *et al.* (2008)
Pup-directed behavior	Maternal behavior with pups, including pup retrieval, sniffing, crouching over the pups, and nest-building is observed and recorded.	Maternal behavior	Ricceri *et al.* (2006); Venerosi *et al.* (2008)
Emotion/Anxiety			
Light/dark box test	The light/dark box test consists of a box with two compartments: a small, opaque, dark compartment closed by a red glass and a large, transluscent and illuminated white compartment. The floor of the light compartment is divided into identical squares to allow determination of exploratory behavior. The compartments are connected by an opening to allow mice to freely move between them. A spotlight illuminates the apparatus to provide an aversive stimulus. Animals are individually placed in the center of the light compartment and allowed to explore the compartments. The time spent in each compartment, the time spent in the central square of the light compartment, and the number of crossings between compartments is recorded. This apparatus provides an unfamiliar environment for rodents where a natural conflict occurs between the tendency to explore a novel environment (spontaneous exploratory behavior) and the innate aversion to an open field and brightly illuminated area. As the light, but not the dark, compartment provides stress-like conditions for mice, the percentage of time spent by the mice in the dark compartment is used as a measure of anxiety in mice. In addition, exploratory behavior in the light compartment (time spent in the central square relative to the entire time spent in the light compartment) is also indicative of anxiety. The number of crossings between compartments provides an estimation of the anxiety level if measured without changes of spontaneous locomotor activity.	Anxiety	Venerosi *et al.* (2008, 2010); Braquenier *et al.* (2010)
Ultrasonic vocalization	During the postnatal period, pups emit ultrasonic calls to solicit parental care from the dam. These calls peak at PND 7–9. The most frequently used paradigm to elicit ultrasonic calls is to isolate a pup from the dam for a few minutes. Parameters that are analyzed include number, duration, and frequency of calls.	Homing/Anxiety	Venerosi *et al.* (2009); Laviola *et al.* (2006)

(Continued)

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 7. (Continued).

Test	Description	Psychobiological significance	References
Elevated plus-maze	The elevated plus-maze is a conflict-avoidance test. Rodents prefer a dark, enclosed, small space over a brightly lit, open, large space, but they are also highly exploratory. This maze represents the natural conflict between the tendency of mice to explore a novel environment and the tendency to avoid a brightly lit open field. The behavior is also influenced by the fear of heights. The elevated plus-maze is in the shape of a "+". The four arms extend from a central platform. Two alternate arms are dark and enclosed, while two alternate arms are open, lit, and available with or without ledges. Additionally, the surface of all arms is raised 1 m above the floor. The intensity of the fear drive (anxiety) is measured by the ratio of entries in open and enclosed arms. The open and enclosed arms of the maze generate exploratory behavior and the avoidance of elevated open arms is an indication of the intensity of anxiety. Furthermore, head dipping over the edges of the open arms, risk-assessment, self grooming, and defecation can be recorded.	Anxiety	Braquenier *et al.* (2010); Ricceri *et al.* (2006); Icenogle *et al.* (2004); Aldridge *et al.* (2004)
Forced swimming test	The animal is observed during a swimming session and behavioral parameters such as floating (the animal floats in the water making only those movements necessary to keep its head above water), struggling (vigorous movements of all four limbs with the forelimbs breaking the surface of the water) and swimming (all four limbs move under the surface of the water with the head kept above water) are observed. The frequency and duration of each pattern are scored.	Mood or adaptation/learning	Venerosi *et al.* (2010)
Reflexes and motor activity			
Open field locomotor test	Assessment takes place in either a circular platform arena or square arena. The animal is placed in the center of the open field arena and allowed to freely move about while being measured by an automated tracking system. Measurements taken include, but are not limited to, general path length, rearing (standing on hind limbs) behaviors, and exploratory behaviors.	Locomotor activity (general activity levels, gross locomotor activity, and exploration habits. Assessment over several days allows habituation to the environment to be evaluated.)	Carr *et al.* (2001); Ricceri *et al.* (2003, 2006)
Geotaxic response	The animal is placed head-downward and given a period of time to turn itself 180 degrees. Animals are scored for completing the 180 degree turn.	Sensorimotor reflexes	Dam *et al.* (2000)
Cliff avoidance	This test assesses pup withdrawal from the edge of a flat surface when its snout and forepaws are placed over a real or virtual cliff; time of aversion can be recorded.	Sensorimotor reflexes	Venerosi *et al.* (2009)
Acoustic startle test	The animal is subjected to a loud sound or click. A subsequent freezing reaction and blink indicate functional hearing abilities.	Auditory reflexes (hearing ability)	Icenogle *et al.* (2004); Johnson *et al.* (2009)
Free-fall and surface righting	Free-fall: The ability of a pup to correct its posture such that it lands on all four feet when released from a supine position above a soft surface.	Posture reflex	Dam *et al.* (2000); Johnson *et al.* (2009)
	Surface: The ability of a pup to right itself to an upright position from a supine position.		
Inclined plane	The animal is placed on a flat plane in the horizontal position with its head facing the side of the board to be raised. The angle at which the animal begins to slip downward is recorded.	Neuromotor performance	Abou-Donia *et al.* (2006)
Grip strength and forepaw grip time	Forelimb strength is measured as tension force using a computerized grip strength meter (GSM). The animal is lifted over the baseplate of the GSM by the tail so that its forepaws are allowed to grasp onto the steel grip. The mouse is then gently pulled backward by the tail until the grip is released. The GSM measures the maximal force before the mouse releases the bar. Forepaw strength is assessed by having animals grip a wooden dowel that is held horizontally and raised so that the animal supports its weight. The time to release grip is recorded in seconds.	Neuromuscular function and muscular strength	Abou-Donia *et al.* (2006)

(Continued)

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 7. (Continued).

Test	Description	Psychobiological significance	References
Beam-walking	This test assesses the animal's performance in crossing a narrow wooden beam.	Motor coordination and balance	Abou-Donia *et al.* (2006)
Rotorod test	This test assesses the number of times pups fall off a rotating rod during a 2-minute period.	Neuromotor performance	Muto *et al.* (1992)
Memory and Learning			
T-maze	The animal is placed in the start arm of the T-maze and is allowed to explore one of the two goal arms. The mouse is then returned to the start for the next trial and will typically choose to explore the alternate arm. Because this behavior is not reinforced it is considered "spontaneous."	Exploratory behavior and/or memory (measures exploratory behavior when the intervals between the trials are short, but also serves a spatial memory test when the interval between trials is lengthened)	Maurissen *et al.* (2000); Icenogle *et al.* (2004); Levin *et al.* (2001, 2002)
Multi-arm radial maze	Animals are trained to visit a pattern of arms in a multi-arm arm maze to receive a reward (*e.g.*, food). To perform well, the animal must keep the egocentric or allocentric/spatial information regarding which arms of the maze it has already visited during the course of the task. Assessment of choice accuracy for information presented within the day's test session and learning that has accumulated across days of testing are often referred to as working memory and reference memory tasks, respectively.	Working and reference memory errors	Johnson *et al.* (2009); Icenogle *et al.* (2004); Levin *et al.* (2001, 2002); Haviland *et al.* (2010)
Morris swim test	Animals are trained to locate a hidden platform within a pool of opaque water using spatial cues around the pool as a guide. The tank is surrounded by curtains that contain contrasting shapes to serve as spatial cues. Free swims (trials in which the platform is removed) are given throughout training to assess learning. Cued or visible platform trials are typically given at the end of training and serve as a sensorimotor index for each animal.	Spatial learning and memory	Jett *et al.* (2001)
Figure-8 apparatus	The apparatus consists of a continuous, enclosed alley in the shape of a Figure-8, and two blind alleys extending from either side. A single photobeam is located in each blind alley and three photobeams are located in the loops of the Figure-8 for a total of eight photobeams. Animals are allowed to roam in the maze for a set time and photobeam breaks are recorded to measure locomotor activity. To assess the speed of habituation, the blocks can be analyzed for linear and quadratic trends.	Motor activity and habituation	Icenogle *et al.* (2004)
Foraging maze	The foraging maze consists of a central platform elevated from the floor on a small movable table, containing eight arms projecting radially. At the end of each radial arm, two choice arms project in a 'Tee' format. At the end of each Tee, a low barrier is placed on both sides that serves as a food cup, and conceals a large bait under a screen to mask olfactory clues. The maze has opaque sides and clear overhead plastic and is housed in a lighted room with multiple extra-maze visual cues overhead. A fasted animal is placed in a cylinder in the central starting area and then allowed to roam freely about the maze for a set time or until all baits have been discovered. Radial arm choice and pre-reward and post-reward arm entries are recorded.	Learning, memory, and foraging behavior	Haviland *et al.* (2010)
Passive avoidance	This task exploits the rodent's natural preference for dark environments. In this task the animal has to choose between responding to obtain a positive reinforcement (entering into a dark chamber) and not responding to avoid an electric footshock (not entering the dark chamber). The latency to refrain from entering into the dark compartment serves as an index of conditioned suppression and the ability to avoid, and allows memory to be assessed.	Conditioned suppression of instrumental behavioral responses (memory)	Ricceri *et al.* (2003)
Acclimatization	The test animal is first placed in a middle chamber and the doorways into side chambers are opened to allow exploration of the whole cage. Animals are observed for time of exploration of the chamber.	Exploration and habituation	Venerosi *et al.* (2008)

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

described above. We distinguish studies by exposure route, as studies with oral exposures are considered more relevant for determining risks to human health and are given more weight than studies with exposures *via* injection. We also distinguish studies by the developmental period of exposure to allow for an assessment of the consistency of findings across studies in which exposures occurred during the same neurodevelopmental window. If chlorpyrifos acts as a neurodevelopmental toxicant in the absence of inhibition of AChE activity in the nervous system, results should be consistent in terms of dose-response relationships and patterns of response over neurodevelopmental windows of exposure, both within and across studies, and indicative of effects at doses below those which induce brain AChE inhibition. Other considerations to bear in mind are the adequacy of the study design (*i.e.*, sufficient number of animals, relevant dose levels, and appropriate methodologies), whether the meaning of the statistical tests become distorted by the multiple comparisons problem when data are analyzed for multiple outcomes, and the biological significance of responses.

3.3.2.1. Social and maternal behavior Three studies conducted by the same research group assessed potential chlorpyrifos-associated effects on social and maternal behavior in mice. One study investigated both prenatal exposure *via* oral dosing of dams during gestation and direct postnatal exposure of pups *via* subcutaneous injection (Riccerri *et al.*, 2006), and two studies examined only postnatal, subcutaneous exposures of pups (Riccerri *et al.*, 2003; Venerosi *et al.*, 2008). The results of these studies are described below and summarized in Table 8.

Socioagonistic behavior: Two studies assessed the effects of chlorpyrifos on socioagonistic behavior in mice. Ricceri *et al.* (2006) exposed mouse dams to 0, 3, or 6 mg/kg-day chlorpyrifos by oral gavage during GD 15–18, and offspring were treated with subcutaneous injection of 0, 1, or 3 mg/kg-day chlorpyrifos during PND 11–14. The authors reported that prenatal exposure with 6 mg/kg-day chlorpyrifos increased offensive upright posture in male offspring during PND 75–80, but showed no effect on the other socioagonistic behaviors examined (defensive or submissive upright posture, attack, aggressive grooming of partner, or tail rattling). There was no postnatal treatment effect nor was there an interaction between prenatal and postnatal treatments for this endpoint. The only socioagonistic effects associated with postnatal exposure were increases in the frequency and duration of attacks by male mice exposed to 3 mg/kg-day chlorpyrifos. In addition, there were no effects on any of the general social interaction behaviors examined, such as rearing, digging, moving around the cage, or self-grooming, with any treatment. The authors also reported that there were no effects of chlorpyrifos exposure during the prenatal, postnatal, or both periods on brain AChE activity on GD 19 or PND 15.

Ricceri *et al.* (2006) reported no effect of chlorpyrifos exposure on rearing, wall rearing, digging, moving around cage, or self-grooming on PND 70 in the offspring of mouse dams exposed prenatally *via* oral gavage to 0, 3, or 6 mg/kg-day chlorpyrifos during GD 15–18 followed by postnatal exposure *via* subcutaneous injection of pups to 0, 1, or 3 mg/kg-day chlorpyrifos during PND 11–14. The authors also reported no effects on brain AChE activity on either GD 19 or PND 15 with any chlorpyrifos exposure.

Ricceri *et al.* (2003) exposed mouse pups to 0, 1, or 3 mg/kg-day chlorpyrifos by subcutaneous injection from PND 1–4 or PND 11–14 and reported an increase in aggressive grooming of social partner in males on PND 45 with both doses during the earlier treatment period. Other markers of agonistic behavior– such as attack, tail rattling, offensive upright posture, and defensive postures– were not individually reported. When all aggressive responses were pooled together for the PND 1–4 exposure period, animals in the 1 mg/kg-day group had an increased aggressive response frequency in the first two (out of four) five-minute observation blocks, while animals in the 3 mg/kg-day group did so only in the last two five-minute blocks of observation. Chlorpyrifos exposures during PND 11–14 resulted in an enhancement of the agonistic behavior observed in animals with exposure during PND 1–4, as greater increases in aggressive response frequency were observed with both doses, but were higher with 1 mg/kg-day *vs.* 3 mg/kg-day chlorpyrifos exposure. Despite these indications of chlorpyrifos-associated agonistic behavior, the authors reported that treatment during PND 1–4 or PND 11–14 had no effect on investigative and affiliative behaviors at 45 days of age. In addition, chlorpyrifos treatment during the earlier exposure period had no effect on soliciting behavior, whereas treatment during the later exposure period produced mixed results, as the authors reported an increased frequency of the "push-under" but not the "crawl" response after treatment with 3 mg/kg-day chlorpyrifos. The authors reported a transient inhibitory effect of chlorpyrifos on brain AChE activity (approximately 20% inhibition relative to controls) on PND 4 in mice exposed to 1 or 3 mg/kg-day chlorpyrifos during PND 1–4 but not in mice exposed to 3 mg/kg-day with treatment during PND 11–14 or PND 32–35.

Maternal behavior: Two studies examined the potential effects of chlorpyrifos on maternal aggression and pup-induced maternal behavior. Ricceri *et al.* (2006) reported neither a main effect of prenatal chlorpyrifos nor an interaction between prenatal and postnatal treatments for foster pup-directed behaviors on PND 90 in virgin female offspring of mouse dams exposed to 0, 3, or 6 mg/kg-day chlorpyrifos *via* oral gavage during GD 15–18, followed by subcutaneous exposure to 0, 1, or 3 mg/kg-day chlorpyrifos during PND 11–14. In addition, there were no effects of chlorpyrifos on latency, frequency, and duration of nest building and pup retrieval to the nest. The authors reported an increase in

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 8. Rodent studies of social and maternal behavior.

Study	Species	Animals per dose	Exposure			Outcome						Effects on AChE activity in brain
			Route	Concentration (mg CPF/ kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	
Ricceri *et al.* (2006)	Mouse, CD-1	10	Oral gavage	0, 3, 6	GD 15–18	Male	PND 75–80	Socioagonistic behavior	Increased offensive upright posture	Defensive or submissive upright posture; frequency and duration of attack response; aggressive grooming of partner, tail rattling	Increase in one marker of socioagonistic behavior with 6 mg/kg-day	Null
			s.c.	0, 1, 3	PND 11–14	Male	PND 75–80	Socioagonistic behavior	Increased frequency and duration of attack response	Offensive upright posture; defensive upright or submissive posture; aggressive grooming of partner, tail rattling	Increase in two markers of socioagonistic behavior with 3 mg/kg-day	
			Oral gavage	0, 3, 6	GD 15–18	Male	PND 75–80	Social interactions	None	Rearing; digging; moving around cage; self-grooming	Null	
			s.c.	0, 1, 3	PND 11–14							
			Oral gavage	0, 3, 6	GD 15–18	Female	PND 90	Maternal behavior	None	Nest building latency, frequency, and duration; pup retrieval to nest; crouch response frequency and duration; pup licking frequency and duration; pup sniffing	Null	
			s.c.	0, 1, 3	PND 11–14	Female	PND 90	Maternal behavior	Increased frequency and duration of crouch response; decreased licking frequency; increased licking duration; decreased sniffing	Nest building latency, frequency, and duration; pup retrieval to nest	Alterations of 5 markers of maternal behavior with 1 and 3 mg/ kg-day	

(Continued)

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 8. (Continued).

Study	Species	Animals per dose	Exposure			Outcome						Effects on AChE activity in brain
			Route	Concentration (mg CPF/ kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	
Ricceri *et al.* (2003)	Mouse, CD-1	7 to 10	s.c.	0, 1, 3	PND 1–4	Male, Female	PND 45	Social interactions	None	Social interactions with an unfamiliar untreated mouse for 20 minutes	Null	~20% inhibition with 1 and 3 mg/kg-day
					PND 11–14							Null
					PND 1–4	Male, Female	PND 45	Social investigative behaviors	None	Anogenital sniffing; nose sniffing; body sniffing; following partner around cage; mutual sniffing of anogenital area)	Null	~20% inhibition with 1 and 3 mg/kg-
					PND 11–14							Null
					PND 1–4	Male, Female	PND 45	Affiliative behaviors	None	Social rest, usually next to partner; grooming partner's body)	Null	~20% inhibition with 1 and 3 mg/kg-day
					PND 11–14							Null
					PND 1–4	Male, Female	PND 45	Soliciting behaviors	None	Push-under response (pushing snout or whole anterior body under partners'); crawl behavior (crawling over partner)	Null	~20% inhibition with 1 and 3 mg/ kg-day
					PND 11–14				Increased frequency of the push-under response	Crawl behavior	Alteration of one marker of soliciting behavior with 3 mg/kg-day	Null
					PND 1–4	Male	PND 45	Socioagonistic behavior	Increased frequency of aggressive grooming of partner; increased frequency of pooled aggressive responses in two of four observation blocks	Individual aggressive behaviors: attack response; tail rattling; offensive upright posture; defensive upright or submissive posture	Increase in one marker of socioagonistic behavior and transient increase in overall socioagonistic behavior with 1 and 3 mg/kg-day	~20% inhibition with 1 and 3 mg/ kg-day

(Continued)

© 2011 Informa Healthcare USA, Inc.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 8. (Continued).

Study	Species	Animals per dose	Exposure			Outcome						Effects on AChE activity in brain
			Route	Concentration (mg CPF/ kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	
					PND 11–14				Increased frequency of pooled aggressive responses in all observation blocks	Individual aggressive behaviors: attack response; aggressive grooming of partner; tail rattling; offensive upright posture; defensive upright or submissive posture	Increased socioagonistic behavior with 1 and 3 mg/ kg-day	Null
Venerosi *et al.* (2008)	Mouse, CD-1	12 to 15	s.c.	0, 3	PND 11–14	Male, Female	PND 40–45	Acclimatization	Decreased distance moved when allowed to freely explore cage in first of three sessions	None	Altered acclimatization with 3 mg/ kg-day	Not examined
							PND 40–45	Social novelty preference and sociability	None	Time spent with partner 1 *vs.* partner 2; time spent with a partner *vs.* empty chamber	Null	
						Female	Birth to PPD 7	Nest-building activity	Longer latency to nest building	Amount of materials used for nest; nest qualitative features	Alteration of one marker of maternal behavior with 3 mg/kg-day	
							PPD 1	Maternal behavior	Shorter latency to licking	Licking duration; licking frequency; time spent with pups; nursing; pup retrieval; pup sniffing	Alteration of one marker of maternal behavior with 3 mg/kg-day	
							PPD 7	Maternal aggression against male intruder	Decreased defensive posture duration; increased digging duration; increase in pooled investigative behaviors	Attack frequency; individual investigative behaviors: circling, social sniffing, and following)	Decrease in 3 markers of maternal aggression with 3 mg/kg-day	

Notes: CPF = chlorpyrifos; GD = gestational day; PND = postnatal day; PPD = postpartum day; s.c. = subcutaneous injection.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

frequency and duration of crouch response (*i.e.*, crouching over the pups), decreased frequency and increased duration of licking, and decreased sniffing with both 1 and 3 mg/kg-day postnatal exposures. These observed behaviors were not associated with altered brain AChE activity. In a subsequent study, Venerosi *et al.* (2008) exposed mice to 0 or 3 mg/kg-day chlorpyrifos by subcutaneous injection during PND 11–14 and reported some changes in maternal behavior on post-partum day (PPD) 1 in chlorpyrifos-treated females that were mated on PND 60. For example, treated females had a shorter latency to licking of pups, but licking duration and frequency were not different from controls. Moreover, time spent with pups, nursing, pup retrieval, and pup sniffing were not affected by chlorpyrifos treatment. In assessing maternal aggressive behavior on PPD 7, the authors reported that treated dams displayed fewer defensive postures, had a longer digging duration, and showed increased investigative behavior, but there were no differences in frequency of attacks compared to untreated controls. In addition, treated dams had a longer latency to build nests, but there was no effect of treatment on either the amount of nest material used or nest quality features on PPD 7.

Summary: The two studies examining social behavior each reported effects on a few isolated markers of socioagonistic behavior in mice, although the same specific markers were not affected in both studies. These effects were observed with prenatal (GD 15–18), oral chlorpyrifos exposure of 6 mg/kg-day or with subcutaneous postnatal (PND 1–4 or PND 11–14) exposures of 1 or 3 mg/kg-day, and AChE inhibition was observed with the postnatal exposures in only one of the studies.

The two studies that assessed maternal behavior reported effects on a few markers with postnatal (PND 11–14) chlorpyrifos exposure *via* subcutaneous injection at doses of 1 or 3 mg/kg-day, with no associated inhibition of AChE activity, but reported no effects on any markers with prenatal (GD 15–18) oral exposures up to 6 mg/kg-day. Similar to the studies of socioagonistic behavior, the same specific markers of maternal behavior that were examined in both studies were not affected in the same way.

3.3.2.2. Emotion and anxiety Nine studies assessed the effects of chlorpyrifos exposure on emotional changes or anxiety in rodents using ultrasonic vocalization monitoring, the elevated plus-maze test, the light-dark box test, and the forced swim test. Three studies examined oral exposures, either prenatally *via* dosing of dams during gestation (Venerosi *et al.*, 2009, 2010) or perinatally *via* gestational exposure that continued after birth through nursing (Braquenier *et al.*, 2010). Six studies investigated subcutaneous exposures either prenatally (Icenogle *et al.*, 2004; Laviola *et al.*, 2006) or postnatally (Aldridge *et al.*, 2005a; Ricceri *et al.*, 2003; Venerosi *et al.*, 2008), with one study including prenatal exposure *via* oral dosing of dams during gestation for a subset of animals (Ricceri *et al.*, 2006). These studies are described below and summarized in Table 9.

Anxiety and distress calling in pups: One study with oral exposure and two studies with subcutaneous exposure examined anxiety and distress calling in rodent pups as measured by ultrasonic vocalization. Venerosi *et al.* (2009) exposed mouse dams to 0 or 6 mg/kg-day chlorpyrifos *via* oral gavage during GD 15–18 and monitored ultrasonic vocalization within 10 days of birth. The authors reported that chlorpyrifos-treated offspring emitted fewer and shorter ultrasonic calls, with longer latency to call and higher frequency of calls, compared to controls.

By contrast, the studies with subcutaneous exposures reported no effects on anxiety or distress calling in mouse pups. Laviola *et al.* (2006) reported that offspring of mouse dams exposed to 5 mg/kg-day chlorpyrifos-oxon during GD 14–16 did not exhibit altered ultrasonic vocalization relative to controls when tested within several days of birth. Consistent with this study, Ricceri *et al.* (2003) reported no effects on ultrasonic vocalization and homing in mouse pups with postnatal exposures to 1 or 3 mg/kg-day chlorpyrifos during PND 1–4 when tested a few days after exposure. Brain AChE activity was transiently inhibited approximately 20% compared to controls on PND 4 with exposures to both doses in these animals.

Anxiety assessed by the elevated plus-maze test: One study with oral exposure and three studies with subcutaneous exposure examined anxiety in adolescent or adult rodents using the elevated plus-maze test. Braquenier *et al.* (2010) assessed effects on anxiety using the elevated plus maze test in 80-day-old female offspring of mouse dams exposed by oral gavage to 0.2, 1, or 5 mg/kg-day chlorpyrifos during GD 15 - PND 14. The authors reported that treated mice were more anxious than controls, as indicated by a decrease in the percent time spent in the open arms of the test and a lower proportion of entries in the open arms. This anxiety was not chlorpyrifos-exposure-dependent, however, as the effects were observed with the 1 mg/kg-day dose but not the 0.2 or 5 mg/kg-day doses, and there is no evidence to support a non-monotonic dose-response curve for this endpoint. In addition, the authors reported a 14% inhibition of brain AChE activity on PND 1 in the offspring of dams exposed orally to 5 mg/kg-day chlorpyrifos from GD 15 until birth compared to controls.

Studies that assessed anxiety with the elevated plus-maze test after subcutaneous exposures of chlorpyrifos reported no increases in anxiety across prenatal and postnatal exposure periods. Icenogle *et al.* (2004) exposed rat dams to 0, 1, or 5 mg/kg-day chlorpyrifos by subcutaneous injection during GD 9–12 and reported no effects on anxiety in 4- to 8-week-old offspring in the elevated plus-maze test, but did report an indication of hyperactivity with the 5 mg/kg-day dose. Similarly, Ricceri *et al.* (2006) reported that after prenatal oral exposure to chlorpyrifos (GD 15–18; 0, 3, or 6 mg/kg-day dosing

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

of dams) followed by postnatal subcutaneous exposure to chlorpyrifos (PND 11–14; 0, 1, or 3 mg/kg-day), there were no prenatal by postnatal exposure interactions in the elevated plus-maze test when mice were four months of age. With prenatal exposure, a decrease in head dipping frequency (indicating less anxiety) was reported for males exposed to 3 mg/kg-day chlorpyrifos. With postnatal exposure, females spent more time in the maze's open arms (also indicating less anxiety) when exposed to 3 mg/kg-day chlorpyrifos. These effects were not associated with altered brain AChE activity. Aldridge *et al.* (2005a) exposed rat pups to 0 or 1 mg/kg-day chlorpyrifos *via* subcutaneous injection during PND 1–4 and reported no treatment-related effects on anxiety in the elevated plus-maze test on PND 52–53 in females and *less* anxious behavior (by spending more time in the open arms) in treated males compared to controls.

Anxiety assessed by the light/dark box test: Two studies with oral exposures and one study with subcutaneous exposure examined anxiety in adult rodents using the light/dark box test. Venerosi *et al.* (2010) exposed mouse dams *via* oral gavage to 0 or 6 mg/kg-day chlorpyrifos during GD 15–18 and offspring were tested on PND 90. The authors reported that chlorpyrifos-treated females, but not males, showed some indication of elevated anxiety by spending more time in the tunnel connecting the light and dark compartments. Treated mice of both sexes showed no effects on several other parameters of this test, however, such as time spent in either the light or dark compartment, latency to enter the dark compartment, and risk assessment and exploratory behavior in the light compartment.

Braquenier *et al.* (2010) assessed effects on anxiety using the light/dark box test in 72-day-old female offspring of mouse dams exposed by oral gavage to 0.2, 1, or 5 mg/kg-day chlorpyrifos during GD 15–PND 14. The authors reported that treated mice were more anxious than controls, indicated by the preference to spend less time in the center of the light compartment and a decreased number of compartment switches. This anxiety was not chlorpyrifos-exposure-dependent, however, as the effects were observed with the 1 mg/kg-day dose but not the 0.2 or 5 mg/kg-day doses. In addition, the authors reported a 14% inhibition of brain AChE activity on PND 1 in the offspring of dams exposed orally to 5 mg/kg-day chlorpyrifos from GD 15 until birth.

In contrast to the oral studies, Venerosi *et al.* (2008) reported that mouse dams that were subcutaneously exposed to 3 mg/kg-day chlorpyrifos during PND 11–14 were *less* anxious than control dams, as they were more likely to enter the light compartment and do it faster than untreated controls on PPD 2 after mating on PND 60.

Mood assessed by the forced swim test: Venerosi *et al.* (2010) exposed mouse dams *via* oral gavage to 0 or 6 mg/kg-day chlorpyrifos during GD 15–18 and reported no treatment-related effects on the performance of offspring in the forced swim test, which assesses mood by monitoring swimming, struggling, and floating in water, on PND 90.

