THE EFFECTS OF SAMPLE DESIGN ON STATISTICAL INFERENCES FROM THE UNIVERSITY OF MICHIGAN DIOXIN EXPOSURE STUDY

Hong B[1], Olson K[2], Lepkowski J[2], Hedgeman E[1], Chen Q[3], Lee S-Y[3], Chang C-W[1], Ward B[2], Ladronka K[2], Gillespie BW[3], Franzblau A[1], Adriaens P[4], Demond A[4] , Garabrant D[1]

[1]Department of Environmental Health Sciences, University of Michigan School of Public Health, 109 S Observatory, Ann Arbor, MI 48109; [2]Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan 48104; [3]Department of Biostatistics, University of Michigan School of Public Health, 109 S Observatory, Ann Arbor, MI 48109; [4]Department of Environmental and Water Resources Engineering, University of Michigan College of Engineering, Ann Arbor, Michigan 48109 USA.

Abstract

Linear regression models were performed to identify factors that explain variation in serum dioxin concentration measured from the 946 participants in the University of Michigan Dioxin Exposure Study (UMDES). All the presented regression analyses accounted for sampling weights, stratification, and clustering to insure the inferences from the regression models were applicable to the population from which participants were selected. Statistical results that do not adjust for sample design effects often generate incorrect inferences. To address the effects of the complex sample design in the UMDES, regression analyses adjusted for sample design and regression analysis treating the survey data as a simple random sample were performed and compared. Most of the covariates related to health, demographics and residential history showed consistent results; however, covariates related to recreational activities and consumption of fish and game from contaminated areas were most likely affected by the sample design. Thus, the variables most affected by ignoring the sample design were those that were strongly correlated with the design variables.

Introduction

The University of Michigan Dioxin Exposure Study (UMDES) was designed to assess exposures to PCDDs, PCDFs and PCBs in adults in five counties of Michigan, and to identify factors that explain variation in serum dioxin concentration. The participants were sampled from five geographically-defined populations[1] from areas in Midland, Saginaw, part of Bay, Jackson and Calhoun Counties in Michigan, using a two-stage area probability household sample design.[2] Linear regression models, in which the outcome variable was the $\log_{10}$ serum dioxin concentration, were performed to determine the significant predictors. Sample design features including weights reflecting selection and non-response probabilities, stratification, and clustering were used in the regression analyses to insure that inferences were applicable to the population from which participants were selected.

Statistical results that do not adjust for sample design effects may generate incorrect inferences. The goal of this paper is to present the difference in inferences based on a regression model accounting for the sample design vs. a regression model treating the survey data as a simple random sample (SRS), and to address the effects of sample design on the serum samples in our study. The sample was stratified by region, with unequal sample probabilities across the regions. Given the variation of the weights across regions and the use of regions as sample strata, we expect variables that are highly correlated with region to be most affected by ignoring the sample design.

Materials and Methods

A total of 946 serum samples were collected from the population in the five studied regions[1]. All serum results were lipid adjusted, and results below the limit of detection (LOD) were substituted with LOD/$\sqrt{2}$ for analysis. Linear regression models with $\log_{10}$ serum dioxin concentration as the outcome variable were performed to identify significant predictors of serum dioxin concentration (all regression analyses were performed using SAS version 9.1). Potential predictors were derived from the UMDES questionnaire (i.e. basic demographic and health variables,

residential history, property use, work history, recreational activities in the contaminated areas, food consumptions including meat, fish, game meat, eggs, milk, other dairy products, and vegetables), and household dust and soil dioxin concentrations. The missing values in the survey questionnaire and the dust, and soil samples were imputed by using a sequential regression imputation procedure.[2] The identified predictors from the final linear regression models, which used a multistage backward selection strategy on five imputed data sets and regression analyses accounting for sample design effects, are reported elsewhere.[4] To address the effects of sample design on the serum samples, two sets of regression analyses were performed on the final identified predictors and used one imputed data set: (1) Linear regression models, with the outcome variable of $\log_{10}$ serum toxic equivalency (TEQ, calculated from WHO 2005 TEFs[3]) or 2,3,7,8-TCDD concentration, adjusted for sample design; (2) models identical to (1) except that the survey samples were treated as SRS, ignoring the unequal weights, clustering and stratification.

In the regression analyses adjusted for sample design, both the parameter estimates and the standard errors of the estimates are affected by sampling weights, stratification and clustering. The design effect (DEFF) of an estimate, which is the ratio of the variance accounting for the complex sample design to the variance computed under the assumption of simple random sampling, [5] measures the impact of the sample design on the variance of the survey estimates. The DEFF (which is specific to each covariate in the model) indicates the extent to which the variability of the parameter estimate derived under an assumption of simple random sampling is underestimated or overestimated, compared to when the sample design effects are properly considered. Thus, we calculated DEFF for each predictor after two sets of regression analyses were performed.

Results and Discussion

Sampling weights for serum samples in the study were developed to compensate for unequal probabilities of selection and unit non-response, and varied across the regions. The sampling weights (Table 1) in the M/S floodplain and near floodplain regions were quite small (mean sampling weights ranged from 7 to 9) compared to the sampling weights in M/S out of floodplain and Jackson/Calhoun regions (mean sampling weights ranged from 344 to 374).

