Dealing with the non-detected and non-quantified data

The Example of the serum dioxin data in the French Dioxin and Incinerators Study

A. Zeghnoun[1], M. Pascal[1], N. Fréry[1], H. Sarter[1], G. Falq[1], J.F. Focant[2], G. Eppe[2]
1/French Institute for Public Health Surveillance - 12, rue du Val d'Osne, 94415 Saint-Maurice, France.
2/CART Mass Spectrometry Laboratory, Chemistry Department, University of Liège, Allée de la Chimie 3, B-6c Sart-Tilman, B-4000 Liège, Belgium.

Introduction

The French Dioxin and Incinerators Study

The FDIS study was carried out in **8 different areas in France** around municipal solid waste incinerators to study whether serum dioxin levels were higher in people living in the vicinity of incinerators compared to referent people. In addition, some factors, such as local food consumption, were also studied to check whether they could influence their serum levels.

Sampling

1030 adults (30-65 years) selected through a **stratified two stage random sampling**, provided blood serum samples for dioxin measurements. Exposure was assessed by concentrations of dioxins, furans and PCBs in human **blood serum**.

Censored data

Data could be censored either by the limit of detection (LOD) or the limit of quantification (LOQ). Several methods are available for handling environmental exposure data in the presence of censored values:

- The **substitution methods** substitute a single value for each censored value. Summary statistics and regression parameters are then estimated using the data above the LOD together with these substituted values.
- The **Tobit regression** model is based on maximum likelihood method. This method does not require substitution for values less than the LOD or LOQ and uses maximum likelihood to estimate regression parameters.

Objective

We investigate the impact of the two different methods on the assessment of:

- distribution of dioxin concentrations,
- regression coefficients of the blood serum dioxin concentrations on several covariates.

Methods

Two methods were selected to handle censored data:

Substitution method

The first method substitutes a censored value by:

- zero,
- LOD if the censored value is less than LOD or LOQ if it is less than LOQ,
- LOD/2 if the censored value is less than LOD or LOQ/2 if it is less than LOQ,
- LOD/$\sqrt{2}$ if the censored value is less than LOD or LOQ/$\sqrt{2}$ if it is less than LOQ,
- LOD/2 when the censored value falls between 0 and LOD, and (LOD+LOQ)/2 when the censored value falls between LOD and LOQ.

Tobit model

- The Tobit model is designed for data with censored values,
- This model makes it possible to distinguish between left-censored data when values are less than the LOD and interval censored data when values are between the LOD and the LOQ,
- We study the association between the logarithm of dioxin concentrations and some selected covariates (age, sex, BMI, social and occupational category, recent change of body weight, smoking status, background/local food consumption, etc.) using a Tobit regression model already developed in the FDIS study to explain the variability of the serum dioxin concentrations.

Software used

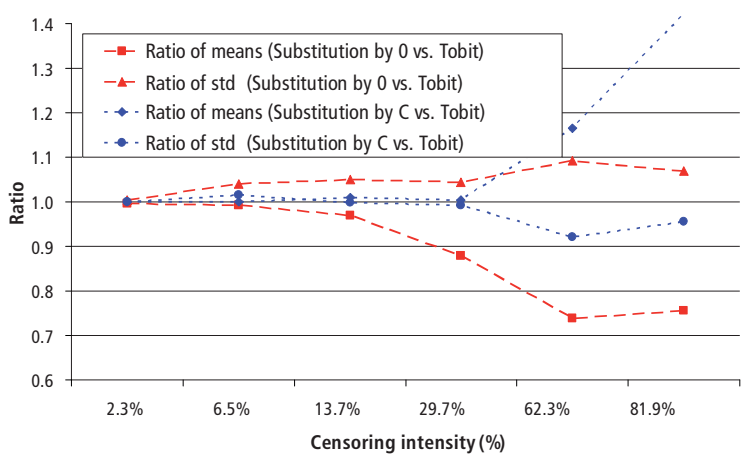
Maximum likelihood estimates are obtained using LIFEREG procedure implemented in SAS which can be used with data that have both interval-censored data along with left-censored data.

The GLM procedure is used when the censored values were substituted.

