DOE Review Release 07/08/2015

External Dose Coworker Methodology		ORAUT-RPRT-0071	Rev. 00
		Effective Date:	07/02/2015
		Supersedes:	None
Subject Expert(s):	Thomas R. LaBone and Nancy M. Chalmers		
Document Owner Approval:	Signature on File Thomas R. LaBone, Document Owner	Approval Date:	05/04/2015
Concurrence:	Signature on File John M. Byrne, Objective 1 Manager	Concurrence Date:	05/04/2015
Concurrence:	Signature on File Edward F. Maher, Objective 3 Manager	Concurrence Date:	05/06/2015
Concurrence:	Vickie S. Short Signature on File for Kate Kimpan, Project Director	Concurrence Date:	05/04/2015
Approval:	Signature on File James W. Neton, Associate Director for Science	Approval Date:	07/02/2015

☒ New ☐ Total Rewrite ☐ Revision ☐ Page Change

FOR DOCUMENTS MARKED AS A TOTAL REWRITE, REVISION, OR PAGE CHANGE, REPLACE THE PRIOR REVISION AND DISCARD / DESTROY ALL COPIES OF THE PRIOR REVISION.

PUBLICATION RECORD

EFFECTIVE DATE	REVISION NUMBER	DESCRIPTION
07/02/2015	00	New report initiated to describe the multiple imputation method to be used for external coworker models. This method will be used to replace censored <LOD external coworker dosimetry results which will replace the current practice of substituting <LOD readings with LOD/2. Incorporates formal internal and NIOSH review comments. Training required: As determined by the Objective Manager. Initiated by Thomas R. LaBone.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

DOE	U.S. Department of Energy
GM	geometric mean
GSD	geometric standard deviation
LOD	limit of detection
NIOSH	National Institute for Occupational Safety and Health
ORAU	Oak Ridge Associated Universities
ROS	regression on order statistics
SRDB Ref ID	Site Research Database Reference Identification (number)

1.0 INTRODUCTION

External dose is measured with a dosimeter that indicates the dose to an individual over the length of time the dosimeter was worn (e.g., a month or calendar quarter). Multiple measurements over a calendar year are summed to give the annual dose for the individual. Dosimeter readings below a given dose, which is often referred to as a limit of detection (LOD), are usually reported as less than that dose, or “less-than” doses. For example, the 29 neutron dosimeter readings in Table 1-1 were reported by a site for one individual over the course of a year. The Reported column contains the doses in rem that were reported by the site. The LOD for the dosimeter was 0.050 rem and, as was the practice at the time, any dose below this LOD was reported as “<LOD”. In statistics, the <LOD data are said to be “censored.” Censored data are problematic because something as simple as the annual dose (the sum of the dosimeter readings) is not unambiguously defined if some of the data are censored (the annual dose is <1.507 rem in this case).

All of the doses that were reported by the site, including those in Table 1-1, were reconstructed to eliminate the censoring. These uncensored doses are given in the Actual column in Table 1-1. The actual doses can be positive or negative because a background dose was subtracted from each dosimeter reading. If only background dose was recorded on each dosimeter, the expected (or mean) dose would be 0 rem with random scatter, which creates the positive and negative doses. The unbiased estimator of the annual dose is therefore the sum of the individual dosimeter readings, both positive and negative. The estimated annual dose in this case is 0.572 rem. Negative doses[1] can be somewhat disconcerting to some individuals. One popular workaround is to set all negative doses to 0, which was done in the Positive column in Table 1-1. This results in a positive bias in the annual dose,[2] which can be seen in Table 1-1 where the Positive column has a sum of 0.625 rem.

Although the Actual or Positive doses are seldom if ever known, it is useful to know the right answer as we apply different methods to make the Reported doses usable. One of these methods is to replace the censored doses with an estimate of the true value of the dose. This process is referred to as “imputing” a dose. Once the censored doses are imputed, they are treated as real doses, combined with the uncensored data, and summed to give the annual dose. Two simple and commonly used imputation[3] methods are:

- Replace censored doses with a dose equal to 0 (Impute A column in Table 1-1).
- Replace censored doses with a dose equal to LOD/2 (Impute B column). This is the current approach (ORAUT 2011) for handling censored dosimeter readings.

Imputing zero for censored results is based on the faulty notion that <LOD results have no physical significance and are indeed equal to 0. This approach cannot provide an unbiased estimator of the annual dose. In the example, replacing the 25 <0.050 rem doses with doses of 0 rem biases the annual dose toward a lower value (0.257 rem versus the best estimate of 0.572 rem).