Summary: One of the three studies examining anxiety and distress calling in pups reported alterations in calling, with oral prenatal exposure of 6 mg/kg-day chlorpyrifos. By contrast, the two studies of this endpoint with subcutaneous exposure did not report effects with chlorpyrifos doses up to 3 mg/kg-day, or with chlorpyrifos-oxon at 5 mg/kg-day.

Of the four studies that assessed anxiety using the elevated plus-maze test, only the study with oral exposure spanning the prenatal and postnatal periods reported increased anxiety (Braquenier *et al.*, 2010). This effect was only studied in females and was not dose-dependent, as it was observed with exposure to 1 mg/kg-day, but not with 0.2 or 5 mg/kg-day, and inhibition of AChE activity in the brain was observed only with exposure to the highest dose. By contrast, two studies with subcutaneous exposure during prenatal periods reported no effects on anxiety in the elevated plus-maze test at doses up to 6 mg/kg-day, one study with prenatal oral exposure reported decreased anxiety in males at a dose of 3 mg/kg-day, and two studies with subcutaneous exposure during postnatal periods reported decreased anxiety in females dosed with 3 mg/kg-day or in males dosed with 1 mg/kg-day.

Consistent with the studies using the elevated plus-maze test, the two studies that assessed anxiety with the light/dark box test after oral exposures during prenatal or both prenatal and postnatal periods reported increased anxiety in females at doses of 1 or 6 mg/kg-day, but not at doses of 0.2 or 5 mg/kg-day. The specific measures of increased anxiety in this test were not the same in both studies, however. By contrast, the one study that assessed the effects of postnatal subcutaneous chlorpyrifos exposure in the light/box test reported *decreased* anxiety in females at a dose of 3 mg/kg-day. Only one study assessed mood in the forced swim test, and no effects were reported in this study after prenatal oral exposure to 6 mg/kg-day chlorpyrifos.

Overall, increased anxiety was only observed in studies with oral chlorpyrifos exposures of 6 mg/kg-day during GD 15–18 or of 1 mg/kg-day during GD 15–PND 14, although effects on anxiety with the latter exposure were not dose-dependent, as no effects were observed with a dose of 5 mg/kg-day that inhibited AChE activity in the brain.

3.3.2.3. Motor function Seventeen studies assessed the effects of chlorpyrifos exposure on motor function in rodents, using tests of locomotor activity, neuromuscular or neuromotor function, and sensorimotor reflexes. Seven studies examined oral exposures, either prenatally *via* dosing of dams during gestation (Venerosi *et al.*, 2009), with one study including postnatal exposure *via* subcutaneous injection for a subset of animals (Ricceri *et al.*, 2006), perinatally *via* gestational exposure that continued after birth through nursing (Maurissen *et al.*,

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 9. Rodent studies of effects on emotion and anxiety.

Study	Species	Animals per dose	Exposure			Outcome						Effects on AChE activity in brain
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	
Venerosi *et al.* (2009)	Mouse, CD-1	15 per sex	Oral gavage	0, 6	GD 15-18	Male, Female	PND 4, 7, 10	Anxiety *via* ultrasonic vocalization	Fewer calls, shorter call duration, longer latency to calls, and higher peak frequency of calls (PND 10 only)	Peak amplitude of calls	Increased anxiety on PND 10 with 6 mg/kg-day	NR
Venerosi *et al.* (2010)	Mouse, CD-1	10 to 12	Oral gavage	0, 6	GD 15-18	Male, Female	PND 90	Anxiety *via* the light/ dark box test	Increased percentage of time spent in the tunnel connecting the two compartments of the apparatus (females only)	Latency to enter dark compartment; time spent in dark or light compartments; risk assessment and exploratory behavior in light compartment	Increase in one marker of anxiety for females only, with 6 mg/ kg-day	NR
		5 to 7					PND 90	Mood *via* the forced swim test	None	Swimming, struggling, and floating in water	Null	
Braquenier *et al.* (2010)	Mouse, CD-1	8 to 10	Oral gavage	0, 0.2, 1, 5	GD 15 - PND 14	Female	PND 80	Anxiety *via* the elevated plus-maze	Decreased percentage of time spent in open arms; decreased proportion of entries in open arms	None	Increased anxiety with 1 mg/kg-day	14% inhibition with 5 mg/ kg-day
							PND 72	Anxiety *via* the light/ dark box test	Decreased percentage of time spent in center of light compartment; decreased number of compartment switches	Percentage of time in dark compartment	Increased anxiety with 1 mg/kg-day	
Ricceri *et al.* 2006	Mouse, CD-1	10	Oral gavage	0, 3, 6	GD 15-18	Male, Female	4 months	Anxiety *via* the elevated plus-maze	Decreased head dipping frequency (males only)	Frequency of arm entries; duration of time spent in open, central, and closed arms; frequency and duration of risk assessment posture	Decreased anxiety in males with 3 mg/kg-day	Null

(Continued)

© 2011 Informa Healthcare USA, Inc.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 9. (Continued).

Study	Species	Animals per dose	Exposure			Outcome						
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
			s.c.	0, 1, 3	PND 11-14	Male, Female	4 months	Anxiety *via* the elevated plus-maze	Increased percentage of time spent in open arms (females only)	Frequency of arm entries; duration of time spent in central and closed arms; head dipping frequency; frequency and duration of risk assessment posture	Decreased anxiety in females with 3 mg /kg-day	
Icenogle *et al.* (2004)	Rat, Sprague-Dawley	10	s.c.	0, 1, 5	GD 9-12	Male, Female	4-8 weeks	Anxiety and hyperactivity *via* the elevated plus-maze	Increased number of center crosses	Duration of time spent in open and closed arms	Increased hyperactivity with 5 mg/kg-day	NR
Laviola *et al.* (2006)	Mouse, C57/B16	4 to 6	s.c.	0, 5 (chlorpyrifos-oxon)	GD 14-16	Male, Female	PND 3, 7, 11	Anxiety *via* ultrasonic vocalization	None	Number of calls	Null	NR
Aldridge *et al.* (2005)	Rat, Sprague-Dawley	9 per sex	s.c.	0, 1	PND 1-4	Male, Female	PND 52-53	Anxiety *via* the elevated plus-maze	Increased percentage of time spent in open arms and increased number of center crosses (males only)	Duration of time spent in closed arms	Decreased anxiety and increased hyperactivity in males with 1 mg/kg-day	NR
Ricceri *et al.* (2003)	Mouse, CD-1	10	s.c.	0, 1, 3	PND 1-4	Male, Female	PND 5, 8, 11	Anxiety *via* ultrasonic vocalization	None	Number of calls	Null	~20% inhibition with 1 and 3 mg/kg-day
							PND 10	Homing	None	Latency to reach nest area; time spent in nest area; locomotor activity during test	Null	

(Continued)

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 9. (Continued).

Study	Species	Animals per dose	Exposure			Outcome					
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
Venerosi *et al.* (2008)	Mouse, CD-1	9 to 11	s.c.	0, 3	PND 11-14	Female	PPD 2 (after mating on PND 60)	Anxiety *via* the light/ dark box test	Increased proportion of dams entered light compartment; decreased latency to enter light compartment	Percentage of time in dark or light compartment; number of transitions across tunnel connecting the two compartments	Decreased anxiety in dams with 3 mg/kg-day	NR

Notes: CPF = chlorpyrifos; GD = gestational day; NR = not reported; PND = postnatal day; PPD = postpartum day; s.c. = subcutaneous injection.

2000; Braquenier *et al.*, 2010), or postnatally by direct oral exposure to pups (Moser *et al.*, 1998; Carr *et al.*, 2001; Johnson *et al.*, 2009). Eight studies investigated subcutaneous exposures administered either prenatally (Chanda and Pope, 1996; Icenogle *et al.*, 2004; Levin *et al.*, 2002; Laviola *et al.*, 2006) or postnatally (Ricceri *et al.*, 2003; Dam *et al.*, 2000; Levin *et al.*, 2001; Chakraborti *et al.*, 1993). Finally, one study examined intraperitoneal exposure (Muto *et al.*, 1992) and another study examined dermal exposure (Abou-Donia *et al.*, 2006) during gestation. These studies are described below and summarized in Table 10.

Locomotor activity: Six studies assessed locomotor activity in rodents after oral exposure to chlorpyrifos. Only one of these studies examined exposure during the prenatal period alone. Venerosi *et al.* (2009) reported no effects on locomotor activity in the 12-day-old offspring of mouse dams exposed to 6 mg/kg-day chlorpyrifos *via* oral gavage during GD 15–18.

Three studies monitored locomotor activities after oral chlorpyrifos exposures that spanned the prenatal and postnatal periods. In a study that complied with US EPA Pesticide Assessment Guidelines and Good Laboratory Practice (GLP) regulations, Maurissen *et al.* (2000) exposed rat dams *via* oral gavage to 0, 0.3, 1, or 5 mg/kg-day chlorpyrifos during GD 6– PND 10 and reported no effects on locomotor activity in offspring assessed at various time points between PND 13 and PND 60. Consistent with this study, Braquenier *et al.* (2010) reported no effects on locomotor activity in the offspring of mouse dams orally exposed to 0, 0.2, 1, or 5 mg/kg-day chlorpyrifos during GD 15– PND 14. The authors also assessed brain AChE activity on PND 1 and reported statistically significant inhibition of brain AChE activity only in the offspring of dams exposed to the highest chlorpyrifos dose (5 mg/kg-day).

Ricceri *et al.* (2006) examined motor activity in 70-day-old offspring of mouse dams treated with 0, 3, or 6 mg/kg-day chlorpyrifos by oral gavage during GD 15–18 and then subsequently treated with subcutaneous injection of 0, 1, or 3 mg/kg-day chlorpyrifos during PND 11–14. The authors reported increased motor activity in animals prenatally treated with 6, but not 3 mg/kg-day chlorpyrifos and postnatally treated with either vehicle or 1, but not 3 mg/kg-day chlorpyrifos compared to control offspring of dams that received only vehicle. These results demonstrate a lack of an exposure-response relationship. The authors reported a three-way interaction between prenatal treatment, postnatal treatment, and five-minute activity-monitoring blocks, however. In addition, the authors reported a hyperactivating effect of chlorpyrifos exposure in the first of four five-minute sessions of the motor activity test in mice with postnatal exposure to 3 mg/kg-day chlorpyrifos, but this was limited to the offspring of dams that received either vehicle or 3 mg/kg-day chlorpyrifos, but not 6 mg/kg-day chlorpyrifos, demonstrating transient effects with no exposure-response relationship. In contrast to Braquenier *et al.* (2010),

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 10. Rodent studies of effects on sensorimotor reflexes and motor activity.

			Exposure			Outcome						
Study	Species	Animals per dose	Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
Venerosi *et al.* (2009)	Mouse, CD-1	13 to 18 per sex	Oral gavage	0, 6	GD 15-18	Male, Female	PND 12	Spontaneous motor behavior	Decreased frequency and duration of pivoting; increased duration of immobility	Frequency and duration of crossing (locomotor activity), head movement, wall climbing, and grooming	Decrease in three markers of spontaneous motor behavior with 6 mg/kg-day	NR
							PND 3, 6, 9, 12, 15	Sensorimotor maturation	None	Righting reflex; grasping reflex; cliff aversion; forelimb and hind limb reflexes; pole grasping	Null	
Ricceri *et al.* (2006)	Mouse, CD-1	10	Oral gavage	0, 3, 6	GD 15-18	Male	PND 70	Locomotor activity *via* open field test	Increased crossing frequency	Crossing duration; rearing; wall rearing; self-grooming	Increase in one marker of locomotor activity with prenatal exposure to 6 mg/kg-day followed by postnatal exposure to 0 or 1 mg/kg-day, and with prenatal exposure to 0 or 3 mg/kg-day followed by postnatal exposure to 3 mg/kg-day	Null
			s.c.	0, 1, 3	PND 11-14							
Maurissen *et al.* (2000)	Rat, Sprague-Dawley	20 per sex	Oral gavage	0, 0.3, 1, 5	GD 6 - PND 10	Male, Female	PND 13, 17, 21, 60	Locomotor activity	None	Locomotor activity	Null	NR
							PND 22, 61	Auditory reflexes *via* acoustic startle test	None	Auditory reflexes	Null	
Braquenier *et al.* (2010)	Mouse, CD-1	8 to 10	Oral gavage	0, 0.2, 1, 5	GD 15 - PND 14	Female	> PND 21 for 8 days	Locomotor activity	None	Locomotor activity	Null	14% inhibition with 5 mg/kg-day
Moser *et al.* (1998)	Rat, Long Evans hooded	5 per sex	Oral gavage	0, 5, 20	PND 17	Male, Female	3.5 or 6.5 hours post-exposure	Locomotor activity *via* Figure-8 apparatus	Decreased locomotor activity	None	Decreased locomotor activity with 20 mg/kg-day	50–90% inhibition with 5 and 20 mg/kg-day
								Sensorimotor reflexes	Righting deficits; gait changes	None	Deficits in sensorimotor reflexes with 20 mg/kg-day	

(Continued)

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 10. (Continued).

Study	Species	Animals per dose	Exposure: Route	Exposure: Concentration (mg CPF/kg b.w./day)	Exposure: Developmental period	Outcome: Sex	Outcome: Test period	Outcome: Endpoint examined	Outcome: Statistically significant results	Outcome: Null endpoints	Outcome: Implication of effects	Outcome: Effects on AChE activity in brain
Carr *et al.* (2001)	Rat, Sprague-Dawley	10	Oral gavage	Low: 3 (PND 1–21)	PND 1–21 (every other day)	Male, Female	PND 10, 12, 14, 16, 18, 20, 25, 30	Locomotor activity *via* open field test	Decreased locomotor activity (PND 25 and 30 only; medium and high dose groups only)	None	Decreased locomotor activity with 6 and 12 mg/kg-day	17–70% inhibition with 3, 6, and 12 mg/kg-day
				Medium: 3 (PND 1–5), 6 (PND 7–21)								
				High: 3 (PND 1–5), 6 (PND 7–13), 12 (PND 15–21)								
Johnson *et al.* (2009)	Rat, Sprague-Dawley	9 to 14 per sex	Oral gavage	Low: 1 (PND 1–5), 1 (PND 6–13), 1 (PND 14–20)	PND 1–20	Male, Female	NR	Sensorimotor reflexes	None	Surface righting; free-fall righting; negative geotaxis; cliff avoidance; auditory reflexes	Null	14–53% inhibition with 1 to 6 mg/kg-day
				Medium: 1 (PND 1–5), 2 (PND 6–13), 4 (PND 14–20)								
				High: 1.5 (PND 1–5), 3 (PND 6–13), 6 (PND 14–20)								
Chanda and Pope (1996)	Rat, Sprague-Dawley	7 to 8	s.c.	25	GD 12–19	Male, Female	PND 1, 3	Sensorimotor reflexes	Decreased cliff avoidance; surface righting deficits	None	Deficits in sensorimotor reflexes with 25 mg/kg-day	60% inhibition with 25 mg/kg-day
Icenogle *et al.* (2004)	Rat, Sprague-Dawley	10 per sex	s.c.	0, 1, 5	GD 9–12	Male, Female	4–6 weeks	Locomotor activity *via* Figure-8 apparatus	Decreased locomotor activity (2 of 12 test sessions only)	None	Transient decrease in locomotor activity with 5 mg/kg-day	NR
							NR	Auditory reflexes *via* acoustic startle test	None	Auditory reflexes	Null	
Levin *et al.* (2002)	Rat, Sprague-Dawley	10 per sex	s.c.	0, 1,	GD 17–20	Male, Female	4–6 weeks	Locomotor activity *via* Figure-8 apparatus	None	Locomotor activity	Null	NR
Laviola *et al.* (2006)	Mouse, C57/B16	4 to 6	s.c.	0, 5 (chlorpyrifos-oxon)	GD 14–16	Male, Female	PND 3,7,11	Neuromuscular function *via* grasping reflex test	Increased fall angle (PND 3 only)	None	Increased maturation of grasping reflex at 5 mg/kg-day chlorpyrifos-oxon	NR
								Sensorimotor reflexes	None	Righting reflex	Null	

(Continued)

© 2011 Informa Healthcare USA, Inc.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 10. (Continued).

Study	Species	Animals per dose	Exposure			Outcome						
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
Ricceri *et al.* (2003)	Mouse, CD-1	10	s.c.	0, 1, 3	PND 1-4	Male, Female	PND 25	Locomotor activity *via* open field test	None	Locomotor activity	Null	~20% inhibition with 1 and 3 mg/kg-day
					PND 11-14				Increased locomotor activity	None	Increased locomotor activity with 1 and 3 mg/kg-day	Null
					PND 1-4	Male, Female	PND 45	Motor behavior	Decreased self-grooming	Exploring; rearing; digging	Decrease in one marker of motor behavior at 1 and 3 mg/kg-day	~20% inhibition with 1 and 3 mg/kg-day
					PND 11-14				None	Exploring; rearing; digging; self-grooming	Null	Null
Dam *et al.* (2000)	Rat, Sprague-Dawley	23 to 24 per sex	s.c.	0, 1	PND 1-4	Male, Female	PND 3, 4	Sensorimotor reflexes	Deficits in reflex righting (females only)	None	Deficit in one marker of sensorimotor reflexes in females with 1 mg/kg-day	20-60% inhibition with 1 mg/kg-day
		23 to 24 per sex	s.c.	0, 1	PND 1-4	Male, Female	PND 5, 8	Sensorimotor reflexes	Deficits in geotaxic response (females only)	None	Deficit in one marker of sensorimotor reflexes in females with 1 mg/kg-day	20-60% inhibition with 1 mg/kg-day
		5 to 7 per sex	s.c.	0, 1	PND 1-4	Male, Female	PND 21, 30	Locomotor activity *via* open field test	Decreased locomotor activity and rearing (males only)	Grooming	Decrease in two markers of locomotor activity in males with 1 mg/kg-day	20-60% inhibition with 1 mg/kg-day
				0, 5	PND 11-14				Increased rearing (males on PND 30 only)	Locomotor activity; grooming	Increase in one marker of locomotor activity in males with 5 mg/kg-day	20-60% inhibition with 5 mg/kg-day

(Continued)

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 10. (Continued).

Study	Species	Animals per dose	Exposure			Outcome						
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
Levin *et al.* (2001)	Rat, Sprague-Dawley	10	s.c.	0, 1	PND 1-4	Male, Female	4-6 weeks	Locomotor activity *via* Figure-8	None	Locomotor activity	Null	NR
				0, 5	PND 11-14							
Chakraborti *et al.* (1993)	Rat, Sprague-Dawley	10	s.c.	0, 40	PND 7-10	Male, Female	2, 4, 6, 8 weeks post-exposure	Locomotor activity *via* open field test	None	Locomotor activity	Null	55-60% inhibition with 40 mg/kg-day
Muto *et al.* (1992)	Rat, species unspecified	12 to 20	i.p.	0, 0.03, 0.1, 0.3 (as Dursban)	GD 0-7	Not specified	PND 16	Neuromotor function *via* the rotorod test	Increased number of falls	None	Decreased neuromotor function with 0.03 to 0.3 mg/kg-day	
					GD 7-21			Neuromotor function *via* the rotorod test	Increased number of falls	None	Decreased neuromotor function with 0.3 mg/kg-day	
					PND 3, 10, or 12			Neuromotor function *via* the rotorod test	Increased number of falls	None	Decreased neuromotor function with 0.1 and 0.3 mg/kg-day	NR
								General motor behavior	None	General motor behavior	Null	
								Neuro muscular function *via* inclined plane test	None	Neuro muscular function	Null	
Abou-Donia *et al.* (2006)	Rat, Sprague-Dawley	10	Dermal	0, 1	GD 4-20	Male, Female	PND 90	Motor coordination/balance *via* beam-walking test	None	Beam-walking score	Null	
								Neuro muscular function *via* inclined plane test	Decreased incline angle (females only)	None	Deficits in one marker of neuromuscular function in females with 1 mg/kg-day	25% increase in females with 1 mg/kg-day
								Neuro muscular function	Decreased forepaw grip time	None	Deficits in one marker of neuromuscular function with 1 mg/kg-day	

Notes: CPF = chlorpyrifos; GD = gestational day; NR = not reported; PND = postnatal day; s.c. = subcutaneous injection.

© 2011 Informa Healthcare USA, Inc.

Ricceri *et al.* (2006) reported no chlorpyrifos-associated effects on brain AChE activity even at a higher prenatal exposure concentration (6 mg/kg-day).

Two studies examined locomotor activity after oral exposure to chlorpyrifos during postnatal periods. Moser *et al.* (1998) reported no effects on motor activity in rats exposed to 5 mg/kg chlorpyrifos *via* oral gavage on PND 17 and tested a few hours post-exposure, although 50–60% inhibition of brain AChE was observed with this exposure. A higher concentration of chlorpyrifos (20 mg/kg) in this study produced both brain AChE inhibition (70–90%) and decreased motor activity. Similarly, Carr *et al.* (2001) reported decreased locomotor activity in 25- and 30-day-old rats exposed to 6 or 12 mg/kg-day chlorpyrifos *via* oral gavage during PND 1–21, but not with exposure to 3 mg/kg-day. Dose-dependent inhibition of brain AChE activity, ranging from 17–70%, was observed at all exposure concentrations used in this study.

Six studies examined effects of subcutaneous exposure to chlorpyrifos on locomotor activity, two of which involved exposures during prenatal periods. Icenogle *et al.* (2004) reported a decrease in locomotor activity in offspring of rat dams exposed during GD 9–12 *via* subcutaneous injection to chlorpyrifos at a dose of 5 mg/kg-day, but not 1 mg/kg-day, when tested during 4–6 weeks of age. By contrast, Levin *et al.* (2002) reported no change in locomotor activity in the offspring of rat dams exposed on GD 17–20 to the same doses of chlorpyrifos by the same route and tested at the same age.

Four studies assessed the effects of subcutaneous exposure to chlorpyrifos during postnatal periods. Ricceri *et al.* (2003) reported an increase in locomotor activity in 25-day-old mice after subcutaneous exposure to 1 or 3 mg/kg-day chlorpyrifos during PND 11–14 but not during PND 1–4. Reduced self-grooming was observed after exposure to both doses during PND 1–4, but not during PND 11–14. The authors also assessed brain AChE activity after exposure but found contrasting results to the assessment of locomotor activity, in that an inhibitory effect on AChE of approximately 20% was observed after exposure to either 1 or 3 mg/kg-day chlorpyrifos during PND 1–4, but not after exposure during PND 11–14 or PND 32–35.

Dam *et al.* (2000) exposed rat pups *via* subcutaneous injection to 0 or 1 mg/kg-day chlorpyrifos during PND 1–4 or to 0 or 5 mg/kg-day chlorpyrifos during PND 11–14 and assessed locomotor activity on PND21 and PND30. The authors reported decreased locomotor activity and decreased rearing in males exposed to 1 mg/kg-day chlorpyrifos during PND 1–4 and no effects on locomotor activity in rats of either sex with exposure to 5 mg/kg-day chlorpyrifos on PND 11–14, although increased rearing was observed in males on PND 30. There were no alterations in grooming for either dose group. Dam *et al.* (2000) also exposed rats to 1 mg/kg chlorpyrifos on PND 1 or 5 mg/kg on PND 11 and assessed brain AChE activity two and four hours post-exposure. AChE inhibition of 20–60% was observed two hours after both exposures, with greater effects (60% inhibition) for males exposed to 1 mg/kg chlorpyrifos on PND 1, but the effects either diminished or disappeared by four hours post-exposure.

In contrast to the study by Dam *et al.* (2000), a subsequent study from the same research group reported no effects on motor activity in 4- to 6-week-old rats exposed to either 1 mg/kg-day chlorpyrifos during PND 1–4 or to 5 mg/kg-day during PND 11–14 *via* subcutaneous injection (Levin *et al.*, 2001). Consistent with this study, rat pups exposed to a much higher concentration of chlorpyrifos (40 mg/kg-day) *via* subcutaneous injection during PND 7–10 showed no changes in motor activity during an 8-week follow-up period, although brain AChE activity was inhibited by 55–60% four days after exposure cessation and by 20–32% two weeks after exposure cessation (Chakraborti *et al.*, 1993).

Neuromuscular and neuromotor function: Three studies examined effects on neuromuscular and/or neuromotor function, using either intraperitoneal injection or dermal application of chlorpyrifos, or subcutaneous injection of chlorpyrifos-oxon. Muto *et al.* (1992) assessed neuromotor function in the rotorod test with the 16-day-old offspring of rat dams exposed to 0 (saline), 0.03, 0.1, or 0.3 mg/kg-day chlorpyrifos as Dursban pesticide (1% chlorpyrifos, 6% xylene, 93% water, according to the authors) *via* intraperitoneal injection during GD 0–7 or GD 7–21. The authors reported deficits in neuromotor function, as evidenced by an increased number of falls in the rotorod test, in offspring exposed to all doses during GD 0–7 and only the highest dose during GD 7–21. This study also examined the effects of direct postnatal exposure of rat pups to 0, 0.1, or 0.3 mg/kg-day chlorpyrifos (as Dursban) *via* intraperitoneal injection on PND 3, 10, or 12. Rats exposed to the higher dose at each of the three time points had an increased number of falls in the rotorod test on PND 16, as did rats in the lower dose group, but only when exposed on PND 12, but not on PND 3 or 10. Postnatal exposure to either dose produced no effects on general motor behavior or in the incline plane test of neuromuscular function.

Abou-Donia *et al.* (2006) exposed rat dams to 1 mg/kg-day chlorpyrifos during GD 4–20 by dermal application and reported that motor coordination and balance were not affected in the beam walking test in offspring of treated dams compared to controls on PND 90. The authors also reported that females, but not males, showed deficits in neuromuscular function in the incline plane test, however, and both sexes showed deficits in forepaw grip time, another test of neuromuscular function. In addition, the authors reported a 25% *increase* in brain AChE activity on PND 90 in treated females relative to controls. Given the expected recovery of AChE activity after exposure to chlorpyrifos from the synthesis of new AChE molecules, it is expected that any inhibition of AChE activity would have been observed shortly after this exposure, so the observed increase in activity 90 days post-exposure is likely unrelated to chlorpyrifos exposure.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Laviola *et al.* (2006) exposed mouse dams to 0 or 5 mg/kg-day chlorpyrifos-oxon *via* subcutaneous injection during GD 14–16 and assessed neuromuscular function in the grasping reflex tests in offspring on PND 3, 7, and 11. The authors reported that treated mice showed an increase in fall angle (*i.e.*, they held on longer when the supporting surface was tilted) compared to controls, but only on PND 3.

Sensorimotor reflexes: Four studies examined the effects of oral chlorpyrifos exposure on measures of sensorimotor reflexes. Venerosi *et al.* (2009) reported no effects on sensorimotor maturation (grasping, righting, and cliff avoidance) in offspring of mouse dams exposed *via* oral gavage to 6 mg/kg-day chlorpyrifos during GD 15–18 when assessed within the first two weeks of life. The authors also reported that chlorpyrifos treatment resulted in shorter and less frequent pivoting behavior and increased immobility, but no changes in other spontaneous motor behavior such as crossing, head moving, wall climbing, or grooming.

Maurissen *et al.* (2000) reported no effect on auditory reflexes (as assessed using the acoustic startle test) of 22- and 61-day-old offspring of rat dams orally exposed to 0.3, 1, or 5 mg/kg-day chlorpyrifos beginning on GD 6 and continuing through PND 10.

Johnson *et al.* (2009) orally exposed rat pups to one of three treatments: 1 mg/kg-day chlorpyrifos during PND 1–20 (low exposure); 1 mg/kg-day chlorpyrifos during PND 1–5, 2 mg/kg-day chlorpyrifos during PND 6–13, and 4 mg/kg-day chlorpyrifos during PND 7–20 (medium exposure); or 1.5 mg/kg-day chlorpyrifos during PND 1–5, 3 mg/kg-day chlorpyrifos during PND 6–13, and 6 mg/kg-day chlorpyrifos during PND 7–20 (high exposure). The authors reported no effects on sensorimotor reflexes (surface and free-fall righting, negative geotaxis, cliff avoidance, and auditory reflexes) with any of the three treatments. They also reported a dose-dependent (14–53%) inhibition of AChE activity in the brain immediately after exposure in each treatment group. Approximately 20% inhibition of AChE activity persisted for up to 20 days post-exposure in the medium and high exposure groups.

