ADJUSTING FOR DISEASE MISCLASSIFICATION DUE TO LOSS TO FOLLOW-UP IN A HISTORICAL COHORT OF TRICHLOROPHENOL WORKERS

Scott LLF*, Maldonado G

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

Introduction

Retrospective mortality studies are one of the most common tools used by epidemiologists, particularly in occupational settings, to evaluate associations between exposure and disease. In general, this type of study offers numerous benefits in that they typically require less time to complete, are inexpensive compared to other types of studies, and are better suited for evaluating multiple outcomes and occurrences of rare diseases. However, historical cohort studies are also prone to loss to follow-up, with a common method for addressing this being to assume that the individuals lost are alive at the end of the study. Occasionally, sensitivity analyses are conducted to assess the impact of such an assumption, but rarely is the effect of any potential bias from disease misclassification due to loss to follow-up thoroughly evaluated or quantified. In a recent study,[1] we described how the application of a method that is relatively new to epidemiology, probabilistic uncertainty analysis (PUA, also known as probabilistic bias analysis), could be used to quantify the effect of disease and exposure misclassification on the odds ratio for occupational exposure to TCDD-contaminated chemicals and ischemic heart disease mortality in a historical cohort of New Zealand trichlorophenol (TCP) workers. Here, we have developed a PUA model to quantify and adjust for disease misclassification from loss to follow-up in the same study.

Materials and Methods

Data from McBride et al.[2] were used to calculate a crude odds ratio and 95% confidence interval (Table 1) for the association between TCDD exposure and ischemic heart disease mortality. In the analysis presented here, we chose to compare those TCP production workers that had the highest cumulative TCDD exposure (n = 162) with those at the facility that were considered never-exposed (n = 465) because 1) using an internal referent group lessens the potential for confounding and healthy worker bias and 2) comparing those workers with the highest exposures to those who have no occupational exposure helps prevent diluting any effect, which could result from combining all exposed workers, including those with much lower exposures, into one group.[3-5]

Table 1: Cell Counts, Odds Ratio and 95% Confidence Limits for the Association between TCDD Exposure and Ischemic Heart Disease Mortality Using Data from McBride et al.[2]

	TCDD Exposure			
IHD Outcome	*>2085.8 ppt TCDD-mo*	*Never*	**CRUDE ODDS RATIO**	**95% CONFIDENCE LIMITS**
IHD Cases	14	14	3.05	1.42, 6.54
IHD Non-cases	148	451		

The mathematical relationship between a causal odds ratio, an observed odds ratio, and error terms for study bias has been previously described by Maldonado.[6] Equation 1 is a modification of this relationship, where $OR_{DM\text{-}LTF}$ is the odds ratio adjusted for disease misclassification due to loss to follow up, $OR_{observed}$ is the observed crude odds ratio, and E_i are the error terms describing the impact of systematic study error. Here, only one error is being evaluated such that the denominator may be simplified to $E_{DM\text{-}LTF}$, the error term for disease misclassification resulting from loss to follow up.

$$OR_{DM-LTF} = \frac{OR_{observed}}{\prod_{i=1}^{n} E_i} = \frac{OR_{observed}}{E_{DM-LTF}} \quad (1)$$

In McBride et al.,[2] a total of 338 individuals were loss to follow-up. To estimate the number of workers lost to follow-up that could have died from IHD for each exposure category, we used a multi-step process to identify the relevant classification parameter at each step and to specify an appropriate uncertainty distribution. First, a distribution for the total number of those lost to follow-up that may have died from any cause was defined using the proportions of known deaths observed for the cohort. We then used 2008 mortality data for the New Zealand population[7] to determine probability distributions for the total number of deaths from IHD for all exposure levels combined. Finally, distributions for the number of IHD deaths for the never-exposed group and the number of deaths for the workers with the highest exposure were specified. Each distribution was constructed using expert judgment and was varied to account for potential differences in mortality due to exposure status (all-cause and IHD mortality), gender (IHD mortality) and ethnicity (IHD mortality), resulting in 18 scenarios (Tables 2A and 2B). The components of Tables 2A and 2B were then used to generate adjusted counts of cases and non-cases by exposure group. For each of the 18 scenarios, 10,000 trials were sampled with Crystal Ball software (version 11), which uses Monte Carlo simulation techniques. Each analysis was conducted separately to generate a frequency distribution for $OR_{DM\text{-}LTF}$. The lower 2.5 and upper 97.5 percentiles of each frequency distribution were used to estimate 95% credible intervals for the adjusted odds ratios.

