DOE Review Release 09/25/2025

Applications of Regression in Dose Reconstruction	ORAUT-RPRT-0087	Rev. 01
	Effective Date:	09/17/2025
	Supersedes:	Revision 00

Subject Expert(s): Daniel M. Sheanshang, Matthew H. Smith, and Keith A. McCartney

Document Owner Approval:	Signature on File Nancy Chalmers, Document Owner	Approval Date:	09/15/2025
Concurrence:	Signature on File Wade C. Morris, Objective 1 Manager	Concurrence Date:	09/16/2025
Concurrence:	Signature on File Scott R. Siebert, Objective 3 Manager	Concurrence Date:	09/15/2025
Concurrence:	Signature on File Jennifer L. Hoff, Project Director	Concurrence Date:	09/15/2025
Approval:	Signature on File Brant Ulsh, Acting Associate Director of Science	Approval Date:	09/17/2025

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

☐ New ☒ Total Rewrite ☐ Revision

PUBLICATION RECORD

EFFECTIVE DATE	REVISION NUMBER	DESCRIPTION
03/13/2018	00	New report initiated to introduce ordinary least squares regression and quantile regression as alternative methods for estimating beta dose based on gamma dose for workers whose dosimeters were relatively insensitive to beta radiation. Initiated by Thomas R. LaBone. Incorporates formal internal and NIOSH review comments. Training required: As determined by the Objective Manager. Initiated by Thomas R. LaBone.
09/17/2025	01	Revised to remove Section 4.0 and other ratio-related content to bring this report in line with ORAUT-RPRT-0106, *Applications of Regression Models and Ratios*, and to incorporate a technique to apply quantile regression in the Interactive RadioEpidemiological Program (IREP). Removed “External” from the document title to reflect the broadened applicability in dose reconstruction. Incorporates formal internal and NIOSH review comments. Constitutes a total rewrite of the document. Training required: As determined by the Objective Manager. Initiated by Wade C. Morris and authored by Nancy Chalmers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE **TITLE** **PAGE**

ACRONYMS AND ABBREVIATIONS

CDF	cumulative distribution function
DOE	U.S. Department of Energy
GM	geometric mean
GSD	geometric standard deviation
IREP	Interactive RadioEpidemiological Program
mrem	millirem
NIOSH	National Institute for Occupational Safety and Health
OLS	ordinary least squares
ORAU	Oak Ridge Associated Universities
ORAUT	ORAU Team
q-q	quantile-quantile
QR	quantile regression
SRDB Ref ID	Site Research Database Reference Identification (number)

1.0 INTRODUCTION

NOTE: This report is intended as a reference for statisticians who use the methods described in the report. The details of its application and necessary justifications are provided in the site-specific reports where the methods are used.

In general, regression analysis aims to predict a variable based on one or more other variables. This report addresses the case where the aim is to predict one variable *Y* (the dependent variable or response) based on one other variable *X* (the independent variable or predictor). This is known as a bivariate analysis because it involves two variables.

Historically, in the dose reconstruction program, the standard method for estimating *Y* from *X* has been fitting a lognormal model to the ratio *Y*/*X* using regression on order statistics [Helsel 2012]. This historical method is a univariate method, meaning it involves one variable, the ratio of *Y* to *X*. ORAUT-RPRT-0106 Rev. 00, *Application of Regression Models and Ratios*, shows that collapsing bivariate (or paired) data to univariate data results in a loss of information and gives incorrect results, except in one specific case [Oak Ridge Associated Universities (ORAU) Team (ORAUT) 2023].

This report introduces ordinary least squares (OLS) regression [Weisberg 2005] and quantile regression (QR) [Koenker 2005; Cade and Noon 2003] as bivariate methods for predicting *Y* given *X*. Regardless of method, the goal is to calculate either:

- A point estimate of *Y* given a point estimate of *X*, or
- A distribution for *Y* given a point estimate of *X*.

This discussion of OLS and QR is made more concrete by using the two regression methods to analyze data from a U.S. Department of Energy (DOE) facility. The example in this report uses paired external dosimetry data, but the regression methods can also be used for other types of paired data. Section 2.0 introduces the dosimetry data. Sections 3.0 and 4.0 present OLS and QR, respectively. Section 5.0 details a method to apply regression in dose reconstruction tools. Section 6.0 presents summary and conclusions. The dataset and R code used in the preparation of this report are available in ORAUT [2025].

2.0 DOSIMETRY DATA

Either by design or nature, external dosimeters can have different sensitivities to different types and energies of radiation. For example, some dosimeters worn by workers are sensitive to both electrons and photons and others are sensitive to photons but relatively insensitive to electrons. To estimate electron dose for workers who wore dosimeters that are insensitive to electrons, health physicists can use the dosimeter readings to develop a model for estimating the electron dose as a function of the photon dose. Problems of this type are not uncommon in external dosimetry.[1]

The dataset used to illustrate the regression analysis methods consists of reported monthly paired photon and electron doses from a DOE site over a 7-year period (1966 to 1972). The censoring level was 30 mrem for photon dose and 50 mrem for electron dose (i.e., a photon dose of less than 30 mrem was reported as <30 mrem). Only paired photon and electron doses that are both uncensored are considered in the analyses. There are more than sufficient uncensored pairs to illustrate the methods, and using only the uncensored pairs is favorable to claimants. In future site-specific analyses, the team analyzing the data might choose to include censored doses, especially if

[1] Neutron vs. photon doses, extremity vs. whole-body doses, etc., can also be addressed in the same way described here.

the model is for annual dose, by using imputation or other methods, but that complication is not necessary for this report.

There are 1,159 uncensored photon and electron pairs in the dataset. Figure 2-1 shows those pairs in a scatterplot with linear scales. In Figure 2-1, there does not appear to be much of a relationship between the paired photon and electron doses. In the dose reconstruction program, logarithmic transformations of the variables are common.[2] Figure 2-2 is a scatterplot of these pairs on log scales. With the log transform, there now appears to be a relationship between the paired photon and electron doses, so the log transform is used in the remainder of this report.

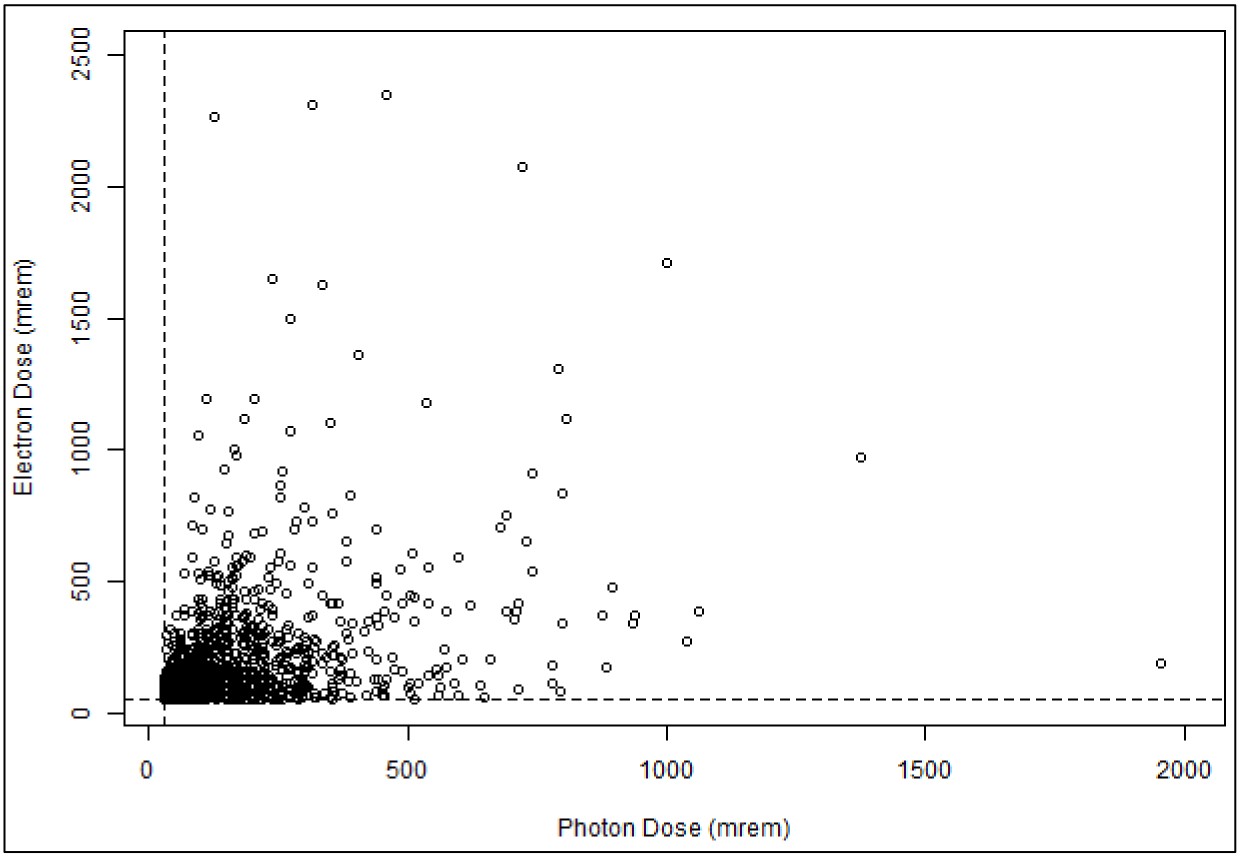

Figure 2-1. Scatterplot of electron versus photon dose with linear scales; the electron dose censoring level is 50 mrem (horizontal dashed line) and the photon dose censoring level is 30 mrem (vertical dashed line). Attachment A contains an extended description.