Moser *et al.* (1998) treated 17-day-old rats with 0, 5, or 20 mg/kg chlorpyrifos *via* oral gavage and reported that exposure to 20 mg/kg was associated with altered sensorimotor reflexes compared to controls when animals were tested within a few hours of exposure. Further, the authors found that exposure to both doses of chlorpyrifos was associated with AChE inhibition in the brain (50–60% inhibition for the 5 mg/kg dose group and 70–90% inhibition for the 20 mg/kg dose group).

Four studies examined effects on sensorimotor reflexes with subcutaneous exposure to chlorpyrifos. Icenogle *et al.* (2004) reported no effects on auditory reflexes in the offspring of rats subcutaneously exposed to 1 or 5 mg/kg-day chlorpyrifos during GD 9–12. In a study with a much higher exposure, Chanda and Pope (1996) reported deficits in sensorimotor reflexes on PND 1 and PND 3 in the offspring of rat dams exposed subcutaneously to 25 mg/kg-day chlorpyrifos during GD 12–19. The effects appeared to decline rapidly with age, as they were decreased on PND 3 *vs.* PND 1. A 60% inhibition of brain AChE activity was observed in treated animals on GD 20. One study used prenatal subcutaneous exposure to chlorpyrifos-oxon instead of chlorpyrifos. Laviola *et al.* (2006) reported that offspring of mouse dams exposed to 5 mg/kg-day chlorpyrifos-oxon via subcutaneous injection during GD 14–16 were not affected in terms of righting reflex on PND 3, 7, or 11.

Dam *et al.* (2000) examined effects on sensorimotor reflexes in rat pups exposed *via* subcutaneous injection to 0 or 1 mg/kg-day chlorpyrifos during PND 1–4 or to 0 or 5 mg/kg-day chlorpyrifos during PND 11–14. The authors reported that female pups exposed to 1 mg/kg-day during PND 1–4 showed deficits in reflex righting, which was assessed on PND 3–4, and in geotaxic response, which was tested on PND 5–8. Brain AChE activity was transiently inhibited by 20–60% with exposure to 1 mg/kg-day chlorpyrifos on PND 1 or by 20–30% with exposure to 5 mg/kg-day chlorpyrifos on PND 11 in this study, with approximately three-fold greater inhibition in males compared to females.

Summary: Three of four studies that examined oral chlorpyrifos exposures that began during gestation reported no effects on locomotor activity at doses up to 6 mg/kg-day. One study reported increased locomotor activity without inhibition of brain AChE activity after prenatal oral exposure to 6, but not 3, mg/kg-day chlorpyrifos and postnatal subcutaneous exposure to either vehicle or 1, but not 3, mg/kg-day chlorpyrifos, demonstrating a lack of an exposure-response relationship. The two studies that examined postnatal oral exposures reported decreased locomotor activity at doses of 6 mg/kg-day and higher, and inhibition of brain AChE activity with these doses. One study with prenatal subcutaneous exposure reported decreased locomotor activity with 5 mg/kg-day, whereas the other study, later in gestation, reported no effects with the same dose. Results were largely null in the four studies that used subcutaneous postnatal exposures, although one study reported increased locomotor activity with exposure during PND 11–14 but not PND 1–4 with 1 and 3 mg/kg-day chlorpyrifos, and another reported decreased locomotor activity with exposure to 1 mg/kg-day during PND 1–4 but not with exposure to 5 mg/kg-day during PND 11–14. Brain AChE activity was inhibited with all the exposures in which these effects were observed.

Two studies reported deficits in neuromuscular and neuromotor function with chlorpyrifos exposure. One of these studies reported effects with intraperitoneal exposure to chlorpyrifos (as Dursban) at concentrations as low as 0.03 mg/kg-day prenatally and 0.1 mg/kg-day postnatally, but it is possible that the effects were attributable to xylene, which was contained in the Dursban mixture and was not controlled for in the study. The other study reported deficits in neuromuscular function with dermal exposure to 1 mg/kg-day chlorpyrifos. A third

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

study reported effects on one parameter of neuromuscular function with subcutaneous exposure to 5 mg/kg-day chlorpyrifos-oxon.

Two of the four studies of sensorimotor reflexes that used oral chlorpyrifos exposures reported alterations in measures of this endpoint with a prenatal dose of 6 mg/kg-day and postnatal doses ≥ 20 mg/kg-day. Inhibition of brain AChE activity was observed with postnatal doses of ≥ 5 mg/kg-day in one oral exposure study and with doses as low as 1 mg/kg-day in another oral study. Two of four studies examining sensorimotor reflexes with subcutaneous chlorpyrifos exposures reported deficits in measures of this endpoint with a prenatal dose of 25 mg/kg-day and a postnatal dose of 1 mg/kg-day, and inhibition of AChE activity in the brain was also observed with these effects.

Overall, studies with postnatal exposures often indicated that effects on motor function are larger at a younger age, require higher exposure concentrations to produce effects with advancing age, and tend to be transient, as they usually persisted from a few hours to several days. In almost every study that examined AChE activity and reported effects on motor function, inhibition of brain AChE was observed at the same chlorpyrifos doses as those associated with the effects on motor function.

3.3.2.4. Cognitive function Ten studies examined the effects of chlorpyrifos exposure on cognitive function in rodents *via* one or more tests that assessed exploratory behavior, learning, or memory. Two of these studies investigated oral exposures, either perinatally *via* gestational exposure that continued after birth through nursing (Maurissen *et al.*, 2000) or postnatally (Johnson *et al.*, 2009). Eight studies examined subcutaneous exposures administrated either prenatally (Icenogle *et al.*, 2004; Levin *et al.*, 2002; Haviland *et al.*, 2010) or postnatally (Aldridge *et al.*, 2005a; Levin *et al.*, 2001; Ricceri *et al.*, 2003; Venerosi *et al.*, 2008; Jett *et al.*, 2001). These studies are described below and summarized in Table 11.

Learning and memory assessed by the T-maze test: One study examined effects on learning and memory in the T-maze test with oral exposure to chlorpyrifos. Maurissen *et al.* (2000) exposed rat dams *via* oral gavage to 0, 0.3, 1, or 5 mg/kg-day chlorpyrifos during GD 6– PND 10, and offspring were assessed in the T-maze during PND 22–24 or PND 61–91. The authors reported no treatment-related effects on learning and memory.

Two studies assessed rodents using the T-maze test after prenatal subcutaneous exposure to chlorpyrifos. Icenogle *et al.* (2004) reported transient effects (only in the first of five observation sessions) on exploratory behavior, as reflected by shorter spontaneous alternation latency (*i.e.*, hyperactivity), in 4- to 8-week-old offspring of rat dams exposed to 1 or 5 mg/kg-day chlorpyrifos *via* subcutaneous injection during GD 9–12. Similarly, Levin *et al.* (2002) reported a transient decrease in alternation latency that resolved with repeated trials of the test in the offspring of rat dams exposed subcutaneously to 1 or 5 mg/kg-day chlorpyrifos during GD 17–20. Alternation between T-maze arms, as a measure of exploratory behavior, was not associated with chlorpyrifos exposure in this study.

One study examined the effects of subcutaneous exposure to chlorpyrifos during the postnatal period on performance in the T-maze test. Levin *et al.* (2001) exposed rat pups subcutaneously to 1 mg/kg-day chlorpyrifos during PND 1–4 or to 5 mg/kg-day chlorpyrifos during PND 11–14 and performance in the T-maze was assessed when the rats were four to six weeks of age. The authors reported no treatment-related effects on alternation frequency but found that males responded with a longer alternation latency compared to controls in the third of 12 test sessions when exposed to 5 mg/kg-day during PND 11–14.

Memory assessed by the radial arm maze: One study assessed memory using the radial arm maze after oral exposure to chlorpyrifos. Johnson *et al.* (2009) reported no effects on working memory errors in female rats orally exposed during PND 1–21 to chlorpyrifos concentrations as high as 6 mg/kg-day in any of the four weeks of testing (four days/week) that began on PND 36, but observed fewer working memory errors in female rats exposed to the "medium" concentration range of 1–4 mg/kg-day chlorpyrifos during PND 1–21 (but not to the "low" exposure of 1 mg/kg-day or to the "high" exposure range of 1.5–6 mg/kg-day) when all the days in the four weeks were averaged. By contrast, male rats in the high exposure group made more working memory errors during all weeks and those in the low and medium exposure groups made more working memory errors during the fourth week of testing. With the medium and high chlorpyrifos exposures, female rats made fewer reference memory errors relative to controls and males made more such errors during week two of testing. The authors also reported a dose-dependent (14–53%) inhibition of AChE activity in the brain immediately after exposure in each treatment group. Inhibition of AChE activity (approximately 20%) persisted for up to 20 days post-exposure in the medium and high exposure groups.

Three studies examined the effects of subcutaneous exposure to chlorpyrifos during the postnatal period on performance in the radial arm maze. Icenogle *et al.* (2004) assessed 8- to 13-week-old offspring of rat dams exposed to 1 or 5 mg/kg-day chlorpyrifos *via* subcutaneous injection during GD 9–12. The authors reported an indication of increased working and reference memory errors with the 5 mg/kg-day exposure, although in most sessions (12 of 18), there were no differences between treated rats and controls. Levin *et al.* (2002) exposed rat dams subcutaneously to 1 or 5 mg/kg-day chlorpyrifos during GD 17–20 and reported no treatment-related effects on error frequency in their 8- to 13-week-old offspring when each of the 18 sessions were considered, but the mean number of both working and reference memory errors for all 18 sessions taken together was higher in females of the 1 mg/kg-day dose group compared to controls. These effects were not observed in females of the higher dose group or in males at

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 11. Rodent studies of effects on cognitive function.

Study	Species	Animals per dose	Exposure			Outcome						
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
Maurissen *et al.* (2000)	Rat, Sprague–Dawley	8 per sex	Oral gavage	0, 0.3, 1, 5	GD 6 – PND 10	Male, Female	PND 22–24, PND 61–91	Learning and memory *via* the T-maze test	None	Spatial delayed alternation (learning acquisition; short-term retention; non-mnemonic factors)	Null	NR
Johnson *et al.* (2009)	Rat, Sprague–Dawley	9 to 14 per sex	Oral gavage	Low: 1 (PND 1–5), 1 (PND 6–13), 1 (PND 14–20)	PND 1–20	Male, Female	PND 36–60	Memory *via* the 12-arm radial maze	Increased number of working memory errors (males only; one of four testing weeks)	Reference memory errors	Transient deficit in one marker of memory in males with 1 mg/kg-day	14–53% inhibition with 1 to 6 mg/kg- day
				Medium: 1 (PND 1–5), 2 (PND 6–13), 4 (PND 14–20)					Fewer working and reference memory errors (females only); increased number of working and reference memory errors (males only; one of four testing weeks)	None	Transient memory deficits in males and memory gains in females and with 1 to 4 mg/kg-day	
				High: 1.5 (PND 1–5), 3 (PND 6–13), 6 (PND 14–20)					Increased number of working memory errors (males only); fewer reference memory errors (females only); increased number of reference memory errors (males only; one of four testing weeks)	None	Memory deficits in males and memory gains in females with 1.5 to 6 mg/kg-day	
Icenogle *et al.* (2004)	Rat, Sprague–Dawley	10 per sex	s.c.	0, 1, 5	GD 9–12	Male, Female	4–8 weeks	Exploratory behavior *via* the T-maze test	Shorter spontaneous alternation latency (first of five sessions only)	None	Increased hyperactivity with 1.5 mg/kg-day	NR
							8–13 weeks	Memory *via* the 16-arm radial maze	Increased working memory errors (two of six sessions); increased reference memory errors (one of six sessions)	None	Transient memory deficits with 5 mg/kg-day	

(Continued)

© 2011 Informa Healthcare USA, Inc.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 11. (Continued).

Study	Species	Animals per dose	Exposure			Outcome						
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
							4–8 weeks	Habituation *via* Figure-8 apparatus	Faster linear trend of habituation	None	Faster habituation with 5 mg/kg-day	
Levin *et al.* (2002)	Rat, Sprague–Dawley	10 per sex	s.c.	0, 1, 5	GD 17–20	Male, Female	4–6 weeks	Exploratory behavior *via* the T-maze test	Decrease in alternation latency (one or two of five test trials)	Alternation between maze arms	Transiently increased hyperactivity with 1 and 5 mg/kg-day	NR
							8–13 weeks	Memory *via* the 16-arm radial maze	Increased mean number of working and reference memory errors across 18 sessions (females only)	Memory error frequency	Memory deficits in females with 1 mg/kg-day	
							4–6 weeks	Habituation *via* Figure-8 apparatus	Slower linear trend of habituation (females only)	Habituation time	Slower habituation in females with 1 and 5 mg/kg-day	
Haviland *et al.* (2010)	Mouse, Swiss Webster	8	s.c.	0, 1, 5	GD 17–20	Male, Female	PND 60–81	Learning *via* the foraging maze	Decreased food recognition learning (two or three of nine sessions; females only); increased food recognition learning (four of nine sessions; males only); decreased food position learning (two of nine sessions; females only); increased food position learning (two of nine sessions; males only)	Foraging rate	Transiently decreased foraging ability in females with 1 and 5 mg/kg-day; transiently increased foraging ability in males with 1 and 5 mg/kg-day	NR
								Memory *via* the 8-arm radial maze	Increased reference memory errors (two of nine sessions); decreased reference memory errors (two of nine sessions; females only)	None	Transient memory deficits in males with 1 or 5 mg/kg-day; transient memory deficits in females with 1 mg/kg-day; transient memory gains in females with 1 or 5 mg/kg-day	

(Continued)

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 11. (Continued).

Study	Species	Animals per dose	Exposure			Outcome						
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
Aldridge *et al.* (2005)	Rat, Sprague-Dawley	9 per sex	s.c.	0, 1	PND 1-4	Male, Female	PND 64-99	16-arm radial maze	Increased mean number of reference memory errors across 18 sessions (males only)	Working memory errors	Memory deficits in males with 1 mg/kg-day	NR
Levin *et al.* (2001)	Rat, Sprague-Dawley	10	s.c.	0, 1	PND 1-4	Male, Female	4-6 weeks	Exploratory behavior *via* the T-maze test	None	Alternation frequency; alternation latency	Null	NR
				0, 5	PND 11-14				Increase in alternation latency (one of 12 sessions; males only)	Alternation frequency	Transiently decreased hyperactivity in males with 5mg/kg-day	
				0, 1	PND 1-4		8-13 weeks	Memory *via* the 16-arm radial maze	Increased number of working and reference memory errors (over first three of 18 sessions; males only); increased mean number of working and reference memory errors (females)	None	Memory deficits with 1 mg/kg-day	
				0, 5	PND 11-14				None	Working and reference memory errors	Null	
				0, 1	PND 1-4		4-6 weeks	Habituation *via* Figure-8 apparatus	None	Habituation	Null	
				0, 5	PND 11-14				Slower linear trend of habituation	None	Slower habituation with 5 mg/kg-day	
Ricceri *et al.* (2003)	Mouse, CD-1	10	s.c.	0, 3	PND 1-4	Male, Female	PND 35-38	Novelty-seeking behavior	Increased activity rate in novel compartment (one of five sessions)	Novelty preference; latency to enter novel compartment; activity rate in familiar compartment	Transient alteration in one marker of novelty-seeking behavior with 3 mg/kg-day	~20% inhibition with 1 and 3 mg/kg-day
					PND 11-14				Increased activity rate in novel compartment (two of five sessions)	Novelty preference; latency to enter novel compartment; activity rate in familiar compartment		Null

(Continued)

© 2011 Informa Healthcare USA, Inc.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Table 11. (Continued).

Study	Species	Animals per dose	Exposure			Outcome						
			Route	Concentration (mg CPF/kg b.w./day)	Developmental period	Sex	Test period	Endpoint examined	Statistically significant results	Null endpoints	Implication of effects	Effects on AChE activity in brain
				0, 1, 3	PND 1-4	Male, Female	PND 60-81	Passive avoidance learning	None	Memory	Null	~20% inhibition with 1 and 3 mg/kg-day
					PND 11-14							Null
Venerosi *et al.* (2008)	Mouse, CD-1	9 to 11	s.c.	0, 3	PND 11-14	Male, Female	PND 40-45	Acclimatization	Decreased exploration (one of three sessions)	None	Transient decrease in exploration with 3 mg/kg-day	NR
Jett *et al.* (2001)	Rat, Long Evans	8 to 10 per sex	s.c.	0, 0.3, 7	PND 7, 11, 15 (pre weaning)	Male, Female	PND 24-28	Cognitive function *via* the Morris swim test	Increased escape latency (two of four testing days); learning deficiency in probe test	Swimming speed	Transient learning deficits with 7 mg/kg-day	Null
		3 to 4 per sex			PND 22, 26 (post-weaning)				Increased escape latency (two of four testing days); learning deficiency in probe test	Swimming speed	Transient learning deficits with 0.3 and 7 mg/kg-day	Null

Notes: CPF = chlorpyrifos; GD = gestational day; NR = not reported; PND = postnatal day; s.c. = subcutaneous injection.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

either dose. Haviland *et al.* (2010) exposed mouse dams to 0, 1, or 5 mg/kg-day chlorpyrifos *via* subcutaneous injection during GD 17–20 and assessed their offspring in the radial arm maze during PND 60–81. The authors reported an increase in reference memory errors compared to controls for treated males and females in only two of nine sessions over the 21-day testing period, but these errors were not exposure-dependent, and treated females made fewer errors compared to controls in two of the sessions.

Two studies assessed memory using the radial arm maze after subcutaneous exposure to chlorpyrifos during postnatal periods. Levin *et al.* (2001) exposed rat pups to 1 mg/kg-day chlorpyrifos *via* subcutaneous injection during PND 1–4 or to 5 mg/kg-day chlorpyrifos during PND 11–14. When assessed at eight to 13 weeks of age, males exposed to 1 mg/kg-day chlorpyrifos during PND 1–4 made more working memory errors than controls when the first three sessions were averaged, but not in any of the other 15 sessions. With this same exposure, female rats made more working memory errors only when all sessions were averaged together. A similar response was observed for reference memory errors with exposure during PND 1–4, and no effects were reported with exposure during PND 11–14. A subsequent study by the same research group (Aldridge *et al.*, 2005a) reported an association between exposure to 1 mg/kg-day chlorpyrifos *via* subcutaneous injection during PND 1–4 and increased reference memory errors in 9-week-old male rats when all 18 observation sessions of the radial arm maze were averaged together, with no effects in females.

Habituation assessed by the Figure-8 apparatus: Two studies investigated habituation with the Figure-8 apparatus after prenatal exposure to chlorpyrifos *via* subcutaneous injection. Icenogle *et al.* (2004) reported faster habituation in the Figure-8 apparatus for the 4- to 8-week-old offspring of rat dams exposed to 5 mg/kg-day chlorpyrifos *via* subcutaneous injection during GD 9–12, but not with exposure to 1 mg/kg-day. By contrast, Levin *et al.* (2002) reported no associations between chlorpyrifos exposure and motor activity level or habituation time in all 12 of the five-minute blocks of the Figure-8 apparatus test in the offspring of rat dams exposed subcutaneously to 1 or 5 mg/kg-day chlorpyrifos during GD 17–20. Exposure to both doses was associated with lower linear trends of habituation (*i.e.*, slower) *vs.* controls in females, but not males, however.

One study assessed habituation after postnatal subcutaneous exposure to chlorpyrifos. Levin *et al.* (2001) reported a decreased motor activity linear trend in the Figure-8 apparatus, an indication of slower habituation, in 4- to 6-week-old rats exposed to 5 mg/kg-day chlorpyrifos during PND 11–14, but not in rats exposed to 1 mg/kg-day chlorpyrifos during PND 1–4.

Other tests of cognitive function: One study with subcutaneous exposure during the prenatal period assessed learning in the nine-session foraging maze. Haviland *et al.* (2010) exposed mouse dams to 0, 1, or 5 mg/kg-day chlorpyrifos *via* subcutaneous injection during GD 17–20 and assessed their offspring during PND 60–81. Treated females showed decreased food recognition learning in sessions 5 and 6 with both chlorpyrifos doses and also in sessions 4, 8, and 9 with the high dose. In males, food recognition learning was *increased* in sessions 3, 4, 5, and 7 in the low, but not the high, dose group. Food position learning was decreased in females in sessions 5 and 8, only in the high dose group, and was *increased* in males in sessions 2 and 7, only in the low dose group. In addition, chlorpyrifos exposure at either dose was not associated with foraging activity.

Three studies with postnatal exposure *via* subcutaneous injection assessed various markers of cognitive function. Ricceri *et al.* (2003) reported that 60-day-old mice that received subcutaneous injections of 1 or 3 mg/kg-day chlorpyrifos during PND 1–4 or PND 11–14 did not show an effect in passive avoidance learning, a test that assesses memory *via* observing conditioned suppression of behavioral responses. In assessing the novelty-seeking behavior of mice in the high dose group on PND 35–38, the authors reported no treatment effect on novelty preference or latency to enter the novel compartment. The authors reported an increased activity rate in the novel compartment with treatment during both postnatal time periods, but this was limited to one of five sessions performed after the earlier treatment and two of five sessions performed after the later one. In addition, the authors reported approximately 20% inhibition of brain AChE in mice exposed to either 1 or 3 mg/kg-day during PND 1–4, but not during PND 11–14 or PND 32–35. In a subsequent experiment from the same laboratory, Venerosi *et al.* (2008) reported that mice exposed to 3 mg/kg-day chlorpyrifos *via* subcutaneous injection during PND 11–14 explored a new cage less than control mice on PND 40–45, although this difference in exploration was limited to the first of three sessions of a five-minute observation period (*i.e.*, the first ~ 1.7 minutes).

Jett *et al.* (2001) exposed rats to 0, 0.3, or 7 mg/kg chlorpyrifos *via* subcutaneous injection on PND 7, 11, and 15 (pre-weaning group) or on PND 22 and 26 (post-weaning group) and tested cognitive function with the Morris swim test on PND 24 through 28. The authors reported that rats exposed to 7 mg/kg chlorpyrifos in the pre-weaning group took longer to find the platform in the Morris swim test on PND 24 and PND 28 than did controls, but no effects were observed on the other testing days. Rats exposed to both chlorpyrifos doses in the post-weaning group also took longer to find the platform, but only on PND 26 and 28. The authors also administered the probe version of the Morris swim test on the last testing day (PND 28). In this test, the platform is removed and the degree of learning is determined by the amount of time spent in close proximity to the learned platform position. Pre-weaning treatment with 7 mg/kg and post-weaning treatment with both chlorpyrifos doses were associated with learning deficiencies in the probe test, but the effect magnitude did not change with exposure in the post-weaning group. Swimming speed

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

was not altered by chlorpyrifos treatment in either the pre- or post-weaning group. The authors also assessed brain AChE activity within a few days of exposure (PND 28) and observed no inhibition of this activity with any chlorpyrifos treatment.

Summary: One study assessing learning and memory in the T-maze test after oral exposure to chlorpyrifos during gestation and lactation reported no effects. Each of the three studies with subcutaneous exposure that used this measure of cognitive function reported transient effects on the latency of alternation. Alternation latency was transiently decreased with prenatal exposure to 1 or 5 mg/kg-day and was transiently increased with postnatal exposure to 5 mg/kg-day.

In studies that assessed memory in the radial arm maze, any treatment-related effects were usually transient, as they were observed in only a few sessions of this test and often only reached statistical significance when all test sessions were averaged together. One study with postnatal oral exposure reported fewer working and reference memory errors in female rats and more such errors in male rats in certain sessions of the test at doses ranging from 1 to 6 mg/kg-day chlorpyrifos, and AChE activity in the brain was also inhibited in each dose group. All three studies of prenatal subcutaneous exposure reported increases in working and reference memory errors in some test sessions with chlorpyrifos exposures of 1 or 5 mg/kg-day, but these did not always show an exposure-response relationship, and one of the studies reported fewer working memory errors in females in some test sessions. Both studies with subcutaneous chlorpyrifos exposure during postnatal time periods reported increases in reference and memory errors when some or all test sessions were averaged with a dose of 1 mg/kg-day during PND 1–4, but not during PND 11–14.

Habituation was tested in the Figure-8 apparatus only after subcutaneous exposures to chlorpyrifos. Prenatal exposure during GD 9–12 was associated with faster habituation at a dose of 5 mg/kg-day in one study. By contrast, exposure to 1 or 5 mg/kg-day chlorpyrifos during GD 17–20 was associated with a trend of slower habituation in another study. One study examined postnatal chlorpyrifos exposure and reported slower habituation after exposure to 5 mg/kg-day chlorpyrifos during PND 11–14, but no effects with exposure to 1 mg/kg-day during PND 1–4.

Four studies assessed cognitive function in rodents after subcutaneous exposure to chlorpyrifos using various other tests, and effects were observed in only a few test sessions in each study. One study assessed learning in the foraging maze after prenatal exposure and reported decreased food recognition and position learning compared to controls in females in some test sessions with exposure to 1 or 5 mg/kg-day chlorpyrifos, and increased food recognition and position learning in males in certain test sessions with exposure to 1 mg/kg-day, but not 5 mg/kg-day chlorpyrifos. In the novelty seeking test, mice had an increased activity rate in the novel compartment in one or two test sessions compared to controls when treated with 3 mg/kg-day chlorpyrifos during PND 1–4 or PND 11–14, and inhibition of AChE activity in the brain was also observed after treatment with this dose during PND 1–4, but not during PND 11–14. Postnatal exposure to 3 mg/kg-day chlorpyrifos was also associated with less exploratory behavior in a new cage in one of three observation sessions. Postnatal exposure to 7 mg/kg-day chlorpyrifos prior to weaning was associated with learning difficulties in the Morris swim test, as was exposure to 0.3 or 7 mg/kg-day administered after weaning. These effects were observed only on two of the four testing days and were not associated with inhibition of AChE activity.

Overall, when chlorpyrifos-associated cognitive effects were reported in the above studies, there was no clear trend associated with the exposure route or the developmental phase during which exposure occurred. Effects were almost always transient, as they were observed only during a few sessions of each test administered, and were often observed to be in the opposite direction in the same test (*e.g.*, more memory errors *vs*. fewer memory errors compared to controls in the radial arm maze). Inhibition of AChE activity in the brain was only assessed in a few of the studies, all of which examined effects of chlorpyrifos exposures during postnatal periods. While some of these studies reported cognitive effects in conjunction with AChE inhibition, others reported effects in different cognitive tests in the absence of AChE inhibition.

3.3.3. Analysis of animal data

In the following sections, we critically examine the animal data as a whole to assess whether the evidence supports the hypothesis that chlorpyrifos is associated with neurodevelopmental effects. This evaluation considers the exposure route, the adequacy of study design, the consistency of reported outcomes and exposure-response relationships within and across studies, and the biological significance of responses.

3.3.3.1. Adequacy of study design The most rigorous studies have a sufficient number of animals, use dose levels and routes of exposure that are relevant to human exposures, and use appropriate laboratory and statistical methodologies. In our evaluation, we considered the rigor of all the animal studies. The outcome of a study has no bearing on how rigorous the study is, and null results should not be dismissed if they are obtained from a robust and well-conducted study. Regardless of their results, the most rigorous studies carry the most weight.

Studies with a sufficient number of animals carry more weight, as they have more statistical power to detect small differences in outcomes among treatment groups, such that a null finding is more likely to represent a true lack of an effect rather than a failure to detect a true difference. While many of the neurodevelopmental studies of chlorpyrifos used at least 20 animals per dose group, several used very small numbers of animals, including the study by Ricceri *et al*. (2006), which used only 10 animals per

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

dose group for examining effects on social and maternal behavior and motor function; the study by Chanda and Pope (1996), which used only 7–8 animals per dose group to examine effects on motor function; and the study by Laviola *et al.* (2006), which used only 4 to 6 animals per dose group to study effects on anxiety, neuromuscular function, and sensorimotor reflexes. Using a sufficient number of animals of both sexes is also important, as sex differences in effects are often observed with studies in rodents.