However, imputing a value of LOD/2 for all <LOD results implicitly assumes that the censored results are evenly distributed along a line that runs from 0 rem to the LOD (0.050 rem in this case). These linearly imputed doses are given in the Impute C column in Table 1-1. The sum of the 25 linearly imputed doses is the same as the sum of 25 doses equal to 0.050/2 (0.025 rem each), which is 0.625 rem. Therefore, the sums of Impute B and Impute C are the same. In general, the

---

1 The true doses are not negative, the measured doses are negative. However, this technicality has no further bearing on this discussion.

2 The bias of an estimate can be known only if the true value is known, which it is in this case.

3 In the literature these are usually referred to as “substitutions” and the term “imputation” is reserved for cases where the dose is estimated using a model that incorporates other ancillary information about the doses. For simplicity, we will not make such a distinction in this report.

imputation of the LOD/2 for censored results is not recommended [for example see discussions in Helsel (2012)

Table 1-1. Table of dosimeter results for a worker.

Date	Actual	Reported	Positive	Impute A	Impute B	Impute C
01/04/1965	0.026	<0.050	0.026	0.000	0.025	0.000
01/18/1965	0.009	<0.050	0.009	0.000	0.025	0.002
02/01/1965	0.004	<0.050	0.004	0.000	0.025	0.004
02/15/1965	0.056	0.056	0.056	0.056	0.056	0.056
03/01/1965	0.012	<0.050	0.012	0.000	0.025	0.006
03/15/1965	0.031	<0.050	0.031	0.000	0.025	0.008
03/29/1965	0.024	<0.050	0.024	0.000	0.025	0.010
04/12/1965	0.033	<0.050	0.033	0.000	0.025	0.012
04/26/1965	0.083	0.083	0.083	0.083	0.083	0.083
05/10/1965	0.020	<0.050	0.020	0.000	0.025	0.015
05/24/1965	0.017	<0.050	0.017	0.000	0.025	0.017
06/07/1965	0.023	<0.050	0.023	0.000	0.025	0.019
06/21/1965	–0.019	<0.050	0.000	0.000	0.025	0.021
06/29/1965	0.006	<0.050	0.006	0.000	0.025	0.023
07/06/1965	–0.019	<0.050	0.000	0.000	0.025	0.025
07/06/1965	0.036	<0.050	0.036	0.000	0.025	0.027
07/19/1965	0.040	<0.050	0.040	0.000	0.025	0.029
08/02/1965	0.012	<0.050	0.012	0.000	0.025	0.031
08/16/1965	–0.000	<0.050	0.000	0.000	0.025	0.033
08/30/1965	0.051	0.051	0.051	0.051	0.051	0.051
09/13/1965	0.021	<0.050	0.021	0.000	0.025	0.035
09/27/1965	0.023	<0.050	0.023	0.000	0.025	0.038
10/11/1965	0.002	<0.050	0.002	0.000	0.025	0.040
10/25/1965	0.005	<0.050	0.005	0.000	0.025	0.042
10/25/1965	0.067	0.067	0.067	0.067	0.067	0.067
11/08/1965	–0.008	<0.050	0.000	0.000	0.025	0.044
11/08/1965	0.013	<0.050	0.013	0.000	0.025	0.046
11/22/1965	–0.007	<0.050	0.000	0.000	0.025	0.048
12/06/1965	0.011	<0.050	0.011	0.000	0.025	0.050
Totals	**0.572**	**<1.507**	**0.625**	**0.257**	**0.882**	**0.882**

and in NCRP (2008)] because it will typically give positively biased parameter estimates for right-skewed data (0.882 rem in this example in comparison to a best estimate of 0.572 rem). However, linear imputation introduces the idea of using a distribution model to impute values for censored data. A more statistically accurate imputation model is necessary to improve the estimated annual dose.

2.0 PURPOSE

This report supports use of a new method, referred to as multiple imputation, to be used to replace censored (<LOD) external dosimeter readings with estimates of the dose to facilitate the calculation of annual external dose. This method will replace the practice of substituting <LOD readings with the LOD/2.