Results

Six congeners were selected for this work representing **different censoring intensities**: 1,2,3,4,6,7,8 HpCDD (2.3%<LOQ), 2,3,7,8 TCDD (6.5%<LOQ), 2,3,4,6,7,8 HCDF (13.7%<LOQ), 1,2,3,4,6,7,8 HpCDF (29.7%<LOQ), 2,3,7,8 TCDF (62.3%<LOQ), 1,2,3,7,8,9 HCDF (81.9%<LOQ).

Figure 1 Ratio between means (standard errors) obtained by the substitution method and the Tobit model



C = LOD/2 if censored value is less than LOD and C = (LOD+LOQ)/2 if censored value fall between LOD and LOQ

As the censoring intensity increases, the discrepancy between estimates of means (standard errors) obtained by substitution methods and Tobit regression model increases, particularly when the censoring intensity is greater than 30%.

Table 1 Distribution of blood serum dioxin concentrations

	2,3,4,6,7,8 HCDF (13.7%<LOQ)							1,2,3,7,8,9 HCDF (81.9%<LOQ)						
	Mean	Standard error	Percentiles					Mean	Standard error	Percentiles				
Substitution by:			5%	25%	50%	75%	95%			5%	25%	50%	75%	95%
zero	1.61	1.35	0	1	1.45	1.99	3.52	0.23	0.58	0	0	0	0	1.43
LOD/2 or LOQ/2	1.66	1.3	0.28	1	1.45	1.99	3.52	0.42	0.52	0.15	0.18	0.21	0.38	1.43
LOD/$\sqrt{2}$ or LOQ/$\sqrt{2}$	1.68	1.28	0.4	1	1.45	1.99	3.52	0.49	0.5	0.21	0.25	0.29	0.53	1.43
LOD/2 if x<LOD & (LOD+LOQ)/2 if LOD <x< LOQ	1.68	1.28	0.38	1	1.45	1.99	3.52	0.44	0.52	0.15	0.18	0.21	0.57	1.43
LOD or LOQ	1.7	1.26	0.55	1.03	1.45	1.99	3.52	0.6	0.47	0.29	0.36	0.42	0.73	1.43
Tobit model	1.66	1.28	0.51	0.92	1.39	2.11	3.83	0.31	0.55	0.03	0.09	0.2	0.46	1.52

The distribution of dioxin concentrations is sensitive to the values used to substitute the censored values. Substitution by zero produces estimates which are low, while substitution by LOD or LOQ results in large estimates. The substitution by the other values produces intermediate estimates.

Table 2 Comparison of parameter estimates* and their standard errors obtained by the substitution method** and the Tobit model for two selected congeners***

	2,3,4,6,7,8 HCDF (13.7%<LOQ)				1,2,3,7,8,9 HCDF (81.9%<LOQ)			
	Subst. method		Tobit method		Subst. method		Tobit method	
	Coeff.	Std	Coeff.	Std	Coeff.	Std	Coeff.	Std
Intercept	0.672	0.736	0.643	0.658	-1.712	0.913	-3.433	1.828
Male	-0.091	0.300	-0.102	0.268	-0.415	0.371	-0.613	0.733
Age	-0.015	0.014	-0.014	0.012	0.014	0.017	0.043	0.034
Male*Age	-0.001	0.004	0.000	0.004	0.006	0.005	0.007	0.011
BMI	-0.031	0.028	-0.029	0.025	-0.005	0.035	0.022	0.070
BMI*Male	0.004	0.008	0.003	0.008	0.008	0.010	0.017	0.020
BMI*Age	0.001	0.001	0.001	0.001	0.000	0.001	-0.001	0.001
Recent body weight change (ref: stable Body weight)								
Body weight increase	**-0.098**	**0.045**	**-0.096**	**0.040**	**-0.125**	**0.056**	**-0.276**	**0.114**
Body weight loss	0.054	0.056	0.062	0.050	0.054	0.069	0.002	0.130
Smoking (ref: no smoking)								
Daily smoker	**-0.805**	**0.053**	**-0.751**	**0.049**	-0.062	0.066	-0.121	0.133
Ex-smoker	**-0.227**	**0.043**	**-0.214**	**0.038**	-0.017	0.053	-0.076	0.103
Urbanisation (ref: rural)								
Urban	-0.063	0.060	-0.064	0.053	-0.068	0.074	-0.153	0.147
Suburban	**-0.150**	**0.048**	**-0.149**	**0.043**	-0.064	0.060	-0.132	0.111
No chimney use	-0.064	0.044	-0.066	0.039	0.028	0.054	0.030	0.101
No dioxin activity	-0.070	0.047	-0.065	0.042	-0.072	0.058	-0.108	0.110
Exposed area (EXP)	-0.119	0.104	-0.100	0.093	-0.057	0.128	-0.204	0.252
Backgound animal consumption	-0.001	0.001	-0.001	0.001	0.001	0.002	0.000	0.003
Background animal consum.*EXP	0.002	0.001	0.002	0.001	0.000	0.002	0.002	0.004
Local animal consumption	0.000	0.003	0.000	0.003	-0.001	0.004	-0.002	0.007
Local animal consum.*EXP	**0.012**	**0.004**	**0.011**	**0.003**	0.004	0.005	0.005	0.009