The use of relevant dose levels is also an important consideration, particularly when examining neurodevelopmental effects of chlorpyrifos which are hypothesized to occur at very low doses that are not associated with systemic toxicity and inhibition of AChE activity in the brain. Several of the studies of neurodevelopmental effects used high exposures that have been shown to cause inhibition of brain AChE activity in many other studies, and these should carry less weight than those studies that examined chlorpyrifos doses below those known to cause AChE inhibition. It should be noted that the lowest observed adverse effect level (LOAEL) of 1 mg/kg-day for AChE inhibition in the brain in the animal studies reviewed here is five orders of magnitude higher than the estimated chlorpyrifos exposures of the mothers in the Columbia and CHAMACOS cohorts.

Oral exposures to chlorpyrifos *via* dietary dosing are the most relevant to current human exposures in non-occupational settings. Inhalation is also a relevant exposure route for humans, mainly for occupational exposures since the restriction of the residential use of chlorpyrifos in 2001. The majority of studies examining neurodevelopmental effects in rodents after chlorpyrifos exposure used oral gavage or subcutaneous injection as the exposure route, and none of the studies used dietary or inhalation exposures. Subcutaneous injection is not a relevant exposure route for humans, although it is similar to inhalation or dermal routes in that it avoids the extensive first-pass detoxifying metabolism that occurs in the liver after oral exposure and, thus, could produce higher systemic doses of chlorpyrifos compared to oral exposures. Because the oral exposure route is more relevant to humans, studies with oral exposures are more applicable to determining risks to human health and are given more weight than studies using exposure *via* injection methods.

Three studies that examined oral exposures to chlorpyrifos and used numbers of animals in the high range across studies (10–18 per sex, per dose) reported mostly null effects. Maurissen *et al.* (2000) reported no effects on locomotor activity or auditory reflexes in groups of 20 male and 20 female rats exposed orally to chlorpyrifos at concentrations ranging from 0.3–5 mg/kg-day during the entire perinatal period (GD 6 through PND 10). This is the only rodent study of neurodevelopmental effects in our evaluation that was conducted under US EPA Pesticide Assessment Guidelines and GLP regulations, which means there were specific control measures taken to help ensure the consistency and reliability of the results. Johnson *et al.* (2009) exposed rat pups (9–14 per sex, per dose) *via* oral gavage to chlorpyrifos doses ranging from 1–6 mg/kg-day during various postnatal periods and reported no effects on sensorimotor reflexes, *decreases* in working and reference memory errors in females in the radial arm maze, and increases in working and reference memory errors in males. The effects on memory were observed when doses spanned the range of 1–6 mg/kg-day, but not when the dose was consistently 1 mg/kg-day, and inhibition of AChE activity in the brain was observed after treatment with all doses. Venerosi *et al.* (2009) reported no effects on anxiety, sensorimotor maturation, or locomotor activity in groups of mice (13–18 per sex, per dose) orally exposed to the relatively high dose of 6 mg/kg-day chlorpyrifos during GD 15–18.

Two studies that also used a relatively high number of animals, but the less relevant exposure route of subcutaneous injection, reported some effects on certain neurodevelopmental markers, but most were also observed in conjunction with inhibition of AChE activity. Ricceri *et al.* (2003) subcutaneously exposed mouse pups (7–17 per sex, per dose) to 1 or 3 mg/kg-day chlorpyrifos during PND 1–4 or PND 11–14 and reported increases in some markers of socioagonistic behavior with both doses during both exposure periods, increased locomotor activity with both doses during the latter exposure period, but not the former, and an increase in one aspect of novelty-seeking behavior with the higher dose during both exposure periods. Inhibition of AChE activity was observed with both doses, but only in mice exposed during the earlier period. Dam *et al.* (2000) exposed 23–24 rat pups per sex *via* subcutaneous injection to 1 mg/kg-day chlorpyrifos during PND 1–4 and reported deficits in reflex righting and geotaxic response in females only, as well as inhibition of AChE activity in the brain.

Together, studies with the most weight report largely null effects across various neurodevelopmental tests. The studies that do report treatment-related effects often report inhibition of AChE activity in the brain at the same doses associated with the neurodevelopmental effects, suggesting that the effects occur *via* inhibition of AChE activity.

3.3.3.2. Consistency of outcomes within and across studies Many of the studies measured several different endpoints, with some of these being assessed at many timepoints, leading to many statistical comparisons. As noted above, when multiple endpoints are examined in the same study, the probability of finding apparent effects when there are none increases, and it becomes more likely that several results will be statistically significant by chance. This is also true across studies, as the probability of chance findings increases as the number of studies examining the same effects increases. The remedy to this problem is to look for consistency in the impacts of exposure on each specific endpoint and endpoints expected

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

to occur *via* a similar mode of action, both within and across studies. If the majority of studies report no statistically significant results for an effect, but a few studies do report effects, there is a high probability that the positive results are not treatment-related but are in fact due to chance or another factor (Goodman *et al.*, 2010).

The studies examining social and maternal behavior reported effects on a few isolated markers of these outcomes, although the same specific markers were not affected in the same way across studies. Effects on socio-agonistic behavior were observed with both oral prenatal exposure of 6 mg/kg-day chlorpyrifos and subcutaneous postnatal exposures of 1 or 3 mg/kg-day, and those on maternal behavior were observed with subcutaneous postnatal exposure of 1 or 3 mg/kg-day chlorpyrifos but not with oral prenatal exposures up to 6 mg/kg-day.

The studies examining anxiety-related outcomes did not report consistent results across studies. Only one of three studies examining anxiety and distress calling in pups reported alterations in calling, with oral prenatal exposure to 6 mg/kg-day chlorpyrifos, whereas the other two reported no effects with subcutaneous exposure with chlorpyrifos doses up to 3 mg/kg-day, or with chlorpyrifos-oxon at 5 mg/kg-day. Of the four studies that assessed anxiety in the elevated plus-maze test, only one study with oral exposure spanning the prenatal and postnatal periods reported increased anxiety in females, and the effects were not dose-dependent. By contrast, one study with prenatal oral exposure reported decreased anxiety in males dosed with 3 mg/kg-day, two studies with prenatal subcutaneous exposure reported no effects at doses up to 6 mg/kg-day, and two studies with postnatal subcutaneous exposure reported decreased anxiety in females dosed with 3 mg/kg-day or in males dosed with 1 mg/kg-day. The two studies that assessed anxiety with the light/dark box test after oral exposures during prenatal or both prenatal and postnatal periods reported increased anxiety in females at doses of 1 or 6 mg/kg-day, but not at doses of 0.2 or 5 mg/kg-day, and the specific measures of increased anxiety were not the same in both studies. By contrast, the one study that examined postnatal subcutaneous exposure reported *decreased* anxiety in females at a dose of 3 mg/kg-day. The one study that assessed mood in the forced swim test reported no effects after prenatal oral exposure to 6 mg/kg-day chlorpyrifos.

The results of studies examining effects of chlorpyrifos on motor function were largely null, with certain effects observed in only a few studies and usually at very high doses. Three of four studies that examined oral chlorpyrifos exposures that began during gestation reported no effects on locomotor activity at doses up to 6 mg/kg-day, whereas the fourth study reported increased locomotor activity after prenatal oral exposure to 6, but not 3, mg/kg-day chlorpyrifos and postnatal subcutaneous exposure to either vehicle or 1, but not 3, mg/kg-day chlorpyrifos, demonstrating no exposure-response relationship. The two studies that examined the effects of postnatal oral exposures reported decreased locomotor activity at doses of 6 mg/kg-day and higher. One study with prenatal subcutaneous exposure reported decreased locomotor activity with 5 mg/kg-day, whereas the other study, later in gestation, reported no effects with the same dose. Of the four studies that used subcutaneous postnatal exposures, two studies reported no effects, one study reported increased locomotor activity with exposure to 1 and 3 mg/kg-day chlorpyrifos during PND 11–14, but not PND 1–4, and another reported decreased locomotor activity with exposure to 1 mg/kg-day during PND 1–4 but not with exposure to 5 mg/kg-day during PND 11–14. Of the two studies examining neuromuscular and neuromotor function associated with chlorpyrifos exposure, one reported deficits in neuromotor function in the rotorod test with intraperitoneal exposure to chlorpyrifos (as Dursban) at concentrations as low as 0.03 mg/kg-day prenatally and 0.1 mg/kg-day postnatally, but no effects on general motor behavior or in the incline plane test of neuromuscular function. It is possible that the effects observed in this study were attributable to the uncontrolled exposure to xylene in the Dursban mixture. The other study reported deficits in neuromuscular function with dermal exposure to 1 mg/kg-day chlorpyrifos, but motor coordination and balance were not affected. Two of the four studies of sensorimotor reflexes that used oral chlorpyrifos exposures reported no effects, whereas the other two reported alterations in measures of this endpoint with a prenatal dose of 6 mg/kg-day and postnatal doses ≥ 20 mg/kg-day. Two of five studies examining sensorimotor reflexes with subcutaneous chlorpyrifos exposures reported no effects, whereas three others reported deficits in measures of this endpoint with a prenatal dose of 25 mg/kg-day and postnatal doses ranging from 1 to 5 mg/kg-day.

In studies of chlorpyrifos-associated cognitive effects, there was no clear trend associated with the exposure route or the developmental phase during which exposure occurred. Effects were almost always transient, as they were observed only during a few sessions of each test administered, and they were often observed to be in the opposite direction in the same test. No effects were reported in one study of learning and memory in the T-maze test after oral exposure to chlorpyrifos during gestation and lactation, and the three studies with subcutaneous exposure reported a transient decrease in alternation latency with prenatal exposure to 1 or 5 mg/kg-day and a transient *increase* in this endpoint with postnatal exposure to 5 mg/kg-day during PND 11–14, but not during PND 1–4. In studies that assessed memory in the radial arm maze, treatment-related effects were usually observed in only a few sessions of this test and often only reached statistical significance when all test sessions were averaged together. One study with postnatal oral exposure reported *fewer* working and reference memory errors in female rats and more such errors in male rats in certain sessions of the test at doses ranging from 1 to 6 mg/kg-day chlorpyrifos. All three studies of

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

prenatal subcutaneous exposure reported increases in working and reference memory errors in some test sessions with chlorpyrifos exposures of 1 or 5 mg/kg-day, but these did not always show an exposure-response relationship, and one of the studies reported *fewer* working memory errors in females in some test sessions. Both studies with subcutaneous chlorpyrifos exposure during postnatal time periods reported increases in reference and memory errors when some or all test sessions were averaged with a dose of 1 mg/kg-day during PND 1–4, but not during PND 11–14. Three studies assessed habituation with the Figure-8 apparatus after subcutaneous exposures to chlorpyrifos. Prenatal exposure during GD 9–12 was associated with faster habituation at a dose of 5 mg/kg-day in one study, whereas exposure to 1 or 5 mg/kg-day chlorpyrifos during GD 17–20 was associated with a trend of slower habituation in another study. The third study reported slower habituation after exposure to 5 mg/kg-day chlorpyrifos during PND 11–14, but no effects with exposure to 1 mg/kg-day during PND 1–4.

Four studies examined other tests of cognitive function after subcutaneous exposures, and effects were observed in only a few test sessions in each study. One study assessed learning in the foraging maze and reported decreased food recognition and position learning in females compared to controls in some test sessions with exposure to 1 or 5 mg/kg-day chlorpyrifos, and *increased* food recognition and position learning in males in certain test sessions with exposure to 1 mg/kg-day, but not 5 mg/kg-day chlorpyrifos. In the novelty seeking test, mice had an increased activity rate in the novel compartment compared to controls in one or two test sessions when treated with 3 mg/kg-day chlorpyrifos during PND 1–4 or PND 11–14, and postnatal exposure to this same dose was also associated with less exploratory behavior in a new cage in one of three observation sessions in another study. Postnatal exposure to 7 mg/kg-day chlorpyrifos prior to weaning was associated with learning difficulties in the Morris swim test on two of four testing days, as was exposure to 0.3 or 7 mg/kg-day administered after weaning.

In addition to inconsistencies of the same effects across studies, there were also inconsistencies across related endpoints at similar doses. It is assumed that if chlorpyrifos causes adverse neurodevelopmental effects, one should see increases in adverse social and maternal behaviors, anxiety, hyperactivity (including increased locomotor activity), and adverse cognitive effects such as memory errors, but this was not always the case within or across studies. For example, although Ricceri *et al.* (2006) reported treatment-related effects on socioagonistic and maternal behavior, as well as increased locomotor activity, the authors reported either no effects or *decreased* anxiety depending on the period of exposure. Icenogle *et al.* (2004) reported cognitive deficits but also reported null effects on anxiety and auditory reflexes and *decreased* locomotor activity. Aldridge *et al.* (2005a) reported increased reference memory errors in male rats but *less* anxiety in these animals. Finally, two studies by Levin *et al.* (2001, 2002) reported slower habituation in the Figure-8 apparatus, which is indicative of increased locomotor activity, but no effects in a direct test of locomotor activity at the same doses.

Taken together, the studies assessing potential neurodevelopmental effects of chlorpyrifos in rodents indicate that it is usually a few, isolated markers of certain behaviors that were determined to be statistically significant in pair-wise comparisons with controls, but these were generally contradicted by other studies reporting no effects in similar dose ranges using similar routes of exposure, and sometimes showed effects in the other direction. In general, for each endpoint we did not find a pattern of an effect that was consistent enough over doses and time points within studies or consistent enough across studies to constitute a repeatable finding. This indicates that the reported effects may likely be due to chance, or, at the very least, need better corroboration before they can be considered compelling results.

3.3.3.3. Exposure-response It is important to keep in mind that many of the reported neurodevelopmental effects in the animal studies were observed at doses that are much lower than those at which the established neurotoxic effects of chlorpyrifos, acting through inhibition of AChE activity in the nervous system, have been reported to occur. If chlorpyrifos is a causal factor for neurodevelopmental effects at low doses, one would still expect to see a relationship between the dose and any reported effects, both within and among studies.

The studies examining the neurodevelopmental effects of chlorpyrifos in rodents do not demonstrate consistent exposure-response relationships either within or across studies. Ricceri *et al.* (2003) reported a larger effect on socioagonistic behavior in mice after exposure to 1 *vs.* 3 mg/kg-day chlorpyrifos, and Ricceri *et al.* (2006) reported some indication of socioagonistic behavior only after exposure to 3, but not 1 mg/kg-day chlorpyrifos. Braquenier *et al.* (2010) reported a chlorpyrifos-associated increase in anxiety in female mice that occurred at an exposure concentration of 1 mg/kg-day chlorpyrifos, but not at the lower or higher exposures in that study (0.2 and 5 mg/kg-day chlorpyrifos, respectively). Ricceri *et al.* (2006) reported increased locomotor activity associated with prenatal exposure to 6 mg/kg-day chlorpyrifos when followed by postnatal exposure to 1 mg/kg-day chlorpyrifos but not to 3 mg/kg-day chlorpyrifos. Levin *et al.* (2002) reported that female rats had a greater number of working and reference memory errors than controls in the radial arm maze after exposure to 1 mg/kg-day chlorpyrifos, but not 5 mg/kg-day. In the foraging maze, male mice showed increased food recognition and position learning after exposure to 1 mg/kg-day chlorpyrifos, but not at the higher dose of 5 mg/kg-day (Haviland *et al.*, 2010).

Across studies, anxiety was increased in female mice assessed in the elevated plus-maze after oral exposure of their dams to 1 mg/kg-day chlorpyrifos during GD

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

15– PND 14 (Braquenier *et al.*, 2010), but not in offspring of mouse dams exposed orally to doses of 3 or 6 mg/kg-day chlorpyrifos during GD 15–18 followed by subcutaneous exposure of the offspring to 1 or 3 mg/kg-day chlorpyrifos during PND 11–14 (Ricceri *et al.*, 2006). Ricceri *et al.* (2003) reported increased locomotor activity in mice exposed to 1 or 3 mg/kg-day by subcutaneous injection during PND 11–14, but both Dam *et al.* (2000) and Levin *et al.* (2001) reported no effects on locomotor activity in rats exposed to 5 mg/kg-day chlorpyrifos by the same route during the same postnatal period. Transient decreases in alternation latency in the T-maze test were reported in two studies after subcutaneous exposure to 1 or 5 mg/kg-day chlorpyrifos during prenatal periods (Icenogle *et al.*, 2004; Levin *et al.*, 2002), but another study reported no effects in the T-maze test after oral exposure to 0.3, 1, or 5 mg/kg-day chlorpyrifos during the perinatal period (Maurissen *et al.*, 2000). The studies that assessed neuromotor function and sensorimotor reflexes did not provide sufficient data to support an exposure-dependent effect of chlorpyrifos on these endpoints, except perhaps at exposures above 6 mg/kg-day chlorpyrifos. Together, the animal data do not demonstrate clear exposure-response relationships between chlorpyrifos exposure and neurodevelopmental effects.

3.3.3.4. Biological significance of responses Addressing the question of causation at doses well below those traditionally recognized as causing biological responses is a challenge because the magnitude of the putative responses may be only marginally detectable. Even if an effect is due to the treatment, one must evaluate the biological relevance of the effect. Statistical significance may be overruled by a lack of biological relevance, such as if the magnitude of response is small or the observed change is not interpretable as an adverse response (Goodman *et al.*, 2010). Many of the behavioral changes associated with chlorpyrifos treatment were mild and/or transient, and several were observed to be in the wrong direction for adversity (*e.g.*, fewer memory errors). Many could also not be replicated under the same or similar conditions in other studies and, thus, may be chance fluctuations or due to another factor. The biological significance of neurodevelopmental effects of chlorpyrifos in animals also depends on whether they are observed at doses above or below the threshold for inhibition of AChE activity in the brain. If chlorpyrifos causes neurodevelopmental effects at very low doses that are presumed to occur through an alternative mechanism besides AChE inhibition, they should be consistently observed at doses below this threshold. Although AChE activity was not assessed in many of the studies, particularly those with exposure during prenatal periods, often the reported effects of chlorpyrifos exposure were observed in conjunction with AChE inhibition or at doses shown to be associated with AChE inhibition in other studies (*i.e.*, ≥ 1 mg/kg-day). This indicates that even if the reported neurodevelopmental effects are real, they do not likely act through a mechanism that only operates at doses below the threshold for AChE inhibition in the brain.

3.3.4. Conclusions for animal data

As a whole, the studies examining potential associations between chlorpyrifos exposure and neurodevelopmental effects in rodents indicate that only a few, isolated markers of certain behaviors are associated with low exposures at a given developmental stage. The specific changes in these markers are not necessarily in agreement across studies, are often transient or in the wrong direction for an adverse effect, are often observed in conjunction with AChE inhibition or at doses shown to be associated with AChE inhibition in other studies, and do not demonstrate consistent exposure-response relationships, except perhaps at very high exposure levels. The majority of studies used subcutaneous injection as the exposure route, which is not as relevant to humans, and studies using the more relevant oral exposure route do not appear to be more likely to report associations with neurodevelopmental effects. In fact, studies with the most weight report largely null effects across various neurodevelopmental tests, and those that do report treatment-related effects often report inhibition of AChE activity in the brain at the same doses. Overall, the animal data are not sufficiently robust to support the hypothesis that chlorpyrifos exposure causes neurodevelopmental effects at exposure levels below those associated with systemic toxicity or AChE inhibition.

3.4. Evaluation of mechanistic data

3.4.1. Introduction

It is well-established that the MoA for acute neurotoxicity of chlorpyrifos is inhibition of AChE activity in the nervous system *via* chlorpyrifos-oxon, with high doses leading to cholinergic toxicity (as reviewed by Eaton *et al.*, 2008). As described earlier in Section 3.1, there is a threshold for this inhibition, as it requires chlorpyrifos exposures that are high enough to overwhelm detoxification pathways, allowing chlorpyrifos-oxon to reach the brain, and clinical symptoms are evident only when at least 70% or greater inhibition of AChE activity is reached (Clegg and van Gemert, 1999). The studies in rodents reviewed above indicate that the LOAEL for AChE inhibition in the brain is 1 mg/kg chlorpyrifos. Below this level, however, cell damage and loss in the developing brain have been reported (Qiao *et al.*, 2002).

Because AChE inhibition is not observed at low chlorpyrifos exposures in humans, other mechanisms have been proposed for potential neurodevelopmental effects of chlorpyrifos at exposures below the threshold for cholinesterase inhibition. The proposed mechanisms involve the action of chlorpyrifos itself, rather than chlorpyrifos-oxon. Evidence for these mechanisms comes mainly from *in vitro* studies, as there are little *in vivo* data available. Below, we review the studies assessing these mechanisms and discuss their relevance to the neurodevelopmental outcomes assessed in epidemiology and animal studies.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

3.4.2. Analysis of mechanistic data

3.4.2.1. Neuronal differentiation Cholinesterases play an important role in early development of the nervous system. Several studies suggest that in addition to its enzymatic activity, AChE has morphogenic activity in the developing nervous system. If chlorpyrifos causes effects at low doses during neurodevelopment, perturbation of this morphogenic activity is one potential mechanism that has been proposed to lead to adverse effects on neuronal differentiation and synaptic function.

Several studies have demonstrated that in noncholinergic neuronal cells *in vitro*, chlorpyrifos inhibits neurite and axonal outgrowth and enhanced dendritic growth at concentrations lower than those which inhibit cholinesterase activity in the brain. Chlorpyrifos inhibited nerve growth factor (NGF)-induced neurite outgrowth in rat pheochromocytoma (PC12) cells at 1 μg/mL (2.85 μM) without inhibiting cholinesterase activity, which occurred at 10 μg/mL (28 μM) (Das and Barone, 1999). PC12 cells are immature neuronal precursors that differentiate into postganglionic sympathetic neuron-type cells with high AChE activity upon NGF stimulation. Chlorpyrifos inhibited axon outgrowth in primary cultures of embryonic rat sympathetic neurons at exposures ≥ 0.001 μM in the presence of a CYP450 inhibitor, indicating the effect is likely attributable to chlorpyrifos and not chlorpyrifos-oxon (Howard *et al.*, 2005). Chlorpyrifos also enhanced BMP-induced dendritic growth at concentrations between 1 and 10 μM. AChE was inhibited only at 1 μM and above. The authors noted that these effects on axonal and dendritic growth are likely both independent of AChE inhibition because, even though the dendritic effects were observed at same concentrations as AChE inhibition, TCPy (which does not inhibit AChE) also enhanced dendritic growth (Howard *et al.*, 2005). Yang *et al.* (2008) also reported decreased axonal length in primary sensory neurons from embryonic rat dorsal root ganglia at concentrations from 0.001 μM to 10 μM. Inhibition of AChE activity was observed at concentrations of 0.1 μM and above. These effects on axon length required the presence of AChE, however, as they were not observed in AChE knockout neurons.

Both *in vitro* and *in vivo* studies have reported effects of chlorpyrifos on cholinergic neuronal development and function in the absence of cholinesterase inhibition. Jameson *et al.* (2006) reported a reduction in the activity of choline acetyltransferase (ChAT), a marker for the cholinergic phenotype, in PC12 cells exposed to 5 μM chlorpyrifos at the beginning of differentiation. Reduced ChAT activity was not observed in undifferentiated cells or in cells exposed during mid-differentiation, whereas increased activity of tyrosine hydroxylase (TH), a marker for the catecholamine phenotype, was observed in both of these cell types. These results indicate that the start of differentiation is a critical period for chlorpyrifos to impair the development of the cholinergic phenotype, whereas promotion of the expression of the catecholaminergic phenotype occurs in both undifferentiated and differentiated cells.

Dam *et al.* (1999) exposed neonatal rats subcutaneously to 1 mg/kg-day chlorpyrifos on PND 1–4 or to 5 mg/kg-day chlorpyrifos on PND 11–14 and examined effects on the development of cholinergic neuronal function in the brainstem, forebrain, and cerebellum, using indices of synaptic proliferation (ChAT activity) and synaptic activity (hemicholinium-3 [HC-3] binding). The brainstem and forebrain develop prominent cholinergic inputs, whereas the cerebellum is sparse in cholinergic projections. Early treatment decreased synaptic proliferation without affecting synaptic activity in the forebrain. Neither measure was affected in the brainstem. Effects of chlorpyrifos were observed on the catecholamine pathways as well: early or late treatment increased the synaptic activity of the neurotransmitters norepinephrine and dopamine, with the greatest effects in the cerebellum. Effects on catecholamine systems were unrelated to the magnitude or temporal pattern of cholinesterase inhibition. The observed deficient cholinergic synaptogenesis and increased catecholaminergic synaptic activity are consistent with the results of the *in vitro* study by Jameson *et al.* (2006).

Subcutaneous exposure of rat dams on GD17–20 with 1 mg/kg-day chlorpyrifos induced small changes in synaptic proliferation and marked suppression of synaptic activity in the forebrain of neonatal offspring. This reduction in synaptic activity returned to normal by weaning, but deficits were again apparent in the regions of the forebrain involved in learning and memory (cerebral cortex, hippocampus, and striatum) in adolescence and adulthood (Qiao *et al.*, 2003). Similar exposure on GD 9–12 increased synaptic proliferation and decreased synaptic activity in the hippocampus and striatum in adolescence and adulthood (Qiao *et al.*, 2004).

Together, these studies indicate that chlorpyrifos exposure at concentrations as low as 1 nM *in vitro* or at doses of 1 or 5 mg/kg-day during the prenatal or neonatal period in rodents can induce effects on neuronal differentiation and function. Some of these effects may be independent of cholinesterase inhibition in neuronal cells *in vitro*, but *in vivo* they are observed at doses shown in other studies to inhibit AChE in the brain.

3.4.2.2. Oxidative stress Several studies have suggested that chlorpyrifos can induce oxidative stress in various neuronal cell types *in vitro* and *in vivo* at concentrations below the threshold for cholinesterase inhibition. If chlorpyrifos is a developmental neurotoxicant at these concentrations, production of oxidative stress in the developing nervous system, leading to oxidative neuronal cell damage, has been proposed as an underlying mechanism (Crumpton *et al.*, 2000).

Crumpton *et al.* (2000) reported a dose-dependent increase in the generation of reactive oxygen species (ROS) measured concurrently with acute treatment (10 minutes duration) of PC12 cell suspensions with 0.5–50 μg/mL (1.4–142 μM) chlorpyrifos. Acute treatment

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

with 10 μg/mL (28 μM) chlorpyrifos-oxon had no effect on ROS generation. The effects with chlorpyrifos were transient, however, as no increases in ROS generation were observed immediately following prolonged (24–72 hours) chlorpyrifos exposure in either undifferentiated or NGF-differentiated PC12 cells.

Qiao *et al.* (2005) reported increased lipid peroxidation, an effect of oxidative stress, in both undifferentiated and NGF-differentiated PC12 cells after exposure to chlorpyrifos at concentrations of 1 μM and above. The authors noted that the lack of enhancement of sensitivity to this effect by differentiation (which increases AChE activity) is consistent with a noncholinergic mechanism.

Oxidative stress was also examined in oligodendrocytes, which are glial cells that are essential to neuronal differentiation, myelination, impulse propagation, and homeostatic maintenance. Disruption of oligodendrocyte function can manifest as motor, cognitive, or behavioral dysfunction. Saulsbury *et al.* (2009) exposed CG-4 cells (oligodendrocyte progenitors) to chlorpyrifos at concentrations between 15 and 120 μM. Chlorpyrifos induced a dose-dependent increase in cell death at concentrations of 30 μM and above. Reactive oxygen species (ROS) and intermediates were observed in cells with 15, 60, and 120 μM chlorpyrifos, but not 30 μM chlorpyrifos. Superoxide generation was induced at 30 and 60 μM, and pretreatment with diethyl maleate (DEM), which reduces intracellular levels of the antioxidant glutathione, enhanced chlorpyrifos-induced cell toxicity at 15 and 30 μM. Addition of a nitric oxide synthase inhibitor did not fully reverse chlorpyrifos-induced toxicity, suggesting that the toxicity is partly caused by production of nitric oxide. Vitamin E, a nonspecific antioxidant, completely spared cells from the toxic effects of chlorpyrifos, further indicating that chlorpyrifos exposure leads to generation of ROS.