3.0 IMPUTATION MODELS AND MULTIPLE IMPUTATION

The main problem with the linear imputation model is that it uses only the LOD and the number of censored results to impute values for the censored results. The model does not utilize the information contained in the distribution of the observed uncensored data. The uncensored data can help to define the censored data and so be very useful in efforts to predict the distribution of the censored results. For example, the plot in Figure 3-1 shows the fit of a lognormal model to the order statistics of 3,736 dosimeter readings – 3,508 of which were <0.050 rem – for 732 workers over a year. The fit was performed with the regression on order statistics (ROS) method (Helsel 2012). Assuming that all of the doses (censored and uncensored) are lognormally distributed means the uncensored data can be used to determine the parameters of the lognormal probability model that describes the distribution of all of the data.

The geometric mean (GM) of the lognormal distribution in Figure 3-1 is 0.00851, and the geometric standard deviation (GSD) is 3.1. Given that all of the dosimeter readings are lognormally distributed with this GM and GSD, random draws from this distribution that are <LOD can be substituted for the reported <LOD readings (Impute 1 column in Table 3-1. The estimated annual dose is the sum of the

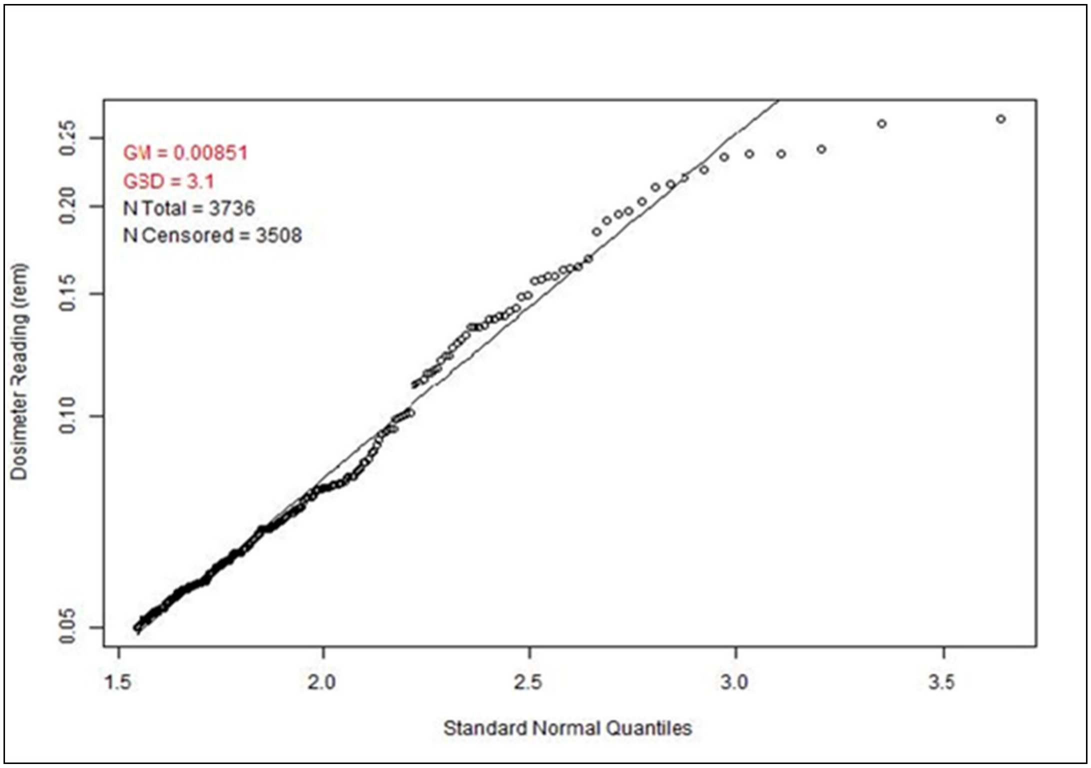

Figure 3-1. Lognormal probability plot of uncensored dosimeter results in year.

values in the Impute 1 column, which is 0.538 rem. If this process is repeated, different values will be imputed for the <LOD results (because the imputed values are random draws from the lognormal distribution) and a different annual dose will be obtained. This can be seen in the Impute 2 to Impute 5 columns. The best estimate of the annual dose is taken to be the mean of the five annual doses, which is 0.531 rem and compares very well to the best estimate of 0.572 rem obtained from the uncensored data. The standard deviation of the five measurements, 0.038 rem, gives a rough idea[4] of how much uncertainty is introduced into the annual dose estimate by the multiple imputations. This approach was inspired by Krishnamoorthy et al. (2009) and Lubin et al. (2004). Krishnamoorthy used maximum likelihood methods rather than ROS to obtain the imputation model for censored environmental contaminant measurements. This analysis used ROS because of its extensive application in internal dosimetry coworker modeling.

