COMMON METHODOLOGICAL ERRORS IN DIOXIN RESEARCH: ASSUMPTIONS ABOUT CONFOUNDING, MISCLASSIFICATION, AND STATISTICAL ANALYSES

Scott LLF*

Division of Environmental Health Sciences, University of Minnesota School of Public Health, Minneapolis, MN, USA

Introduction

Over the last decade novel advances in epidemiologic methods have revitalized attempts to disentangle intricate causal relationships between environmental exposures and disease. Regrettably, few of these methodological improvements have been applied to human dioxin research, with most studies utilizing classical epidemiological methods and failing to address statistical and model assumptions that are typically not satisfied. Here, we reviewed the human dioxin literature to identify the most common epidemiological mistakes in dioxin research and discuss how current thinking about epidemiologic methods can be used to improve the quality of these studies.

Materials and Methods

A search of the peer-reviewed literature was conducted to identify epidemiological studies evaluating the association between exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and/or dioxin-like compounds and health outcomes. Various combinations of keywords and phrases – including but not limited to 'epidemiology,' 'health effects,' 'mortality,' 'cancer,' 'incidence', 'exposure,' 'dioxin,' and 'TCDD' – were used to identify relevant articles published in English and indexed in online databases (i.e. PubMed, Medline, Toxline, etc). Applicable studies were also identified by conducting general internet searches and reviewing secondary references from relevant studies and annual indices of selected journals. The Methods and Discussion/Conclusion sections of each article were reviewed to ascertain study limitations/methodological errors and any acknowledgement and/or evaluation of the impact of these imperfections on the results.

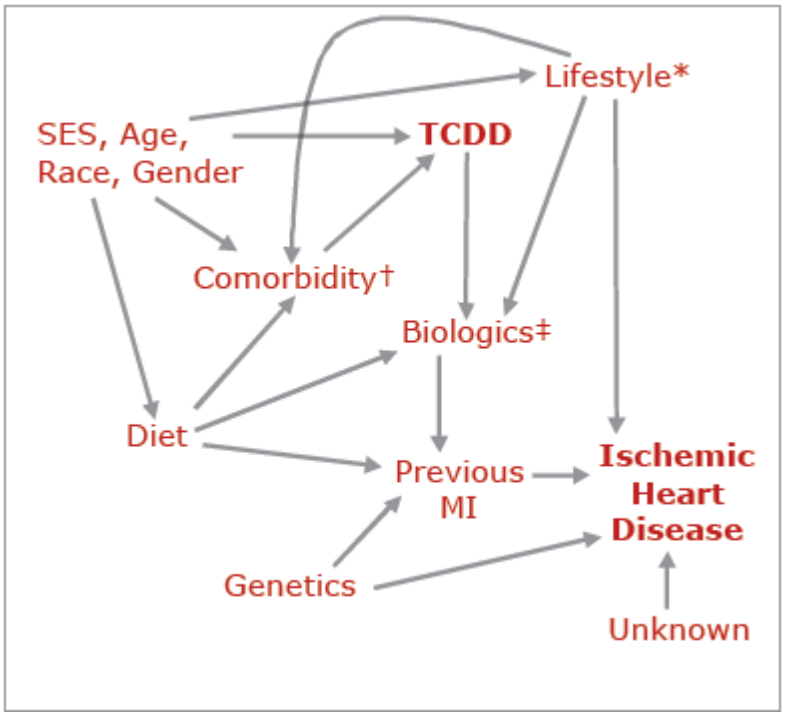
Results and Discussion

A prominent shortcoming of the epidemiological dioxin literature is the presence, and often lack of acknowledgement, of confounding. More specifically, studies of dioxin exposure and health outcomes do not, and frequently cannot, isolate the true effect of exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) from the effects of other dioxin-like compounds and/or mixtures of other chemical exposures. Although more recent studies have begun to examine the effect of dioxin mixtures on health outcomes,[1-7] a large majority of epidemiological studies have focused only on TCDD as the chemical exposure of interest.[8] For instance, of the 13 studies evaluating cardiovascular disease (CVD) mortality in occupational cohorts, only five assessed the combined impact of multiple congeners on disease and just three of these estimated exposure using total toxic equivalencies (TEQ), posing a critical problem when making causal inferences about the TCDD-CVD relationship.