Slotkin *et al.* (2005) reported no increases in lipid peroxidation in the developing brain of rats exposed to chlorpyrifos by subcutaneous injection on GD 17–20 or PND 1–4, even at doses well above the threshold for cholinesterase inhibition. Lipid peroxidation increased in the forebrain and cerebellum of males, but not females, with exposure to 5 mg/kg-day on PND 11–14, but no increases were reported in the brainstem. The authors stated that cholinergic hyperstimulation is not responsible for the oxidative damage, as the cerebellum, which, as noted above, is sparse in cholinergic projections, was affected more than the brainstem, which has major cholinergic inputs. These data are not consistent with the *in vitro* data of Qiao *et al.* (2005), who reported lipid peroxidation in undifferentiated cells which correspond to earlier stages of neurodevelopment. By contrast, Slotkin and Seidler (2009) interpreted the Qiao *et al.* (2005) data as indicating that co-exposure to NGF enhanced the lipid peroxidation induced by chlorpyrifos, which is consistent with increased sensitivity to oxidative stress during neurodifferentiation, and that the study by Slotkin *et al.* (2005) confirms the Qiao *et al.* (2005) results by showing greater lipid peroxidation in vulnerable brain regions during peak periods of axonogenesis and synaptogenesis.

Slotkin and Seidler (2009) examined the effects of 30 μM chlorpyrifos on mRNA levels of genes involved in oxidative stress responses and genes encoding receptors of glutamate (a neurotransmitter) in PC12 cells. They reported larger and more widespread transcriptional changes in genes related to oxidative stress response in differentiating cells compared to undifferentiated cells. In undifferentiated cells, they reported more robust effects on the expression of genes for ionotropic glutamate receptors, which mediate excitotoxic cell death in the developing brain, than on genes for oxidative stress or for metabotropic glutamate receptors, which are not involved in excitotoxic cell death. These results suggest a greater role for excitotoxicity than oxidative stress in the undifferentiated state (earlier stages of neurodevelopment) and an increasing role for oxidative stress as cells undergo differentiation (later stages of neurodevelopment).

Together, these studies indicate that chlorpyrifos exposure induces oxidative stress in neuronal cells *in vitro* at concentrations of at least 1 μM, particularly when these cells undergo differentiation, and induction of oxidative stress is observed *in vivo* at doses shown in other studies to inhibit AChE in the brain.

3.4.2.3. cAMP-related cell signaling Perturbations of the adenylyl cyclase (AC) signal transduction pathway have been proposed as a mechanism for the potential neurodevelopmental toxicity of chlorpyrifos at exposures below the threshold for cholinesterase inhibition in several studies. Stimulation of the AC pathway catalyzes the synthesis of cyclic AMP (cAMP), which is involved in the control of cell replication and differentiation. Higher levels of cAMP increase cAMP-dependent kinase (PKA)-mediated phosphorylation of several proteins, including the transcription factor CREB. CREB is critical for synaptic plasticity and transcription-dependent forms of memory and has a role in cell survival and differentiation during brain development. Perturbation of this pathway during development would be expected to have an impact on brain cell development and cognitive function.

Schuh *et al.* (2002) reported that chlorpyrifos increased pCREB levels (suggesting a stimulation of the AC pathway) in primary rat cortical neurons in culture with an EC50 of 60 pM. AChE activity was not affected at concentrations up to 100 nM chlorpyrifos, but was observed at 1 and 10 μM. Chlorpyrifos also increased pCREB levels in rat hippocampal neurons with an EC50 in the range of 1–10 nM, but not in astrocytes at concentrations up to 10 μM. The CYP450 inhibitor SKF-525A did not alter the effects of chlorpyrifos, indicating that metabolism to chlorpyrifos-oxon is not necessary for these changes in pCREB levels.

Slotkin *et al.* (2007) reported that exposure to 30 μM chlorpyrifos decreased basal, fluoride-stimulated, and

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

forskolin-stimulated AC activity in differentiating PC12 cells. Acetylcholine receptor inhibitors did not protect cells from these effects, indicating no contribution of cholinesterase inhibition. Addition of vitamin E worsened the effects on AC signaling, indicating that oxidative stress is also not involved. Theophylline, which prevents breakdown of cAMP, restored AC activity to normal or supranormal levels.

Adigun *et al.* (2010) also examined effects of chlorpyrifos on AC signaling in PC12 cells. Treatment with 50 μM chlorpyrifos had no effect on AC signaling in undifferentiated PC12 cells, but treatment of differentiating cells produced deficits in all AC measures (basal activity and response to fluoride, forskolin, and manganese) when exposure started at the onset of differentiation. If chlorpyrifos exposure was continued for six days, or if cells were exposed for two days and then examined four days later, there was complete reversal of the inhibitory effects on AC signaling. Effects on cell signaling were distinct from those on indices of cell number and neurite outgrowth, which showed progressively greater effects at six days than at two days. This indicates that early exposure reprograms the function of the AC signal transduction pathway.

Song *et al.* (1997) subcutaneously injected neonatal rats with 1 or 5 mg/kg-day chlorpyrifos on PND 1–4 or with 5 mg/kg-day chlorpyrifos on PND 11–14 and examined effects on components of the AC cascade in brain regions enriched (forebrain) or sparse (cerebellum) in cholinergic innervation. Rats exposed to 5 mg/kg-day chlorpyrifos on PND 1–4 exhibited > 50% mortality during the exposure period. The authors measured inhibition of brainstem AChE activity 24 hours after the last dose and reported 25% inhibition for the 1 mg/kg-day group and 65% inhibition for the 5 mg/kg-day group exposed on PND 11–14, with substantial recovery from inhibition during the next five days. In the forebrain, 1 mg/kg chlorpyrifos induced deficits of 25–35% in basal and stimulated AC activities by PND 10 in rats exposed on PND 1–4. There were no effects on AC activity at PND 5, but effects worsened over the next five days after treatment cessation, the period in which recovery from cholinesterase inhibition occurred. The effects were not as pronounced (5–15% deficits) in animals exposed to 5 mg/kg chlorpyrifos on PND 11–14. The deficiencies in AC activity were also reported in the cerebellum and the heart, indicating that cholinergic overstimulation alone cannot account for these changes. Raising the dose to induce systemic toxicity (*i.e.*, 5 mg/kg-day on PND 1–4) did not further enhance the effects on AC activity, suggesting that these effects may also occur at lower concentrations in the absence of cholinesterase inhibition.

Meyer *et al.* (2004) subcutaneously injected rat dams with 1 or 5 mg/kg-day chlorpyrifos on GD 9–12 or GD 17–20, or rat pups with 1 mg/kg-day chlorpyrifos on PND 1–4 or 5 mg/kg-day on PND 11–14, and examined the function of the AC signaling pathway in several different brain regions during adulthood (PND 60). Effects on the AC pathway in rats exposed on GD 9–12 required the higher dose of 5 mg/kg-day chlorpyrifos. Exposures on GD 17–20 and later produced sex-specific alterations in the AC pathway. The effects were either stimulatory or inhibitory to the pathway, depending on the time of exposure, sex, and brain region. The authors noted that this rules out the possibility that chlorpyrifos interacts directly with the neurotransmitter receptors or proteins of the AC signaling cascade, because otherwise the alterations would have been similar in every region, for both sexes, and for each dosing regimen. Instead, they note that their results suggest that chlorpyrifos disrupts the program for development of cell signaling, with targeting of specific brain regions for each sex that depend upon the maturational phases of vulnerability of various neural cell populations.

Together, these studies indicate that chlorpyrifos exposure can lead to a disruption and reprogramming of signaling cascades related to the AC signal transduction pathway, but *in vivo*, these effects are only observed at doses shown in other studies to inhibit AChE activity in the brain.

3.4.2.4. Serotonergic dysfunction Neurodevelopmental alterations from chlorpyrifos are not confined to cholinergic systems and may involve a wide variety of neurotransmitters, such as serotonin (5HT) or dopamine (DA). Several studies have reported effects of chlorpyrifos on the functioning of 5HT synapses in rats after subcutaneous exposure during different stages of early development. These effects involved alterations in levels of the 5HT presynaptic transporter (5HTT), a biomarker for the concentration of 5HT nerve terminals that is responsible for regulating the concentration of 5HT in the synapse, levels of 5HT receptors that control cell signaling, or activities of 5HT or DA.

Gestational exposure to 1 or 5 mg/kg-day chlorpyrifos on GD 9–12 increased 5HT activity in the cerebral cortex, but not the midbrain or brainstem, in both sexes during adolescence (PND 30) (Slotkin and Seidler, 2007). Similar effects were observed for DA activity, but at lower magnitude. In adulthood, elevations in levels of 5HTT and the 5HT receptors, 5HT1A and 5HT2, were observed in the cerebral cortex, midbrain, and brainstem of both sexes (Aldridge *et al.*, 2004).

Exposure to 1 or 5 mg/kg-day chlorpyrifos on GD 17–20 increased 5HT activity in brain regions with either 5HT projections or cell bodies in males at both doses and in females only at the higher dose (above the threshold for cholinesterase inhibition) during adolescence (Slotkin and Seidler, 2007). Similar effects were reported for DA activity, but at lower magnitude and with no sex preference (Slotkin and Seidler, 2007). This exposure scenario also induced larger effects on elevations of 5HT1A, 5HT2, and 5HTT during adulthood compared to those with exposure during GD 9–12, with selectivity for brain regions with 5HT nerve terminals and preferential effects in males (Aldridge *et al.*, 2004). Also during adulthood,

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

decreased 5HT levels were reported in animals exposed to 5 mg/kg-day and a net increase in 5HT activity was reported with both doses, with no preference for these effects in either sex (Aldridge *et al.*, 2005b). DA content was unaffected in most brain regions, but large deficits were observed in the hippocampus with both doses of chlorpyrifos in both sexes (Aldridge *et al.*, 2005b). DA turnover was increased in the cerebral cortex, striatum, and midbrain in both sexes with the 5 mg/kg dose (Aldridge *et al.*, 2005b).

Exposure to 1 mg/kg-day chlorpyrifos on PND 1–4 resulted in an increase in levels of 5HT receptors during adulthood, with larger effects in males and in regions with 5HT cell bodies (Aldridge *et al.*, 2004). Levels of 5HTT were increased in both sexes in the brainstem and decreased in all other brain regions examined in females (Aldridge *et al.*, 2004). In behavioral tests conducted during adulthood, treated animals showed abnormalities related to 5HT deficiencies. The normal sex differences for the elevated plus maze and 16-arm radial maze tests were ablated, as the behavior of treated males was "feminized," resulting in similar scores to those observed for control females (Aldridge *et al.*, 2005a). The levels of 5HT were unchanged in males and slightly decreased in females, whereas 5HT activity was increased in both sexes (Aldridge *et al.*, 2005b). DA content and activity were decreased in the cerebrocortical area and increased in the striatum, and DA activity was increased in the midbrain (Slotkin *et al.*, 2002).

Following exposure to 5 mg/kg-day chlorpyrifos on PND 11–14, smaller increases in 5HT1A and 5HT2 were observed in adulthood, and 5HTT levels were decreased with the same regional and sex selectivity as observed in animals exposed on PND 1–4 (Aldridge *et al.*, 2004). There were no effects on the content or activity of either 5HT (Aldridge *et al.*, 2005b) or DA (Slotkin *et al.*, 2002).

Together, these *in vivo* data indicate that the immediate perinatal period has the greatest sensitivity and sex-selectivity to effects of chlorpyrifos on indices of serotonergic activity in rodents. The windows of GD 17–20 and PND 1–4 encompass the peak period of sexual differentiation in the brain (MacLusky and Naftolin, 1981), and exposures prior to this period did not produce male-female differences in outcomes. Aldridge *et al.* (2005b) suggested that the effects on 5HT indices observed during the perinatal period are indicative of deficient synaptic communication that is consistent with a "miswiring" of 5HT circuits and that effects on 5HT may be one component of a larger spectrum of chlorpyrifos-induced disruption of synaptic development and function that can ultimately contribute to behavioral anomalies. One hypothesized example could be through disruption of 5HT-mediated cell signaling, which includes AC signaling (Aldridge *et al.*, 2004); another is that deficiencies in 5HT systems could create a situation of increased reliance on 5HT mechanisms for cognitive function that aren't normally called into play (Aldridge *et al.*, 2005a). Alternatively, chlorpyrifos may not specifically target serotonergic systems, as effects could be secondary to those on neuronal differentiation (Aldridge *et al.*, 2004). It should be noted, however, that these effects are only observed with chlorpyrifos exposures that are associated with inhibition of AChE activity in the brain in other studies.

3.4.3. Conclusions for mechanistic data

Several mechanisms that presumably do not involve inhibition of AChE activity in the nervous system have been explored to determine whether chlorpyrifos can act as a neurodevelopmental toxicant *via* other mechanisms. The potential mechanisms include perturbation of the morphogenic, rather than enzymatic, activity of AChE; neuronal cell damage caused by induction of oxidative stress; disruption of the AC signal transduction pathway; and dysfunction of serotonergic systems. These mechanisms have been suggested to be involved in a large spectrum of effects such as chlorpyrifos-induced neuronal cell damage or disruption of systems controlling neuronal differentiation and synaptic function, although serotonergic dysfunction is involved in appetite and affective (depression) disorders and it is unclear how this would be relevant to the neurodevelopmental outcomes assessed in the epidemiology and animal studies described above. Evidence for the action of the proposed mechanisms at doses not affecting AChE activity comes mainly from *in vitro* studies, so their relevance to potential outcomes in children with very low exposures to chlorpyrifos is unclear. In addition, the chlorpyrifos concentrations used in the *in vitro* studies must be quantitatively considered for their relevance to human systemic concentrations. Chlorpyrifos concentrations in the blood of subjects in the Columbia cohort averaged 4 pg/g, which is equivalent to 0.01 nM, but were also measured as high as 0.1 nM (Eaton *et al.*, 2008). Using a blood/brain partition coefficient of 33 calculated by Timchalk *et al.* (2002) for chlorpyrifos in rats, brain concentrations would be estimated to range from 0.33–3.3 nM (Eaton *et al.*, 2008). Almost all of the *in vitro* studies reported potential mechanistic effects at micromolar concentrations, although three studies reported effects at concentrations ≤ 1 nM (Schuh *et al.*, 2002; Howard *et al.*, 2005; Yang *et al.*, 2008). AChE inhibition was also reported at micromolar concentrations in the *in vitro* studies (Das and Barone, 1999; Howard *et al.*, 2005; Schuh *et al.*, 2002). Overall, these data indicate that chlorpyrifos exposures at which potential mechanistic effects and AChE inhibition were observed in most of the *in vitro* studies are 1,000-fold higher than the estimated exposures to chlorpyrifos in the epidemiology studies.

3.5. HBWoE evaluation of the potential neurodevelopmental toxicity of chlorpyrifos

A general hypothesis that has been put forth in the scientific literature is that chlorpyrifos causes adverse neurodevelopmental effects in humans at exposures below

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

the threshold for cholinesterase inhibition in the brain by chlorpyrifos-oxon, and that chlorpyrifos itself is acting as the neurotoxicant at low exposures by one or more proposed mechanisms during critical periods in the developing brain. The concern for chlorpyrifos-induced neurodevelopmental effects stems from a few epidemiology studies reporting potential associations between low-dose chlorpyrifos exposure and effects on infant neurobehavior, as well as cognitive and motor development and behavior outcomes in children, although other studies do not show these effects. The epidemiology studies have shortcomings, however, such as being subject to substantial confounding variables including low SES, maternal smoking and alcohol use, and exposures to other pesticides, as well as a lack of specific, reliable biomarkers of exposure. The studies with more robust factors, such as reliable exposure metrics or larger sample sizes, do not appear to be more likely to report associations with adverse neurodevelopmental effects.

In addition to the epidemiology studies, many rodent studies have been conducted to examine the potential neurodevelopmental effects of chlorpyrifos. These studies indicate that only a few isolated markers of certain behaviors are associated with low exposures (compared to those which cause systemic toxicity) at a given developmental stage. The specific changes in these markers are not necessarily in agreement across studies, are often transient or in the wrong direction for an adverse effect, are often observed in conjunction with AChE inhibition or at concentrations shown to be associated with AChE inhibition in other studies, and do not demonstrate consistent exposure-response relationships, except perhaps at very high exposure levels. While some investigators have proposed potential mechanisms for the effects of chlorpyrifos at doses below those associated with cholinesterase inhibition, the evidence for these comes mainly from *in vitro* studies, and the animal data for neurodevelopmental outcomes do not provide strong support for chlorpyrifos neurotoxicity at doses below this threshold.

Below, we evaluate the scientific data relevant to examining whether there is a causal association between exposure to chlorpyrifos and adverse neurodevelopmental effects in humans using the structured HBWoE approach (Rhomberg *et al.*, 2010, 2011). This approach weighs all of the data from epidemiology, animal toxicity, and mechanistic studies in terms of quality and relevance to humans, allowing each of these data sets to inform one another. All of the data are then evaluated together to determine whether a causal relationship between chlorpyrifos at low exposures and neurodevelopmental effects in humans is plausible. This evaluation considers the uncertainties and inconsistencies in the data sets, as well as any *ad hoc* assumptions that may be required for some of the hypotheses put forth. The key outcome of this approach is an evaluation and comparison of alternative accounts (or hypotheses) of all the available data. If data of poor quality are used as a basis to support one of the accounts, the logic of how these data have been interpreted and the *ad hoc* assumptions needed to fit these data to the proposed hypothesis are discussed. Various competing accounts are weighed by comparing the *ad hoc* assumptions needed for each, with more credence given to the hypothesis that requires the least amount of assumptions.

We consider two hypotheses for MoAs that have been put forth in the literature to explain the existence of human risks for adverse neurodevelopmental effects of chlorpyrifos. These hypotheses are based on the human, animal, and mechanistic data from studies assessing whether there are adverse neurodevelopmental effects of chlorpyrifos at doses below those at which other effects have been observed, and rely on several lines of evidence regarding potential mechanisms for low-dose effects.

3.5.1. Hypotheses under consideration

We reiterate that stating the hypotheses in the HBWoE approach requires more than just putting forth the question of whether chlorpyrifos causes neurodevelopmental effects in human populations at the levels to which they are exposed. It is important to articulate the logical basis one is invoking to consider the available studies as evidence that is relevant to the potential for risk in the target human population. It is this articulation of a logical basis that identifies what is asserted as being in common between each studied situation and the others, and between each studied situation and the target human population, such that the relevance of findings and the ways of accounting for similarities and differences and for consistencies and inconsistencies can be considered as one brings the results to bear on the motivating question.

The first hypothesis is that chlorpyrifos induces adverse effects on the developing nervous system at doses below those which inhibit the activity of AChE in the brain. Moreover, any such effects are hypothesized to apply similarly across mammalian species, such that effects observed in animal experiments would be expected to apply across species, including in humans. Low-exposure human studies can be evaluated to determine whether they show indications of this presumed ability to affect neural development, which would be consistent with this hypothesis. Several specific mechanisms for effects at doses below those which inhibit brain AChE activity have been proposed by various research groups, each involving the action of chlorpyrifos itself rather than chlorpyrifos-oxon. These potential mechanisms include perturbation of the morphogenic, rather than the enzymatic, activity of AChE; neuronal cell damage caused by induction of oxidative stress; disruption of the AC signal transduction pathway; and dysfunction of serotonergic systems. It has been suggested that these mechanisms may be involved in a large spectrum of chlorpyrifos-induced neuronal cell damage or disruption of systems

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

controlling neuronal differentiation and synaptic function. Evidence for these mechanisms comes mainly from *in vitro* studies, as there are little *in vivo* data available, so their relevance to potential neurodevelopmental outcomes in children is unclear.

Under the first hypothesis, it would need to be presumed that the various kinds of measured outcomes across studies are all manifestations of a common underlying neurotoxic mechanism that is being hypothesized to apply across settings. Thus, it would be presumed that the observation of somewhat different particular results in each study and in each species can be taken as evidence of the operation of the same mechanism of action. Under this view, differences in particular outcomes across studies are attributed to the way investigators have chosen to measure the observable manifestation of this common underlying mechanism of toxicity.

The second hypothesis is that chlorpyrifos induces neurodevelopmental toxicity only through the established mechanism of inhibition of AChE activity in the nervous system by its metabolite, chlorpyrifos-oxon. This mechanism requires sufficient doses of chlorpyrifos for chlorpyrifos-oxon to reach the brain. If neurodevelopmental effects are observed in animal studies, this mechanism potentially underlies these effects if observed with doses that inhibit AChE activity to some extent, even if this inhibition is not high enough for systemic toxicity (*i.e.*, > 70% inhibition). This hypothesis is consistent with a lack of neurodevelopmental effects in humans because exposures in humans are far below those that are associated with AChE inhibition.

3.5.2. Evaluation of hypotheses for each line of evidence

We next considered the two hypotheses in the context of each line of evidence (epidemiology, animal, and mechanistic) and evaluated how well the hypotheses are in agreement with the available data, how well they would explain patterns in the data if they were true, what other events or processes should be observed if they are true, and whether these processes, in fact, are observed. For each hypothesis, the following questions become evident:

1. Are the data from the epidemiology studies compelling? Are there alternative explanations for the few positive associations observed in these studies?
2. What is the evidence that chlorpyrifos is associated with neurodevelopmental effects in animals, and are these effects observed in the absence of cholinesterase inhibition in the brain?
3. What is the evidence that the candidate mechanisms act only through chlorpyrifos and not chlorpyrifos-oxon?

1. Are the data from the epidemiology studies compelling? Are there alternative explanations for the few positive associations observed in these studies?

There are few studies of each specific neurodevelopmental outcome examined in each cohort, limiting the ability to look for consistency of outcomes across studies or cohorts. The studies that carry more weight, such as those with more reliable exposure metrics or larger sample sizes, do not appear to be more likely to report associations with adverse neurodevelopmental effects.

The outcome examined in the largest number of studies was newborn head circumference (Perera *et al.*, 2003; Whyatt *et al.*, 2004; Barr *et al.*, 2010; Eskenazi *et al.*, 2004; Berkowitz *et al.*, 2004; Wolff *et al.*, 2007). Regardless of the weight of each study, null results were reported across all studies of this outcome, increasing the likelihood that the overall findings are robust and that there is no association between chlorpyrifos exposure and decreased newborn head circumference.

The two studies that examined infant neurobehavior (Engel *et al.*, 2007; Young *et al.*, 2005) reported associations between chlorpyrifos exposure and abnormal reflexes in the BNBAS, but these studies carry less weight because of their potential for exposure measurement error from the use of urinary DEPs as the exposure metric and misclassification of outcome from a single assessment of neurobehavior.

Four studies examined cognitive and motor development, and three of these studies used the BSID-II and reported no associations between chlorpyrifos exposure and MDI or PDI scores up to 24 months of age in the Columbia (Rauh *et al.*, 2006), CHAMACOS (Eskenazi *et al.*, 2007), and Mount Sinai (Engel *et al.*, 2011) cohorts. Only Rauh *et al.* (2006) examined these outcomes in children at 36 months of age, using a single measure of chlorpyrifos in cord blood as the exposure metric. At this age, associations with lower PDI scores and with mental and motor delays were reported, although it is unclear whether these effects are clinically significant, as the mean scores for children in the highest exposure group were well within the normal range. Of the two studies that assessed scores on the WISC-IV, one reported an association between cord blood chlorpyrifos and a decrement of 0.35 to 0.81 points in working memory scores at age seven in the Columbia cohort (Rauh *et al.*, 2011), whereas the other reported no association with changes in scores on the WISC-IV in children between seven and 9 years of age in the Mount Sinai cohort (Engel *et al.*, 2011). Rauh *et al.* (2011) did not report the mean or range of WISC-IV scores, so it is unclear whether the modest decrement reported for the working memory index is clinically significant.

Two studies conducted a single assessment of behavioral outcomes at one age per study using the CBCL. This may lead to outcome misclassification, particularly when the measure is based on reporting of behavior by mothers, which is subject to reporting bias. The study using the more robust exposure metric at only one time point (Rauh *et al.*, 2006) reported associations with attention problems, ADHD, and PDD, whereas the study with the larger sample size but a less-reliable exposure metric at two time points (Eskenazi *et al.*, 2007) only reported an association with PDD. In both studies, very few children

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

scored in the clinical ranges of the CBCL, limiting the clinical significance of the results. The lack of confirmation of the associations, and the methodological issues in these studies increases the likelihood that there are alternative explanations for the observed outcomes other than exposure to chlorpyrifos.

There are many alternative explanations for the few positive associations with neurodevelopmental outcomes reported in some of the cohort studies. One explanation is that exposure measurement error increases the uncertainty that any associations are specifically attributable to chlorpyrifos exposure. Direct measurement of the parent compound more accurately reflects the chlorpyrifos dose in the brain than do measurements of its metabolites in urine. Urinary TCPy originates from exposure to not only chlorpyrifos, but to chlorpyrifos-methyl and to TCPy itself, and urinary DEPs originate from diazinon and disulfoton in addition to chlorpyrifos, so measurements of these biomarkers are not specific to chlorpyrifos and can overestimate exposure to chlorpyrifos. Most of the exposure metrics used in the cohort studies were measured at only one point in time, but there is large intraindividual variability in exposures to chlorpyrifos, so a single measurement may not represent average exposure over time or exposure at some earlier or later time. In addition, because of the rapid elimination of chlorpyrifos and its metabolites from the body, any measure of chlorpyrifos or its metabolites in blood or urine at a single time point reflects exposure during the brief period of time prior to measurement and may not accurately reflect exposure throughout the entire critical period of neurodevelopment.

Another explanation for the positive findings in some of the cohort studies stems from the lack of adequate adjustment for several confounding factors, decreasing the likelihood that any observed effects are attributable to chlorpyrifos exposure. Each cohort was exposed to multiple types of pesticides besides chlorpyrifos, but these exposures were not controlled for in any of the studies reporting associations. Maternal smoking and alcohol use have been associated with adverse neurodevelopmental outcomes, but these factors were likely underestimated in the few cohorts in which they were considered. The cohorts in which positive associations were reported all come from populations with low SES, which has also been associated with effects on neurodevelopment. Of the cohort studies reporting associations with neurodevelopmental outcomes, all examined at least some confounding factors related to SES (*e.g.*, mother's education; household income; quality of home environment) and those that were associated with outcomes were included as covariates in the final models. Although there was adjustment for several different confounders in each cohort, other factors that could affect the results may not have been accounted for, which increases the likelihood that there are alternative explanations for the observed outcomes other than exposure to chlorpyrifos.

Finally, the positive associations reported in some of the cohort studies could actually be statistical anomalies rather than actual associations with chlorpyrifos exposure. This is because in each study that reported associations between chlorpyrifos exposure and neurodevelopmental outcomes, many different analyses were conducted and there was no adjustment for multiple comparisons. This increases the likelihood that several results were statistically significant by chance. In addition, if studies with smaller sample sizes report associations but studies with larger sample sizes do not, it is more likely that the reported associations are statistical anomalies or that publication bias occurred.

Overall, we found that the epidemiology data are not sufficiently robust to support the hypothesis that chlorpyrifos is a causal factor for neurodevelopmental effects. The cohort studies do not report consistent results, and there is a lack of clear exposure-response information. Because of the many uncertainties in these studies, the few positive results may also support alternative explanations that factors other than chlorpyrifos are causal for the reported outcomes, or that the observed associations are statistical anomalies.

2. What is the evidence that chlorpyrifos is associated with neurodevelopmental effects in animals, and are these effects observed in the absence of cholinesterase inhibition in the brain?