4 This estimate of the variance in the mean can be refined (Gelman and Hill 2006, p. 542), but it is not necessary because the Project does not use the variance in the parameters for anything in the coworker model.

In practice, there are several ways that an imputation model can be developed. For example:

- Non-negative, right skewed distributions other than the lognormal (e.g., Gamma or Weibull) can be used to model the dosimeter readings. However, in practice a lognormal distribution will be used by default unless the site subject-matter-expert has reason to believe that a distribution other than the lognormal is to be expected or the statistician performing the fit of the model to the data decides that there is a significant lack-of-fit with the lognormal model.

- The data for the imputation model can be collected over different periods (e.g., multiple quarters or multiple years). Probably the best approach is to select a period during which the exposure conditions, dosimetry equipment (i.e., type of dosimeter), and LOD are fairly constant and to use the data from that period to construct the imputation model for censored data in that period.

- Subsets of the dosimeter readings can be used to develop the imputation model. For example, exclusion from the dataset of dosimeter readings from individuals who had essentially no potential for occupational exposure, or development of different imputation models for weekly, monthly, and annual monitoring periods.

Table 3-1. Results of multiple imputation models.

Date	Impute 1	Impute 2	Impute 3	Impute 4	Impute 5
01/04/1965	0.004	0.003	0.003	0.002	0.001
01/18/1965	0.001	0.022	0.008	0.006	0.017
02/01/1965	0.011	0.011	0.003	0.008	0.008
02/15/1965	0.056	0.056	0.056	0.056	0.056
03/01/1965	0.018	0.002	0.005	0.012	0.029
03/15/1965	0.011	0.006	0.002	0.010	0.002
03/29/1965	0.032	0.010	0.009	0.005	0.040
04/12/1965	0.002	0.046	0.022	0.003	0.044
04/26/1965	0.083	0.083	0.083	0.083	0.083
05/10/1965	0.014	0.006	0.002	0.001	0.005
05/24/1965	0.016	0.018	0.003	0.006	0.007
06/07/1965	0.004	0.022	0.002	0.002	0.002
06/21/1965	0.011	0.003	0.005	0.024	0.013
06/29/1965	0.015	0.014	0.004	0.035	0.019
07/06/1965	0.002	0.007	0.005	0.009	0.003
07/06/1965	0.008	0.012	0.040	0.007	0.002
07/19/1965	0.007	0.006	0.017	0.003	0.011
08/02/1965	0.014	0.004	0.008	0.008	0.008
08/16/1965	0.013	0.004	0.027	0.010	0.010
08/30/1965	0.051	0.051	0.051	0.051	0.051
09/13/1965	0.034	0.021	0.013	0.011	0.014
09/27/1965	0.012	0.004	0.008	0.005	0.002
10/11/1965	0.005	0.029	0.011	0.023	0.012
10/25/1965	0.008	0.009	0.006	0.007	0.002
10/25/1965	0.067	0.067	0.067	0.067	0.067
11/08/1965	0.005	0.030	0.008	0.002	0.002
11/08/1965	0.025	0.031	0.001	0.005	0.001
11/22/1965	0.003	0.012	0.014	0.027	0.008
12/06/1965	0.006	0.004	0.027	0.006	0.003
Totals	**0.538**	**0.593**	**0.510**	**0.494**	**0.522**

The analysis used multiple imputation to estimate annual doses that contain censored results. Other methods that have been used for this purpose include expectation-maximization (Datta et al. 2008) and Bayesian methods (Xue and Shore 2003). The Project chose the multiple imputation method because it tends to give reasonable results for the types of external monitoring data the Project commonly encounters, and it can be used with techniques like ROS that the Project has used extensively in internal dose coworker modeling, and because multiple imputation is more intuitive than expectation-maximization and Bayesian methods.

4.0 COWORKER MODELS

Up to this point the discussion has focused on methods for imputing values for censored individual dosimeter readings to obtain an unbiased estimate[5] of an individual's annual dose. Once this is performed for all individuals in a given population of monitored workers, a coworker model can be developed from the dataset to estimate annual doses for unmonitored workers who are assumed to be from the same population. The general procedure is:

- Fit a lognormal model to all of the individual dosimeter readings (censored and uncensored) for a given year using ROS. The resulting model, as discussed in the previous section, is referred to as the "lognormal imputation model."

- Draw random values below the censoring level from the imputation model and assign these values to the censored dosimeter readings during the year. This is in contrast to the previous practice of imputing a constant one-half of the censoring level (i.e., the LOD/2) for each censored result.