2 In this report, "log" refers to the natural logarithm.

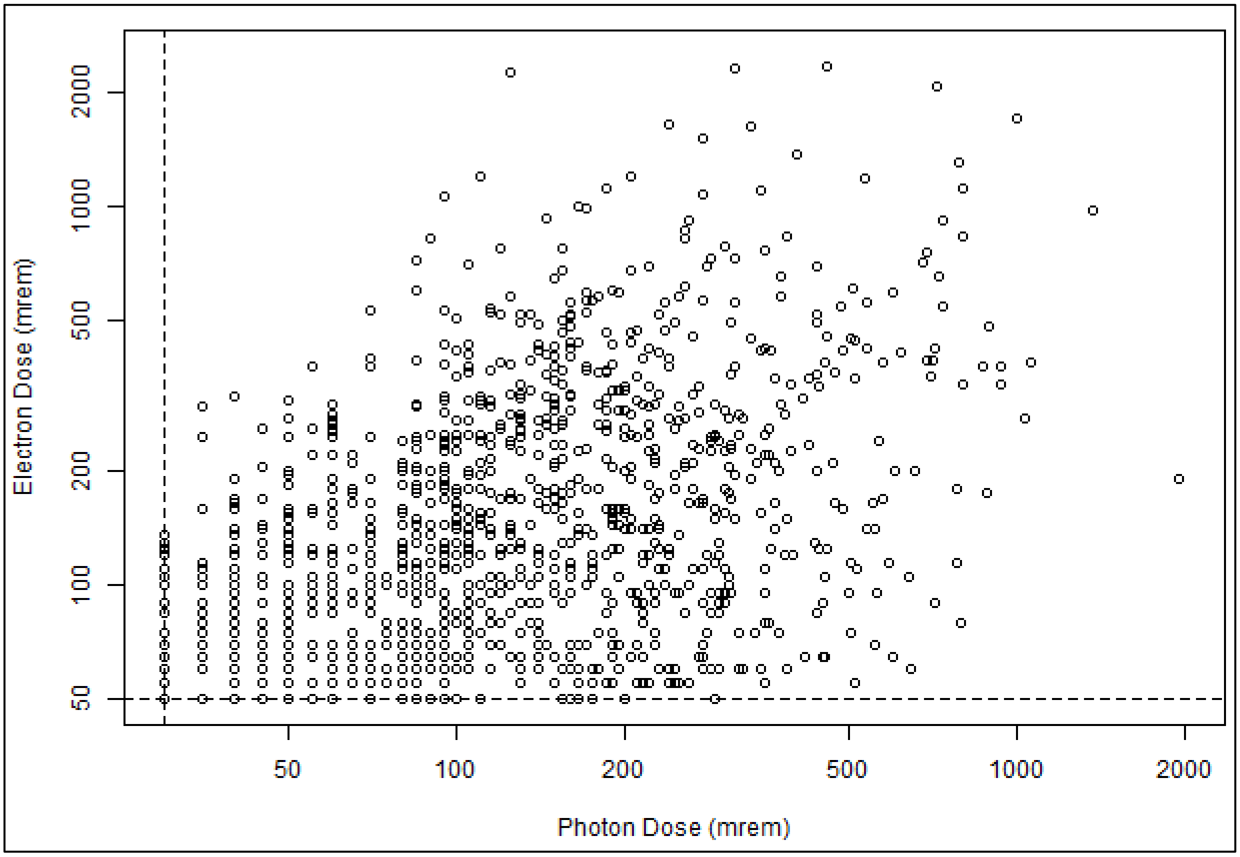

Figure 2-2. Scatterplot of electron versus photon dose with log scales; the electron dose censoring level is 50 mrem (horizontal dashed line) and the photon dose censoring level is 30 mrem (vertical dashed line). Attachment A contains an extended description.

This report fits the paired photon and electron data with both types of regression to illustrate the two methods. In practice, the dataset, assumptions, and limitations of each method should be evaluated by the statistician to choose which of the two regression methods (OLS or QR) is most appropriate.

3.0 ORDINARY LEAST SQUARES

The scatterplot of the log electron dose versus the log photon dose in Figure 2-2 is fairly linear, which suggests that a linear regression model could be fit to the data as follows:

$$E\left[\log\left(H_e\right)\right] = \beta_0 + \beta_1 \log\left(H_\gamma\right) \quad (3\text{-}1)$$

where

- $E[\log(H_e)]$ is expectation (mean) of the log of the electron dose for a given photon dose
- H_e is observed electron doses (mrem)
- H_γ is observed photon doses (mrem)
- β_0 is intercept in the model
- β_1 is slope in the model

3.1 ASSUMPTIONS AND LIMITATIONS

Note that OLS regression is not limited to a linear function with only an intercept and slope. For example, it allows addition of a quadratic term to the regression. The linear function is used here for simplicity.

As described in Section 3.3, OLS regression has a major limitation in that, for any given photon dose, the percentiles for electron dose are completely specified by the normal distribution. This means that the percentiles are symmetric around the mean (all the percentile lines must have the same slope) and, for example, there cannot be one relationship (e.g., linear) for the 95th percentile and another relationship (e.g., quadratic) for the 5th percentile. Therefore, the percentiles that are calculated with OLS regression might not accurately model what is occurring in the entire dataset.

The assumptions of OLS regression deal with the errors of the model. Error is the vertical distance from each point to the true relationship between the two variables. OLS regression assumes [Weisberg 2005]:

1. Independence of errors,
2. Errors have a mean of zero across all values of the independent variable,
3. Errors have constant variance across all values of the independent variable, and
4. Errors are normally distributed.

Errors are unknowable, so the assumptions can be evaluated with the residuals, the vertical distance from each point to the fitted relationship. Therefore, residual evaluations cannot be performed until a model has been fit. Ideally, the choice between OLS and QR would be made before a fit is done. A trained eye can evaluate a scatterplot like the one in Figure 2-2 and draw conclusions about the OLS assumptions, especially violations. Here, plots to evaluate the OLS assumptions are presented and evaluated, but this is usually not necessary.

Figure 3-1 is the residuals versus predicted (or fitted) values plot. Figure 3-2 is a normal quantile-quantile (q-q) plot of the residuals.

1. The plot of residuals versus predicted values should show random scatter (no pattern) of the points around the zero line. Based on Figure 3-1, this assumption seems to be reasonable.
2. The plot of residuals versus predicted values should be evenly distributed around the zero line. Based on Figure 3-1, this assumption seems to be violated because of the “diagonal floor”[3] at the lower left side of the plot. This diagonal floor stems from the censoring level cutoffs for the doses and can be identified in the scatterplot in Figure 2-2.
3. The plot of residuals versus predicted values should show an even amount of scatter from left to right. Based on Figure 3-1, this assumption seems to be violated partially because of the diagonal floor but also because lower doses tend to have less scatter in the residuals. This violation can be identified in the scatterplot in Figure 2-2 because lower values of photon dose tend to have less scatter in their paired electron dose than do higher values of photon dose.

---

[3] “Diagonal floor” is not a formal term. It is used here to describe the behavior of the negative residuals that appear to be cut off at a diagonal in the lower left portion of Figure 3-1.

4. The q-q plot of the residuals should show the points falling fairly close to the line determined by the normal distribution. Based on Figure 3-2, this assumption seems to be violated, especially in the tails.