Table 2A: Description of Probability Distributions Used to Determine the Number of Total All-cause Deaths and the Number of Total IHD Deaths

	Total All-cause Deaths		**Total IHD Deaths**	
Scenario	*Likeliest N (%)*	*Distribution (Parameters)*	*Likeliest N (%)*	*Distribution (Parameters)*
1	169 (50.0)	Negative Binomial[a] (0.018, 4)	34 (20.4)	BetaPERT[b] (0, 34, Total # Deaths)
2	169 (50.0)	Negative Binomial (0.018, 4)	34 (20.4)	BetaPERT (0, 34, Total # Deaths)
3	169 (50.0)	Negative Binomial (0.018, 4)	34 (20.4)	BetaPERT (0, 34, Total # Deaths)
4	169 (50.0)	Negative Binomial (0.018, 4)	23 (13.9)	BetaPERT (0, 23, Total # Deaths)
5	169 (50.0)	Negative Binomial (0.018, 4)	23 (13.9)	BetaPERT (0, 23, Total # Deaths)
6	169 (50.0)	Negative Binomial (0.018, 4)	23 (13.9)	BetaPERT (0, 23, Total # Deaths)
7	104 (30.9)	Negative Binomial (0.02, 3)	21 (20.4)	BetaPERT (0, 21, Total # Deaths)
8	104 (30.9)	Negative Binomial (0.02, 3)	21 (20.4)	BetaPERT (0, 21, Total # Deaths)
9	104 (30.9)	Negative Binomial (0.02, 3)	21 (20.4)	BetaPERT (0, 21, Total # Deaths)
10	104 (30.9)	Negative Binomial (0.02, 3)	14 (13.9)	BetaPERT (0, 14, Total # Deaths)
11	104 (30.9)	Negative Binomial (0.02, 3)	14 (13.9)	BetaPERT (0, 14, Total # Deaths)
12	104 (30.9)	Negative Binomial (0.02, 3)	14 (13.9)	BetaPERT (0, 14, Total # Deaths)
13	37 (11.0)	Negative Binomial (0.027, 2)	8 (20.4)	BetaPERT (0, 8, Total # Deaths)
14	37 (11.0)	Negative Binomial (0.027, 2)	8 (20.4)	BetaPERT (0, 8, Total # Deaths)
15	37 (11.0)	Negative Binomial (0.027, 2)	8 (20.4)	BetaPERT (0, 8, Total # Deaths)
16	37 (11.0)	Negative Binomial (0.027, 2)	5 (13.9)	BetaPERT (0, 5, Total # Deaths)
17	37 (11.0)	Negative Binomial (0.027, 2)	5 (13.9)	BetaPERT (0, 5, Total # Deaths)
18	37 (11.0)	Negative Binomial (0.027, 2)	5 (13.9)	BetaPERT (0, 5, Total # Deaths)

[a]Negative binomial distribution (probability, shape)
[b]BetaPERT distribution (minimum, likeliest, maximum)

Table 2B: Description of Probability Distributions Used to Determine the Number of IHD Deaths by Exposure Status

	IHD Deaths by Exposure Status	
Scenario	*Distribution (Parameters) - Never*	*Distribution (Parameters) - >2085.8 ppt TCDD-mo*
1	BetaPERT[a] (0, 1/3 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed[b])
2	BetaPERT (0, 1/2 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
3	BetaPERT (0, 1/4 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
4	BetaPERT (0, 1/3 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
5	BetaPERT (0, 1/2 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
6	BetaPERT (0, 1/4 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
7	BetaPERT (0, 1/3 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
8	BetaPERT (0, 1/2 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
9	BetaPERT (0, 1/4 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
10	BetaPERT (0, 1/3 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
11	BetaPERT (0, 1/2 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
12	BetaPERT (0, 1/4 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
13	BetaPERT (0, 1/3 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
14	BetaPERT (0, 1/2 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
15	BetaPERT (0, 1/4 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
16	BetaPERT (0, 1/3 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
17	BetaPERT (0, 1/2 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)
18	BetaPERT (0, 1/4 Total # IHD Deaths, Total # IHD Deaths)	BetaPERT (0, 1/2 # IHD Deaths for Ever Exposed, # IHD Deaths for Ever Exposed)

[a]BetaPERT distribution (minimum, likeliest, maximum)
[b]Up to 148, which is the number of individuals in the highest exposure group (i.e. >2085.8 ppt-mo) that were classified as non-cases.