Additionally, most epidemiological studies of dioxin exposure examine the mortality experience of exposed populations. While these types of studies can be useful, caution should be exercised when interpreting the study results, particularly if mortality rates in the population of interest are compared solely to the rates of an external referent population. In particular, comparison of workers to the general population can result in confounding since employed workers are often healthier and have differing characteristics compared to the general population. Consequently, the use of nested case-control, Poisson or time-dependent proportional hazards regression models is suggested for mortality cohort studies as these methods utilize an internal referent population. Notable examples of these approaches have been employed in only a few dioxin studies,[5,9-15] but suggest these methods may be useful techniques for managing specific aspects of confounding in dioxin research, particularly if the internal referent population is a suitable substitute population with which to compare the exposed individuals.

In general, human dioxin studies over-adjust for a multitude of covariates or, more frequently, neglect to adjust for confounders specific to the disease of interest. A potential solution to this problem, and a new technique with which to help identify possible confounders, is the use of directed acyclic graphs (DAGs). DAGs are diagrams based on graph theory that do not contain directed cycles and can be used as a tool to determine which variables to include in explicit exposure-disease epidemiological models. DAGs are constructed based on *a priori* assumptions about the causal relationships of identified characteristics and can help refine epidemiologic methods based on specific study goals by guiding data collection, analysis and modeling methods and answering causal questions for different causal models. An example of a DAG describing the relationship between exposure to TCDD, ischemic heart disease, and a multitude of other factors that may affect exposure and disease incidence is presented in Figure 1.

Figure 1: Directed Acyclic Graph for Identifying Possible Confounders of the Relationship between Exposure to TCDD and Ischemic Heart Disease



SES, socioeconomic status; MI, myocardial infarction
*Smoking, stress, physical activity
†Diabetes, hypertension, and other comorbid conditions
‡High density lipoprotein (HDL), low density lipoprotein (LDL), triglycerides, and other biometric screening tests

While DAGs provide a new approach for identifying potential confounders of exposure-disease relationships, they can be challenging to draw and inclusion of latent (i.e. unmeasured) variables can pose additional problems to their application. More importantly, adjustment for covariates identified as confounders using DAGs may not be sufficient to control for confounding. For example, Table 1 presents the distribution for a disease risk factor by exposure and population type, where doomed indicates that a population of individuals will become ill regardless of exposure and immune represents a population of individuals who, despite exposure status, will never develop the disease (i.e., exposure has no effect on disease outcome). Although only the "Total" and "Incidence" data would be observable in an actual study, the counts and proportions presented demonstrate that the risk factor is related to exposure and predicts disease among unexposed individuals. Yet, if the risk factor is not controlled for, the proportion of diseased individuals in the exposed and unexposed groups is equal, as it should be since exposure does not affect the outcome. On the contrary, adjusting for the risk factor produces estimates that erroneously suggest the risk factor inhibits the development of disease in both strata.