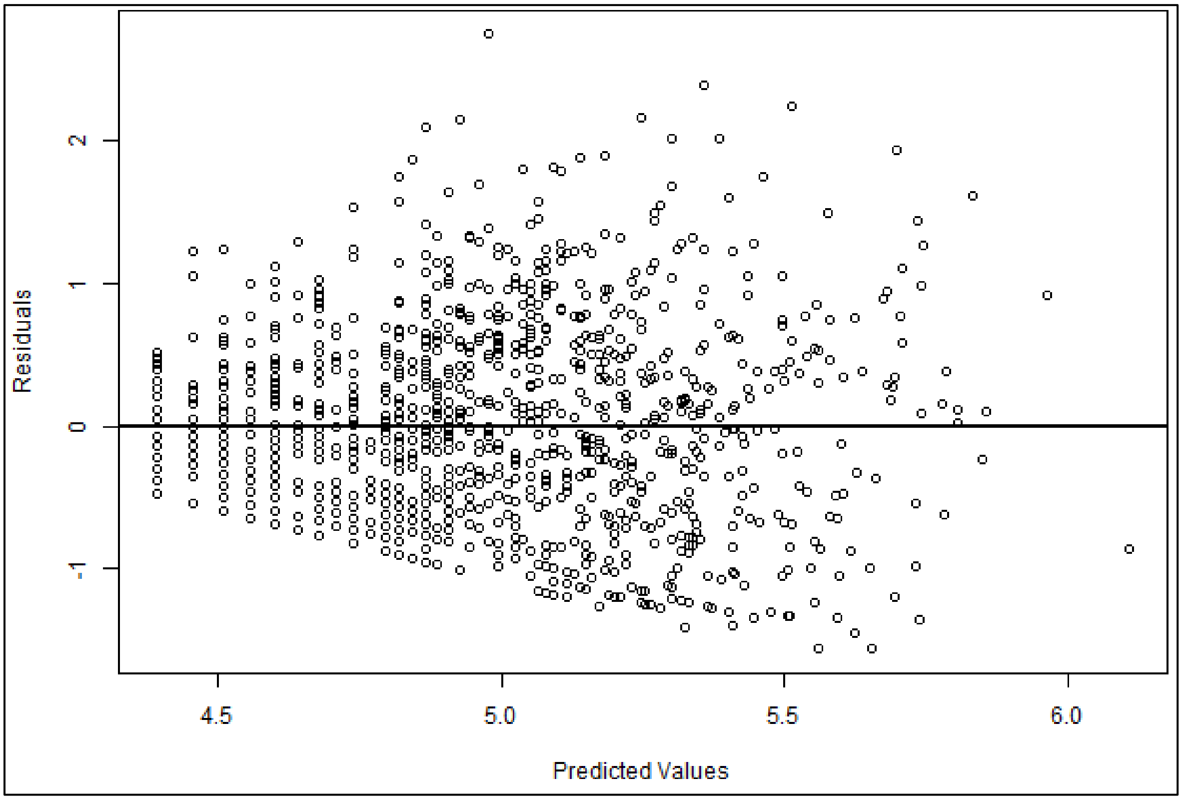

Figure 3-1. Residuals versus predicted (or fitted) values plot, with horizontal line at zero. Attachment A contains an extended description.

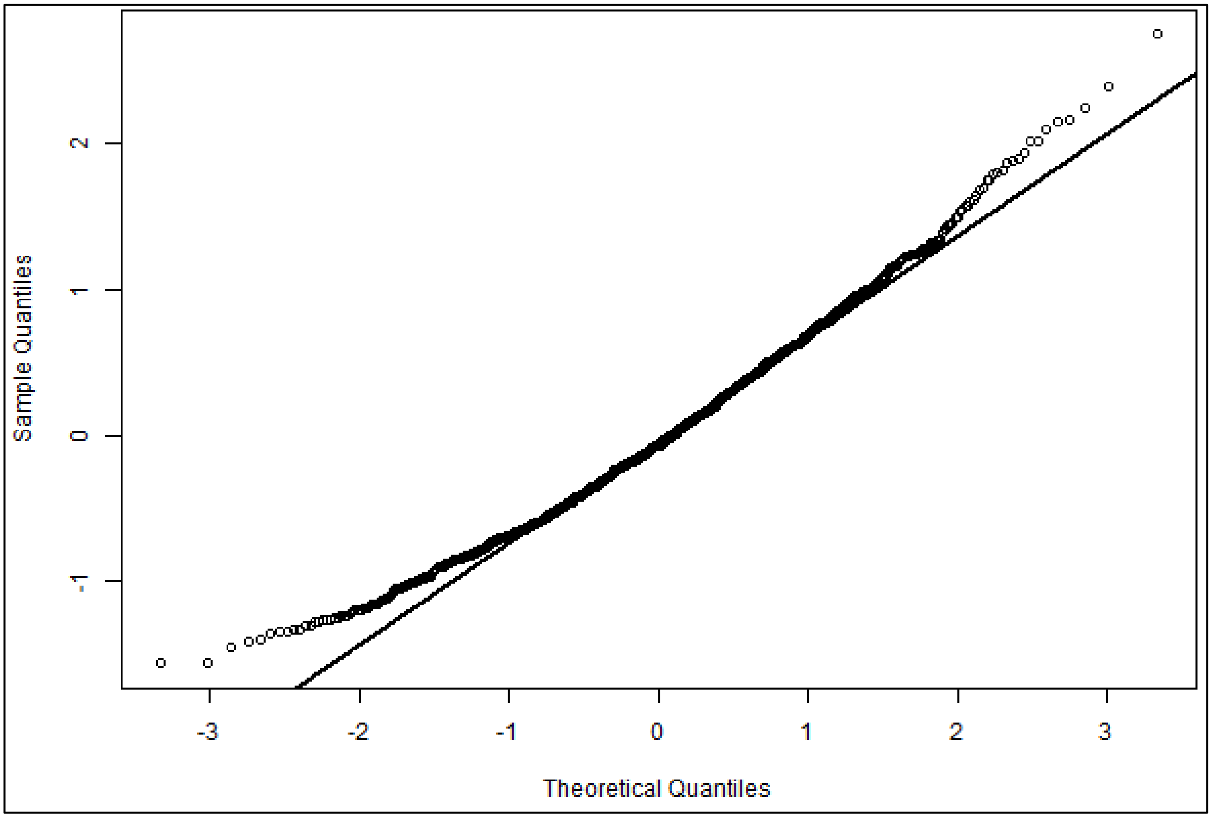

Figure 3-2. Normal q-q plot of the residuals. Attachment A contains an extended description.

In summary, an evaluation of the scatterplot in Figure 2-2 shows violations of assumptions 2 and 3 and Figure 3-2 shows a violation of assumption 4, so OLS with log transformation is probably not the most appropriate regression technique; it is presented here for illustrative purposes.

3.2 FITTED MODEL AND POINT ESTIMATE (MEAN)

OLS regression of the log electron dose on the log photon dose gives a fitted model of:

$$\log\left(\widehat{H_e}\right) = b_0 + b_1 \log\left(H_\gamma\right) \tag{3-2}$$

where

$\log\left(\widehat{H_e}\right)$	is	log of predicted value of electron dose for a given log photon dose
b_0	is	2.9912, fitted intercept from the regression
b_1	is	0.4114, fitted slope from the regression

Figure 3-3 shows the fitted model as the solid line added to the scatterplot.

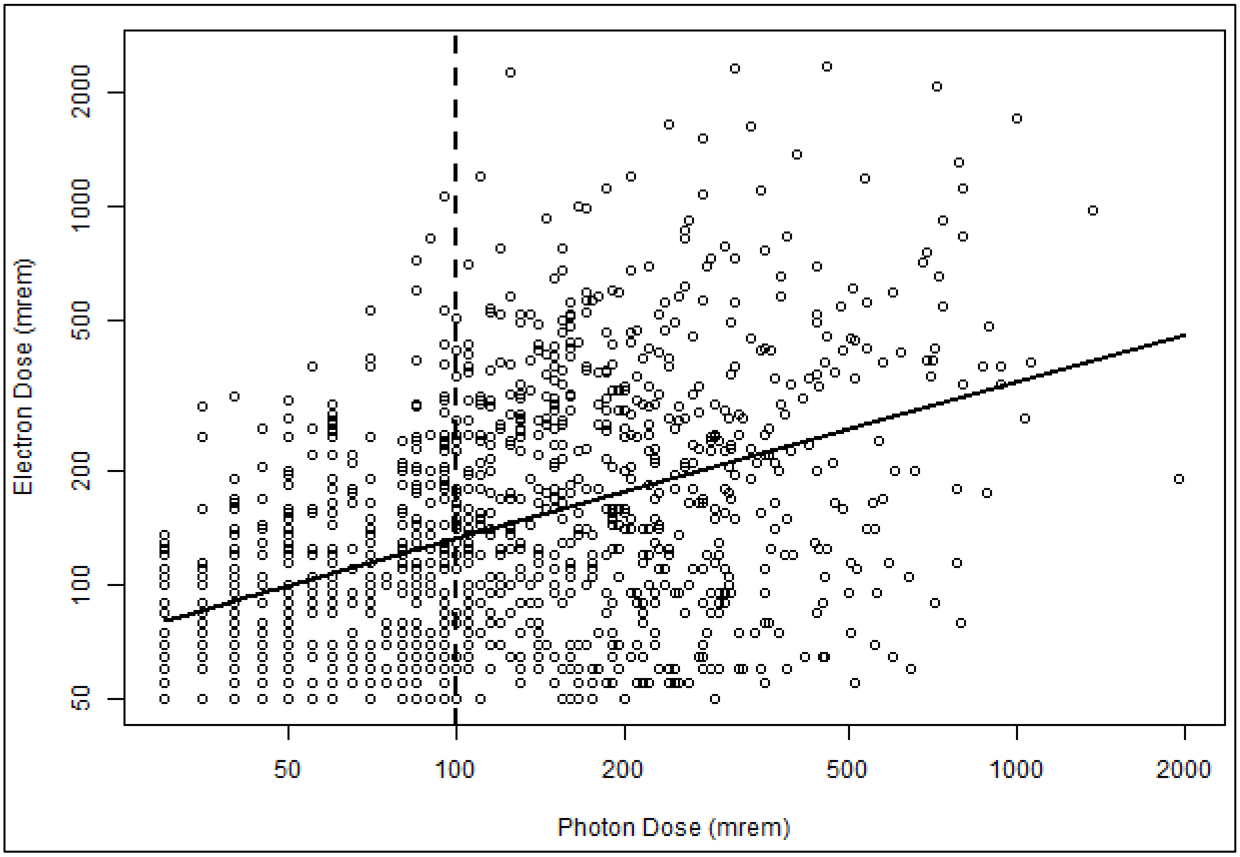

Figure 3-3. Log-log scatterplot of electron doses versus photon doses. The solid line is the fitted model using OLS regression.

