Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 18-19, 2017

The Impact of Below Detection Limit Samples in (NPP) Decommissioning Residual Risk Assessment

Jung Hwan Kim, Carol Hornibrook, Man-Sung Yim*
Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701
**Corresponding author: msyim@kaist.ac.kr*

1. Introduction

Decommissioning is an emerging international issue in the nuclear industry. The term 'decommissioning' refers to the administrative and technical actions taken in order to allow the removal of some or all regulatory controls (e.g. operating license) from a nuclear facility. In order to terminate the nuclear power plant (NPP) license, the owner must demonstrate compliance with the regulatory controls for restricted or unrestricted future use of the site. However, some technical issues associated with finalizing regulations related to the site release remain. For example, the representativeness of the contaminated soil samples related to hot spots, the lack of information on the radionuclide distribution, and observations reported as below detection limits (BDL). The techniques currently used for correctly assessing hot spots and the availability of radionuclide distribution are taken directly from the literature.

The regulatory guidelines for site reuse after decommissioning are commonly challenged because the majority of the activity in the soil is at or below the limit of detection [1]. Observations reported as below the detection limits are caused by the inherent limitations of the measurement methods, i.e. detectors have detection limits. For example, in the National Human Exposure Assessment Survey database, 30% to 70% of the observations are below the detection limits for many pollutants [2]. Multiple detection limits arise from different sampling procedures or different sampling volumes. Although a DL might be insignificantly low, it is dangerous to ignore DLs because the dose is the result of functions such as dose conversion factors, daughter nuclides, types of radionuclide, etc. If the data are not treated correctly, there can be a significant affect, usually an overestimation of the health risk to the public and overestimation of the volume of soil removal and associated costs.

Conventional methods that are currently used for treating the data below the detection limits that either ignore or simply replace the detection limit with zero, a fraction of the detection limit, or the detection limit itself are statistically biased [3]. For example, ignoring or replacing the data with a DL overestimates the mean and replacing it with zero underestimates the mean. In order to resolve these issues, statistical techniques were evaluated.

The three methods used to estimate the summary statistics (e.g. mean, standard deviation) are the Kaplan-Meier method, robust regression on order statistics (ROS) method, and maximum likelihood estimation (MLE) method. These techniques are used by numerous researchers in environmental science and technology, but have not been widely used in nuclear evaluations of risk.

Analyzing the long term environmental health risk and costs can be significantly affected by relatively small variations in the mean value. The small variation in the mean value can affect the dose, cancer risk, and volume of soil removal. To identify the impact of including nondetects, it is necessary to calculate the dose, cancer risk, and volume of soil removal with and without censored data values. To release site for regulatory control, site should be cleaned to a certain level to reduce the potential negative health effects due to the radioactive contamination. It is necessary to determine the amount of soil cleanup required to achieve the cleanup goals. The maximum total dose should be reduced to 0.15mSv/yr for unrestricted use in the future. The RESRAD code helps to estimate the volume of soil needed to be remediated. The goal of this research is to develop more precise methods for risk assessment, estimation of volume of soil removal, and examination of cost savings.

Without a basic understanding of uncertainty, it is difficult to understand how and why site specific soil data is analyzed. Analyses of the environmental issues such as exposure assessments are related to uncertainty [4]. Although only limited data is available for analysis, regulatory decisions should be based on the entire data set. When the data set is limited, some model parameter uncertainties can be represented using probability density functions. The number of model input parameters are simulated simultaneously in order to determine their combined effect on the model outputs. Latin hypercube sampling has been used for this type of uncertainty analysis in probabilistic risk assessment.

In order to compare the strengths of the three statistical methods investigated in this research, errors between the actual mean and the statistical methods are estimated. Then, these errors form the basis for selecting the appropriate the statistical approaches. The smaller the error, the better matched the statistical approach is to the data. Each analysis addresses changes in the amount of radioactivity, types of distributions, censoring percentages, and numbers of detection limits used.

2. Methods and Results

2.1 Proposed methods

The Kaplan-Meier method is a nonparametric technique. Nonparametric techniques describe data that does not follow a specific parametric distribution such as a normal, lognormal, or Weibull distributions. The Kaplan-Meier (KM) method calculates the probability distribution in order to estimate the summary statistics such as the mean, the standard deviation, the percentiles, etc., using censored data. Because it is a nonparametric technique, it is well-suited for many environmental data sets [5].