Table 1: Hypothetical Distribution for a Disease Risk Factor by Exposure and Individual Type*

	Risk Factor Present		Risk Factor Absent		Crude	
Individual Sub-type	Exposed	Unexposed	Exposed	Unexposed	Exposed	Unexposed
Doomed[†]	60	70	40	180	100	250
Immune[‡]	40	30	60	220	100	250
Total	100	100	100	400	200	500
Incidence	0.60	0.70	0.40	0.45	0.50	0.50

*From Greenland and Robbins (1986)[16]
[†]Become ill regardless of exposure
[‡]Are immune to disease regardless of exposure

Misclassification of both exposures and health outcomes also provide an ever-present challenge when interpreting the dioxin epidemiological literature. Misclassification of even a few individuals in studies with small numbers of exposed (or unexposed) cases can result in substantial changes to the estimate of effect. Such an example is demonstrated in Tables 2A and 2B. Based on the incorrect classification of exposure (Table 2A), the observed OR is OR = (61*451)/(1073*14) = 1.83. If three of the cases that were actually ever-exposed were misclassified as never-exposed, the true effect of dioxin exposure on ischemic heart disease would be (64*451)/(1073*11) = 2.45, a larger effect than the misclassified OR.

Table 2A: Misclassification of Exposure by Disease Status*

	Ischemic Heart Disease +	Ischemic Heart Disease -	Total
TCDD Ever-Exposed	61	1073	1134
TCDD Never-Exposed	14	451	465
Total	75	1524	1599

*Adapted from McBride et al. (2009)[15]

Table 2B: True Classification of Exposure by Disease Status

	Ischemic Heart Disease +	Ischemic Heart Disease -	Total
TCDD Ever-Exposed	64	1073	1137
TCDD Never-Exposed	11	451	462
Total	75	1524	1599

Additionally, the common assertion that non-differential misclassification causes estimates of effect to be biased toward the null is erroneous. Indeed, exact non-differential misclassification only results in bias toward the null under exceedingly precise conditions[17-18] and, even so, does not necessitate that the measure of association underestimates the true effect of exposure on disease (i.e. the reported effect could be biased toward the null and still overestimate the true effect).

Undoubtedly, one of the more conspicuous errors in the epidemiological dioxin literature is the excessive reliance on statistical significance. An over-emphasis on reporting significant associations and associations with the largest effects in peer-reviewed articles has resulted, generally, in systematic publication bias and outcome reporting bias.[19-21] When reporting study findings, consideration must be given to the number of outcomes evaluated and both significant and non-significant results. Reporting only statistically significant findings can result in type I errors. For example, if 100 endpoints are examined and a study uses a type I error probability (i.e. alpha) of 0.05, then we can expect five significant results due to chance alone. Another concern with reporting only statistically significant results is the manipulability of statistical methods. Often the probability of making a type I error is set at 0.05 or 0.10. When alpha = 0.10, a p-value of 0.06 would be statistically significant. A p-value of 0.06 would not be statistically significant, however, when alpha = 0.05. De-emphasizing the weight given to statistical significance is essential considering that it often cannot truly be detected since a good mathematical equation for random error has not been defined for non-randomized (i.e. observational) studies and the calculation of a p-value assumes no

systematic error is present. Rather, attention should focus on the size of the estimate of effect and uncertainty resulting from misclassification, bias, and confounding. This is particularly important since a measure of effect can be elevated but not significant and still represent an important finding that needs to be explored further, whereas some statistically significant results may not be meaningful at all (e.g., OR = 1.03 and 95% Confidence Interval = 1.02 – 1.04).

As with most epidemiologic research, there are several limitations of the published literature on dioxin exposure and human disease occurrence. While it is important to acknowledge the shortcomings of these studies, the benefit of quantifying how study imperfections impact measures of effect is becoming more evident as the desire to identify biologically relevant causal relationships intensifies within the field. Conducting sensitivity and/or uncertainty analyses are possible solutions to address study weaknesses in addition to using novel methods for identification of confounders and to manage confounding. Moreover, it is vital that interpretation and pertinence of epidemiological dioxin research is not contingent on statistical significance alone, but rather focuses on causal inference.