The line is a conditional mean that identifies the mean log electron dose for a given log photon dose. Solving Equation 3-2 for the predicted value of electron dose gives:

$$\widehat{H_e} = \exp\left[b_0 + b_1 \log\left(H_\gamma\right)\right] \tag{3-3}$$

Therefore, the predicted mean electron dose for a 100-mrem photon dose (the vertical dashed line in Figure 3-3) is:

$$\widehat{H_e} = \exp\left[2.9912 + 0.4114 \log\left(100\text{ mrem}\right)\right] = 132.3825\text{ mrem} \tag{3-4}$$

3.3 QUANTILES

The OLS regression in Figure 3-3 assumes that log electron doses are normally distributed (assumption 4 from Section 3.1) about the mean regression line and that the variance of the normal distribution is constant for all photon doses (assumption 3 from Section 3.1). In the fitted model from the previous section, the variance of the normal distribution is estimated along with the slope and intercept parameters. The square root of that estimated variance (the standard deviation, s is 0.6836) can be used to estimate quantiles. Any percentile line can be constructed by simply shifting the mean line upward or downward by the appropriate number of standard deviations, keeping the same slope:

$$\log\left(\widehat{H_{e,x}}\right) = b_0 + b_1 \log\left(H_\gamma\right) + s Z_x \tag{3-5}$$

where Z_x is the xth percentile of the standard normal distribution. Solving Equation 3-5 for the predicted value of electron dose gives:

$$\widehat{H_{e,x}} = \exp\left[b_0 + b_1 \log\left(H_\gamma\right) + s\, Z_x\right] \qquad (3\text{-}6)$$

In particular, the 95th percentile line is:

$$\widehat{H_{e,95}} = \exp\left[b_0 + b_1 \log\left(H_\gamma\right) + s\, Z_{95}\right] \qquad (3\text{-}7)$$

where Z_{95} is 1.645, the 95th percentile of the standard normal distribution. The predicted 95th-percentile electron dose for a 100-mrem photon dose is:

$$\widehat{H_{e,95}} = \exp\left[2.9912 + 0.4114 \log\left(100 \text{ mrem}\right) + 0.6836 \times 1.645\right] = 407.5087 \text{ mrem} \qquad (3\text{-}8)$$

In Figure 3-4 the dashed diagonal lines denote the percentiles: the highest dashed line is the 99th-percentile electron dose as a function of the photon dose, the lowest dashed line is the 1st-percentile electron dose, and the dashed lines in between go from the 5th percentile to the 95th percentile in steps of 5. These percentile lines are symmetrically distributed about the mean line and, while each percentile line has its own unique intercept, they have the same slope as the mean line. Table 3-1 lists the intercepts and slopes for all the percentile lines in Figure 3-4.

The regression equations and quantiles presented here can be applied using the method in Section 5.0.

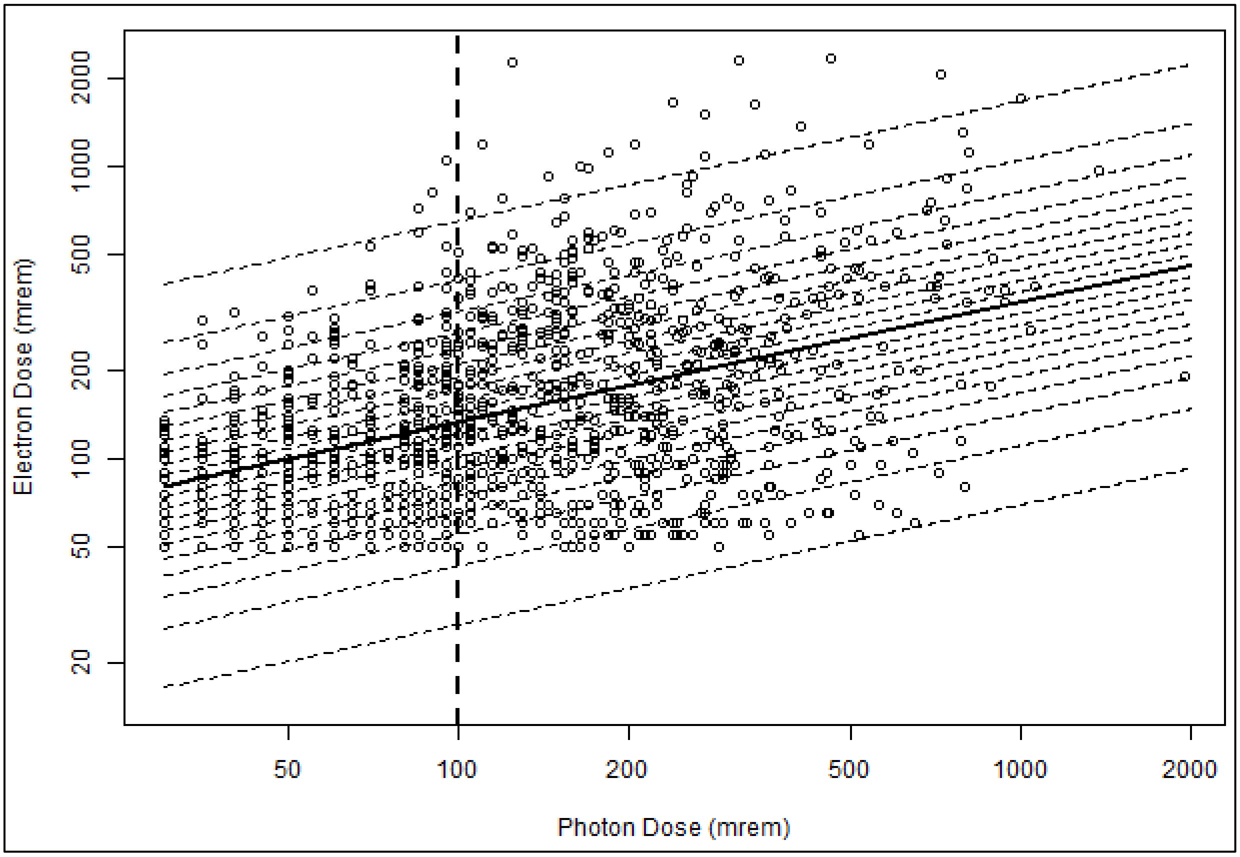

Figure 3-4. Log-log scatterplot of electron doses versus photon doses. The solid line is the OLS regression mean line and the dashed diagonal lines are the 1st through 99th percentiles based on the OLS model. Attachment A contains an extended description.

Table 3-1. Intercepts and slopes for percentile lines in Figure 3-4.

Percentile	Intercept	Slope
1	1.4010	0.4114
5	1.8669	0.4114
10	2.1152	0.4114
15	2.2827	0.4114
20	2.4159	0.4114
25	2.5302	0.4114
30	2.6328	0.4114
35	2.7278	0.4114
40	2.8180	0.4114
45	2.9053	0.4114
50	2.9912	0.4114
55	3.0771	0.4114
60	3.1644	0.4114
65	3.2546	0.4114
70	3.3497	0.4114
75	3.4523	0.4114
80	3.5665	0.4114
85	3.6997	0.4114
90	3.8672	0.4114
95	4.1156	0.4114
99	4.5814	0.4114

4.0 QUANTILE REGRESSION

The formal linear model used to fit the example dataset with QR is:

$$Q_x\left[\log\left(H_e\right)\right] = \beta_{0,x} + \beta_{1,x}\log\left(H_\gamma\right) \tag{4-1}$$

where

$Q_x[\log(H_e)]$	is xth percentile of the log electron dose for a given log photon dose
H_e	is observed electron doses (mrem)
H_γ	is observed photon doses (mrem)
$\beta_{0,x}$	is intercept of the model for the xth percentile
$\beta_{1,x}$	is slope of the model for the xth percentile

4.1 ASSUMPTIONS AND LIMITATIONS

Just as with OLS, QR is not limited to a linear function with only an intercept and slope. For example, it allows addition of a quadratic term to the regression. The linear function is used here for simplicity.

As described in Section 4.3, QR allows for different slopes and intercepts for different percentiles, whereas OLS assumes that all the percentile lines must have the same slope. In QR, there can be one relationship (e.g., linear) for the 95th percentile and another (e.g., quadratic) for the 50th percentile. This is not possible with OLS. This makes QR more flexible than OLS and able to better fit the quantiles because the QR lines can take forms different from each other as needed. Because QR is more flexible than OLS, it is possible for QR lines to cross. For example, if QR lines for the 50th and 60th percentiles cross, for some values of X, the predicted value for the 50th percentile would be larger than the predicted value for the 60th percentile. The statistician must take care to make sure that does not happen for the quantiles of interest.