Table 1: Descriptive Statistics of sampling weights for serum samples in UMDES study.

Region	N	**Mean**	Median	75th percentile	Standard Deviation
M/S FP	243	**7**	6	9	5
M/S NearFP	205	**9**	7	9	9
M/S OutFP	204	**374**	342	458	283
M/S Plume	43	**126**	77	118	143
Jackson/Calhoun	251	**344**	271	439	249

Two different approaches, linear regression analyses adjusted for sample design (Reg-SampleDesign) and linear regression analyses treating the survey data as SRS (Reg-SRS), were performed, and the results (parameter estimates and p-values) from the two approaches were compared (Table 2). There are three main findings: (1) Most predictors related to health, demographics, and residential history (for example, age, BMI, BMI loss in the past 12 months, gender, the number of months the first child was breast-fed, pack-years of smoking, etc.) had consistent results between Reg-SampleDesign and Reg-SRS. This was expected since these variables were distributed similarly across the five regional populations. (2) Region, predictors related to recreational activities (including water activities, fishing and hunting activities in or near the contaminated areas) and consumption of food from contaminated areas were frequently affected by ignoring the sample design. This was not unexpected since the effects of ignoring the sample design were most likely to occur with variables that are correlated with the region, such as eating local fish and game and recreation near the Tittabawassee River. (3) The DEFF ranged from 0.1 to 4.6 and varied by predictors, with an average DEFF of 1.3 for TEQ-2005 and 1.5 for 2,3,7,8-TCDD model. For example, the maximal soil TEQ concentration was associated with the region and the DEFF = 0.1, meaning that the Reg-SRS overestimated the variance of the parameter estimate for soil by a factor of 10. In contrast, the number of years eating fish caught from the Tittabawassee River from 1980 to the present had a DEFF = 2.4, meaning the Reg-SRS underestimated the

variance of the parameter estimate for these fish consumptions by a factor of 2. In both cases, the statistical inferences based on Reg-SRS would be incorrect because they do not account for the effects of the study design on the variance of the parameter estimates.

Although the common belief is that accounting for the sample design will lead to increased variances, this analysis has shown that both increased variances and decreased variances can occur. When the correlation between the sampling probabilities, stratification, and clustering variables and the survey variables of interest is strong, then the variability of estimates can be decreased.

Table 2: Different inferences between Reg-SampleDesign vs. Reg-SRS, with the outcome variable of $\log_{10}$ serum dioxin concentration (results that had similar statistical inferences are not shown). Blue-shaded values indicate significant positive estimates and yellow shaded values indicate significant negative estimates ($p \leq 0.05$).

Total sample size=946	**TEQ-2005**					**2,3,7,8-TCDD**				
	SampleDesign[1]		**SRS[2]**		**DEFF[3]**	**SampleDesign[1]**		**SRS[2]**		**DEFF[3]**
Final identified predictors:	**Est.[4]**	**Pv.[5]**	**Est.[4]**	**Pv.[5]**		**Est.[4]**	**Pv.[5]**	**Est.[4]**	**Pv.[5]**	
Health/Demographic										
M/S Floodplain vs. J/C						-0.0216	0.637	0.0846	0.021	1.6
M/S Plume vs. J/C	-0.0376	0.171	-0.0633	0.026	0.9					
At least high school: Y vs. N						-0.1094	0.001	-0.0424	0.196	1.0
White: Y vs. N						-0.1197	0.004	-0.0748	0.111	0.8
Property use										
Lived in a property damaged by fire (yrs in 60-79)	0.0296	0.000	0.0195	0.063	0.3	0.0259	0.003	0.0130	0.469	0.2
Worked in a flower garden: Y vs. N	-0.0277	0.045	-0.0189	0.087	1.6					
Maximum soil concentration in ppt	9E-06	0.009	2E-05	0.053	0.1					
Working history										
Worked as an emergency responder (yrs after 80)	-0.0059	0.020	0.0012	0.536	1.7	-0.0173	0.000	-0.0062	0.068	1.1
Worked at jobs with chemicals: chlorophenol, herbicides, etc. (yrs in 40-59)						0.0149	0.003	0.0149	0.211	0.2
Stationed in Vietnam (yrs)						0.0621	0.000	0.0312	0.227	0.4
Worked in waste disposal, metal scrap yards, water treatment facility, etc. (yrs in 60-79)						0.0168	0.002	0.0113	0.164	0.4
Worked or lived with a worker at other chemical co. (yrs in 60-79)	-0.0080	0.017	0.0008	0.800	1.2					
Water Activities										
Water activities near Kalamazoo R. in 60-79: Y vs. N.						-0.1265	0.001	-0.0565	0.529	0.2
Water activities near Saginaw R./Bay after 80: >=1 per mo. vs. never						-0.2769	0.000	-0.0262	0.651	1.2
Water activities near Tittabawassee R. in 60-79: >=1per mo. vs. never	0.2676	0.000	0.0447	0.240	4.0	0.2955	0.004	0.0583	0.429	1.9
Water activities near Tittabawassee R. in 60-79: <1 per mo. but ever did vs. never						-0.0591	0.482	-0.0779	0.046	4.6
Water activities near Tittabawassee R. after 80: >=1 per mo. vs. never						0.2489	0.019	0.0553	0.283	4.3
Water activities near Tittabawassee R. after 80: < 1 per mo. but ever did vs. never						0.1108	0.000	0.0248	0.380	1.2
Water activities near other rivers in 60-79: >=1 per mo. vs. never	-0.0418	0.003	-0.0055	0.715	0.9					
Food and related activities										
Ate fish caught from Tittabawassee R.						0.0074	0.001	0.0026	0.067	2.4