The studies assessing potential neurodevelopmental effects of chlorpyrifos in rodents indicate that usually a few, isolated markers of certain behaviors were determined to be statistically significant in pair-wise comparisons with controls, but these were generally contradicted by other studies reporting no effects in similar dose ranges using similar routes of exposure, and sometimes showed effects in the other direction. In addition, many of the behaviors were observed in conjunction with AChE inhibition or at concentrations shown to be associated with AChE inhibition in other studies, and they did not demonstrate consistent exposure-response relationships, except perhaps at very high exposure levels. For each endpoint, we did not find a pattern of an effect that was consistent over doses and time points within studies or consistent across studies to constitute a repeatable finding, increasing the likelihood that the reported effects may be due to chance. This was not only true for the animal data as a whole, but also for the most rigorous animal studies, which should carry the most weight. When we separately examined the studies that used dose groups with a relatively high number of animals of each sex, considering both oral and subcutaneous exposures and including the one study that complied with US EPA Pesticide Assessment Guidelines and GLP regulations, we found that these studies reported largely null effects across various neurodevelopmental tests, and those that did report treatment-related effects often reported inhibition of AChE activity in the brain at the same doses, suggesting that effects occur *via* this pathway.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Several animal studies examined AChE activity in the brain after both oral and subcutaneous chlorpyrifos exposures during postnatal periods, and inhibition of AChE was observed in these studies at doses as low as 1 mg/kg-day when assessed within one day of exposure cessation. Very few of the neurodevelopmental studies assessed AChE inhibition after prenatal chlorpyrifos exposure, however, and only at doses of 3 mg/kg-day and greater, with the exception of one study with dermal exposure to 1 mg/kg-day that reported *increased* AChE activity 90 days after exposure cessation (Abou-Donia *et al.*, 2006). Given the rapid recovery from AChE inhibition after exposure to chlorpyrifos, it is expected that inhibition would have been observed shortly after this exposure. In addition, a study by Qiao *et al.* (2002) reported a no observed adverse effect level (NOAEL) of 1 mg/kg-day and a LOAEL of 2 mg/kg-day for AChE inhibition in the brain when assessed 24 hours after exposure of fetal rats *via* subcutaneous injection of dams on GD17–20. It remains possible that the 1 mg/kg-day dose could have inhibited brain AChE activity if assessed within a few hours of exposure cessation. Regardless of the limited data for AChE inhibition at doses below 3 mg/kg-day, none of the studies with prenatal oral exposures reported neurodevelopmental effects at doses below 6 mg/kg-day, with the exception of two studies reporting different markers of increased anxiety in female mice. In addition, in almost all studies that reported effects with prenatal subcutaneous exposures, they were observed at doses of at least 5 mg/kg-day, with the exception of a few effects observed at a dose of 1 mg/kg-day. These effects included a slower trend of habituation in female animals in one study, but not another study with the same doses also administered during a prenatal period, and a few transient effects in some of the cognitive tests that were also transiently observed in the opposite direction.

Together, the animal data do not provide clear evidence that chlorpyrifos is associated with neurodevelopmental effects at doses that are below the threshold for inhibition of AChE in the brain. Most of the observed effects occurred at doses high enough to inhibit AChE, but not to induce systemic toxicity, which is hypothesized to occur when the extent of AChE inhibition is above 70% (Clegg and van Gemert, 1999). Because of this, it cannot be ruled out that exposures to concentrations of chlorpyrifos that induce a smaller degree of AChE inhibition are associated with certain neurodevelopmental effects in rodents.

3. What is the evidence that the candidate mechanisms act only through chlorpyrifos and not chlorpyrifos-oxon?

The specific mechanisms proposed to support the hypothesis that chlorpyrifos induces adverse neurodevelopmental effects at doses below those which inhibit the activity of AChE in the brain involve the action of chlorpyrifos itself rather than chlorpyrifos-oxon. In some of the *in vitro* studies, evidence for these mechanisms was observed in cells that were exposed to chlorpyrifos in the presence of a CYP450 inhibitor, which prevented metabolism of chlorpyrifos to chlorpyrifos-oxon (Howard *et al.*, 2005; Schuh *et al.*, 2002). Because of this, it is unlikely that any of the observed effects were attributable to chlorpyrifos-oxon unless it was present in the experiment through contamination. In addition, one of the *in vitro* studies reported ROS generation in cell suspensions after exposure to chlorpyrifos but not chlorpyrifos-oxon (Crumpton *et al.*, 2000).

The *in vivo* evidence for the proposed mechanisms is less clear, as some studies reported effects in both cholinergic and noncholinergic brain regions, as well as in serotonergic systems. These effects were all observed in the presence of AChE inhibition or at doses associated with AChE inhibition in other studies, with the possible exception of effects on synaptic proliferation and activity observed in rats exposed to 1 mg/kg-day chlorpyrifos during GD 17–20 (Qiao *et al.*, 2003), a treatment that was not associated with AChE inhibition in the brain in another study, at least when assessed 24 hours after exposure cessation (Qiao *et al.*, 2002). Although all of the *in vivo* mechanistic studies used subcutaneous injection as the exposure route to avoid first-pass metabolism of chlorpyrifos to chlorpyrifos-oxon, there is evidence for extrahepatic metabolism to chlorpyrifos-oxon in the brain (Chambers and Chambers, 1989). Because the *in vivo* studies do not provide evidence for an absence of AChE inhibition, chlorpyrifos-oxon is presumed to be present in the nervous system in sufficient amounts to inhibit AChE in these studies and could be driving the specific mechanisms, regardless of whether they act through noncholinergic processes. Although chlorpyrifos itself can also inhibit AChE in the brain, this requires much higher concentrations compared to chlorpyrifos-oxon. If the proposed mechanisms involve only the action of chlorpyrifos because they are presumed to occur at chlorpyrifos exposures below those that induce AChE inhibition in the brain from chlorpyrifos-oxon, then these mechanisms should also operate at exposures below those that allow chlorpyrifos itself to inhibit AChE. This is clearly not the case, however, because AChE activity is inhibited at the chlorpyrifos exposures used in the *in vivo* mechanistic studies.

3.5.3. Evaluation of hypotheses for all lines of evidence together

Now that we have considered the two hypotheses in the context of each line of evidence, we now consider all lines of evidence as a whole and how they inform the interpretation of each other. For each hypothesis, we ask the following questions:

1. How do the neurodevelopmental effects examined in the animal studies relate to those examined in the cohort studies? Are the rodent behavioral tests sufficient to detect the subtle effects asserted to be caused by chlorpyrifos exposure in humans?

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

2. How do the doses used in the animal studies compare to human exposures from the cohort studies and to the doses used in the mechanistic studies?
3. What is the evidence for the *in vivo* operation of the candidate mechanisms for neurodevelopmental effects at doses below those which induce brain cholinesterase inhibition? What consequences would these mechanisms have in humans, and is there any evidence for this in the literature?

1. How do the neurodevelopmental effects examined in the animal studies relate to those examined in the cohort studies? Are the rodent behavioral tests sufficient to detect the subtle effects asserted to be caused by chlorpyrifos exposure in humans?

The rodent tests for neurobehavioral assessment used in the animal studies were designed to measure similar functions as neurobehavioral tests conducted for humans but they cannot match the complexity of human behavior (Ulbrich and Palmer, 1996). The test batteries include tests of locomotor activity and crude assessments of learning, sensory, and motor integration, and these are typically parallel tests, in that they are conducted in a different manner in humans but it is believed that the same functions are being measured (Sharbaugh *et al.*, 2003). Ideally, homologous tests, which follow the same procedure in both animals and humans, would provide a more accurate measure of the same cognitive function, but such tests are not available (Sharbaugh *et al.*, 2003). For example, although a common endpoint for children is a score on a standardized test of intelligence or IQ, there are no standardized intelligence tests for animals that can identify subtle cognitive dysfunction (Rice, 2005; Winneke, 1992). The rodent test batteries are not capable of thoroughly characterizing the types of impairment produced in specific domains or in identifying the domains affected, so extrapolation from their results to specific deficits in children is problematic (Rice, 2005). In addition, humans are often not evaluated to the same extent as rodents after exposures to potential neurotoxicants, so generalizations from studies with rodent tests may be unfounded (Ulbrich and Palmer, 1996). Thus, current rodent models are only conceptual analogs to human studies, and results in animals models can be used to qualitatively characterize neurotoxic effects in humans (Winneke, 1992; Rice, 2005; Bellinger; 2005).

2. How do the doses used in the animal studies compare to human exposures from the cohort studies and to the doses used in the mechanistic studies?

Regarding the hypothesis that chlorpyrifos induces adverse effects on the developing nervous system in humans at doses below those associated with inhibition of AChE activity in the brain, if the animal studies report neurodevelopmental effects in conjunction with AChE inhibition or at doses in the range of those associated with AChE inhibition in other studies, then they should not be relevant to the human situation.

As noted in section 3.2.1.2, Eaton *et al.* (2008) estimated the daily intake of chlorpyrifos for mothers in the Columbia cohort as 0.008 μg/kg-day and for the CHAMACOS cohort as 0.007 μg/kg-day. These exposures are four orders of magnitude lower than the dose (100 μg/kg-day) that has been shown to significantly inhibit BuChE in plasma, which is the most sensitive *in vivo* biological effect of chlorpyrifos (Coulston *et al.*, 1972), and five orders of magnitude lower than the LOAEL for AChE inhibition in the brain in the animal studies reviewed here. Because of this, it seems highly unlikely that exposures in these cohorts could produce inhibition of AChE in the maternal or fetal brain.

The chlorpyrifos exposures in the epidemiology studies are also much lower than those used in the *in vivo* and *in vitro* mechanistic studies. Chlorpyrifos concentrations in the brain of subjects in the Columbia cohort were estimated to range from 0.33 to 3.3 nM (Eaton *et al.*, 2008). By contrast, Marty *et al.* (2007) reported that five-day-old rat pups exposed to 1 mg/kg chlorpyrifos had maximum blood concentrations (at two hours post-dosing) ranging from 16 to 140 nM, depending on the exposure route, which resulted in estimated brain concentrations of 0.5–4.6 μM. Almost all of the *in vitro* studies reported potential mechanistic effects at micromolar concentrations, although three studies reported effects at concentrations ≤ 1 nM (Schuh *et al.*, 2002; Howard *et al.*, 2005; Yang *et al.*, 2008). AChE inhibition was also reported at micromolar concentrations in the *in vitro* studies (Das and Barone, 1999; Howard *et al.*, 2005; Schuh *et al.*, 2002). These data indicate that estimated exposures to chlorpyrifos in the cohort studies are 1,000-fold lower than those used in the animal studies or those at which effects, including AChE inhibition, were observed in most of the *in vitro* mechanistic studies.

3. What is the evidence for the in vivo operation of the candidate mechanisms for neurodevelopmental effects at doses below those which induce brain cholinesterase inhibition? What consequences would these mechanisms have in humans, and is there any evidence for this in the literature?

The underlying data used to support the various candidate mechanisms were mainly examined in animal cell lines *in vitro*. Herein, we summarize the limited *in vivo* evidence for these mechanisms and discuss whether there is evidence for their consequences in humans.

One proposed mechanism is the perturbation of the morphogenic, rather than enzymatic, activity of AChE in the developing nervous system. A few *in vivo* studies reported certain effects on synaptic activity and/or proliferation in brain regions that are either dense or sparse with cholinergic inputs after exposure to 1 mg/kg-day chlorpyrifos. It is not clear that these effects occurred in absence of AChE inhibition, however. There is also no

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

evidence that they are attributable only to chlorpyrifos and not chlorpyrifos-oxon, although some of the *in vitro* studies provide evidence to rule out chlorpyrifos-oxon, such as the presence of effects after co-exposure to CYP450 inhibitors and an absence of AChE inhibition. The inhibition of neurite outgrowth observed in the *in vitro* studies is a plausible mechanism for neurological deficits, as there is evidence in animals that neurite growth during brain development is essential for neuronal connectivity, and disruption of this process can lead to cognitive deficits (Berger-Sweeney and Hohmann, 1997; Webb *et al.*, 2001); however, it has not been shown that chlorpyrifos can induce such effects *in vivo*.

For the proposed mechanism involving neuronal cell damage caused by induction of oxidative stress, there is only one *in vivo* study available. This study reported increased lipid peroxidation only with postnatal exposure occurring after the period in which neuronal cell differentiation takes place, which is not consistent with results from *in vitro* studies showing similar amounts of lipid perioxidation in both undifferentiated as well as differentiated neurons. The effects were observed at a dose that has been shown to inhibit brain AChE in other studies (5 mg/kg-day), but the authors noted that lipid peroxidation was greater in the cerebellum, a brain region with sparse cholinergic innervations, suggesting that these effects do not involve cholinergic hyperstimulation. Although oxidative stress is a known mechanism for neuronal cell damage, including during human development (Ikonomidou and Kaindl, 2011), it has not been clearly shown that chlorpyrifos induces oxidative stress in the developing brain *in vivo* at concentrations that do not inhibit the activity of AChE.

Two *in vivo* studies examined the proposed mechanism of disruption of the AC signal transduction pathway. One study examined postnatal exposures and reported effects on this pathway only at doses that were also associated with AChE inhibition in the brain (Song *et al.*, 1997). The authors stated that because the effects were observed in several brain regions, including the cerebellum, and also in the heart, the effects cannot be attributable to cholinergic hyperstimulation alone. The other *in vivo* study used prenatal and postnatal exposures at doses that cause AChE inhibition and reported sex-specific alterations in the AC pathway (Meyer *et al.*, 2004). The effects were either stimulatory or inhibitory to the pathway, depending on the exposure period, sex, and brain region, and the authors hypothesized that they are indicative of a disruption and reprogramming of the AC signaling cascade during neurodevelopment. Perturbation of this pathway during development would be expected to have an impact on brain cell development and cognitive function, but the available evidence does not support effects on this pathway *in vivo* after exposure to chlorpyrifos at doses below the threshold for AChE inhibition in the brain.

The final proposed mechanism is dysfunction of serotonergic systems. All studies examining this mechanism were conducted *in vivo* using chlorpyrifos doses that have been shown to inhibit AChE in the brain in other studies. The effects are assumed to be noncholinergic because serotonin is not a neurotransmitter for cholinergic systems. Serotonergic dysfunction is involved in appetite and affective (depression) disorders, so it is unclear how this mechanism is relevant to the neurodevelopmental outcomes assessed in the epidemiology and animal studies.

3.5.4. Evaluation of alternative accounts

An HBWoE evaluation comes down to an evaluation of alternative "accounts," which are proposed sets of explanations for the observed phenomena across the body of relevant lines of evidence. These competing accounts should be evaluated to determine how the evidence supports them, what is necessary to assume for their support, and how the overall weight of the evidence for each suggests how compelling the account is. An account is most compelling when it is not only supported by the factual record, but also helps explain the data by finding common reasons for sets of observations and, moreover, achieves this ability much more readily than any competing account.

For chlorpyrifos, there are two competing accounts that need consideration:

1. The epidemiology evidence is sufficiently compelling that, even in the face of inconsistent evidence in animals with much higher doses, one or more of the proposed mechanisms for low-dose neurodevelopmental effects of chlorpyrifos must be right, and the biological implausibility of these mechanisms is mistaken (*i.e.*, because an effect appears, it must have a causal explanation).
2. Doubts about the potential mechanisms have merit, and the few apparent associations from epidemiology studies do not indicate a causal connection (*i.e.*, the appearance of some associations is due to chance or to shortcomings of the studies and should be deemed false positive results).

Acceptance of the first account is associated with many unanswered questions and *ad hoc* explanations for how the data should be interpreted as supporting it. This account requires that one dismiss the many alternative explanations for the few positive associations observed in the epidemiology studies, despite the plausibility of these explanations. Exposure measurement error from the use of unreliable exposure metrics and the presence of a wide variety of confounding factors that were not adequately adjusted for are important considerations when evaluating the epidemiology data for chlorpyrifos. In addition, this account requires that, although there were many statistical analyses in the epidemiology studies, one chooses to focus only on the few statistically significant findings, regardless of their clinical significance or lack of confirmation in other studies. This account also

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

requires that one accepts the existence of an exposure-response relationship, despite the lack of consistently observed exposure-response relationship within and among the epidemiology and animal studies. It requires that the human studies that failed to show an increase in risk for neurodevelopmental effects did so for plausible reasons, such that the lack of effects in these studies does not contradict the asserted neurodevelopmental toxicity of chlorpyrifos.

This account also requires the inclusion of an explanation for why the animal studies failed to show consistent neurodevelopmental effects in the absence of AChE inhibition. That is, what is being argued to be happening in humans must for some reason not be happening in experimental animals, or that a sufficient number of studies with concentrations low enough to not perturb AChE activity have not been conducted, such that further research in this area is needed. It requires that one rely heavily on the truth of the mechanistic hypotheses that permit a biologically plausible MoA in the absence of AChE inhibition in the brain, even though effects at doses below the threshold for AChE inhibition were only observed in a few *in vitro* studies and have not been validated *in vivo*. These mechanisms are *ad hoc* rather than *a priori*, making it necessary to find independent, positive evidence of their operation in humans.

To accept this account as true, one must accept that somehow chlorpyrifos can enter the brain of the fetus or child after exposure to doses in the range of background levels in the general population and induce neurotoxicity in the developing brain *via* mechanisms that are independent of AChE inhibition, even though there is a large body of evidence that does not support this *in vivo*. The proposed mechanisms appear to have been chosen to fit the low-dose hypothesis already put forth in the epidemiology studies, and they are not based purely on an evaluation of the WoE as a whole. Because of this, the alternative account should be considered.

The alternative, competing account is that the few apparent associations from epidemiology studies do not indicate a causal connection and there is not adequate support for a biologically plausible mechanism for neurodevelopmental effects of chlorpyrifos in the absence of AChE inhibition in the brain. This account is supported by the totality of the data, which provides plausibility for the few associations observed in the epidemiology studies to be deemed false positive results. This account requires that one accept that the animal data indicate that a few neurodevelopmental effects are observed in the presence of some degree of AChE inhibition in the brain, but not necessarily at the level required for systemic toxicity, as numerous animal studies reported neurodevelopmental effects only in the presence of AChE inhibition or at concentrations shown to be associated with AChE inhibition in other studies. If this account is true, a causal relationship between chlorpyrifos exposure and neurodevelopmental effects in the absence of AChE inhibition in the brain would be understood as not plausible for humans, and the few positive associations observed in epidemiology studies would be attributed to alternative explanations.

When assessing the weight of the available evidence in support of the competing accounts, it is clear that the first account requires far more *ad hoc* assumptions and is not adequately supported by the data as a whole. Because of this, the WoE for this account is weak in comparison to the more substantial WoE supporting the lack of a causal association at chlorpyrifos doses below the threshold for inhibition of AChE activity in the brain.

4. US EPA, ECETOC, and HBWoE frameworks

US EPA and ECETOC have proposed frameworks specifically as guidance for weighing evidence in the context of evaluating potential human disease causation (US EPA, 2010; ECETOC, 2009). Below, we provide perspective on our approach compared with these frameworks, by describing and evaluating the US EPA and ECETOC frameworks and contrasting their rationales with that of the HBWoE approach.

4.1. US EPA framework for incorporating human epidemiology and incident data in health risk assessment

US EPA's OPP has proposed a Draft "Framework for Incorporating Epidemiologic and Incident Data in Health Risk Assessment" (US EPA, 2010). The framework is designed to incorporate epidemiology and human incident data into human health risk assessments specifically for pesticides and to be consistent with the NRC report of 21st Century Toxicity Testing (NRC, 2007). The NRC report promotes shifting toxicity testing away from apical toxicity endpoints to toxicity pathways (cellular response pathways) to inform potential adverse effects in humans and, ultimately, risk decision making. The US EPA framework proposes to use the Bradford Hill Criteria as modified in the MoA framework (US EPA, 1999, 2005; Sonich-Mullin *et al.*, 2001; Meek *et al.*, 2003; Seed *et al.*, 2005).

The US EPA framework first describes strengths and weaknesses of different types of epidemiology studies (*e.g.*, case-control, cohort, longitudinal, cross-sectional, ecologic) and important factors to consider when evaluating epidemiology data (*e.g.*, exposure assessment, confounding factors, statistical analysis, potential bias in observational research, interpretation of null studies, external validity). The framework describes the benefits and uses of epidemiology data, stating, "Epidemiology studies have the potential to help inform multiple components of the risk assessment in a variety of ways. High quality studies with robust exposure assessment may be used to estimate risk quantitatively. However, often due to resource constraints, most epidemiology studies suffer some limitations in size, scope, exposure assessment, or

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

data analysis which prevent their use in quantitative risk assessment (Caulderon [sic], 2000)." The framework describes how human studies are expected to play a significant role in the new vision of toxicity testing in the 21st century (NRC, 2007) in that human chemical exposure information can help guide *in vitro* and *in vivo* studies that are focused on investigating toxicity pathway *vs*. apical effect dose-response endpoints. Further, the framework describes how potential sources of uncertainty in animal data can be informed by human studies, emphasizing species extrapolation and population variability and "real-world exposures" *vs*. high-dose animal studies. US EPA notes that, while epidemiology studies can pose a challenge in interpretation, the evaluation of multiple routes and multiple chemical exposures may be very informative.

The US EPA framework next describes the utility of human incident data, including case reports and surveillance studies of acute pesticide poisoning incidents. These studies are often high dose exposures of short-term duration (frequently one-time exposures), and effects are often reversible. The framework indicates that human incident studies are often valuable because they can identify novel health effects potentially associated with a specific chemical (*e.g.*, pesticide) exposure, and can be compared to effects from acute animal studies. The US EPA framework describes how human incident data are used broadly to evaluate trends over time and patterns of severity and frequency of pesticide exposure, and further to inform risk assessment and risk management decisions. The framework describes the strengths and weaknesses of sources of human incident data: OPP Incident Data System (IDS); American Association of Poison Control Centers (PCC); National Pesticide Information Center (NPIC); Sentinel Event Notification System for Occupational Risk (SENSOR); California Pesticide Illness Surveillance Program (PISP).

The US EPA framework then describes how to conduct a WoE evaluation of the epidemiology and human incident data so that the conclusions are based on all of the available data rather than on any one study, and introduces the idea that multiple lines of evidence should be considered in addition to the epidemiology and human incident data (*i.e.*, *in vitro*, *in vivo*, and *in silico* data). The framework describes the specific steps involved in the WoE analysis, including, as a first step, a written review of each epidemiology and human incident study that describes the study design, results, conclusions, strengths and weaknesses, and the quality of the exposure assessment. The second step of the US EPA framework relies on the modified Bradford Hill Criteria as used in the MoA framework (US EPA, 1999, 2005; Sonich-Mullin *et al.*, 2001; Meek *et al.*, 2003; Seed *et al.*, 2005) which includes the following steps for organizing and integrating information: postulated MoA; key events; dose-response relationships; temporal association; strength, consistency, and specificity of association of key events and the toxic effect; biological plausibility and coherence; other potential MoAs. The framework briefly describes each step and the important factors to consider in the MoA analysis so that areas of uncertainty and areas of future research may be identified. The final step of the US EPA framework is the "Overall conclusions, statement of areas of confidence and uncertainty, and recommendation for risk assessment." The framework briefly states that this section should discuss the overall conclusions based on the WoE; that is, identify new areas of research, provide recommendations for source data for regulatory values, and extrapolate from animals to humans, if necessary.

The US EPA framework lays out the necessary elements that are important to consider for determining the strength and limitations of each epidemiologic or human incident data set, and suggests a written summary of each study. Although the framework provides a small discussion regarding how the overall conclusions and WoE should be presented, this discussion simply states that this should be done, but provides no guidance on how one should actually weigh the evidence. The framework is put forth as being designed to incorporate epidemiology and human incident data specifically into human health risk assessments for pesticides, but there is no discussion about how to actually *integrate* the epidemiologic data into a risk assessment or how to weigh this evidence with other types of data (*e.g.*, animal, pharmacokinetic, exposure, and MoA studies). The framework generally discusses how potential sources of uncertainty in animal data can be informed by human studies, and states that the framework is designed to include multiple lines of evidence (*i.e.*, epidemiology, toxicology, exposure, pharmacokinetic, and MoA data), but provides no real framework for how one should systematically consider all of the evidence.

As shown in Table 1, the US EPA framework provides guidance on key WoE aspects 1 and 2, as described above in Section 2, with a focus on epidemiology data. That is, the framework focuses mostly on the intrinsic quality of individual epidemiology studies and how to evaluate the body of epidemiology data across studies, but provides little guidance on how to integrate these data with animal studies or MoA data. Therefore, the framework is not really a framework *per se* for *integrating* epidemiology data with other data, but more a conceptual guidance on how to *evaluate* epidemiology studies. Although guidance on human data evaluation and interpretation is necessary, it is only part of what is required for integrating these data with other relevant data in evaluating human health risk and disease causation.

4.2. ECETOC framework for integration of human and animal data in chemical risk assessment

ECETOC has proposed a Framework for Integration of Human and Animal Data in Chemical Risk Assessments (ECETOC, 2009). The framework addresses quality aspects of both animal and human data, strongly encourages the use of both types of data in a combined approach, and suggests that human and animal data ideally should be "complimentary and should confirm each other (*i.e.*,

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

both indicate excess risk, or both indicate the absence of risk)." The framework indicates that when there are apparent contradictions, efforts should be made to better understand the biological basis of the contradictory evidence which will often further inform the risk assessment process.

Overall, the ECETOC framework involves three steps, which are discussed in more detail in the next sections:

1. Assessment of collective WoE of human data with regard to quality;
2. Assessment of collective WoE of animal data with regard to quality and relevance to humans;
3. Integration of the available evidence.

4.2.1. Human data quality in the ECETOC framework

The ECETOC framework provides an overview of different types of human data, including observational epidemiology (cross-sectional, cohort, and case-control) and controlled experimental studies, describing the quality aspects of human data (study design, exposure information, health outcome data, and other data quality issues similar to those discussed above in the US EPA framework), and how to evaluate the strengths and limitations of a single human study. The framework acknowledges that all data quality requirements are rarely met in epidemiology studies, which complicates their interpretation. The framework describes the criteria for evaluating the quality of human and animal data, the relevance of the animal data in evaluating human risk, and the integration of the two data sets.

The ECETOC framework describes Human Data Quality Criteria (based on the Bradford Hill Criteria), emphasizing that the criteria should be applied to human data in an integrated fashion and should generally consider the stage of the risk assessment. The framework provides techniques that can be used to arrive at a judgment about causal interpretations of each risk assessment stage. That is, the framework describes how Bradford Hill guidelines or meta-analyses can be used for hazard identification; how comparisons of LOAELs and NOAELs of different effects can be used to determine the critical or lead effect; and how different exposure levels should be considered to determine the appropriate dose-response curve to arrive at a LOAEL or NOAEL. The ECETOC framework emphasizes that determining whether an effect is critical depends on "the severity of the effect, its reversibility or whether it is deemed to be 'adverse.' "

The ECETOC framework provides a scheme for scoring the quality of human data. The framework first describes two prerequisites for human data that must be satisfied: (1) exposures must have occurred, and (2) the health effect should be determined adequately. If these criteria are not met, then the study is considered to provide "no information" and is not considered further in the evaluation. According to the framework, if a study meets these prerequisites, then the intrinsic quality of the human data are assessed as "High," "Good," "Compromised," "Poor," or "No Information." The framework provides fairly prescriptive guidance (checklists) on how to assess the quality of the study in terms of these five categories. Once the quality has been determined, the next step in the framework is to determine the nature of the health effect in the study (*i.e.*, chronic, sub-chronic, or acute, and whether the effect is specific or non-specific). The basis of the scoring considers that an effect that is chronic requires a stronger data set than an acute effect, with the order being (from less to more data required): acute specific effect < acute non-specific effect < sub-chronic or chronic specific effect < sub-chronic or chronic non-specific effect. Based on the combined nature and intrinsic quality of a study, the framework assigns a human data quality score of A, B, C, D, or X (*e.g.*, a high quality study with an acute specific effect would have a score of "A," and a poor quality study with a non-specific sub-chronic or chronic effect would have a score of "D," and various combinations would fall in between). The framework discusses how adjustments to the scoring scheme are possible, noting that identical considerations should go into evaluating both positive and null data, but that there are some exceptions (*e.g.*, size of population and confidence intervals of null studies) that should be considered that could change the scoring for null studies. The ECETOC framework states that, "A small positive study may be all that is needed to unequivocally suggest causality or, perhaps potency (*e.g.*, consider a lethal concentration of a substance); in contrast, the absence of effects usually requires larger population sizes." We address this point further in Section 4.2.5. The framework discusses that the stage in the risk assessment for which the data are applicable should affect the scoring because each stage has different requirements (*e.g.*, observation of an effect is important for hazard identification, but at this point in the assessment, exposure level is not important because dose-response is not being assessed).

4.2.2. Animal data quality in the ECETOC framework

The ECETOC framework describes quality aspects of the animal data, emphasizing that animal data need to be integrated with the human data, where good human and animal data are available. The three criteria ECETOC uses to describe the quality of animal data in human risk assessment are reliability, relevance, and adequacy, based on the criteria put forth by Klimisch *et al.* (1997) to determine whether data are sufficient or if more studies are needed in the context of the Organization for Economic Co-operation and Development (OECD) Existing Chemicals Program that is intended to ensure sufficient quality data for high production chemicals.

1. **Reliability** refers to the quality of the test, and takes into account whether standardized methodologies are used. The framework applies the four reliability

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

categories proposed by Klimisch *et al.* (1997) to determine reliability of the animal studies.