- Sum all of the dosimeter readings, uncensored and imputed, for each individual in the year to obtain the annual dose for each individual.

- Repeat this procedure *k* times (e.g., 10 times) to create *k* complete datasets.

- Fit a separate lognormal model to each of the *k* complete datasets to obtain *k* values of the log(GM) and *k* values of the log(GSD). In principle, the coworker model can be based on any appropriate probability distribution supported by IREP. The lognormal model is used by default unless the statistician performing the analysis decides that there is significant lack-of-fit.

- Average the *k* values of the log(GM) to obtain the final estimate of the log(GM), average the *k* values of the log(GSD) to obtain the final estimate of the log(GSD), and exponentiate both final estimates to get the final estimates of the GM and GSD.

- Calculate the parametric 50th-percentile (GM) and 95th-percentile ($GM \times GSD^{1.645}$) doses, which are assigned to unmonitored workers.

A coworker model based on this analysis was developed for 732 annual neutron doses (the data for the individual in Table 1-1 are from this dataset). Figure 4-1 shows the $k = 10$ sets of data. External dose datasets tend to be large enough (like this one) to allow good parameter estimates to be made. The statistician performing the analysis will make the judgment as to whether or not a given dataset is large enough to provide usable parameter estimates. Results in pink consist entirely of imputed dosimeter results, whereas the results in gold have at least one uncensored result. The red line is the final lognormal model having parameters that are the means of the parameters that were obtained

---

5 Technically speaking, because the true value is unknown, we cannot know the bias of the estimate. Nevertheless, an unbiased estimate, as it is used in this discussion, is taken to be that of an unbiased estimator.

from fits to the 10 datasets: the final estimate of the GM is 0.02799, and the final estimate of the GSD is 4.673. Because the actual uncensored annual doses are known, it is useful to compare them with the predictions of this coworker model (Figure 4-2). The first thing to notice is that the actual annual doses (pink dots) are not lognormal – they cannot be because there are negative doses. Datasets like this can often be modeled as a mixture of a normal distribution and a lognormal distribution (ORAUT 2009). The predictions of the lognormal coworker model that was developed from the multiply imputed annual doses are the black line in Figure 4-2. The lognormal model fits the actual annual doses about as well as can be expected given that the data are not lognormal. The 50th- and 95th-percentile doses (vertical red lines), which are the only predicted doses that are assigned to workers, match the actual doses very well. Specifically, the ratio of the predicted 50th-percentile dose (0.028 rem) to the observed 50th-percentile dose (0.015 rem) is 1.848, and the ratio of the predicted 95th-percentile dose (0.354 rem) to the observed 95th-percentile dose (0.326 rem) is 1.085.