QR is less sensitive to outliers than OLS regression. With QR, the four assumptions (from Section 3.1) of OLS are not necessary, so the censoring level cutoff that contributes to the violations of the OLS assumptions is not an issue with QR.

Historically, QR was more computationally intensive than OLS regression, but that is no longer an issue. QR is not readily available in spreadsheet software, but it is in most statistical software.

The biggest limitation of QR is that more data are usually needed to perform QR than for OLS, especially for extreme percentiles (close to 0 or 100). For example, if a dataset only has 30 pairs of data spread across a variety of values of the independent variable, a QR using the 95th percentile would be very suspect. OLS does not have this sample size limitation because it makes assumptions that (if met) compensate for the smaller sample size. The dataset used in this analysis has 1,159 uncensored pairs, so the sample size limitation is not an issue here.

Based on the evaluation of OLS assumptions in Section 3.1 and the QR assumptions in this section, QR is the more appropriate regression technique for these photon and electron data, and the analysis is presented in the following sections.

4.2 FITTED MODEL AND POINT ESTIMATE (MEDIAN)

QR of the median (50th-percentile) log electron dose on the log photon dose gives a fitted model of:

$$\log\left(\widehat{H_{e,50}}\right) = b_{0,50} + b_{1,50}\log\left(H_\gamma\right) \tag{4-2}$$

where

$\log\left(\widehat{H_{e,50}}\right)$	is	log of predicted value of electron dose for a given log photon dose
$b_{0,50}$	is	2.8636, fitted intercept from the median regression
$b_{1,50}$	is	0.4269, fitted slope from the median regression

Figure 4-1 shows the fitted model as the solid line added to the scatterplot. Note that the median QR line in Figure 4-1 is similar to the mean regression line from OLS in Figure 3-3.

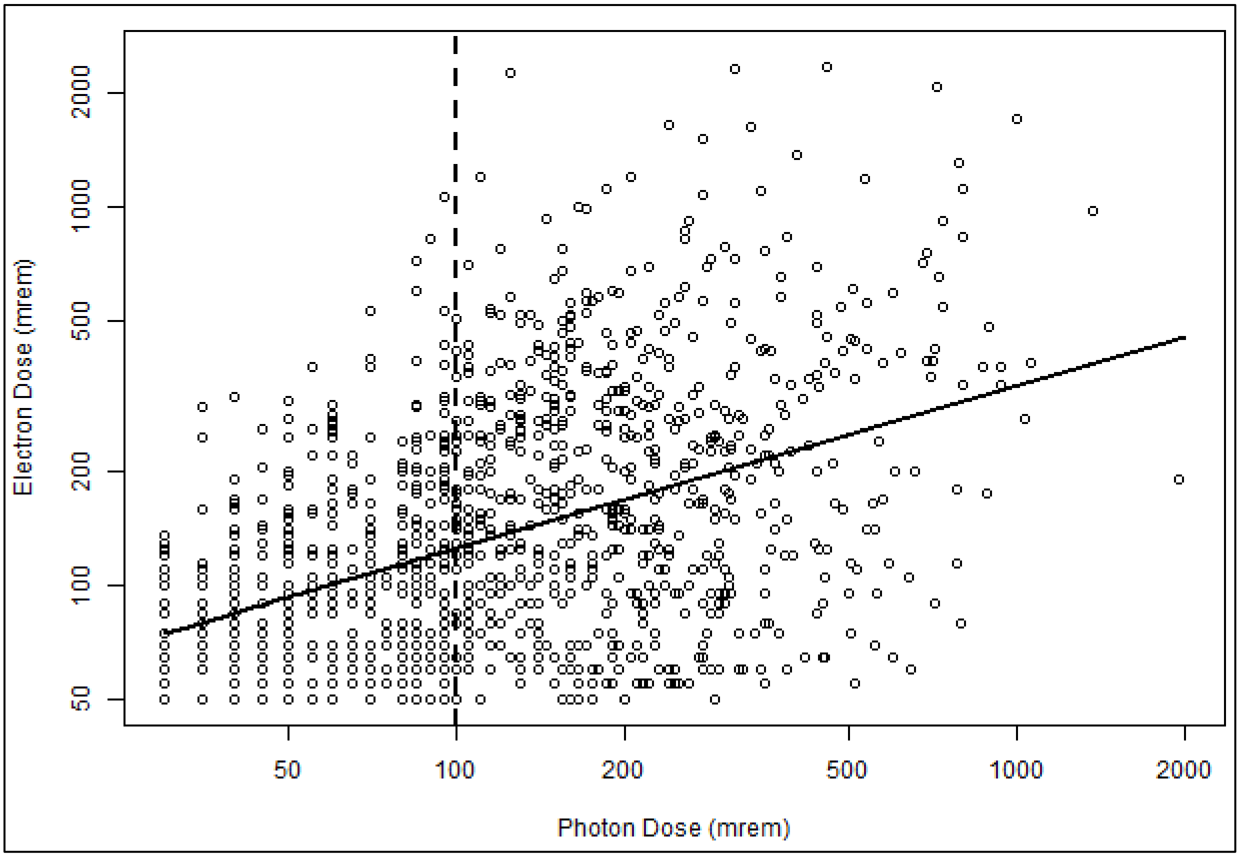

Figure 4-1. Log-log scatterplot of electron doses versus photon doses. The solid line is the fitted model using median QR.

Solving Equation 4-2 for the predicted value of median electron dose gives:

$$\widehat{H_{e,50}} = \exp\left[b_{0,50} + b_{1,50}\log\left(H_{\gamma}\right)\right] \quad (4\text{-}3)$$

Therefore, the predicted median electron dose for a 100-mrem photon dose (the vertical dashed line in Figure 4-1) is:

$$\widehat{H_{e,50}} = \exp\left[2.8636 + 0.4269\log\left(100\text{ mrem}\right)\right] = 125.1652\text{ mrem} \quad (4\text{-}4)$$

Note that in Equation 4-4 the predicted median electron dose for a 100-mrem photon dose using QR is 125.1652 mrem. The predicted mean electron dose for the same 100-mrem photon dose using OLS (Equation 3-4) is 132.3825 mrem.

4.3 OTHER QUANTILES

With OLS, all the percentile lines have the same slope as the mean line. That is not the case for QR (hence the *x* subscript on the intercept and slope terms in Equation 4-1), so each percentile must be fit separately using Equation 4-1. In Figure 4-2, the dashed diagonal lines denote the percentiles: the highest dashed line is the 99th-percentile electron dose as a function of the photon dose, the lowest dashed line is the 1st-percentile electron dose, and the dashed lines in between go from the 5th percentile to the 95th percentile in steps of 5. Typically, the statistician would not model all 21 of these percentiles, only the percentiles of interest. All 21 percentiles are included here for illustrative purposes. The intercepts and slopes for all the percentile lines in Figure 4-2 are given in Table 4-1.

