Chemosphere 144 (2016) 2044–2051

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

On the consequence of substituting maximum likelihood estimates for the observations below the limit of detection☆

Ram B. Jain

2959 Estate View Court, Dacula, GA 30019, USA

HIGHLIGHTS

- Multiple imputations do not adversely affect estimates of statistical parameters.
- Variations in regression slopes over multiple analyses are within statistical noise.
- P-values for pairwise contrasts are not impacted over multiple analyses of same data.

ARTICLE INFO

Article history:
Received 7 May 2015
Received in revised form 8 October 2015
Accepted 26 October 2015
Available online 13 November 2015

Handling editor: Ian Cousins

Keywords:
Maximum likelihood estimation
Regression models
Polyfluorinated compounds
Polybrominated diphenyl ethers

ABSTRACT

Use of maximum likelihood estimation procedures with multiple imputations to replace observations below the limit of detection (LOD) has been recommended. There is concern that the use of multiple imputations may introduce variability in the data resulting in different conclusions every time the same data are statistically analyzed. We analyzed data from National Health and Nutrition Examination Survey for 7 perfluorinated and 7 polybrominated diphenyl ethers to address these concerns. Data for each variable were subjected to 10 different iterations of statistical analysis. All observations below LOD were replaced by maximum likelihood estimation procedures with 5 imputations. The maximum variation in computing unadjusted geometric means over 10 iterations of analysis was about 2.5%. Unless the percent observations below LOD was more than 40%, maximum variation in computing adjusted geometric means was less than 1.5%. Maximum variation for computing adjusted geometric standard deviation was less than 6%. Except for border line comparisons, significance probabilities for pairwise comparisons did not vary enough to render contrasts from being statistically significant to statistically non-significant or vice versa. Similar conclusions applied to significance probabilities for regression slopes. The use of more than one multiply imputed variable in a regression model was not found to be of concern. The results show that the use of multiple imputations does not generate additional variabilities in the estimates of these statistics beyond tolerable statistical noise. However, when the percent observations in the data are relatively high, there is some possibility of obtaining disparate results.

1. Introduction

When a sample has total concentration of the chemical of interest below a certain minimum, the instrument used to analyze the sample fails to generate a measurable signal. The minimum concentration of the chemical required to generate a measurable signal is called the limit of detection (LOD). Methods to estimate the LOD for a specified sample volume are standard and can be found in any basic text on analytical chemistry and elsewhere, for example, at http://www.und.edu/dept/chromatography/Docs/Determination%20of%20LODs.pdf or http://www.chem.utoronto.ca/coursenotes/analsci/StatsTutorial/LimDetect.html. Let the estimated LOD computed for a sample of volume y be X. If certain samples in the batch of samples to be analyzed have volumes other than y, then those samples will have LODs different than X. If volume of a sample is 2y, then its LOD is X/2. If the volume of the sample is y/2, then the LOD for that sample will be 2X. Thus, in practice, a dataset to be analyzed may have same or fixed LOD for all observations for the chemical of interest for whom measurable signal were not obtained. Or the dataset may have different LODs, called variable LODs for different observations for which measurable signals were not obtained. In summary, chemical dataset may have certain observations for which exact measurements are not available. Instead, these observations may have measurements anywhere between zero and the LOD. Of course these observations may be

☆ Author declares that he has no financial or other conflicts that could have affected the conclusions arrived at in this communication. Author received no funds to conduct this research. All data used in this research are available free of charge from www.cdc.gov/nchs.

E-mail address: Jain.ram.b@gmail.com

http://dx.doi.org/10.1016/j.chemosphere.2015.10.113
0045-6535/

deleted and analysis can be done on the remaining observations in the data. This obviously is not an option in many cases particularly when these below the LOD observations form a substantial part of the total data. How to incorporate these observations in the overall statistical analysis has been a subject of interest for many years. Earliest methods to handle these observations were called fixed substitution methods. Hornung and Reed (1990) recommended substituting all observations below the LOD as LOD/$\sqrt{2}$ or LOD/2 depending up on the underlying geometric standard deviation and/or the percent observations below the LOD. Substituting all nondetectables by LOD/$\sqrt{2}$ (or about 0.7LOD) or LOD/2 means all values below the LOD are assumed to be at least half of the LOD. This seems to defy logic and depending upon the percent observation below the LOD, it is more than likely to overestimate measures of central tendency and distort estimates of variance. In fact, this act of substituting fixed values for all observations below the LOD dilutes variability in the data. The analysis of variability being the basic foundation of statistical sciences, artificially diluting variability in the data is akin to attacking the very foundation of statistical sciences. This manipulation of variability will distort relationships, for example, correlation coefficients, regression coefficients, and inferences drawn among two or more groups (Helsel, 2006). In fact, Helsel (2006) considers such substitutions as data fabrication. It is also more than obvious that better methods have been proposed for a long time but have been slow to be used (Helsel, 2005a). Many of these alternate methods have been described by Helsel (2005b).