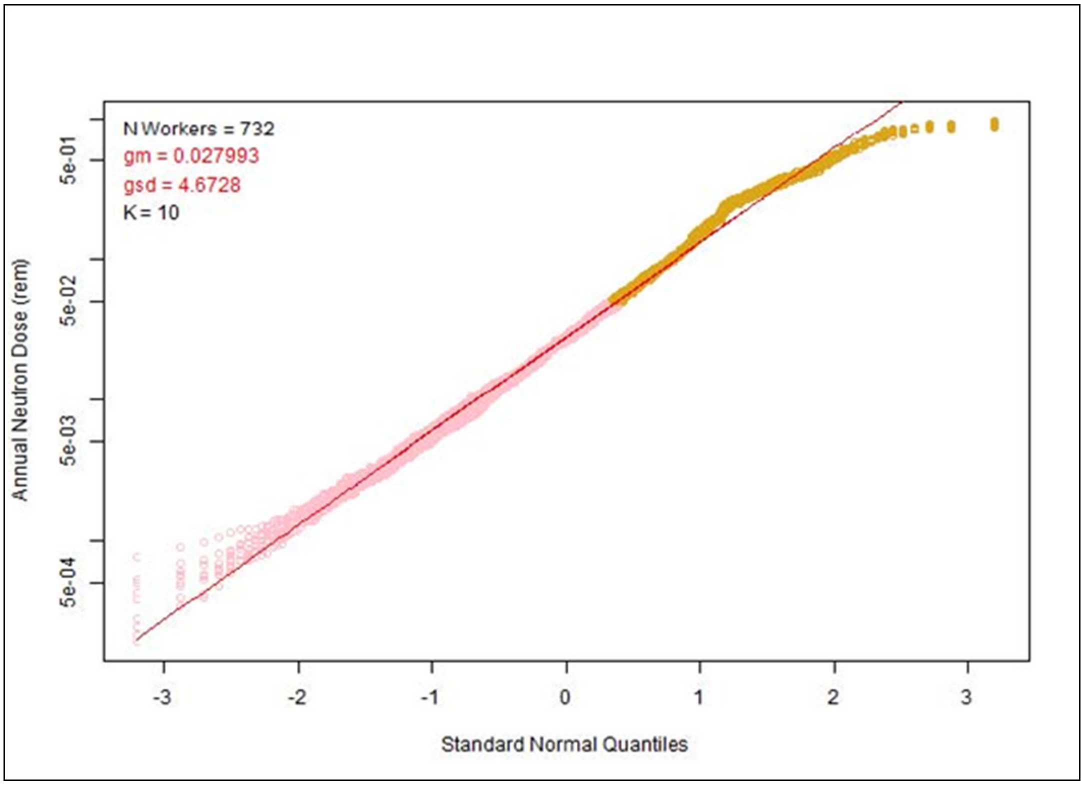

Figure 4-1. Multiple imputation coworker model.

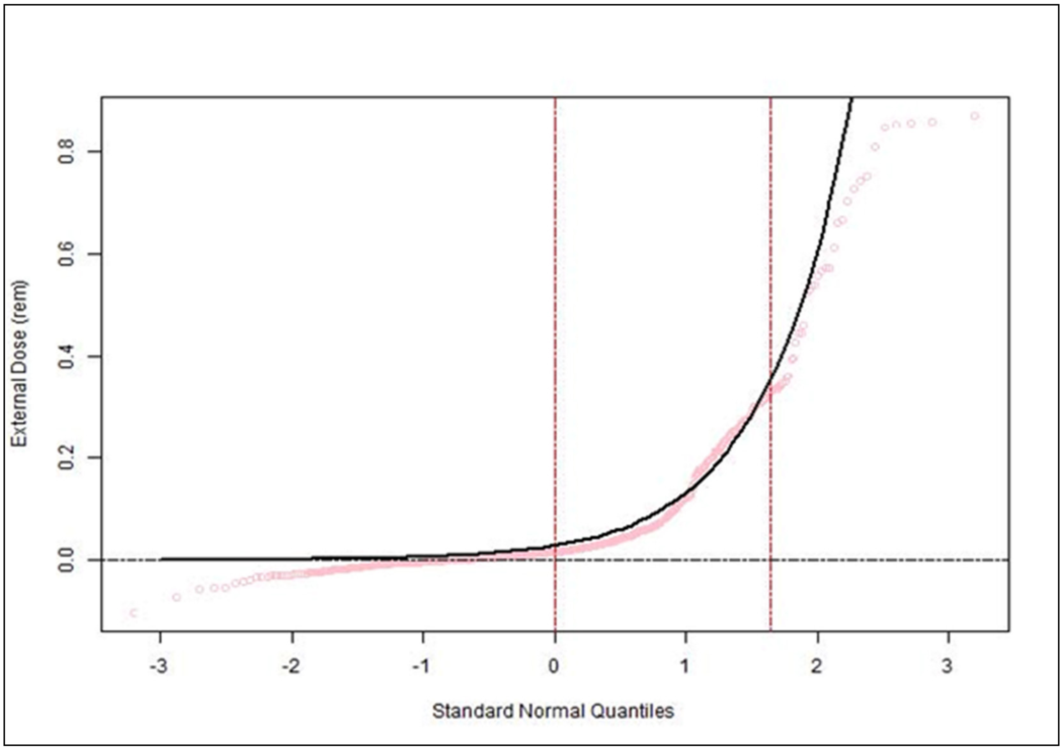

Figure 4-2. Multiple imputation coworker model (black line) versus actual annual doses (pink points). The vertical red lines are the 50th- and 95th-percentile doses.

5.0 LOD/2 IMPUTATION MODEL

For comparison purposes, a coworker model was developed by imputing the LOD/2 for all <LOD results (Figure 5.1). For these data the GM of the lognormal distribution is 0.05553 and the GSD is 4.558. The GSDs of this imputation model and the multiple imputation model are very similar but, as expected, the GM of model using the LOD/2 imputed data is greater than the GM of the model using the multiply imputed data. A number of factors contribute to the magnitude of the differences in parameters the two methods produce, including the fraction of data that are censored and the details of the multiple imputation model. In general, the overestimation of the LOD/2 imputation model increases as the degree of censoring goes up. The degree to which the LOD/2 imputation model overestimates the annual doses is clearly shown in Figure 5-2, where the predictions of the lognormal coworker model developed from the LOD/2 imputed data are the black line (compare this plot to the one in Figure 4-2). The ratio of the predicted 50th-percentile dose (0.056 rem) to the observed 50th-percentile dose (0.015 rem) is 3.665, and the ratio of the predicted 95th-percentile dose (0.673 rem) to the observed 95th-percentile dose (0.326 rem) is 2.066.