2. **Relevance** is defined as the extent to which data and tests are appropriate for a particular hazard, and how closely toxicity in a test species predicts toxicity in humans. Studies that may lack relevance are atypical species, *in vitro* studies with no *in vivo* confirmation of effect, or exposure routes that may not be relevant. The framework provides general guidelines for determining relevance.
3. **Adequacy** is defined as the usefulness of the data for hazard and/or risk assessment purposes. The framework discusses the importance of considering statistical significance, the types of effects observed (*i.e.*, adaptive *vs*. adverse; reversibility and severity as discussed by Lewis *et al.*, (2002)), whether effects could be due to chance, and consideration of concordance in deciding whether a study is adequate or not. The framework provides general guidelines for determining adequacy and guidance on what action to take when the data are discordant.

In reference to adequacy of an animal study, the ECETOC framework states that if there is discordance between animal species then the study of higher quality should be used, and if the studies are of equal quality, then the worst-case study should be used. The framework further states with regard to this point that, "If the discordance between animal studies cannot be rationalized in terms of MoA and the animal studies are both Klimisch category 1 or 2 (considered reliable without or with restrictions, respectively), then the worst-case data should be used in the assessment, while also taking quality aspects into account." We discuss this point further below.

4.2.3. Relevance of animal data in human risk assessment

The ECETOC framework next describes the relevance of the body of animal data to human risk assessment. The steps involved in interpreting the body of animal studies in the context of human risk, within the framework, are as follows: forming a MoA hypothesis; dosimetry; relevance to humans; and dose-route extrapolation.

The framework briefly discusses that the MoA hypothesis should be based on considering possible key events reviewed against the modified Bradford Hill Criteria (Seed *et al.*, 2005; Boobis *et al.*, 2006) to determine whether the WoE is sufficient to derive a MoA. The framework further describes how sometimes a MoA hypothesis is not possible based on a lack of data or fundamental understanding of the biology, and sometimes more than one plausible MoA is possible, but as discussed below, provides little guidance on what to do in these cases. The dosimetry step is briefly described within the framework as identification of toxicant, target organ, dose-response, and temporal aspects of dose-response, and that animal data and kinetic modeling techniques should be applied to animal data to analyze dosimetry.

The framework provides a scheme for determining relevance of animal data to human risk, based on considering each MoA key event in animals and its plausibility in humans (Boobis *et al.*, 2008; Seed *et al.*, 2005; Boobis *et al.*, 2006). The framework recommends an extended version of the Human Relevance Framework developed by the International Programme on Chemical Safety (IPCS), including the following scoring criteria for data quality and availability:

A. **Reliable animal data directly relevant to humans** (may also include null findings with confidence that they are applicable to humans)
B. **Reliable animal data relevant to humans** (less confident in how reliable null findings apply to humans)
C. **Reliable animal data, with uncertain but assumed relevance to humans**
D. **Unreliable animal data**
X. **No relevant animal data, or data not relevant to humans**

A flow diagram is provided in the ECETOC framework that incorporates reliability and relevance (in the context of MoA as outlined in the IPCS framework) into evaluating human relevance of animal data, with descriptions of how to determine the various categories A–X. The framework then describes dose-route extrapolation as the next step when the route of human exposure being considered is different from the critical animal study, briefly describing how dosimetry considerations and exposure modeling can assist in this step.

4.2.4. Integration of animal and human data

Finally, the ECETOC framework applies a matrix for placing the body of human and animal data into categories A-D so that it is clear what data (human or animal) are more relevant for the risk assessment. That is, the quality of the human data and the quality and relevance of the animal data are considered together and scored, for example, as A/A, A/B, C/C, *etc.* As such, the outcome will indicate which data should take precedence in the human risk assessment. The framework discusses that when the human and animal data have equivalent scores, the data needing fewer adjustments should be used, and that is typically the human data. Further, when the scoring is equivalent but the data are not concordant, the data suggesting a hazard should generally take precedence. If both suggest a hazard, the one with the lower level should take precedence, considering an upper bound for the other data source. The framework indicates that when the human data are scored as "A," then these data should take precedence regardless of animal data. We discuss this point further below.

ECETOC discusses that it is not possible to construct a matrix that can easily be applied to all situations, and that it is acceptable to deviate from the procedures outlined in the framework as long as they can be scientifically justified. The framework further indicates that if both human

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

and animal data are category "C" (*i.e.*, poor or compromised studies in humans, and reliable animal data but with uncertain relevance to humans), hazard and risk assessments should proceed with caution, particularly if the data are discordant.

The ECETOC framework provides 15 case study examples so that application and implementation of the framework can be more clearly understood.

4.2.5. Practical application of WoE in the ECETOC framework

The ECETOC framework defines the various categories of the animal data and human data and provides some guidance (mostly by example in case studies within which there are brief descriptions of categorization) about how to weigh all of the animal or human evidence in order to assign a category. The flowchart (Figure 3 of the ECETOC framework) and table (Table 1 of the ECETOC framework) of the framework provide some general guidance for categorization. The framework discusses the importance of human relevance when interpreting the animal data, and the category choice relies on what level of relevance is determined, but the framework could provide more guidance (even if by example) on exactly how to weigh all of the animal evidence so that one can assign the appropriate category of relevance to humans. The case studies include very brief descriptions of each step, providing little discussion of the actual process of weighing the evidence.

The scoring for determining human data quality within the framework consists of a checklist of requirements for each category (High, Good, Compromised, Poor, or No Information), with a prescribed number of requirements that need to be met for a study to fall into a given category. To determine whether each requirement is met, however, involves some judgment on the part of the risk assessor because the requirements are very qualitative. "Quality" is judged on intrinsic properties of the study conduct relative to prevailing standards for studies of that type, rather than on how study strengths and shortcomings affect the application of results to the causality questions at hand. Therefore, although some structure is useful, determining data quality should also consist of a narrative discussion of the logic for how the quality was determined. Further, the ECETOC framework discusses that there are some exceptions that should be considered that could change the scoring for null studies that may be different from positive studies (*e.g.*, size of population and confidence intervals of null studies). Identical considerations should go into evaluating both positive and null data, however, and the determination of study quality should be made based on methodology and not study results.

One question presented in the framework flowchart regarding relevance of animal data to humans (Figure 3 of the ECETOC framework) is whether the MoA is established in animals, with a yes or no answer guiding how to categorize the data. The ECETOC framework provides little guidance on how to actually weigh the MoA evidence in animals in order to determine whether an MoA has been established. The answer may not be a simple yes or no; if MoAs are suggested but not yet "established," it is not clear how the unproven but perhaps informative possibilities affect the interpretation of available results. The framework suggests that if the MoA has not been established in animals that it should be assumed that it is relevant to humans and given a data category of "C." There should be an option, however, to modify that categorization based on MoA data that may or may not be relevant to humans. That is, the animal MoA may not be definitively established, but there may be enough data to hypothesize several MoAs, and if so, it would be important to consider how plausible those proposed MoAs are in humans. This would involve considering animal, human, pharmacokinetic, exposure, and MoA data together, and allowing the data sets to inform one another in weighing all of the evidence to determine plausibility. This is often the case for the available data (*i.e.*, proposed MoAs that are being tested for a given causal question, as opposed to definitive MoAs that are already established).

Further, the framework discusses, with regard to discordance within the animal data, that "it should be noted that a lack of concordance between sexes or species, or even between strains of the same species, could provide invaluable information about the mode of action (MoA) of the substance....If the discordance between animal studies cannot be rationalized in terms of MoA and the animal studies are both Klimisch category 1 or 2 [considered reliable], *then the worst-case data should be used in the assessment*, while also taking quality aspects into account" [emphasis added]. The first part of this statement is true, in that lack of concordance between animal data could provide useful information with regard to mode of action in animals and humans, and this point should be emphasized more within the WoE for the framework. The framework does not elaborate on this point, however, but instead indicates that the worst-case data should be used. One should consider that, in a case where there is uncertainty in the MoA because of discordance in animals, particularly if the studies are reliable, this uncertainty implies that relevance to humans is uncertain and proceeding with the risk assessment may not be appropriate. These data should perhaps suggest more studies as the recommended next step (rather than assuming worst-case) to determine what is causing the lack of concordance. Understanding the lack of concordance in animals would provide useful information for potential human relevance. Further, although not mentioned in the ECETOC framework, different animal species may be more or less relevant to humans (*e.g.*, monkey data *vs*. rodent data) and this should be considered in the context of the question of the particular human disease causation (Gray *et al.*, 2005).

The ECETOC framework provides a methodology for integrating the animal and human effect data based on the scoring of each body of data (animal or human), with the outcome being whether precedence should be

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

given to human or animal data for the final risk assessment. The framework, however, does not explicitly consider how all of the data together (*i.e.*, in weighing all of the evidence), both negative and positive, can help inform interpretation of one another (animal, human, exposure, and MoA studies). Instead, it emphasizes choosing one basis for inference (the "best" according to the scoring scheme), taking that study's results as most indicative of the true situation, and not dealing with how discrepancies between this single basis and the remaining body of evidence is to be accounted for. Although it does discuss forming a MoA hypothesis and considering dosimetry as part of determining whether the animal MoA data are relevant to humans (based on the IPCS human relevance framework), it does not explicitly consider what to do when there is more than one plausible MoA, or the importance of how these MoA data may inform interpretation of the epidemiology or animal toxicology data.

The framework discusses, with regard to integrating the animal and human data, that when the scoring is equivalent for humans and animals, but the data are not concordant, the data suggesting a hazard should generally take precedence (apart from a category of A for human data, which should generally be given precedence regardless of whether an effect was observed), and if both suggest a hazard, the one with the lowest level should take precedence, considering an upper bound for the other data source. In this case, it would seem more appropriate to carefully analyze all of the data (animal, human, and MoA) to see how all of the data inform interpretation of each other, so that the logic can be clearly traced with regard to the choice of the dose-response information to use in the assessment. For example, in some cases it may seem more logical to accept category B human data with a less sensitive effect than the more sensitive category B animal data because there may be MoA data (in animals and/or humans) that suggest the proposed animal MoA is not plausible in humans. If nothing else, such a case should suggest further study to determine the plausibility of the animal MoA in humans.

Although intended to be flexible, the ECETOC framework perhaps provides too many steps and checklists so that each piece of the puzzle is dealt with separately, and perhaps eliminated before other data are considered that might have informed interpretation of that particular piece of data. Parts of the framework could be very useful, however, perhaps as tools applied to a more holistic approach to weighing all of the evidence. As noted in the conclusions of the ECETOC framework: "any attempt to systematize reporting, conduct, or classification of data is likely to be criticized. The objection is well-founded; if a classification framework is too rigid it can stifle creativity and if it is too lax, it may only provide the veneer of an evaluation. The Task Force believes that the primary benefit of the proposed ECETOC framework will be an evolving improvement towards the transparent evaluation and integration of human and animal data in the risk assessment process."

Overall, the ECETOC framework provides a very useful first step in ranking human and animal data quality and relevance of animal data to humans. The framework, however, would benefit from more discussion of: (i) importance of carefully weighing all the animal and human data, and guidance on how one should go about this, so that the appropriate data quality category can be determined; (ii) going beyond the categorization of animal and human effect data to integrate other important data sets (exposure, MoA, and pharmacokinetic); and (iii) the importance of considering all of the data together and how all of the data can inform each other (both negative and positive data, and of varying quality) so that the evidence as a whole can be truly weighed. As presented in Table 1, the ECETOC framework provides guidance for key aspects 1 and 2 of a WoE evaluation, and provides some guidance on key aspect 5 with regard to integrating human and animal data.

4.3. Comparison of HBWoE, US EPA, and ECETOC frameworks

Although the methodology varies, the three WoE frameworks described here each include a systematic review of the quality of the individual studies relevant to the question of human disease causation, and each examine the data within a particular line of investigation (*i.e.*, epidemiology, animal toxicology, or MoA studies) for particular endpoints, evaluating consistency, specificity, and reproducibility of outcomes (described in steps 1 and 2 in Table 1). One difference between the frameworks is that the US EPA framework focuses on epidemiology data, the ECETOC framework focuses on epidemiology and animal toxicology data, and the HBWoE framework evaluates all relevant data (*i.e.*, epidemiology, animal toxicology, and MoA data) individually and then within each line of investigation. Within these steps, the frameworks vary in the degree to which the steps are explicitly described, with the US EPA framework providing little prescriptive guidance and the ECETOC framework perhaps providing too much. We think the HBWoE framework falls somewhere in between, incorporating the key aspects of each, with more flexibility than the ECETOC framework, but a bit more guidance than the US EPA framework. Because our framework is intended to be flexible, various aspects of the US EPA and ECETOC frameworks could be applied within the HBWoE framework for the first two steps of the evaluation.

As summarized in Table 1, the US EPA and ECETOC frameworks provide guidance for only the first two key aspects of a WoE evaluation. Although the US EPA and ECETOC frameworks discuss the importance of integrating all of the relevant data, there is little guidance on how to actually do that. The HBWoE framework goes beyond these steps in that it integrates all of the relevant data, within the context of proposed hypotheses, so that each line of evidence can inform interpretation of one

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

another (steps 3–5 in Table 1). The integration includes, but goes well beyond, simply noting the patterns and degrees of concordance and discordance, among studies within a realm (*i.e.*, across human studies or in different sexes and species of animals); it takes the stance that what makes data relevant to human hazard identification is the hypothesized commonality that the source data and the target human population have in how the agent acts to provoke observed effects in the source and, presumably, similar or at least mechanistically related effects in the target population. Under this view, discordant results are possible but require their own (at least tentative) explanations for why the causal process proposed to be common to the key studies and the target human population do not apply to the discordant studies, and the WoE for human hazard is judged by the success and biological plausibility of the set of explanations of concordant as well as discordant results. That is, the HBWoE framework compares the various accounts of the observations at hand, discussing consistencies and inconsistencies within the data and the *ad hoc* assumptions required to support each account, and tracing the logic and reasoning for how the data support (or do not support) each account's hypotheses (step 6). In this way, the HBWoE framework does not seek to prove or disprove any one hypothesis, nor to definitively choose one and reject the others; rather, it seeks to present the lines of reasoning for each account of the observations so that the data will speak for themselves.

As part of comparing various accounts, the HBWoE framework will often require tracing the logic and reasoning for how a poor quality study is used to support a particular line of argument (or hypothesis). In contrast to the ECETOC framework that explicitly describes how poor quality studies should be eliminated early on in the evaluation, a key aspect of the HBWoE framework is that all data, positive and negative, and of varying quality (even poor quality) are maintained and carried through the evaluation. Poor quality studies may have some useful information, and it should not be taken as self-evident that their results are false; rather, such outcomes should have lesser weight. In HBWoE, this lesser weight arises naturally from the consideration of the comparative inability of poor studies (compared to more robust ones) to provide outcomes that differentiate between the generally operating causal factors being evaluated and extraneous, study-specific explanations that could produce spurious outcomes. In a HBWoE evaluation, the logic for how a poor quality study fits (or does not fit) with the available data needs to be considered and articulated as part of one account of the observations at hand, so that it can be compared to other accounts of the available data.

In tracing the logic and reasoning for how certain studies or lines of evidence fit (or do not fit) with the available data, the HBWoE framework necessitates inclusion of all data relevant to the various hypotheses that have been put forth. By contrast, checklists do not work in weighing all of the evidence if they lead one to make certain assumptions about a given study or data set without consideration of all of the data when enough features of the array of criteria seem to fit. In fact, the criteria developed by Bradford Hill (which he called "postulates") were designed to articulate the basis for judgments and facilitate the integration of evaluations across criteria, and were not intended simply as checklists from which causality could be concluded. Hill saw the postulates as guides to thinking rather than as measures of evidence. The HBWoE framework emphasizes the importance of how each piece of information (positive or negative) might inform interpretation of one another, or how studies of varying quality (even poor quality) need to be considered insofar as they have bearing on distinguishing between alternative explanations, particularly if the study is the basis for a particular line of argument that needs to be articulated as part of one account, and in this way integrates all data relevant to questions of potential human disease causation.

In comparison to the other frameworks, there are aspects of more traditional WoE approaches that the HBWoE framework does not have. The outcome of the HBWoE framework is complex and not easily summarized succinctly. The HBWoE framework does not arrive at decisions, but it is used to inform decision makers by characterizing uncertainty and plausibility of alternative conclusions. The HBWoE framework is not readily codified, so can be quite complicated in practice, requiring deep and broad expertise. Although judgments will still be needed, and these judgments will instill scientific debate, the debate can be more clearly focused on the scientific bases of the various lines of argument.

4.4. Comparison of frameworks in the context of chlorpyrifos

Applying the HBWoE framework to evaluate the health effects of chlorpyrifos, we concluded that the most likely account of the epidemiology, animal toxicology, and mechanistic data is that the few apparent associations from epidemiology studies do not indicate a causal connection and there is not adequate support for a biologically plausible mechanism for neurodevelopmental effects in the absence of AChE inhibition in the brain.

The US EPA and ECETOC frameworks provide guidance for reviewing studies systematically and examining the consistency, specificity, and relevance of outcomes across studies. They do not, however, provide guidance on how to integrate all of the relevant data or how to use each line of evidence to inform the integration of other kinds of data. For example, these frameworks do not provide information regarding how the results of chlorpyrifos toxicology and mechanistic studies should inform the interpretation of the epidemiology studies. In this case, the toxicology and mechanistic data

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

indicate a lack of effects at exposures below those which cause AChE inhibition, casting doubt on the validity of positive associations in epidemiology studies. The US EPA and ECETOC frameworks also do not compare the different accounts. That is, while they aim to determine how well the data support a specific hypothesis, they do not consider how well the data support alternative hypotheses, nor do they explicitly address the question of why discordant results exist and how these should be accounted for. By using the HBWoE framework, it is clear that the WoE supports the account of no causation much more than the account of causation, and that more *ad hoc* assumptions are required to support the causation account. This is not as evident using the other frameworks.

Both the US EPA and ECETOC frameworks focus on the "best" studies. That is, while they discuss reviewing all the studies, they both come down to focusing on a specific set of data, sometimes ignoring other data. For example, US EPA might choose the most robust epidemiology study, but not consider whether results from this study are consistent with those of less robust studies, or whether this study is robust enough to draw conclusions. Similarly, ECETOC states that poor quality studies should be eliminated early on. In a HBWoE evaluation, the logic for how a poor quality study fits (or does not fit) with the available data is considered and articulated as part of one account of the observations at hand; poor studies are poor because they fail to discriminate between the causal hypothesis being evaluated and other extraneous explanations of their outcomes, so they do not help to differentiate the relative plausibility of competing accounts. Thus, using either of these frameworks would have resulted in some or several of the chlorpyrifos studies being ignored, and their role in each account would have been missed.

5. Conclusions

As regulatory agencies make greater use of human data in chemical risk assessments, it will be a challenge to determine how to assess all of the data that are relevant to the question of human disease causation. We compared three frameworks that have been proposed to guide risk assessors in this endeavor and assessed how well each framework incorporates key aspects of WoE.

While the three WoE frameworks each include a systematic review of the quality of the individual studies and examine the data within a particular line of evidence (*i.e.*, epidemiology, animal toxicology, or MoA studies), the US EPA and ECETOC frameworks provide little guidance for integrating all of the relevant data. By contrast, the HBWoE framework integrates all of the relevant data within the context of proposed hypotheses, so that each line of evidence can inform the interpretation of one another. Further, the HBWoE framework compares the various accounts of the observations at hand, discussing consistencies and inconsistencies within the data and the *ad hoc* assumptions required to support each account, tracing the logic and reasoning for how the data support (or do not support) each account's hypotheses. The HBWoE framework emphasizes the importance of how each piece of information (positive or negative) might inform interpretation of one another, and how studies of varying quality (even poor quality) need to be considered, characterizing the uncertainty and plausibility of alternative conclusions while integrating all of the data relevant to potential human disease causation questions.

In our application of the HBWoE framework to evaluate the data relevant to examining whether there is a causal association between exposure to chlorpyrifos and adverse neurodevelopmental effects in humans, we found that the epidemiology data are not sufficiently robust to support the hypothesis that chlorpyrifos is a causal factor for neurodevelopmental effects. The available studies do not report consistent results, and there is a lack of clear exposure-response information. Because of the many uncertainties in these studies, the few positive results may also support alternative explanations that other factors are causal for the reported outcomes, or that the observed associations are statistical anomalies. In addition, the animal toxicity data do not provide clear evidence that chlorpyrifos is associated with neurodevelopmental effects at doses that are below the threshold for inhibition of AChE in the brain; this would be relevant to exposures in the epidemiology studies, which are at least 1000-fold lower than those used in the animal studies. Further, the mechanisms proposed to underlie potential neurodevelopmental effects in humans at doses below those associated with inhibition of AChE activity in the brain have not been shown to operate in the developing brain *in vivo* at concentrations that do not inhibit the activity of AChE.

For chlorpyrifos to act as a neurodevelopmental toxicant at the near-background exposure levels in the epidemiology studies, it must be accepted that chlorpyrifos can enter the brain of the fetus or child after exposure to doses in the range of background levels in the general population and induce neurotoxicity in the developing brain *via* mechanisms that are independent of AChE inhibition, even though there is a large body of evidence that does not support this *in vivo*. Rather, the few apparent associations from epidemiology studies are not indicative of a causal connection, and there is not adequate support for a biologically plausible mechanism for neurodevelopmental effects of chlorpyrifos in the absence of AChE inhibition in the brain. The weight of the available evidence more strongly indicates that a causal association between chlorpyrifos exposure and neurodevelopmental effects in the absence of AChE inhibition in the brain is not plausible for humans, and the few positive associations observed in epidemiology studies would be attributed to alternative explanations.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

Declaration of interest

This paper was prepared with financial support to Gradient, a private environmental consulting firm, by Dow AgroSciences. The work reported in this paper was conducted during the normal course of employment. The authors have the sole responsibility for the writing and contents of this paper.

References

Abou-Donia MB, Khan WA, Dechkovskaia AM, Goldstein LB, Bullman SL, Abdel-Rahman A. (2006). In utero exposure to nicotine and chlorpyrifos alone, and in combination produces persistent sensorimotor deficits and Purkinje neuron loss in the cerebellum of adult offspring rats. Arch Toxicol, 80:620–631.

Achenbach T, Rescorla L. (2000). Manual for the ASEBA Preschool Forms and Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Achenbach TM, Dumenci L, Rescorla LA. (2003). DSM-oriented and empirically based approaches to constructing scales from the same item pools. J Clin Child Adolesc Psychol, 32:328–340.

Adigun AA, Ryde IT, Seidler FJ, Slotkin TA. (2010). Organophosphate exposure during a critical developmental stage reprograms adenylyl cyclase signaling in PC12 cells. Brain Res, 1329:36–44.

Agency for Toxic Substances and Disease Registry (ATSDR). (1997). Toxicological profile for chlorpyrifos (Update - Draft for public comment). Research Triangle Park, NC.

Ahdaya SM, Monroe RJ, Guthrie FE. (1981). Absorption and distribution of intubated insecticides in fasted mice. Pestic Biochem Physiol 16:38–46.

Aiuto LA, Pavlakis SG, Boxer RA. (1993). Life-threatening organophosphate-induced delayed polyneuropathy in a child after accidental chlorpyrifos ingestion. J Pediatr, 122:658–660.

Albers JW, Cole P, Greenberg RS, Mandel JS, Monson RR, Ross JH, Snodgrass WR, Spurgeon A, van Gemert M. (1999). Analysis of chlorpyrifos exposure and human health: Expert panel report. J Toxicol Environ Health B Crit Rev, 2:301–324.

Albers JW, Berent S, Garabrant DH, Giordani B, Schweitzer SJ, Garrison RP, Richardson RJ. (2004a). The effects of occupational exposure to chlorpyrifos on the neurologic examination of central nervous system function: A prospective cohort study. J Occup Environ Med, 46:367–378.

Albers JW, Garabrant DH, Schweitzer SJ, Garrison RP, Richardson RJ, Berent S. (2004b). The effects of occupational exposure to chlorpyrifos on the peripheral nervous system: A prospective cohort study. Occup Environ Med, 61:201–211.

Albers JW, Garabrant DH, Schweitzer S, Garrison RP, Richardson RJ, Berent S. (2004c). Absence of sensory neuropathy among workers with occupational exposure to chlorpyrifos. Muscle Nerve, 29:677–686.

Albers JW, Garabrant DH, Mattsson JL, Burns CJ, Cohen SS, Sima C, Garrison RP, Richardson RJ, Berent S. (2007). Dose-effect analyses of occupational chlorpyrifos exposure and peripheral nerve electrophysiology. Toxicol Sci, 97:196–204.

Aldridge JE, Levin ED, Seidler FJ, Slotkin TA. (2005a). Developmental exposure of rats to chlorpyrifos leads to behavioral alterations in adulthood, involving serotonergic mechanisms and resembling animal models of depression. Environ Health Perspect, 113:527–531.

Aldridge JE, Meyer A, Seidler FJ, Slotkin TA. (2005b). Alterations in central nervous system serotonergic and dopaminergic synaptic activity in adulthood after prenatal or neonatal chlorpyrifos exposure. Environ Health Perspect, 113:1027–1031.

Aldridge JE, Seidler FJ, Slotkin TA. (2004). Developmental exposure to chlorpyrifos elicits sex-selective alterations of serotonergic synaptic function in adulthood: Critical periods and regional selectivity for effects on the serotonin transporter, receptor subtypes, and cell signaling. Environ Health Perspect, 112:148–155.

American Psychiatric Association (APA). (2000). Diagnostic and Statistical Manual of Mental Disorders, 4th Edition. Washington, DC: American Psychiatric Association.

Aprea C, Sciarra G, Sartorelli P, Desideri E, Amati R, Sartorelli E. (1994). Biological monitoring of exposure to organophosphorus insecticides by assay of urinary alkylphosphates: Influence of protective measures during manual operations with treated plants. Int Arch Occup Environ Health, 66:333–338.

Bakke JE, Feil VJ, Price CE. (1976). Rat urinary metabolites from O,O-diethyl-O-(3,5,6-trichloro-2-pyridyl) phosphorothioate. J Environ Sci Health B, 11:225–230.

Barr DB, Angerer J. (2006). Potential uses of biomonitoring data: A case study using the organophosphorus pesticides chlorpyrifos and malathion. Environ Health Perspect, 114:1763–1769.

Barr DB, Ananth CV, Yan X, Lashley S, Smulian JC, Ledoux TA, Hore P, Robson MG. (2010). Pesticide concentrations in maternal and umbilical cord sera and their relation to birth outcomes in a population of pregnant women and newborns in New Jersey. Sci Total Environ, 408:790–795.

Barr DB, Barr JR, Maggio VL, Whitehead RD Jr, Sadowski MA, Whyatt RM, Needham LL. (2002). A multi-analyte method for the quantification of contemporary pesticides in human serum and plasma using high-resolution mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci, 778:99–111.

Bellinger DC. (2005). Neurobehavioral assessment in studies of exposures to neurotoxicants. Int Rev Res Ment Retard 30:263–300.

Berger-Sweeney J, Hohmann CF. (1997). Behavioral consequences of abnormal cortical development: Insights into developmental disabilities. Behav Brain Res, 86:121–142.

Berkowitz GS, Wetmur JG, Birman-Deych E, Obel J, Lapinski RH, Godbold JH, Holzman IR, Wolff MS. (2004). In utero pesticide exposure, maternal paraoxonase activity, and head circumference. Environ Health Perspect, 112:388–391.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D, Farland W. (2006). IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit Rev Toxicol, 36:781–792.

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

Bradman A, Whitaker D, Quirós L, Castorina R, Claus Henn B, Nishioka M, Morgan J, Barr DB, Harnly M, Brisbin JA, Sheldon LS, McKone TE, Eskenazi B. (2007). Pesticides and their metabolites in the homes and urine of farmworker children living in the Salinas Valley, CA. J Expo Sci Environ Epidemiol, 17:331–349.

Braquenier JB, Quertemont E, Tirelli E, Plumier JC. (2010). Anxiety in adult female mice following perinatal exposure to chlorpyrifos. Neurotoxicol Teratol, 32:234–239.

Brazelton TB, Nugent JK. (1995). Neonatal Behavioral Assessment Scale (Third Edition). London, United Kingdom: Mac Keith Press.