"True" concentration of each observation below the LOD is somewhere between zero and the LOD as mentioned above. Consequently, irrespective of the methodology used to estimate the "true" concentration of these observations, there will always be certain level of uncertainty associated with these estimates. In order to incorporate this uncertainty, to the degree possible, into estimates, it has been suggested that, instead of a single estimate, multiple estimates of the concentrations be obtained for each observation below the LOD. Multiple imputations (MI) is the name given to obtaining multiple estimates for the concentration of the same observation. Details are presented as Supplemental Material. These multiply imputed estimates can then be combined by using one or the other methods (see Supplemental Material). It has been suggested that 5 imputations are sufficient to account for this uncertainty (Jain et al., 2008; Jain and Wang, 2008).

Among the MI approaches that have been used involve maximum likelihood estimation (MLE) of the values below the LOD (Lynn, 2001; Lubin et al., 2004). Maximum likelihood methodology fits an assumed (for example, normal) distribution to the data "that matches both the values for detected observations, and the proportion of observations falling below each detection limit" (Helsel, 2006). Thus, MLE methods are applicable to the data which have multiple LODs. In addition, MLE can be applied with or without multiple imputations (Helsel, 2006). Jain et al. (2008) and Jain and Wang (2008) conducted a simulation study on the data for four chemicals (serum sodium, serum iron, urine albumin, and blood lead) for which data were publically downloaded from National Health and Nutrition Examinations Survey (NHANES) (http://www.cdc.gov/nchs/nhanes/nhanes2003-2004/lab03_04.htm). These chemicals had less than one percent observations below the LOD. Thus their "true" means and standard deviation were known on the log scale. Observations below the LOD in the range from 10% to 80% were then introduced in each dataset and estimates of mean and standard deviations were estimated by one of the three maximum likelihood methods, namely, as specified by Lynn (2001) and as specified by Lubin et al. (2004) with one and five imputations respectively. These estimates were compared with true means and standard deviations. Lubin et al.'s (2004) procedure with 5 imputations was considered to be the procedure of choice as long as sample size was not less than 40 and percent observation below the LOD was less than 70%. Baccarelli et al. (2005) also concluded that estimates obtained using MLE with multiple imputations provided unbiased estimates even if the underlying distributions show mild/moderate departures from the assumed distribution and even if the percent observations below the LOD are as much as 60–70%. In surveys like NHANES, the sample of size 40 should not be an issue unless we are dealing with very small subgroups and except for chemicals like certain PCBs, dioxins etc., percent observations below the LOD are not likely to be found to be more than 70%. In a preliminary analysis using NHANES 2003–2004 data for blood lead, based on 500 simulations, biases for mean LOD/$\sqrt{2}$ substitution ranged from 23.09% to 36.6% when percent observations below the LOD were $\geq$50% (data not shown). The corresponding biases for Lubin et al.'s (2004) procedure with 5 imputations were below 5%. Similarly, biases for standard deviation for LOD/$\sqrt{2}$ substitution were as much as 10 times more than for Lubin et al.'s (2004) procedure with 5 imputations. In order to use MLE, each dataset need to be considered to consist of two sub-datasets as explained below.

A dataset Q with observations below the LOD can be thought of having two sub-datasets A and B. Sub-dataset A may consist of all observations below the LOD. Since true measurements for each of these observations lies somewhere between zero and LOD, sub-dataset A can be considered to have variable data points for the purpose of applying MLE methods. Sub-dataset B has all observations $\geq$LOD. Application of MLE methods to these observations does not involve changing any of these observations. Once a specific MLE method has assigned randomly determined values to all observations in sub-dataset A, sub-datasets A and B can be integrated and can be used to compute statistics like geometric means (GM) and geometric standard deviations (GSD) etc. Let this integrated dataset be labeled as IDS_1 and let GM and GSD computed from this dataset be GM_1 and GSD_1 respectively. If there were to be five imputations, there will be statistics GM_1, …, GM_5 and statistics GSD_1, …, GSD_5. The final step in the analysis will be to compute overall final statistics GM_F using GM_1, …, GM_5 and overall final statistics GSD_F using GSD_1, …, GSD_5 according to a predefined algorithm, for example, GM_F may be the mean of GSD_1, …, GSD_5 or GM_F may be equal to $(GSD_1{*}GSD_2{*}GSD_3{*}GSD_4{*}GSD_5)^{0.2}$. However, if the analysis was to be repeated M times using the same original dataset Q, there will be M different values of GM_F and GSD_F.

An argument can be made that since the new values of GM_F and GSD_F will be different from the corresponding values GM_F and GSD_F previously calculated, we cannot use MLE method with or without multiple imputations. Getting "different results" every time the same original data are analyzed is not acceptable. The notion of "different results" seems to stem from the fact that this will somehow or the other create variability in the results which may be beyond tolerable statistical noise. These kinds of arguments do not realize the fact that the results from analyzing the same data may also change from the use of one software to another software within tolerable, inconsequential variability, like, for example, p-value varying from 0.50 to 0.60. These arguments also do not realize the fact that if a sample is split into ten subsamples and all of them are analyzed in the same run by the same instrument at the same time, it is almost impossible to get exactly the same measurements of the concentrations for all ten subsamples. Such variations are random statistical noise. They can occur in a chemistry laboratory while analyzing samples and also, when doing statistical analysis.

Variations in results, if any, may occur in estimating adjusted and unadjusted means, geometric means, standard deviations, standard errors, regression slopes and associated p-values, and significance probabilities associated with pair-wise contrasts. In or-