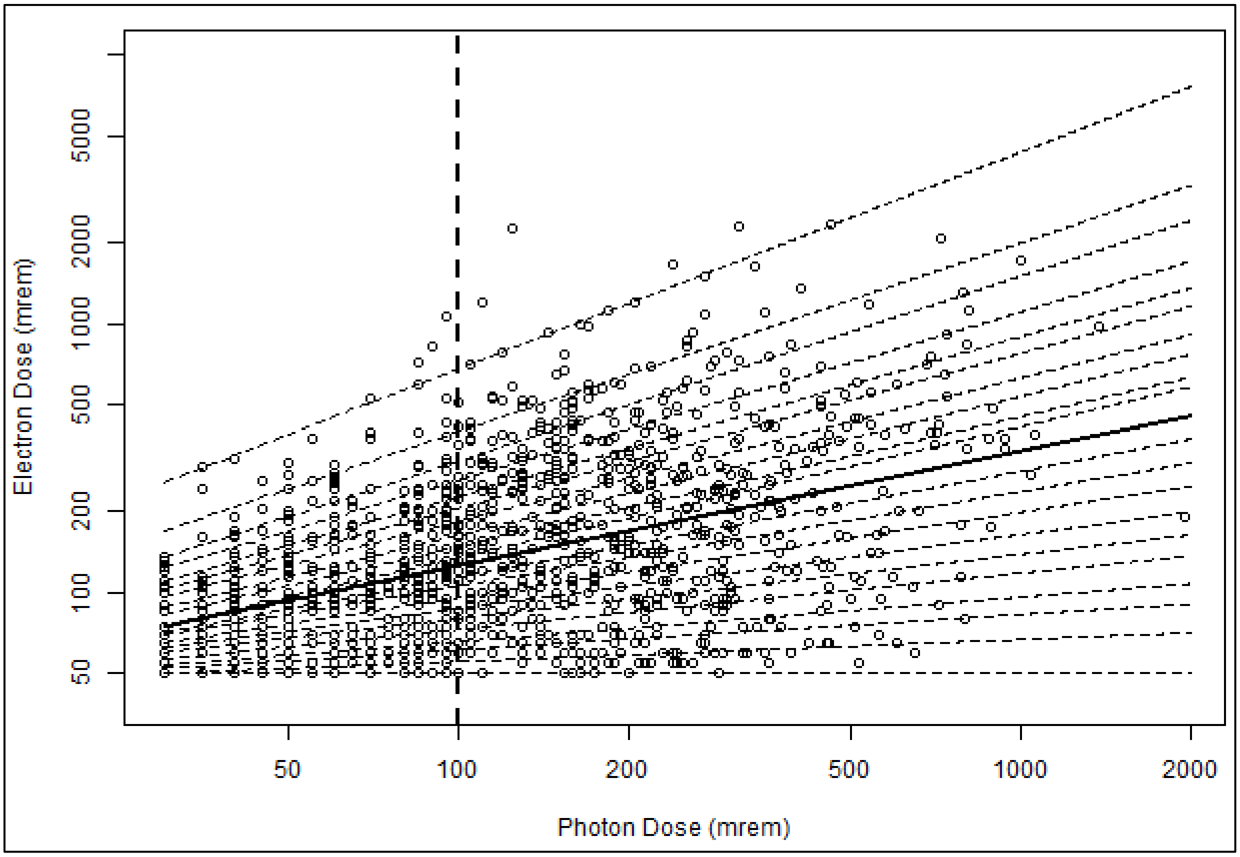

Figure 4-2. Log-log scatterplot of electron doses versus photon doses. The solid line is the median regression line and the dashed diagonal lines are the 1st through 99th percentiles based on the QR model.

Table 4-1. Intercepts and slopes for percentile lines in Figure 4-2.

Percentile	Intercept	Slope
1	3.9120	9.0907E-09
5	3.6308	0.0827
10	3.5060	0.1317
15	3.4186	0.1660
20	3.3179	0.2105
25	3.2816	0.2389
30	3.1905	0.2767
35	3.0885	0.3193
40	3.0464	0.3509
45	2.9213	0.3943

Percentile	Intercept	Slope
50	2.8636	0.4269
55	2.6602	0.4873
60	2.7687	0.4839
65	2.7199	0.5160
70	2.7584	0.5330
75	2.6755	0.5759
80	2.6787	0.5961
85	2.6853	0.6257
90	2.5693	0.6868
95	2.7393	0.7040
99	2.8008	0.8076

To calculate the predicted value of the 95th-percentile electron dose:

$$\widehat{H_{e,95}} = \exp\left[b_{0,95} + b_{1,95}\log\left(H_{\gamma}\right)\right] \tag{4-5}$$

Therefore, the predicted 95th-percentile electron dose for a 100-mrem photon dose (the vertical dashed line in Figure 4-2) is:

$$\widehat{H_{e,95}} = \exp\left[2.7393 + 0.7040\log\left(100\text{ mrem}\right)\right] = 396.0673\text{ mrem} \tag{4-6}$$

Note that in Equation 3-8 the predicted 95th-percentile electron dose for a 100-mrem photon dose using OLS is 407.5087 mrem and using QR (Equation 4-6) is 396.0673 mrem.

The regression equations and quantiles presented here can be applied using the method in Section 5.0.

5.0 APPLICATION IN DOSE RECONSTRUCTION TOOLS

The dose reconstruction tools are not equipped to handle regression equations like those in Equations 3-1 (OLS) and 4-1 (QR). The goal is to assign electron dose given a known photon dose for a worker. If assigning constant electron dose is appropriate for the application, the constant electron dose can be calculated as illustrated in Equations 3-4, 3-8, 4-4, and 4-6. If the application calls for assigning a distribution for electron dose, a more complicated application method must be used. The application method presented here balances appropriate statistical techniques with the capabilities of the dose reconstruction tools.

5.1 SLICE METHOD

The goal of this application method is to calculate the parameters of a distribution of the dependent variable *Y* (electron dose) given a known value of the independent variable *X* (photon dose). The method is called the “slice method” because it takes a vertical slice of the scatterplot (see vertical dashed line in Figures 3-3, 3-4, 4-1, and 4-2) at the known value of the independent variable. This is accomplished using regression lines for two quantiles, modeled with OLS or QR. Normal and lognormal distributions are the most commonly assigned dose distributions in the dose reconstruction project, so the theory is presented below for the normal distribution, and the lognormal is simply an extension [Cook 2010]. The dose reconstruction tools and IREP include the Weibull distribution and Cook [2010] presents theory for that distribution, if it is ever the appropriate distribution to assign.

Figure 5-1 is a plot of the 50th- and 95th-percentile QR lines from Figure 4-2 with a dotted vertical line at 100 mrem. The slice method takes a vertical slice of Figure 5-1 at 100-mrem photon dose, which intersects each QR line at one point, and determines the geometric mean (GM) and geometric standard deviation (GSD) (if lognormal is assumed) that would give those 50th- and 95th-percentile values. Visually, this is like taking the vertical slice (at 100 mrem – the dotted line) of Figure 5-1 and rotating it clockwise 90 degrees to get Figure 5-2. The point at which the dotted line in Figure 5-1 intersects the 50th-percentile line is now the solid vertical line in Figure 5-2. Likewise, the point at which the dotted line in Figure 5-1 intersects the 95th-percentile line is now the dashed vertical line in Figure 5-2. Because lognormal is assumed, a lognormal distribution has been drawn in Figure 5-2. This lognormal has a 50th percentile that is the solid vertical line and a 95th percentile that is the dashed vertical line. The slice method determines the GM and GSD of the lognormal distribution in Figure 5-2.

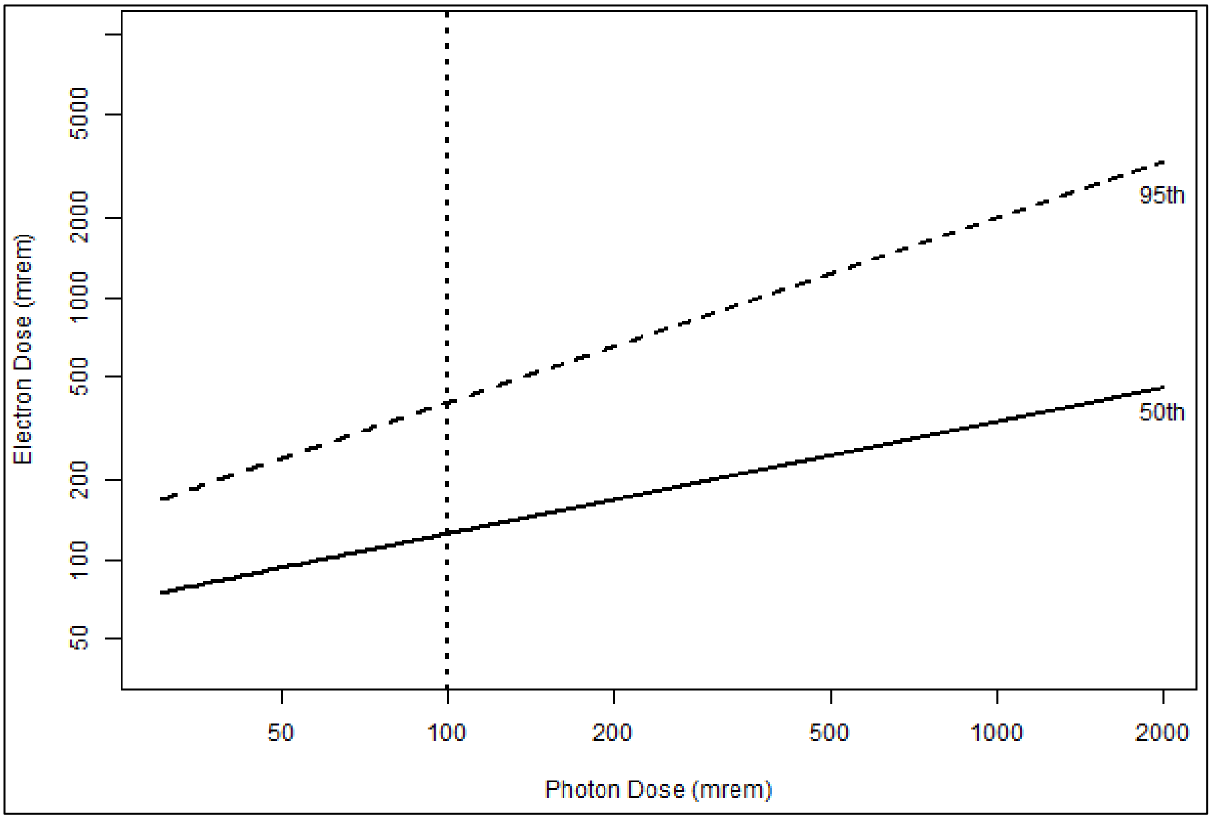

Figure 5-1. Plot of the 50th and 95th percentile QR lines from Figure 4-2. Attachment A contains an extended description.