Results and Discussion

Results for each simulation of the probabilistic uncertainty analysis are summarized in Table 3. The geometric mean and median $OR_{DM\text{-}LTF}$ ranged from 2.19 to 3.37 and 2.26 to 3.23, respectively. Between 24.0% and 55.8% of the simulation trials yielded adjusted ORs greater than the unadjusted $OR_{observed}$. When the never-exposed group was more likely or as likely as the highest exposure group to be misclassified as non-cases, adjustment for study bias due to loss to follow-up resulted in a shift of the $OR_{DM\text{-}LTF}$ frequency distributions toward the null, lessening the effect of occupational TCDD exposure on IHD mortality. In contrast, when the highest exposure group was more likely than the never-exposed to be misclassified as non-cases, the $OR_{DM\text{-}LTF}$ frequency distributions shifted away from the null.

The application of probabilistic uncertainty analysis to the mortality data described by McBride et al.[2] provides insight into the magnitude and direction of disease misclassification resulting from loss to follow-up. Combining this probabilistic uncertainty model with the one we developed to adjust for exposure and disease misclassification is the next step in our comprehensive analysis of the effect of study bias on the findings reported for this cohort of chemical production employees. We believe this technique will be invaluable in determining whether a true causal relationship between occupational TCDD exposure and IHD exists or if the effect observed in these workers is an artifact of systematic error. Such re-assessment of existing data should prove to be an inexpensive alternative to traditional epidemiological data collection and analysis methods. Additionally, the results of the proposed analyses will build on all sources of information available for TCDD exposure and IHD in chemical production workers, making the findings more useful, and appropriate, for identifying research priorities and guiding public health policy decisions.

Table 3: Descriptive Statistics and 95% Certainty Intervals for Probabilistic Uncertainty Results after 10,000 Simulation Trials by Scenario

Scenario	$OR_{DM\text{-}LTF}$ Median	$OR_{DM\text{-}LTF}$ GM	95% Certainty Interval	% of Trials with $OR_{DM\text{-}LTF} > OR_{observed}$
1	2.88	2.91	0.88 – 10.3	45.2
2	2.26	2.19	0.64 – 7.10	25.9
3	3.23	3.37	1.05 – 12.8	55.5
4	2.94	2.93	0.96 – 9.66	45.5
5	2.36	2.26	0.72 – 6.56	25.3
6	3.20	3.35	1.12 – 12.0	55.8
7	2.90	2.89	1.06 – 8.50	43.8
8	2.38	2.29	0.79 – 6.21	25.0
9	3.19	3.29	1.17 – 10.3	55.8
10	2.93	2.92	1.12 – 7.94	44.3
11	2.49	2.36	0.85 – 5.73	24.9
12	3.16	3.27	1.31 – 9.43	55.6
13	2.97	2.93	1.38 – 6.10	43.6
14	2.65	2.50	1.09 – 4.87	24.0
15	3.10	3.18	1.58 – 7.20	54.8
16	2.99	2.94	1.45 – 5.80	43.7
17	2.72	2.56	1.19 – 4.51	24.2
18	3.09	3.16	1.61 – 6.67	54.5

$OR_{DM\text{-}LTF}$, odds ratio adjusted for disease misclassification due to loss to follow-up; GM, geometric mean; $OR_{observed}$, odds ratio for observed data

References

1. Scott LLF. (2012); *Organohalogen Compounds*. 74:1105-8
2. McBride DI, Collins JJ, Humphry NF, Herbison P, et al. (2009); *J Occup Environ Med*. 51(9):1049-56
3. Shore RE, Iyer V, Altshuler B, Pasternack BS. (1992); *Regul Toxicol Pharmacol*. 15(2 Pt 1):180-221
4. Hemon D. (1986); *Rev Epidemiol Sante Publique*. 34(4-5):230-6
5. Parodi S, Gennaro V, Ceppi M, Cocco P. (2007); *Int J Occup Environ Health*. 13:143-52
6. Maldonado G. (2008); *J Epidemiol Community Health*. 62(7):655-63
7. New Zealand Ministry of Health. (2011); Mortality and Demographic Data 2008. Wellington, NZ.