The robust regression on order statistics (ROS) method is a semi-parametric method that can be used to estimate the summary statistics with censored data. The ROS method uses detected values to develop a probability plot and estimate the parameters using a regression line.

The maximum likelihood estimation (MLE) method is a parametric, model-based method that can be used to estimate the summary statistics with censored data. Probability plots and other goodness-of-fit techniques are used to determine the matching distributions. Nondetects are distributed similarly to the detected values. The parametric MLE method assumes a distribution that will closely fit the observed data.

2.2 Software used for calculation of proposed methods

Statistical software is used to compute the estimates of the censored data. The three software packages used in this research are ProUCL, MATLAB, and R.

ProUCL is statistical software for environmental applications for data sets with and without nondetects observations; it is provided by the US Environmental Protection Agency (EPA) [6]. ProUCL assists in computing the upper confidence limit of the population mean based on left-censored data sets containing nondetects observations. It includes goodness-of-fit (GOF) tests and skewness for left-censored data sets. The nonparametric Kaplan-Meier (KM), robust regression on order statistics (ROS), and maximum likelihood estimation (MLE) methods are used to calculate the estimates the summary statistics including the mean, standard deviation, and percentiles. Box plots, histograms, and Q-Q plots can be developed in ProUCL.

MATLAB and R are programming languages and environments for statistical computing and graphics. These languages include effective data handling and storage. It is possible to estimate the parameters of each distribution using MATLAB. The KM, ROS, and MLE methods are analyzed to estimate the summary statistics using the “Nondetect And Data Analysis (NADA)” package in R.

2.3 Case Study for Analyzing Risk, Volume of Soil and Cost

CSMRI stands for Colorado School of Mines Research Institute site in Golden, Colorado. The S.M. Stoller Corporation conducted soil characterization and remediation activities necessary for the termination of the radioactive materials license and free release of the site.

Table I: Comparison of the mean of radioactivity between ignoring the ND and including the ND.

Radionuclides	Mean of radioactivity Ignore ND (pCi/g)	Mean of radioactivity Include ND using KM (pCi/g)
Ra-226	27.96	25.98
Ra-228	3.66	3.43
Th-232	3.47	3.36
Th-230	21.45	20.58
Th-228	3.56	3.33
U-234	19.45	18.63
U-235	1.13	1.12
U-238	19.86	18.95

Table II: The maximum total dose for ignoring ND and including ND.

Case	Maximum total dose (mSv/yr)
Ignore ND	0.2592
Include ND	0.2414

Table I, II shows the maximum total dose for CSMRI. To analyze conservatively, a recreationist scenario which assumes no consumption of plant, milk, and water is considered. Site-specific data of CSMRI is prepared to run the RESRAD code. It is possible to calculate the maximum total dose using the above equation. The difference of the radioactivity mean affects the maximum total dose. According to the Table I, II, the maximum total dose of both cases (ignoring and including the ND) are higher than the 0.15 mSv/yr the general limit for soil cleanup or site decontamination. In this case, site remediation is required to meet the cleanup criteria. It is necessary to reduce the maximum total dose to 0.15 mSv/yr in order to release the site for either restricted or unrestricted use in the future. The maximum total dose was 0.2592 mSv/yr for the ignoring ND case and 0.2414 mSv/yr for the including ND case. There is no big difference in terms of maximum total dose. However, the maximum total dose is higher in case of ignoring ND than including ND. The cost of remediation could be reduced using the proposed methods that helps to include ND, and there will be cost savings.

Treatment of the contaminated soil includes excavation, transportation, and disposal of the contaminated soil. Through calculating the difference of the total volume of soil removal required in order to meet the cleanup criteria between the ignoring nondetects case and the including nondetects case, it is possible to estimate the number of 200-liter drums and the associated cost savings. The total volume of soil removal can be calculated for both ignoring nondetects and including nondetects; thus, it is possible to estimate the volume difference of the soil removal. If all contaminated soil is classified as low level waste (LLW), the cost of managing the contaminated soil can be assumed using the cost of one 200-liter drum. Assume that the cover thickness of 0.1 m will be used and some contaminated soil will be excavated for remediation. The total volume of soil removal can be estimated using the site geometry.