*Adjusted for Social and Occupational Category and for Location of the incinerators **Substitution by LOD/2 if censored value is less than LOD and (LOD+LOQ)/2 if censored value fall between LOD and LOQ. ***Congeners with moderate and maximum censoring intensity. Characters in bold represent significant parameters ($p<0.05$).

Table 2 shows roughly similar regression coefficients when the censoring intensity was less than 30%. The disagreement between the substitution method and the Tobit model is large when the censoring intensity is larger than 30%. Standard errors of the regression coefficients were roughly similar except with heavily censored data.

Conclusion

As in most studies, when the censoring intensity is small, the method for treatment of left-censored data makes little difference. Furthermore, in this study the difference between the substitution method and the Tobit model is acceptable when the censoring intensity is less than 30%. However, when the censoring intensity is larger than 30%, the resulting analysis of this study could be sensitive to the method for handling left and interval-censored data.

The substitution by zero clearly distorts the distribution and we recommend not to use it. Furthermore, as the choice of the values used to substitute the censored data is arbitrary, and as the distribution of dioxin concentrations is sensitive to the values used to substitute the censored data, simple substitution methods are not recommendable except when the censoring intensity is low. The Tobit model uses the censored data and the uncensored data in a regression procedure. It does not require substitution of censored values. We recommend to use it, particularly when the censoring intensity is moderately large.

References

Fréry N, Volatier JL, Zeghnoun A, Falq G, Mouajjah S, Thébault A, Pascal M, Bérat B, Grange D, De Crouy-Chanel P, Sarter H, Eppe G, Heyman C, Guillois-Becel Y, Lucas N, Mathieu A, Noury U, Pouey J, Schmitt M, Salines G. Organohalogen compounds 2007, submitted

Pascal M, Eppe G, Zeghnoun A, Falq G, Fréry N. Organohalogen Compounds 2007 submitted.

Helsel DR. Environ. Sci. Thechnol. 1990; 24; 1766:1774.

Judge G, Griffiths WE, Hill RC, Lutkepohl H and Lee TC. New York, John Wiley 1985.

Helsel DR and Hirsch RM. U.S. Geological Survey, Techniques of Water-Resources Investigations Book 4, Chapter A3 2002.

SAS institute. SAS/STAT User's guide version 9. Cary, NC: SAS Institute Inc. 2004.

Baccarelli A, Pfeiffer R, Consonni D, Pesatori AC, Bonzini M, Patterson DG Jr, Bertazzi PA, LAndi MT. Chemosphere 2005; 898:906.

Acknowledgments

This study was supported by the French Ministry of Health within the framework of the Cancer Plan. The authors would like to acknowledge the scientific committee for its advice, the population for its participation and all the staff from the regional units of InVS and the French National Agency for Blood.

August 2007

Contact : Abdelkrim Zeghnoun (a.zeghnoun@invs.sante.fr) - Institut de veille sanitaire - 12, rue du Val d'Osne 94415 Saint-Maurice Cedex - France - Tel. : +33 (0)1 41 79 69 23 - Fax : +33 (0)1 41 79 67 68