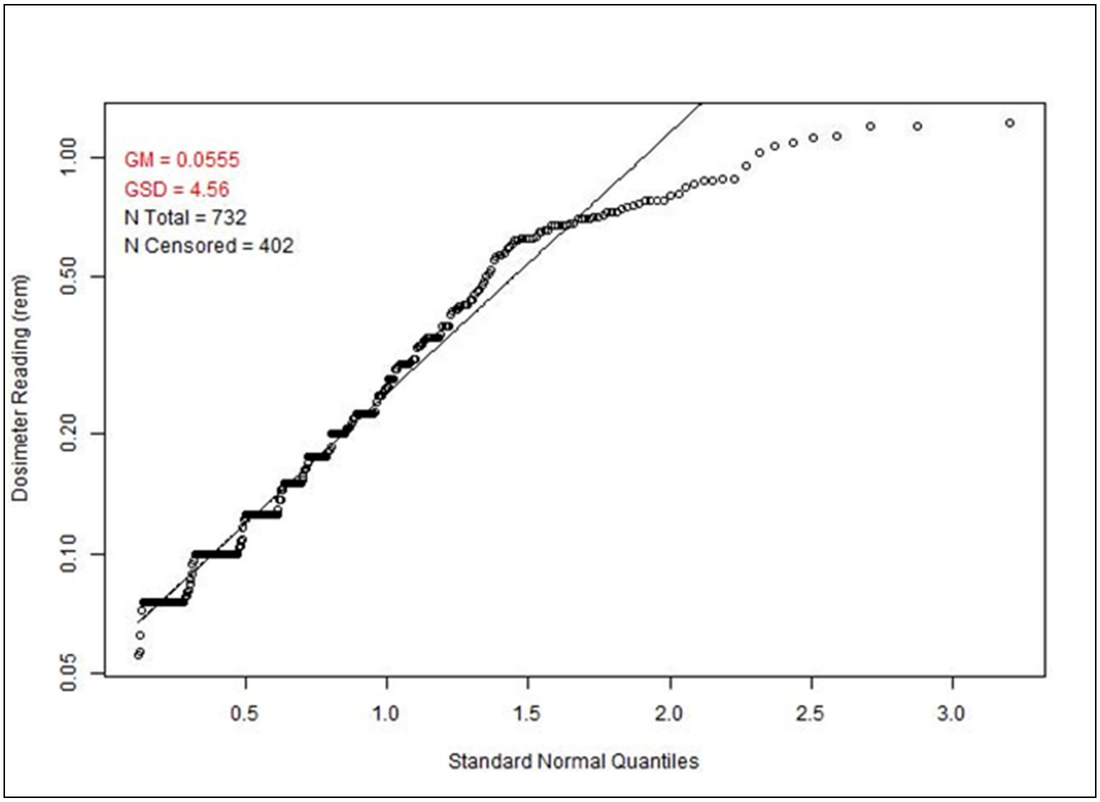

Figure 5-1. LOD/2 imputation coworker model.