Total sample size=946	TEQ-2005					2,3,7,8-TCDD				
	SampleDesign[1]		SRS[2]		DEFF[3]	SampleDesign[1]		SRS[2]		DEFF[3]
Final identified predictors:	Est.[4]	Pv.[5]	Est.[4]	Pv.[5]		Est.[4]	Pv.[5]	Est.[4]	Pv.[5]	
(yrs after 80)										
Ate walleye/perch from the Kalamazoo R., else areas, store or restaurant bought currently[+]: < 1 per mo. but ever ate vs. never	0.0528	0.021	0.0270	0.063	2.5					
Ate walleye/perch caught from Saginaw R./Bay currently[+]: >=1per mo. vs. never						-0.2579	0.001	-0.0784	0.138	2.2
Ate walleye or perch caught from Saginaw R./Bay currently[+]: <1 per mo. but ever ate vs. never						-0.1630	0.011	-0.0631	0.057	3.7
Ate any fish (other than walleye/perch) from Saginaw R./Bay currently[+]:>=1 per mo. vs. never	-0.2865	0.000	-0.0238	0.730	0.3					
Fishing in Saginaw R./Bay after 80: >=1 per mo. vs. never						0.1855	0.002	0.0963	0.061	1.4
Ate the skin of the wild turkey, pheasant, grouse, quail, or woodcock currently[+]: Y vs. N	0.0592	0.034	0.0192	0.452	1.2					
Hunting near Saginaw R./Bay after 80: >=1 per mo. vs. never	-0.2254	0.002	-0.0393	0.587	1.0					
Hunting near Saginaw R./Bay after 80: <1 per mo. but ever did vs. never						-0.1570	0.027	-0.0448	0.346	2.2
Ate eggs & dairy from cows home-raised in the Tittabawassee R. currently[+]: >=1 per mo. vs. never						0.1889	0.016	0.0432	0.571	1.0
Ate eggs & dairy from cows home-raised in else areas, store or restaurant bought currently[+]: >=1 per wk. vs. never						-0.2032	0.000	-0.0577	0.561	0.3

SampleDesign[1]: Regression analyses adjusted for sample design; SRS[2]: Regression analyses with simple random sample assumption;
DEFF[3]: Design Effect; Est.[4]: Parameter Estimates; Pv.[5]: P-value; Currently[+]: in the last 5 years

Acknowledgments

Financial support for this study comes from the Dow Chemical Company through an unrestricted grant to the University of Michigan. The authors acknowledge Ms. Sharyn Vantine for her continued assistance and Drs. Linda Birnbaum, Ron Hites, Paolo Boffetta and Marie Haring Sweeney for their guidance as members of our Scientific Advisory Board.

References

1. Garabrant D, Franzblau A, Lepkowski J, Adriaens P, Demond A, Hedgeman E, Knutson K, Zwica L, Chen Q, Olson K, Ward B, Towey T, Ladronka K, Sinibaldi J, Chang S-C, Lee S-Y, Gwinn D, Sima C, Swan S, Gillespie BW. *Organohalogen Comp* 2006; 68:225
2. Lepkowski J, Olson K, Ward B, Ladronka K, Sinibaldi J, Franzblau A, Adriaens P, Gillespie BW, Chang SC, Chen Q, Demond A, Gwinn D, Hedgeman E, Knutson K, Lee S-Y, Sima C, Swan S, Towey T, Zwica L, Garabrant D. *Organohalogen Comp* 2006; 68:209
3. Van den Berg M, Birnbaum L, Denison M, De Vito M, Farland W, Feeley M, Fiedler H, Hakansson H, Hanberg A, Haws L, Rose M, Safe S, Schrenk D, Tohyama C, Tritscher A, Tuomisto J, Tysklind M, Walker N, Peterson RE. *Toxicological Sciences* 2006; 93: 223.
4. Garabrant D, Hong B, Chen Q, Franzblau A, Lepkowski J, Adriaens P, Demond A, Hedgeman E, Knutson K, Zwica L, Chang C-W, Lee S-Y, Olson K, Towey T, Trin H, Wenger Y, Luksemburg W, Maier M, Gillespie BW. *Organohalogen Comp* 2007 (forthcoming).
5. Kish L. (1965) *Survey Sampling*. John Wiley & Sons, Inc., New York.