Letter to the Editor

Imputation vs. Substitution

Dear Editor:

On page 12 of the December 2012 issue of CSA News magazine, there is an apparent error in the definition of a statistical approach to dealing with values reported below a laboratory detection limit. In the article entitled “Analyzing Spatial Data of PCB Concentrations in Soils,” the third paragraph contains the statement “... (perhaps by replacing all such data by half of the QL, namely an “imputation” approach).”

This approach of substitution of values below the detection limit with one value, such as one-half the QL, is actually termed “substitution.” Dennis Helsel in his book entitled Statistics for Censored Environmental Data Using Minitab and R (second edition, John Wiley & Sons, 2012), emphatically explains the difference between the substitution and imputation method on page xix in the introduction: “Substitution is not imputation, which implies using a model such as a relationship with a correlated variable to impute (estimate) values.”

Thus, imputation implies the application of a statistically based approach to incorporate data below the detection limit into a broader statistical analysis. I normally would not be one to comment on what appears to be a relatively small difference, but I am currently immersed in a similar situation regarding a data set we are analyzing.

Thank you,

John Troiano, Ph.D.
Research Scientist III
Ground Water Protection Program
Department of Pesticide Regulation
California Environmental Protection Agency
Sacramento, CA