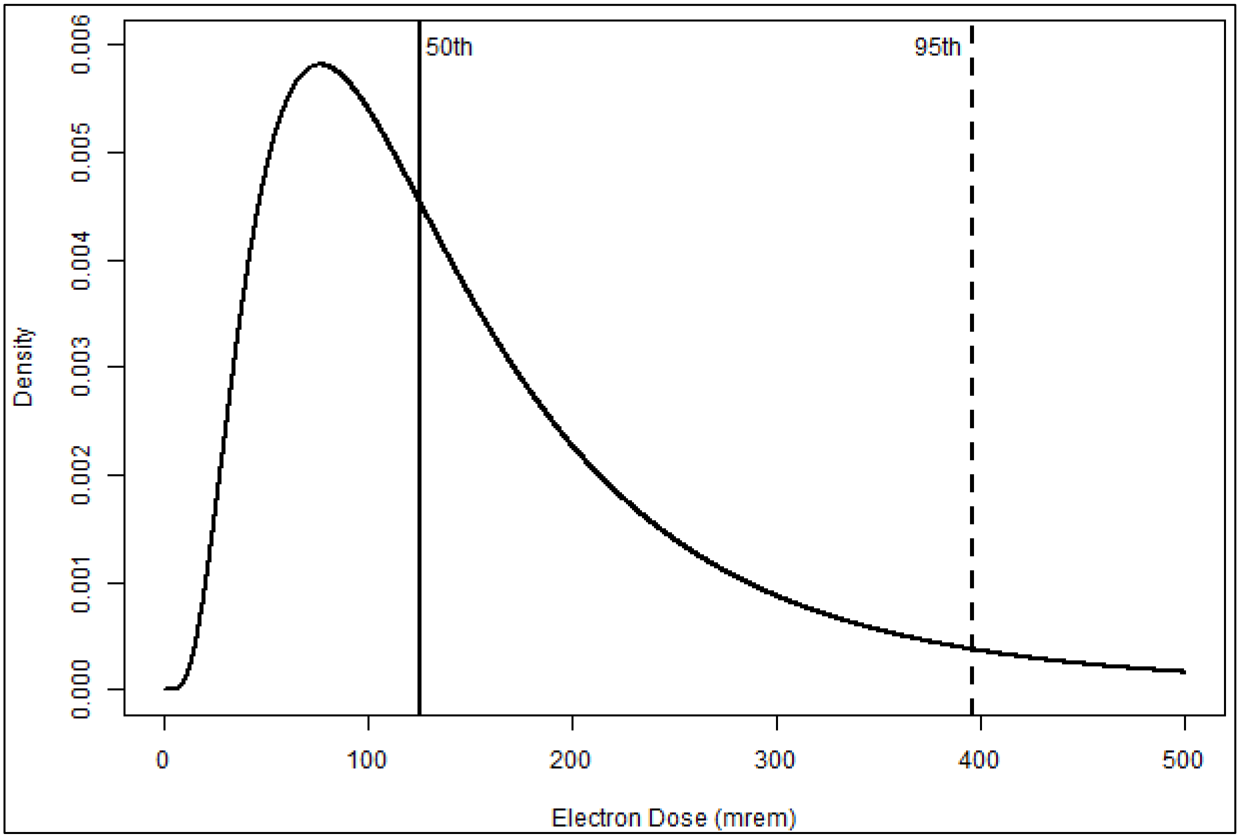

Figure 5-2. Vertical slice of Figure 5-1, rotated clockwise 90 degrees, with a lognormal distribution. Attachment A contains an extended description.

5.1.1 Normal Distribution

For a normal distribution, suppose the two percentiles of interest y_1 and y_2 correspond to probabilities p_1 and p_2, where $y_1 < y_2$ and $p_1 < p_2$. For a random variable Y, that means:

$$P(Y < y_1) = p_1 \quad \text{and} \quad P(Y < y_2) = p_2 \tag{5-1}$$

Inside the parentheses of Equation 5-1, subtracting μ (the mean of the normal distribution) and dividing by σ (the standard deviation of the normal distribution) on both sides of the inequalities:

$$P\left(\frac{Y - \mu}{\sigma} < \frac{y_1 - \mu}{\sigma}\right) = p_1 \quad \text{and} \quad P\left(\frac{Y - \mu}{\sigma} < \frac{y_2 - \mu}{\sigma}\right) = p_2 \tag{5-2}$$

The standard normal distribution (denoted as random variable Z with mean 0 and standard deviation 1) is the shifted and scaled normal defined by the left side of each inequality in Equation 5-2:

$$P\left(Z < \frac{y_1 - \mu}{\sigma}\right) = p_1 \quad \text{and} \quad P\left(Z < \frac{y_2 - \mu}{\sigma}\right) = p_2 \tag{5-3}$$

The cumulative distribution function (CDF) is the probability that a random variable is less than some value. The standard normal CDF is denoted by Φ:

$$\Phi\left(\frac{y_1-\mu}{\sigma}\right)=p_1 \quad \text{and} \quad \Phi\left(\frac{y_2-\mu}{\sigma}\right)=p_2 \tag{5-4}$$

The CDF has an inverse function, so:

$$\frac{y_1-\mu}{\sigma}=\Phi^{-1}(p_1) \quad \text{and} \quad \frac{y_2-\mu}{\sigma}=\Phi^{-1}(p_2) \tag{5-5}$$

$$y_1=\sigma\,\Phi^{-1}(p_1)+\mu \quad \text{and} \quad y_2=\sigma\,\Phi^{-1}(p_2)+\mu \tag{5-6}$$

Solving Equation 5-6 for σ:

$$y_2-y_1=\left[\sigma\,\Phi^{-1}(p_2)+\mu\right]-\left[\sigma\,\Phi^{-1}(p_1)+\mu\right] \tag{5-7}$$

Attachment A contains an extended description.

$$y_2-y_1=\sigma\left[\Phi^{-1}(p_2)-\Phi^{-1}(p_1)\right] \tag{5-8}$$

$$\sigma=\frac{y_2-y_1}{\Phi^{-1}(p_2)-\Phi^{-1}(p_1)} \tag{5-9}$$

Substituting σ from Equation 5-9 into the equation for y_1 in Equation 5-6 and solving for μ:

$$y_1=\sigma\,\Phi^{-1}(p_1)+\mu \tag{5-10}$$

$$y_1=\left[\frac{y_2-y_1}{\Phi^{-1}(p_2)-\Phi^{-1}(p_1)}\right]\Phi^{-1}(p_1)+\mu \tag{5-11}$$

Attachment A contains an extended description.

$$\mu=y_1-\left[\frac{y_2-y_1}{\Phi^{-1}(p_2)-\Phi^{-1}(p_1)}\right]\Phi^{-1}(p_1) \tag{5-12}$$

Attachment A contains an extended description.

$$\mu=\frac{y_1\left[\Phi^{-1}(p_2)-\Phi^{-1}(p_1)\right]-(y_2-y_1)\Phi^{-1}(p_1)}{\Phi^{-1}(p_2)-\Phi^{-1}(p_1)} \tag{5-13}$$

Attachment A contains an extended description.

$$\mu=\frac{y_1\Phi^{-1}(p_2)-y_1\Phi^{-1}(p_1)-y_2\Phi^{-1}(p_1)+y_1\Phi^{-1}(p_1)}{\Phi^{-1}(p_2)-\Phi^{-1}(p_1)} \tag{5-14}$$

Attachment A contains an extended description.

$$\mu = \frac{y_1\Phi^{-1}(p_2) - y_2\Phi^{-1}(p_1)}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)} \tag{5-15}$$

Attachment A contains an extended description.

So, for a normal distribution, if the two percentiles of interest correspond to p_1 and p_2, then the mean and standard deviation of the normal distribution are:

$$\mu = \frac{y_1\Phi^{-1}(p_2) - y_2\Phi^{-1}(p_1)}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)} \tag{5-16}$$

Attachment A contains an extended description.

$$\sigma = \frac{y_2 - y_1}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)} \tag{5-17}$$

5.1.2 Lognormal Distribution

For lognormal data, the logs of the data are normally distributed. That means replacing *y* with log(*y*) in Equations 5-16 and 5-17 gives the “log mean” and “log standard deviation” of the lognormal distribution. The log mean and log standard deviation can be exponentiated to give the GM and GSD of the lognormal distribution:

$$GM = \exp\left[\frac{\log(y_1)\Phi^{-1}(p_2) - \log(y_2)\Phi^{-1}(p_1)}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)}\right] \tag{5-18}$$

Attachment A contains an extended description.

$$GSD = \exp\left[\frac{\log(y_2) - \log(y_1)}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)}\right] \tag{5-19}$$

Attachment A contains an extended description.

Example
This application example focuses on the QR analysis because it is the more appropriate regression method. The slice method can be used for OLS as well.

For a known cycle photon dose of 100 mrem using QR analysis (Equations 4-4 and 4-6):

$$y_1 = 125.1652,\ p_1 = 0.5 \tag{5-20}$$

$$y_2 = 396.0673,\ p_2 = 0.95 \tag{5-21}$$

Using the GM and GSD formulas in Equations 5-18 and 5-19:

$$GM = \exp\left[\frac{\log(125.1652)\Phi^{-1}(0.95) - \log(396.0673)\Phi^{-1}(0.5)}{\Phi^{-1}(0.95) - \Phi^{-1}(0.5)}\right] = 125.1652\ \text{mrem} \tag{5-22}$$

Attachment A contains an extended description.

$$GSD = \exp\left[\frac{\log(396.0673) - \log(125.1652)}{\Phi^{-1}(0.95) - \Phi^{-1}(0.5)}\right] = 2.0144 \tag{5-23}$$

Attachment A contains an extended description.