Table II: Total volume difference between ignoring ND and including ND

Measure	Ignore ND (0.2592 mSv/yr)	Include ND (0.2414 mSv/yr)
Cover thickness needed (m)	0.1	0.1
CZ thickness to be removed (m)	0.611	0.5385
Total volume of soil needed (m^3)	724.4308	650.5613
Volume difference (m^3)	73.8695	

In case of ignoring ND, it is necessary to cover non contaminated soil of 0.1m thickness; the removal of contaminated soil of 0.611m CZ thickness is required. Similarly, in case of including ND, it is necessary to cover non contaminated soil of 0.1m thickness; the removal of contaminated soil of 0.5385m CZ thickness is required. In order to meet the cleanup criteria of 0.15 mSv/yr for unrestricted use in the future, the total volume of soil removal was calculated for both cases of ignoring and including the nondetects. The amounts of soil removal for ignoring the nondetects and including the nondetects were 724.4308 m3 and 650.5613 m3, respectively.

Assume that all contaminated soil is classified as LLW. In order to dispose 73.8695 m3 of contaminated soil, it is necessary to prepare 370 200-liter drums. A 200-liter drum can accommodate 200–500 kg. In Korea, the cost of a 200-liter drum is 12,190,000 won. Therefore, a more precise estimate of the activity in the soil results in a lower volume of soil removal and significant cost savings are achieved through including the BDL data.

There are 44 types of distribution for the LHS technique, each with their own parameters. The peak total dose, peak pathway doses, and peak nuclide doses that result from the set of input variables is analyzed, and the cumulative density function (CDF) is presented graphically. The probabilistic distributions of the most of the parameters are developed from the RESRAD recommended probabilistic distributions and the site-specific data of CSMRI. Specially, the uniform distribution is selected for soil concentration parameters for uncertainty analyses to consider the effect of the including ND. The results provide the correlations and regression coefficients of the doses. These are the parameters of interest for uncertainty analyses.

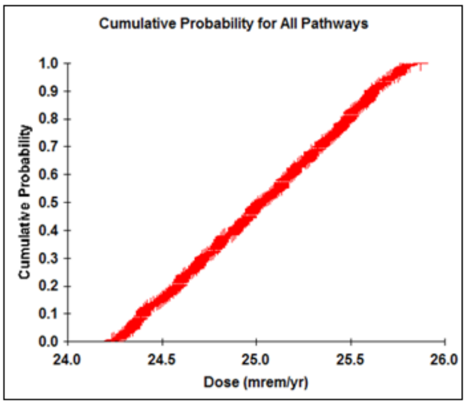

Fig. 1. Sensitivity analysis of soil concentrations CDF changes of the mean of peak dose.

The dose can be significantly affected by soil concentration, contaminated zone, and ingestion non-dietary. In contrast, saturated zone, unsaturated zone, occupancy, and ingestion dietary does not significantly affect the dose. It is prove that analyzing the long term environmental health risk can be significantly affected by the relatively small variation in the soil concentration mean value.

The contribution ranking for the dose was evaluated for 41 input parameters using Partial Correlation Coefficient (PCC) and Partial Ranked Correlation Coefficient (PRCC) approaches. As seen in the Table XII, the radioactivity of the radionuclides contributes significantly to the maximum total dose. Although the mean of radioactivity and the detection limits are insignificantly small in case of Ra-228, it ranks 3rd in contribution for the dose. When the detection limits are small, data below the detection limits is usually ignored. However, it is dangerous to ignore censored data, since the dose and cancer risk can be significantly affected by not only the size of detection limits but also the type of radionuclides, daughter radionuclides, and amount of radioactivity.

2.4 Decision-Making Framework for Selecting the Best Statistical Technique

The first step in the process is to develop a known distribution using the parameters, e.g. the mean and the standard deviation, to set the detection limit and the sample size, and to delete the BDL data. The best-fit statistical technique can be recommended through calculating the bias of the mean from the known distribution. Three types of distributions were considered in order to evaluate the effect of the different types of distributions. The second step is to remove or delete the data below detection limits. The detection limits for 10%, 30%, and 60% cumulative probability of censoring were calculated using the parameters of the distributions and 107 samples. The Kaplan-Meier method, the robust regression on order statistics method, and the maximum likelihood estimation methods were applied to each specified population distribution for each type of distribution, censoring percentage, and sample size to estimate data below the detection limits. It was conducted by simulating 1000 bootstrap samples. Not only the censoring percentage and sample size, but also the amount of radioactivity, the number of detection limits, and the types of distributions can affect the selection of the best statistical technique. However, the study of these influences has not been done.