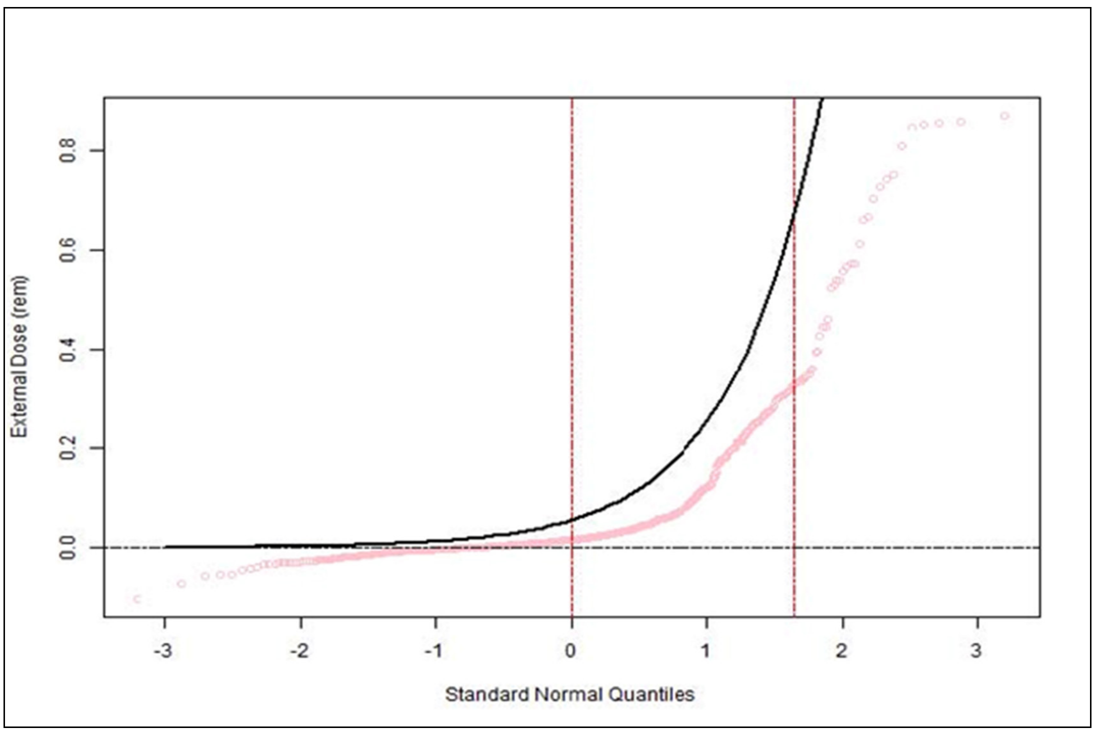

Figure 5-2. LOD/2 imputation coworker model (black line) versus actual annual doses (pink points). The vertical red lines are the 50th- and 95th-percentile doses.

6.0 SUMMARY

This report supports use of a new method, referred to as multiple imputation, to replace censored (<LOD) external dosimeter readings with estimates of the dose to facilitate the calculation of annual external dose. This method replaces the previous practice of substituting <LOD readings with the LOD/2. Multiple imputation can provide more accurate estimates of annual dose than can be obtained with LOD/2 substitution.

REFERENCES

Datta, D., S. Singh, B. E. Johnson, and H. S. Kushwaha, 2008, "Maximum Likelihood Estimates of Mean and Variance of Occupation Radiation Doses Subjected to Minimum Detection Levels," *Radiation Protection Dosimetry*, volume 129, number 4, pp. 411–418 (electronically published December 14, 2007).

Gelman, A., and J. Hill, 2006, *Data Analysis Using Regression and Multi-Level/Hierarchical Models*, Cambridge University Press, Cambridge, Massachusetts, December 18.

Helsel, D. R., 2012, *Statistics for Censored Environmental Data Using Minitab® and R, Second Edition*, John Wiley & Sons, Hoboken, New Jersey.

Krishnamoorthy, K., A. Mallick, and T. Mathew, 2009, "Model-Based Imputation Approach for Data Analysis in the Presence of Non-Detects," *Annals of Occupational Hygiene*, volume 53, number 3, pp. 249–263.

Lubin, J. H, J. S. Colt, D. Camann, S. Davis, J. R. Cerhan, R. K. Severson, L. Bernstein, and P. Hartge, 2004, "Epidemiologic Evaluation of Measurement Data in the Presence of Detection Limits," *Environmental Health Perspectives*, volume 112, number 17, pp. 1691–1696.

NCRP (National Council on Radiological Protection and Measurements), 2008, *Uncertainties in the Measurement and Dosimetry of External Radiation*, Report 158, Bethesda, Maryland, November 19.

ORAUT (Oak Ridge Associated Universities Team), 2009, *Analysis of Bioassay Data with a Significant Fraction of Less-Than Results*, ORAUT-RPRT-0044, Rev. 00, Oak Ridge, Tennessee, August 7.

ORAUT (Oak Ridge Associated Universities Team), 2011, *Use of Coworker Dosimetry Data for External Dose Assignment*, ORAUT-OTIB-0020, Rev. 03, Oak Ridge, Tennessee, November 14.

Xue, X., and R. E. Shore, 2003, "A Method for Estimating Occupational Radiation Doses Subject to Minimum Detection Levels," *Health Physics*, volume 84, number 1, pp. 61–71.