Calderon RL. (2000). Measuring risks in humans: The promise and practice of epidemiology. Food Chem Toxicol, 38:S59–S63.

Carr RL, Chambers HW, Guarisco JA, Richardson JR, Tang J, Chambers JE. (2001). Effects of repeated oral postnatal exposure to chlorpyrifos on open-field behavior in juvenile rats. Toxicol Sci, 59:260–267.

Castorina R, Bradman A, McKone TE, Barr DB, Harnly ME, Eskenazi B. (2003). Cumulative organophosphate pesticide exposure and risk assessment among pregnant women living in an agricultural community: A case study from the CHAMACOS cohort. Environ Health Perspect, 111:1640–1648.

Centers for Disease Control and Prevention (CDC). (2009). Fourth National Report on Human Exposure to Environmental Chemicals. (Online) Available at: http://www.cdc.gov/exposurereport/pdf/FourthReport.pdf. Accessed on 11 December 2009.

Chakrabortí TK, Farrar JD, Pope CN. (1993). Comparative neurochemical and neurobehavioral effects of repeated chlorpyrifos exposures in young and adult rats. Pharmacol Biochem Behav, 46:219–224.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Chambers JE, Chambers HW. (1989). Oxidative desulfuration of chlorpyrifos, chlorpyrifos-methyl, and leptophos by rat brain and liver. J Biochem Toxicol, 4:201–203.

Chanda SM, Mortensen SR, Moser VC, Padilla S. (1997). Tissue-specific effects of chlorpyrifos on carboxylesterase and cholinesterase activity in adult rats: An *in vitro* and *in vivo* comparison. Fundam Appl Toxicol, 38:148–157.

Chanda SM, Pope CN. (1996). Neurochemical and neurobehavioral effects of repeated gestational exposure to chlorpyrifos in maternal and developing rats. Pharmacol Biochem Behav, 53:771–776.

Clegg DJ, van Gemert M. (1999). Determination of the reference dose for chlorpyrifos: Proceedings of an expert panel. J Toxicol Environ Health B Crit Rev, 2:211–255.

Cole TB, Jampsa RL, Walter BJ, Arndt TL, Richter RJ, Shih DM, Tward A, Lusis AJ, Jack RM, Costa LG, Furlong CE. (2003). Expression of human paraoxonase (PON1) during development. Pharmacogenetics, 13:357–364.

Coulston F, Griffin T, Golberg L. (1972). Safety Evaluation of Dowco 179 in Human Volunteers. Albany, NY: Institute of Experimental Pathology and Toxicology, Albany Medical College.

Crumpton TL, Seidler FJ, Slotkin TA. (2000). Is oxidative stress involved in the developmental neurotoxicity of chlorpyrifos? Brain Res Dev Brain Res, 121:189–195.

Dam K, Garcia SJ, Seidler FJ, Slotkin TA. (1999). Neonatal chlorpyrifos exposure alters synaptic development and neuronal activity in cholinergic and catecholaminergic pathways. Brain Res Dev Brain Res, 116:9–20.

Dam K, Seidler FJ, Slotkin TA. (2000). Chlorpyrifos exposure during a critical neonatal period elicits gender-selective deficits in the development of coordination skills and locomotor activity. Brain Res Dev Brain Res, 121:179–187.

Das KP, Barone S Jr. (1999). Neuronal differentiation in PC12 cells is inhibited by chlorpyrifos and its metabolites: Is acetylcholinesterase inhibition the site of action? Toxicol Appl Pharmacol, 160:217–230.

Eaton DL, Daroff RB, Autrup H, Bridges J, Buffler P, Costa LG, Coyle J, McKhann G, Mobley WC, Nadel L, Neubert D, Schulte-Hermann R, Spencer PS. (2008). Review of the toxicology of chlorpyrifos with an emphasis on human exposure and neurodevelopment. Crit Rev Toxicol, 38 Suppl 2:1–125.

Engel SM, Berkowitz GS, Barr DB, Teitelbaum SL, Siskind J, Meisel SJ, Wetmur JG, Wolff MS. (2007). Prenatal organophosphate metabolite and organochlorine levels and performance on the Brazelton Neonatal Behavioral Assessment Scale in a multiethnic pregnancy cohort. Am J Epidemiol, 165:1397–1404.

Engel SM, Wetmur K, Chen J, Zhu C, Barr DB, Canfield RL, Wolff MS. (2011). Prenatal exposure to organophosphates, paraoxonase 1, and cognitive development in childhood. Environ Health Perspect. doi:10.1289/ehp.1003183.

Eskenazi B, Harley K, Bradman A, Weltzien E, Jewell NP, Barr DB, Furlong CE, Holland NT. (2004). Association of in utero organophosphate pesticide exposure and fetal growth and length of gestation in an agricultural population. Environ Health Perspect, 112:1116–1124.

Eskenazi B, Marks AR, Bradman A, Harley K, Barr DB, Johnson C, Morga N, Jewell NP. (2007). Organophosphate pesticide exposure and neurodevelopment in young Mexican-American children. Environ Health Perspect, 115:792–798.

European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC). (2009). Framework of the Integration of Human and Animal Data in Chemical Risk Assessment. ECETOC Technical Report No. 104. p. 130.

Eyer P. (2003). The role of oximes in the management of organophosphorus pesticide poisoning. Toxicol Rev, 22:165–190.

Goodman JE, Dodge DG, Bailey LA. (2010). A framework for assessing causality and adverse effects in humans with a case study of sulfur dioxide. Regul Toxicol Pharmacol, 58:308–322.

Hack M, Breslau N, Weissman B, Aram D, Klein N, Borawski E. (1991). Effect of very low birth weight and subnormal head size on cognitive abilities at school age. N Engl J Med, 325:231–237.

Haviland JA, Butz DE, Porter WP. (2010). Long-term sex selective hormonal and behavior alterations in mice exposed to low doses of chlorpyrifos in utero. Reprod Toxicol, 29:74–79.

Hill AB. (1965). The environment and disease: Association or causation? Proc R Soc Med, 58:295–300.

Howard AS, Bucelli R, Jett DA, Bruun D, Yang D, Lein PJ. (2005). Chlorpyrifos exerts opposing effects on axonal and dendritic growth in primary neuronal cultures. Toxicol Appl Pharmacol, 207:112–124.

Huff RA, Corcoran JJ, Anderson JK, Abou-Donia MB. (1994). Chlorpyrifos oxon binds directly to muscarinic receptors and inhibits cAMP accumulation in rat striatum. J Pharmacol Exp Ther, 269:329–335.

Icenogle LM, Christopher NC, Blackwelder WP, Caldwell DP, Qiao D, Seidler FJ, Slotkin TA, Levin ED. (2004). Behavioral alterations in adolescent and adult rats caused by a brief subtoxic exposure to chlorpyrifos during neurulation. Neurotoxicol Teratol, 26:95–101.

Ikonomidou C, Kaindl AM. (2011). Neuronal death and oxidative stress in the developing brain. Antioxid Redox Signal, 14:1535–1550.

Jett DA, Navoa RV, Beckles RA, McLemore GL. (2001). Cognitive function and cholinergic neurochemistry in weanling rats exposed to chlorpyrifos. Toxicol Appl Pharmacol, 174:89–98.

Johnson FO, Chambers JE, Nail CA, Givaruangsawat S, Carr RL. (2009). Developmental chlorpyrifos and methyl parathion exposure alters radial-arm maze performance in juvenile and adult rats. Toxicol Sci, 109:132–142.

Karanth S, Pope C. (2000). Carboxylesterase and A-esterase activities during maturation and aging: Relationship to the toxicity of chlorpyrifos and parathion in rats. Toxicol Sci, 58:282–289.

Kisicki JC, Seip CW, Combs ML. (1999). A Rising Dose Toxicology Study to Determine the No-Observable-Effect-Levels (NOEL) for Erythrocyte Acetylcholinesterase (AChE) Inhibition and Cholinergic Signs and Symptoms of Chlorpyrifos at Three Dose Levels. Report to Dow AgroSciences LLC. p. 40.

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

Krimsky S. (2005). The weight of scientific evidence in policy and law. Am J Public Health, 95 Suppl 1:S129–S136.

Landrigan PJ, Claudio L, Markowitz SB, Berkowitz GS, Brenner BL, Romero H, Wetmur JG, Matte TD, Gore AC, Godbold JH, Wolff MS. (1999). Pesticides and inner-city children: Exposures, risks, and prevention. Environ Health Perspect, 107 Suppl 3:431–437.

Lasky RE, Klein RE, Yarbrough C, Engle PL, Lechtig A, Martorell R. (1981). The relationship between physical growth and infant behavioral development in rural Guatemala. Child Dev, 52:219–226.

Laviola G, Adriani W, Gaudino C, Marino R, Keller F. (2006). Paradoxical effects of prenatal acetylcholinesterase blockade on neuro-behavioral development and drug-induced stereotypies in reeler mutant mice. Psychopharmacology (Berl), 187:331–344.

Lemons JA, Schreiner RL, Gresham EL. (1981). Relationship of brain weight to head circumference in early infancy. Hum Biol, 53:351–354.

Lester BM, Als H, Brazelton TB. (1982). Regional obstetric anesthesia and newborn behavior: A reanalysis toward synergistic effects. Child Dev, 53:687–692.

Levin ED, Addy N, Baruah A, Elias A, Christopher NC, Seidler FJ, Slotkin TA. (2002). Prenatal chlorpyrifos exposure in rats causes persistent behavioral alterations. Neurotoxicol Teratol, 24:733–741.

Levin ED, Addy N, Nakajima A, Christopher NC, Seidler FJ, Slotkin TA. (2001). Persistent behavioral consequences of neonatal chlorpyrifos exposure in rats. Brain Res Dev Brain Res, 130:83–89.

Lewis RW, Billington R, Debryune E, Gamer A, Lang B, Carpanini F. (2002). Recognition of adverse and nonadverse effects in toxicity studies. Toxicol Pathol, 30:66–74.

Li WF, Matthews C, Disteche CM, Costa LG, Furlong CE. (1997). Paraoxonase (PON1) gene in mice: Sequencing, chromosomal

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

localization and developmental expression. Pharmacogenetics, 7:137–144.

Linkov I, Loney D, Cormier S, Satterstrom FK, Bridges T. (2009). Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. Sci Total Environ, 407:5199–5205.

Lotti M, Moretto A, Zoppellari R, Dainese R, Rizzuto N, Barusco G. (1986). Inhibition of lymphocytic neuropathy target esterase predicts the development of organophosphate-induced delayed polyneuropathy. Arch Toxicol, 59:176–179.

MacIntosh DL, Needham LL, Hammerstrom KA, Ryan PB. (1999). A longitudinal investigation of selected pesticide metabolites in urine. J Expo Anal Environ Epidemiol, 9:494–501.

MacLusky NJ, Naftolin F. (1981). Sexual differentiation of the central nervous system. Science, 211:1294–1302.

Marty MS, Domoradzki JY, Hansen SC, Timchalk C, Bartels MJ, Mattsson JL. (2007). The effect of route, vehicle, and divided doses on the pharmacokinetics of chlorpyrifos and its metabolite trichloropyridinol in neonatal Sprague-Dawley rats. Toxicol Sci, 100:360–373.

Maurissen JP, Hoberman AM, Garman RH, Hanley TR Jr. (2000). Lack of selective developmental neurotoxicity in rat pups from dams treated by gavage with chlorpyrifos. Toxicol Sci, 57:250–263.

McMullin TS, Lowe ER, Bartels MJ, Marty MS. (2008). Dynamic changes in lipids and proteins of maternal, fetal, and pup blood and milk during perinatal development in CD and Wistar rats. Toxicol Sci, 105:260–274.

Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD, Longfellow DG, Pastoor T, Seed J, Patton DE. (2003). A framework for human relevance analysis of information on carcinogenic modes of action. Crit Rev Toxicol, 33:591–653.

Meyer A, Seidler FJ, Aldridge JE, Tate CA, Cousins MM, Slotkin TA. (2004). Critical periods for chlorpyrifos-induced developmental neurotoxicity: Alterations in adenylyl cyclase signaling in adult rat brain regions after gestational or neonatal exposure. Environ Health Perspect, 112:295–301.

Massoulié J, Sussman J, Bon S, Silman I. (1993). Structure and functions of acetylcholinesterase and butyrylcholinesterase. Prog Brain Res, 98:139–146.

Morgan MK, Sheldon LS, Croghan CW, Jones PA, Robertson GL, Chuang JC, Wilson NK, Lyu CW. (2005). Exposures of preschool children to chlorpyrifos and its degradation product 3,5,6-trichloro-2-pyridinol in their everyday environments. J Expo Anal Environ Epidemiol, 15:297–309.

Moretto A, Lotti M. (1998). Poisoning by organophosphorus insecticides and sensory neuropathy. J Neurol Neurosurg Psychiatr, 64:463–468.

Moser VC, Chanda SM, Mortensen SR, Padilla S. (1998). Age- and gender-related differences in sensitivity to chlorpyrifos in the rat reflect developmental profiles of esterase activities. Toxicol Sci, 46:211–222.

Mueller RF, Hornung S, Furlong CE, Anderson J, Giblett ER, Motulsky AG. (1983). Plasma paraoxonase polymorphism: A new enzyme assay, population, family, biochemical, and linkage studies. Am J Hum Genet, 35:393–408.

Muto MA, Lobelle F Jr, Bidanset JH, Wurpel JN. (1992). Embryotoxicity and neurotoxicity in rats associated with prenatal exposure to DURSBAN. Vet Hum Toxicol, 34:498–501.

Namba T, Nolte CT, Jackrel J, Grob D. (1971). Poisoning due to organophosphate insecticides. Acute and chronic manifestations. Am J Med, 50:475–492.

National Research Council (NRC) of the National Academies. (2007). Toxicity Testing in the Twenty-first Century: A Vision and a Strategy. Washington, DC: National Academies Press.

Needham LL. (2005). Assessing exposure to organophosphorus pesticides by biomonitoring in epidemiologic studies of birth outcomes. Environ Health Perspect, 113:494–498.

Nolan RJ, Rick DL, Freshour NL, Saunders JH. (1984). Chlorpyrifos: Pharmacokinetics in human volunteers. Toxicol Appl Pharmacol, 73:8–15.

Ounsted M, Moar VA, Scott A. (1988). Head circumference and developmental ability at the age of seven years. Acta Paediatr Scand, 77:374–379.

Palmer JS, Rowe LD, Crookshank HR. (1980). Effect of age on tolerance of calves to chlorpyrifos. Am J Vet Res, 41:1323–1325.

Perera FP, Rauh V, Tsai WY, Kinney P, Camann D, Barr D, Bernert T, Garfinkel R, Tu YH, Diaz D, Dietrich J, Whyatt RM. (2003). Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population. Environ Health Perspect, 111:201–205.

Pope CN, Chakraborti TK. (1992). Dose-related inhibition of brain and plasma cholinesterase in neonatal and adult rats following sublethal organophosphate exposures. Toxicology, 73:35–43.

Pope CN, Chakraborti TK, Chapman ML, Farrar JD, Arthun D. (1991). Comparison of *in vivo* cholinesterase inhibition in neonatal and adult rats by three organophosphorothioate insecticides. Toxicology, 68:51–61.

Qiao D, Seidler FJ, Padilla S, Slotkin TA. (2002). Developmental neurotoxicity of chlorpyrifos: What is the vulnerable period? Environ Health Perspect, 110:1097–1103.

Qiao D, Seidler FJ, Abreu-Villaça Y, Tate CA, Cousins MM, Slotkin TA. (2004). Chlorpyrifos exposure during neurulation: Cholinergic synaptic dysfunction and cellular alterations in brain regions at adolescence and adulthood. Brain Res Dev Brain Res, 148:43–52.

Qiao D, Seidler FJ, Slotkin TA. (2005). Oxidative mechanisms contributing to the developmental neurotoxicity of nicotine and chlorpyrifos. Toxicol Appl Pharmacol, 206:17–26.

Qiao D, Seidler FJ, Tate CA, Cousins MM, Slotkin TA. (2003). Fetal chlorpyrifos exposure: Adverse effects on brain cell development and cholinergic biomarkers emerge postnatally and continue into adolescence and adulthood. Environ Health Perspect, 111:536–544.

Rauh V, Arunajadai S, Horton M, Perera F, Hoepner L, Barr DB, Whyatt R. (2011). Seven-year neurodevelopmental scores and prenatal exposure to chlorpyrifos, a common agricultural pesticide. Environ Health Perspect, 119:1196–1201.

Rauh VA, Garfinkel R, Perera FP, Andrews HF, Hoepner L, Barr DB, Whitehead R, Tang D, Whyatt RW. (2006). Impact of prenatal chlorpyrifos exposure on neurodevelopment in the first 3 years of life among inner-city children. Pediatrics, 118:e1845–e1859.

Rauh VA, Whyatt RM, Garfinkel R, Andrews H, Hoepner L, Reyes A, Diaz D, Camann D, Perera FP. (2004). Developmental effects of exposure to environmental tobacco smoke and material hardship among inner-city children. Neurotoxicol Teratol, 26:373–385.

Rhomberg LR, Bailey LA, Goodman JE, Hamade AK, Mayfield D. (2011). Is exposure to formaldehyde in air causally associated with leukemia? – A hypothesis-based weight-of-evidence analysis. Crit Rev Toxicol (in press).

Rhomberg LR, Bailey LA, Goodman JE. (2010). Hypothesis-based weight of evidence: A tool for evaluating and communicating uncertainties and inconsistencies in the large body of evidence in proposing a carcinogenic mode of action–naphthalene as an example. Crit Rev Toxicol, 40:671–696.

Ricceri L, Markina N, Valanzano A, Fortuna S, Cometa MF, Meneguz A, Calamandrei G. (2003). Developmental exposure to chlorpyrifos alters reactivity to environmental and social cues in adolescent mice. Toxicol Appl Pharmacol, 191:189–201.

Ricceri L, Venerosi A, Capone F, Cometa MF, Lorenzini P, Fortuna S, Calamandrei G. (2006). Developmental neurotoxicity of organophosphorous pesticides: Fetal and neonatal exposure to chlorpyrifos alters sex-specific behaviors at adulthood in mice. Toxicol Sci, 93:105–113.

Rice DC. (2005). From animals to humans: Models and constructs. Int Rev Res Ment Retard 30:301–337.

Richardson RJ. (1995). Assessment of the neurotoxic potential of chlorpyrifos relative to other organophosphorus compounds: A critical review of the literature. J Toxicol Environ Health, 44:135–165.

Rushton JP, Ankney CD. (2009). Whole brain size and general mental ability: A review. Int J Neurosci, 119:691–731.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK

Saulsbury MD, Heyliger SO, Wang K, Johnson DJ. (2009). Chlorpyrifos induces oxidative stress in oligodendrocyte progenitor cells. Toxicology, 259:1–9.

Schoenemann PT, Budinger TF, Sarich VM, Wang WS. (2000). Brain size does not predict general cognitive ability within families. Proc Natl Acad Sci USA, 97:4932–4937.

Schuh RA, Lein PJ, Beckles RA, Jett DA. (2002). Noncholinesterase mechanisms of chlorpyrifos neurotoxicity: Altered phosphorylation of Ca2+/cAMP response element binding protein in cultured neurons. Toxicol Appl Pharmacol, 182:176–185.

Seed J, Carney EW, Corley RA, Crofton KM, DeSesso JM, Foster PM, Kavlock R, Kimmel G, Klaunig J, Meek ME, Preston RJ, Slikker W Jr, Tabacova S, Williams GM, Wiltse J, Zoeller RT, Fenner-Crisp P, Patton DE. (2005). Overview: Using mode of action and life stage information to evaluate the human relevance of animal toxicity data. Crit Rev Toxicol, 35:664–672.

Shah PV, Fisher HL, Sumler MR, Monroe RJ, Chernoff N, Hall LL. (1987). Comparison of the penetration of 14 pesticides through the skin of young and adult rats. J Toxicol Environ Health, 21:353–366.

Sharbaugh C, Viet SM, Fraser A, McMaster SB. (2003). Comparable measures of cognitive function in human infants and laboratory animals to identify environmental health risks to children. Environ Health Perspect, 111:1630–1639.

Slotkin TA. (2004). Cholinergic systems in brain development and disruption by neurotoxicants: Nicotine, environmental tobacco smoke, organophosphates. Toxicol Appl Pharmacol, 198:132–151.

Slotkin TA, MacKillop EA, Ryde IT, Seidler FJ. (2007). Ameliorating the developmental neurotoxicity of chlorpyrifos: A mechanisms-based approach in PC12 cells. Environ Health Perspect, 115:1306–1313.

Slotkin TA, Oliver CA, Seidler FJ. (2005). Critical periods for the role of oxidative stress in the developmental neurotoxicity of chlorpyrifos and terbutaline, alone or in combination. Brain Res Dev Brain Res, 157:172–180.

Slotkin TA, Seidler FJ. (2007). Prenatal chlorpyrifos exposure elicits presynaptic serotonergic and dopaminergic hyperactivity at adolescence: Critical periods for regional and sex-selective effects. Reprod Toxicol, 23:421–427.

Slotkin TA, Seidler FJ. (2009). Oxidative and excitatory mechanisms of developmental neurotoxicity: Transcriptional profiles for chlorpyrifos, diazinon, dieldrin, and divalent nickel in PC12 cells. Environ Health Perspect, 117:587–596.

Slotkin TA, Tate CA, Cousins MM, Seidler FJ. (2002). Functional alterations in CNS catecholamine systems in adolescence and adulthood after neonatal chlorpyrifos exposure. Brain Res Dev Brain Res, 133:163–173.

Smith GN, Watson BS, Fischer FS. (1967). Investigations on dursban insecticide: Metabolism of [36Cl] O,O-diethyl-O- 3,5,6-trichloro-2-pyridyl phosphorothioate in rats. J Agric Food Chem 15:132–138.

Song X, Seidler FJ, Saleh JL, Zhang J, Padilla S, Slotkin TA. (1997). Cellular mechanisms for developmental toxicity of chlorpyrifos: Targeting the adenylyl cyclase signaling cascade. Toxicol Appl Pharmacol, 145:158–174.

Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J, Fenner-Crisp P, Grant D, Hartley M, Knaap A, Kroese D, Mangelsdorf I, Meek E, Rice JM, Younes M; International Programme on Chemical Safety. (2001). IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul Toxicol Pharmacol, 34:146–152.

Sood B, Delaney-Black V, Covington C, Nordstrom-Klee B, Ager J, Templin T, Janisse J, Martier S, Sokol RJ. (2001). Prenatal alcohol exposure and childhood behavior at age 6 to 7 years: I. dose-response effect. Pediatrics, 108:E34.

Sultatos LG, Murphy SD. (1983). Kinetic analyses of the microsomal biotransformation of the phosphorothioate insecticides chlorpyrifos and parathion. Fundam Appl Toxicol, 3:16–21.

Sultatos LG, Murphy SD. (1983b). Hepatic microsomal detoxification of the organophosphates paraoxon and chlorpyrifos oxon in the mouse. Drug Metab Dispos, 11:232–238.

Timchalk C, Nolan RJ, Mendrala AL, Dittenber DA, Brzak KA, Mattsson JL. (2002). A Physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model for the organophosphate insecticide chlorpyrifos in rats and humans. Toxicol Sci, 66:34–53.

Ulbrich B, Palmer AK. (1996). Neurobehavioral aspects of developmental toxicity testing. Environ Health Perspect, 104 Suppl 2:407–412.

US EPA. (1999). Guidelines for carcinogen risk assessment (SAB Review Draft). Risk Assessment Forum, NCEA-F-0644. [Online] Available at: www.epa.gov/ncea/raf/crasab.htm. Accessed on 27 June 2001.

US EPA. (2002). Interim reregistration eligibility decision for chlorpyrifos. EPA 738-R-01-007. [Online] Available at: http://www.epa.gov/oppsrrd1/REDs/chlorpyrifos_ired.pdf.

US EPA. (2005). Guidelines for Carcinogen Risk Assessment. Washington, DC: Risk Assessment Forum. EPA/630/P-03/001F. p. 166.

US EPA. (2010). Framework for Incorporating Human Epidemiologic & Incident Data in Health Risk Assessment (Draft). Office of Pesticide Programs. p. 68.

Venerosi A, Cutuli D, Colonnello V, Cardona D, Ricceri L, Calamandrei G. (2008). Neonatal exposure to chlorpyrifos affects maternal responses and maternal aggression of female mice in adulthood. Neurotoxicol Teratol, 30:468–474.

Venerosi A, Ricceri L, Rungi A, Sanghez V, Calamandrei G. (2010). Gestational exposure to the organophosphate chlorpyrifos alters social-emotional behaviour and impairs responsiveness to the serotonin transporter inhibitor fluvoxamine in mice. Psychopharmacology (Berl), 208:99–107.

Venerosi A, Ricceri L, Scattoni ML, Calamandrei G. (2009). Prenatal chlorpyrifos exposure alters motor behavior and ultrasonic vocalization in CD-1 mouse pups. Environ Health, 8:12.

Webb SJ, Monk CS, Nelson CA. (2001). Mechanisms of postnatal neurobiological development: Implications for human development. Dev Neuropsychol, 19:147–171.

Weed DL. (2005). Weight of evidence: A review of concept and methods. Risk Anal, 25:1545–1557.

Wessels D, Barr DB, Mendola P. (2003). Use of biomarkers to indicate exposure of children to organophosphate pesticides: Implications for a longitudinal study of children's environmental health. Environ Health Perspect, 111:1939–1946.

Whitney KD, Seidler FJ, Slotkin TA. (1995). Developmental neurotoxicity of chlorpyrifos: Cellular mechanisms. Toxicol Appl Pharmacol, 134:53–62.

Whyatt RM, Camann DE, Kinney PL, Reyes A, Ramirez J, Dietrich J, Diaz D, Holmes D, Perera FP. (2002). Residential pesticide use during pregnancy among a cohort of urban minority women. Environ Health Perspect, 110:507–514.

Whyatt RM, Garfinkel R, Hoepner LA, Andrews H, Holmes D, Williams MK, Reyes A, Diaz D, Perera FP, Camann DE, Barr DB. (2009). A biomarker validation study of prenatal chlorpyrifos exposure within an inner-city cohort during pregnancy. Environ Health Perspect, 117:559–567.

Whyatt RM, Rauh V, Barr DB, Camann DE, Andrews HF, Garfinkel R, Hoepner LA, Diaz D, Dietrich J, Reyes A, Tang D, Kinney PL, Perera FP. (2004). Prenatal insecticide exposures and birth weight and length among an urban minority cohort. Environ Health Perspect, 112:1125–1132.

Winneke G. (1992). Cross species extrapolation in neurotoxicology: Neurophysiological and neurobehavioral aspects. Neurotoxicology, 13:15–25.

Wolff MS, Engel S, Berkowitz G, Teitelbaum S, Siskind J, Barr DB, Wetmur J. (2007). Prenatal pesticide and PCB exposures and birth outcomes. Pediatr Res, 61:243–250.

Yan X, Lashley S, Smulian JC, Ananth CV, Barr DB, Ledoux TA, Hore P, Robson MG. (2009). Pesticide concentrations in matrices collected in the perinatal period in a population of pregnant women and newborns in New Jersey, USA. Hum Ecol Risk Assess 15:948–967.

Yang D, Howard A, Bruun D, Ajua-Alemanj M, Pickart C, Lein PJ. (2008). Chlorpyrifos and chlorpyrifos-oxon inhibit axonal growth by interfering with the morphogenic activity of acetylcholinesterase. Toxicol Appl Pharmacol, 228:32–41.

Young JG, Eskenazi B, Gladstone EA, Bradman A, Pedersen L, Johnson C, Barr DB, Furlong CE, Holland NT. (2005). Association between

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

RIGHTSLINK

in utero organophosphate pesticide exposure and abnormal reflexes in neonates. Neurotoxicology, 26:199–209.

Youngstrom E, LaKind JS, Kenworthy L, Lipkin PH, Goodman M, Squibb K, Mattison DR, Anthony BJ, Anthony LG. (2010). Advancing the selection of neurodevelopmental measures in epidemiological studies of environmental chemical exposure and health effects. Int J Environ Res Public Health, 7:229–268.

Critical Reviews in Toxicology Downloaded from informahealthcare.com by 173.10.127.137 on 01/04/12
For personal use only.

© 2011 Informa Healthcare USA, Inc.

RIGHTSLINK