3. Conclusions

The KM, ROS, and MLE methods were verified to give more precise estimates of the mean compared to the conventional methods in which censored data set is ignored or replaced with the detection limit, half of the detection limit, or zero.

The proposed methods of the Kaplan-Meier, ROS, and MLE methods were performed using the soil samples from the monazite powder manufacturing plant and CSMRI. The KM, ROS, and MLE are flexible and robust methods for analyzing data below the detection limits. The concept of the bootstrap simulation to estimate confidence intervals for the mean was introduced, and the MLE/Bootstrap method was implemented in respect to the various percent confidence intervals for the mean of monazite powder manufacturing plant data set. The preliminary evaluation demonstrated that the proposed methods can be effectively used to provide the best estimated radioactivity levels at a decommissioned NPP site, and it can also estimate the uncertainty in the mean values. The RESRAD code was used to estimate the radiation doses and cancer risks in each case. The risk assessment and volume estimation was performed using the proposed decision-making framework. The amount of remediation of the contaminated soil was estimated and compared with the results of the conventional method. Furthermore, the cost saving difference was analyzed between the conventional method and the proposed methods.

Probabilistic distributions were developed for the RESRAD input parameters to analyze uncertainty. The uncertainty in the maximum total dose for different parameters was analyzed and the contribution rankings were estimated. The number of model input parameters were simulated simultaneously in order to determine their combined effect on the model outputs. Although only limited data was available for analysis, a regulatory decision can be made based on the uncertainty analysis. Sensitivity analysis of RESRAD input parameters was performed for CSMRI data set. The key sources that contribute to the maximum total dose were identified.

If the MDAs are less than 10% of the Derived Concentration Guideline Levels (DCGL), it is possible to ignore data below the detection limits. Although the mean of radioactivity and detection limits are insignificantly small, it is dangerous to ignore data below the detection limits. Since the dose and cancer risk can be significantly affected by not only the size of detection limits but also the type of radionuclides, daughter radionuclides, and amount of radioactivity. The key advantages of the proposed methods are that they are statistically unbiased estimates and can be used for a variety of situations such as different types of distributions, censoring percentages, sample sizes and the number of detection limits.

Through changing the amount of radioactivity, types of distributions, censoring percentages, sample sizes, and number of detection limits, the recommended methods are defined for estimating the summary statistics. Recommended methods are defined to estimate summary statistics, based upon simulations that address lognormal, gamma, and Weibull distributions for different sample sizes of 20, 40, and 100 and censoring percentages of 10%, 30%, 60%, (10,30)%, (30,60)%, (10,60)%, (10,30,60)%. The development of a nondetects analysis framework for decision-making will be provided to the regulators.

Using additional statistical analyses of the contaminated soil before or after decommissioning is expected to provide better and more reliable probabilistic exposure assessments, better economics, and improved communication with the public. Efforts to include nondetects in order to assess risk, estimate volume of soil removal, and examine cost savings more precisely should be made.

REFERENCES

[1] Zheng J, Frey HC. Treatment of Censored Data and Dependency among Sampling Distributions in Probabilistic Human Health Exposure Assessment.

[2] Gordon SM, Callahan PJ, Nishioka MG, Brinkman MC, O'Rourke MK, Lebowitz MD, Moschandreas DJ. Residential environmental measurements in the national human exposure assessment survey (NHEXAS) pilot study in Arizona: preliminary results for pesticides and VOCs. Journal of Exposure Analysis and Environmental Epidemiology. 1999 Oct 1;9(5):456-70.

[3] Zhao Y, Frey HC. Quantification of variability and uncertainty for censored data sets and application to air toxic emission factors. Risk Analysis. 2004 Aug 1;24(4):1019-34.

[4] Baker S, Driver J, McCallum D. Residential Exposure Assessment. In Residential Exposure Assessment 2000 (pp. 9-15). Springer US.

[5] Versus Time TC. STATISTICAL ANALYSIS OF GROUNDWATER MONITORING DATA AT RCRA FACILITIES.

[6] Schumacher B, Barnett F, Singh A, Singh AK. ProUCL Version 4.00. 04 Technical Guide.