Therefore, for a known cycle photon dose of 100 mrem, the distribution of cycle electron dose is assigned as lognormal with GM is 125.1652 mrem and GSD is 2.0144. Note that the GM is exactly the predicted 50th percentile (as expected), no matter the value of the known photon dose, if the 50th percentile is one of the two percentiles used.

6.0 SUMMARY AND CONCLUSIONS

This report has presented OLS and QR as methods to model bivariate data. OLS assumes, among other things, that the electron doses are normally distributed around the mean regression line, which simplifies the calculation of quantiles when this assumption is met. QR does not assume any particular distribution of the data and is therefore more flexible than OLS regression. The cost of this flexibility is added computational complexity and sample size requirements. This report also presents the slice method, which is a method for applying regression in the dose reconstruction tools.

The choice of which regression method to use depends on the specifics of the dataset and the desired results for the particular application. For this reason, a statistician and subject matter expert should jointly determine the best method based on those specifics.

REFERENCES

Cade BS, Noon BR [2003]. A gentle introduction to quantile regression for ecologists. Front Ecol Environ *1*(8):412–420. [SRDB Ref ID: 167471]

Cook JD [2010]. Determining distribution parameters from quantiles. Houston, TX: University of Texas. January 27. [SRDB Ref ID: 206785]

Helsel DR [2012]. Statistics for censored environmental data using Minitab and R. 2nd ed. Hoboken, NJ: John Wiley & Sons. [SRDB Ref ID: 150353]

Koenker R [2005]. Quantile regression. New York: Cambridge University Press. Econometric Society Monographs No. 38. [SRDB Ref ID: 167513]

ORAUT [2023]. Application of regression models and ratios. Oak Ridge, TN: Oak Ridge Associated Universities Team. ORAUT-RPRT-0106 Rev. 00, August 3. [SRDB Ref ID: 198501]

ORAUT [2025]. ORAUT-RPRT-0087 rev 01 support files. Zip file. Oak Ridge, TN: Oak Ridge Associated Universities Team. August 11. [SRDB Ref ID: 206786]

Weisberg S [2005]. Applied linear regression. 3rd ed. Hoboken, NJ: John Wiley & Sons. [SRDB Ref ID: 169440]

ATTACHMENT A
EXTENDED DESCRIPTIONS OF FIGURES AND EQUATIONS

Figure 2-1
The *x*-axis is linear-scaled, labeled “Photon Dose (mrem)”, and ranges from 0 to 2,000. The *y*-axis is linear-scaled, labeled “Electron Dose (mrem)”, and ranges from 0 to 2,500. There is a dashed vertical line at 30 and a dashed horizontal line at 50. The 1,159 points are in a cloud shape, with most of the points concentrated in the lower left corner and becoming more spread out as they move upward and to the right.

Figure 2-2
The *x*-axis is log-scaled, labeled “Photon Dose (mrem)”, and ranges from 30 to 2,000. The *y*-axis is log-scaled, labeled “Electron Dose (mrem)”, and ranges from 50 to 2,500. There is a dashed vertical line at 30 and a dashed horizontal line at 50. The 1,159 points are in a thick diagonal band from the lower left corner of the plot to the upper right corner.

Figure 3-1
The *x*-axis is linear-scaled, labeled “Predicted Values”, and ranges from about 4.5 to 6. The *y*-axis is linear-scaled, labeled “Residuals”, and ranges from about -2 to 3. There is a solid horizontal line at zero. The points are in a cone shape (less spread on the left side than the right), with a pronounced diagonal floor on the lower left side of the plot.

Figure 3-2
The *x*-axis is linear-scaled, labeled “Theoretical Quantiles”, and ranges from about -3 to 3. The *y*-axis is linear-scaled, labeled “Sample Quantiles”, and ranges from about -2 to 3. The points go from the lower left corner to the upper right corner with a slightly concave up shape. There is a solid diagonal line that agrees well with the data from about *x* equals -1 to *x* equals 2.

Figure 3-4
The *x*-axis is log-scaled, labeled “Photon Dose (mrem)”, and ranges from 30 to 2,000. The *y*-axis is log-scaled, labeled “Electron Dose (mrem)”, and ranges from 20 to 2,500. There is a dashed vertical line at 100. The solid line goes from approximately (30, 80) to (2,000, 450). There are 20 diagonal dashed lines, with the same slope as the solid line, with the intercepts listed in Table 3-1.

Figure 5-1
The *x*-axis is log-scaled, labeled “Photon Dose (mrem)”, and ranges from 30 to 2,000. The *y*-axis is log-scaled, labeled “Electron Dose (mrem)”, and ranges from 50 to 10,000. There is a dotted vertical line at 100. The solid line goes from approximately (30, 75) to (2,000, 450). The text beside the solid line is “50th”. The dashed line goes from approximately (30, 250) to (2,000, 1,400). The text beside the dashed line is “95th”.

Figure 5-2
The *x*-axis is linear-scaled, labeled “Electron Dose (mrem)”, and ranges from 0 to 500. The *y*-axis is linear-scaled, labeled “Density”, and ranges from 0 to 0.006. There is a solid vertical line at about 125. There is a dashed vertical line at about 400. There is a lognormal density (right-skewed) with mode at about 80.

Equation 5-7
y sub 2 minus y sub 1 equals open parenthesis sigma times capital phi inverse open parenthesis p sub 2 close parenthesis plus mu close parenthesis minus open parenthesis sigma times capital phi inverse open parenthesis p sub 1 close parenthesis plus mu close parenthesis.

ATTACHMENT A
EXTENDED DESCRIPTIONS OF FIGURES AND EQUATIONS (continued)

Equation 5-11
y sub 1 equals open square bracket the quantity y sub 2 minus y sub 1 divided by the quantity capital phi inverse open parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis close square bracket times capital phi inverse open parenthesis p sub 1 close parenthesis plus mu.

Equation 5-12
mu equals y sub 1 minus open square bracket the quantity y sub 2 minus y sub 1 divided by the quantity capital phi inverse open parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis close square bracket times capital phi inverse open parenthesis p sub 1 close parenthesis.

Equation 5-13
mu equals the quantity y sub 1 times open square bracket capital phi inverse open parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis close square bracket minus open parenthesis y sub 2 minus y sub 1 close parenthesis times capital phi inverse open parenthesis p sub 1 close parenthesis divided by the quantity capital phi inverse open parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis.

Equation 5-14
mu equals the quantity y sub 1 times capital phi inverse open parenthesis p sub 2 close parenthesis minus y sub 1 times capital phi inverse open parenthesis p sub 1 close parenthesis minus y sub 2 times capital phi inverse open parenthesis p sub 1 close parenthesis plus y sub 1 times capital phi inverse open parenthesis p sub 1 close parenthesis divided by the quantity capital phi inverse open parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis.

Equation 5-15
mu equals the quantity y sub 1 times capital phi inverse open parenthesis p sub 2 close parenthesis minus y sub 2 times capital phi inverse open parenthesis p sub 1 close parenthesis divided by the quantity capital phi inverse open parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis.

Equation 5-16
mu equals the quantity y sub 1 times capital phi inverse open parenthesis p sub 2 close parenthesis minus y sub 2 times capital phi inverse open parenthesis p sub 1 close parenthesis divided by the quantity capital phi inverse open parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis.

Equation 5-18
GM equals exp open square bracket the quantity log open parenthesis y sub 1 close parenthesis times capital phi inverse open parenthesis p sub 2 close parenthesis minus log open parenthesis y sub 2 close parenthesis times capital phi inverse open parenthesis p sub 1 close parenthesis divided by the quantity capital phi inverse open parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis close square bracket.

Equation 5-19
GSD equals exp open square bracket the quantity log open parenthesis y sub 2 close parenthesis minus log open parenthesis y sub 1 close parenthesis divided by the quantity capital phi inverse open

ATTACHMENT A
EXTENDED DESCRIPTIONS OF FIGURES AND EQUATIONS (continued)

parenthesis p sub 2 close parenthesis minus capital phi inverse open parenthesis p sub 1 close parenthesis close square bracket.

Equation 5-22
GM equals exp open square bracket the quantity log open parenthesis 125.1652 close parenthesis times capital phi inverse open parenthesis 0.95 close parenthesis minus log open parenthesis 396.0673 close parenthesis times capital phi inverse open parenthesis 0.5 close parenthesis divided by the quantity capital phi inverse open parenthesis 0.95 close parenthesis minus capital phi inverse open parenthesis 0.5 close parenthesis close square bracket equals 125.1652 mrem.

Equation 5-23
GSD equals exp open square bracket the quantity log open parenthesis 396.0673 close parenthesis minus log open parenthesis 125.1652 close parenthesis divided by the quantity capital phi inverse open parenthesis 0.95 close parenthesis minus capital phi inverse open parenthesis 0.5 close parenthesis close square bracket equals 2.0144.