References:

1. Burns JS, Williams PL, Sergeyev O, Korrick S, Lee MM, Revich B, Altshul L, Del Prato JT, Humblet O, Patterson DG Jr, Turner WE, Needham LL, Starovoytov M, Hauser R. (2011) *Pediatrics*. 127(1):e59-68
2. Viel JF, Floret N, Deconinck E, Focant JF, De Pauw E, Cahn JY. (2011) *Environ Int*. 37(2):449-453
3. Cai LY, Izumi S, Suzuki T, Goya K, Nakamura E, Sugiyama T, Kobayashi H. (2011) *Hum Reprod*. 26(1):117-126
4. Chang JW, Ou HY, Chen HL, Guo HR, Liao PC, Lee CC. (2010) *Chemosphere*. 81(8):1027-1032
5. Collins JJ, Bodner K, Aylward LL, Wilken M, Bodnar CM. (2009) *Am J Epidemiol*. 170(4):501-506
6. Konishi K, Sasaki S, Kato S, Ban S, Washino N, Kajiwara J, Todaka T, Hirakawa H, Hori T, Yasutake D, Kishi R. (2009) *Environ Res*. 109(7):906-913
7. Wilhelm M, Wittsiepe J, Lemm F, Ranft U, Krämer U, Fürst P, Röseler SC, Greshake M, Imöhl M, Eberwein G, Rauchfuss K, Kraft M, Winneke G. (2008) *Mutat Res*. 659(1-2):83-92
8. Agency for Toxic Substances and Disease Registry. (1998) Toxicological Profile for Chlorinated Dibenzo-p-dioxins (CDDs). Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service
9. Kogevinas M, Kauppinen T, Winkelmann R, Becher H, Bertazzi PA, Bueno-de-Mesquita HB, Coggon D, Green L, Johnson E, Littorin M, Lynge E, Marlow DA, Mathews JD, Neuberger M, Benn T, Pannett B, Pearce N, Saracci R. (1995) *Epidemiol*. 6(4):396-402
10. Ott MG, Zober A. (1996) *Occup Environ Med*. 53(9):606-612
11. Kogevinas M, Becher H, Benn T, Bertazzi PA, Boffetta P, Bueno-de-Mesquita HB, Coggon D, Colin D, Flesch-Janys D, Fingerhut M, Green L, Kauppinen T, Littorin M,Lynge E, Mathews JD, Neuberger M, Pearce N, Saracci R. (1997) *Am J Epidemiol*. 145(12):1061-1075
12. Vena J, Boffetta P, Becher H, Benn T, Bueno-de-Mesquita HB, Coggon D, Colin D, Flesch-Janys D, Green L, Kauppinen T, Littorin M, Lynge E, Mathews JD, Neuberger M, Pearce N, Pesatori AC, Saracci R, Steenland K, Kogevinas M. (1998) *Environ Health Perspect*. 106 Suppl 2:645-653
13. Steenland K, Piacitelli L, Deddens J, Fingerhut M, Chang LI. (1999) *J Natl Cancer Inst*. 91(9):779-786
14. Consonni D, Pesatori AC, Zocchetti C, Sindaco R, D'Oro LC, Rubagotti M, Bertazzi PA. (2008) *Am J Epidemiol*. 167(7):847-858
15. McBride DI, Collins JJ, Humphry NF, Herbison P, Bodner KM, Aylward LL, Burns CJ, Wilken M. (2009) *J Occup Environ Med*. 51(9):1049-1056
16. Greenland S, Robins JM. (1986) *Int J Epidemiol.* 15(3):413-419
17. Jurek AM, Greenland S, Maldonado G. (2008) *Int J Epidemiol.* 37(2):382-385
18. Jurek AM, Greenland S, Maldonado G, Church TR. (2005) *Int J Epidemiol.* 34:680-687
19. Phillips CV. (2004) *BMC Med Res Methodol.* 4:20-30
20. Kavvoura FK, Liberopoulos G, Ioannidis JPA. (2007) *PLoS Med* 4(3): e79
21. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghersi D, Ioannidis JPA, Simes J, Williamson PR. (2008) *PLoS ONE* 3(8): e3081