ANALYSIS OF AN OUTCOME BEYOND A DETECTION LIMIT: APPLICATION TO HIV RNA DATA

Nuria Perez-Alvarez[1,2], Peter Konings[3], Kristel Van Steen[4,5], Nadia Tahri[6]

[1]Fundació de Lluita contra la SIDA. Badalona, Barcelona. Spain; [2]Universitat Politècnica de Catalunya. Barcelona. Spain; [3]Katholieke Universiteit Leuven, Leuven, Belgium; [4]Universite de Liège, Liège, Belgium;[5]StepGen cvba, Merelbeke, Belgium; [6]AmGen, U.S.A.

A common practice adopted by many analytical laboratories is to report results of biological samples as "less than a specified value". This results in “incomplete” datasets, requiring special care when used as input to statistical analysis. A still popular way to deal with the introduced censoring of observations is to substitute fixed values, such as the detection limit itself or half the detection limit. The literature on how to conduct a proper statistical analysis in this situation is quite dispersed, covering different areas of research including environmental science and medicine.

This work reviews some of the available techniques to handle data below the detection limit. We perform survival analyses (Cox models and parametric models assuming a lognormal distribution) considering the values beyond a detection limit as left censored data. In addition, a variety of multiple imputation models are considered so as to obtain complete data sets. These methods are compared with simple, yet commonly used, imputation methods using the detection limit itself as a substation value, or the detection limit divided by 2 or by the square root of 2. Each approach is discussed on the basis of statistical suitability and validity, and clinical interpretation.

Pros and cons are illustrated using data on plasma viral load in HIV patients, relating viral load to several continuous and dichotomous factors. In particular, for all patients entering the HIV unit in the Hospital Universitari Germans Trias i Pujol (Barcelona) from October 2004 to October 2005, HIV viral load is described in its relation to CD4 nadir (a continuous variable) and the time of initiation of antiretroviral therapy: Pre- and Post-HAART era, before and after 1996 respectively (a dichotomous variable). HAART regimen changes the natural history of HIV-infected-patients and the infection is transformed from a rapidly progressive and nearly uniformly fatal condition to a treatable chronic infection. Of the 2163 patients in the cohort, 712 started medication for HIV infection before 1996 (pre-HAART era) and 1451 after 1996 (post-HAART era). The median CD4 nadir of the cohort was 203 (interquartile range: 89;327). 65.2% of the patients had a viral load below detection limit (HIV RNA< 50).

We emphasize that single substitution can lead to bias and other undesirable effects. The use of survival methodologies (such Cox models and parametric models) or multiple imputation procedures outperform single imputation methods and should be advocated in a larger community.

References

1. Succop PA, Clark S, Chen M, Galke W. (2004) Imputation of data values that are less than a detection limit. J Occup Environ Hyg. Jul;1(7):436-41.
2. Rubin, D.B. (1987). Multiple Imputation for Nonreponse in Surveys. New York: Wiley