DOE Review Release 07/01/2016

Use of Claimant Datasets for Coworker Modeling	ORAUT-OTIB-0075	Rev. 01
	Effective Date:	06/17/2016
	Supersedes:	Revision 00

Subject Expert(s): Thomas R. LaBone

Document Owner Approval:	Signature on File Thomas R. LaBone, Document Owner	Approval Date:	06/06/2016
Concurrence:	Signature on File John M. Byrne, Objective 1 Manager	Concurrence Date:	06/03/2016
Concurrence:	Signature on File Edward F. Maher, Objective 3 Manager	Concurrence Date:	06/03/2016
Concurrence:	Dianne Poncio Signature on File for Kate Kimpan, Project Director	Concurrence Date:	06/06/2016
Approval:	Signature on File James W. Neton, Associate Director for Science	Approval Date:	06/17/2016

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

☐ New ☒ Total Rewrite ☐ Revision ☐ Page Change

PUBLICATION RECORD

EFFECTIVE DATE	REVISION NUMBER	DESCRIPTION
05/25/2009	00	New document to present the results of a study of Y-12 Plant worker and claimant data to determine the feasibility of using claimant data in place of complete coworker data for reconstruction of doses at sites that do not have complete coworker data. Includes Mound and SRS datasets. Incorporates formal internal and NIOSH review comments. Training required: As determined by the Objective Manager. Initiated by Thomas R. LaBone.
06/17/2016	01	Revised document to incorporate time-weighted OPOS bioassay data and update lists of claimants. Incorporates formal internal and NIOSH review comments. Constitutes a total rewrite of the document. 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

CER	Center for Epidemiologic Research
CSV	comma-separated values (file type)
d	day
DOE	U.S. Department of Energy
dpm	disintegrations per minute
EEOICPA	Energy Employees Occupational Illness Compensation Program Act of 2000
GM	geometric mean
GSD	geometric standard deviation
HPRED	Health Protection Radiation Exposure Database
NIOSH	National Institute for Occupational Safety and Health
ORAU	Oak Ridge Associated Universities
SRDB Ref ID	Site Research Database Reference Identification (number)
SRS	Savannah River Site
TIB	technical information bulletin
TWOPOS	time-weighted one person – one statistic
U.S.C.	United States Code
§	section or sections

1.0 INTRODUCTION

Technical information bulletins (TIBs) are not official determinations made by the National Institute for Occupational Safety and Health (NIOSH) but are rather general working documents that provide historical background information and guidance to assist in the preparation of dose reconstructions at particular sites or categories of sites. They will be revised in the event additional relevant information is obtained about the affected site(s). TIBs may be used to assist NIOSH staff in the completion of individual dose reconstructions.

In this document, the word “facility” is used as a general term for an area, building, or group of buildings that served a specific purpose at a site. It does not necessarily connote an “atomic weapons employer facility” or a “Department of Energy (DOE) facility” as defined in the Energy Employees Occupational Illness Compensation Program Act of 2000 [42 U.S.C. § 7384l(5) and (12)].

2.0 PURPOSE

In the dose reconstruction process, coworker models are used to assign doses to workers who were not monitored at certain times during their employment but probably should have been (ORAUT 2005). Coworker models are typically developed from essentially complete datasets of worker monitoring data. For example, the collection of all uranium urine bioassay data from 1950 to 1988 for the workers at the Oak Ridge Y-12 Plant (Y-12) was used to develop a coworker model for unmonitored workers at Y-12 (ORAUT 2012). There are sites where complete worker monitoring datasets are not available to develop coworker models. Complete datasets are not available at these sites because not all of the necessary monitoring results have been transcribed from hard-copy records to an electronic format suitable for analysis. Because of the considerable cost and effort that would be involved with creating complete datasets, there is interest in developing coworker models from a sample of the complete dataset.

An ideal sample would consist of randomly selected workers from the entire monitored worker population. However, random sampling is resource intensive in this situation. One readily obtainable sample of the complete dataset for a site is the bioassay data of the workers (energy employees) who are named in a claim under the Energy Employees Occupational Illness Compensation Program Act of 2000 (EEOICPA). These workers are referred to in this paper as the claimants for a site. Strictly speaking, claimant datasets are samples of convenience rather than random samples. Nevertheless, claimant datasets are random enough for this purpose in the same sense that pseudo-random computer-generated numbers are adequate for most applications that require random numbers. Therefore, the purpose of this document is to present the results of a study to determine if claimant datasets can be treated as random samples from the complete datasets from which they were drawn for the purpose of developing coworker models. The study consisted of taking sites where well-defined complete datasets are available and comparing these datasets with the claimant datasets. Table 2-1 lists the available complete datasets.

Table 2-1. Available complete datasets.

Site	Data type	Period
Y-12 Plant	Uranium urine bioassay	1950–1988
Mound Laboratory	Plutonium urine bioassay	1960–1990
Savannah River Site (SRS)	Tritium dose	1991–2000

The first step in the study was a direct, descriptive comparison of annual 50th and 84th percentiles that were obtained from slopes and intercepts of lognormal probability models for the complete and claimant datasets[1]. Then a bootstrap analysis was used to determine if the claimant dataset behaves like a random sample from the complete dataset in relation to the values of the slope and intercept of the fits to the data. If there is no significant difference between the claimant and random datasets, it is assumed that the same coworker model would be obtained, on average, from the claimant dataset as from an actual random sample.

Sections 3.0 to 5.0 explain this logic in detail using the Y-12 data and present the Y-12 results. Sections 6.0 and 7.0 briefly discuss the data and present results for the Mound Laboratory and SRS, respectively.

3.0 DESCRIPTION OF THE Y-12 DATA

Two Y-12 datasets are discussed in this section. The first is the complete dataset, which is all of the uranium urine bioassay data for all monitored Y-12 workers from 1950 to 1988. This dataset[2] was developed by the Oak Ridge Institute for Science and Engineering Center (ORISE) for Epidemiologic Research (CER) for use in epidemiology studies (Watkins et al. 1993) and has undergone detailed verification of its integrity. A member of the complete dataset is referred to as a worker. There are 7,537 workers[3] in the complete dataset who submitted a total of 467,749 uranium urine samples. The year-by-year breakdowns of the total number of submitted samples and total number of monitored workers are shown in Figures 3-1 and 3-2, respectively.

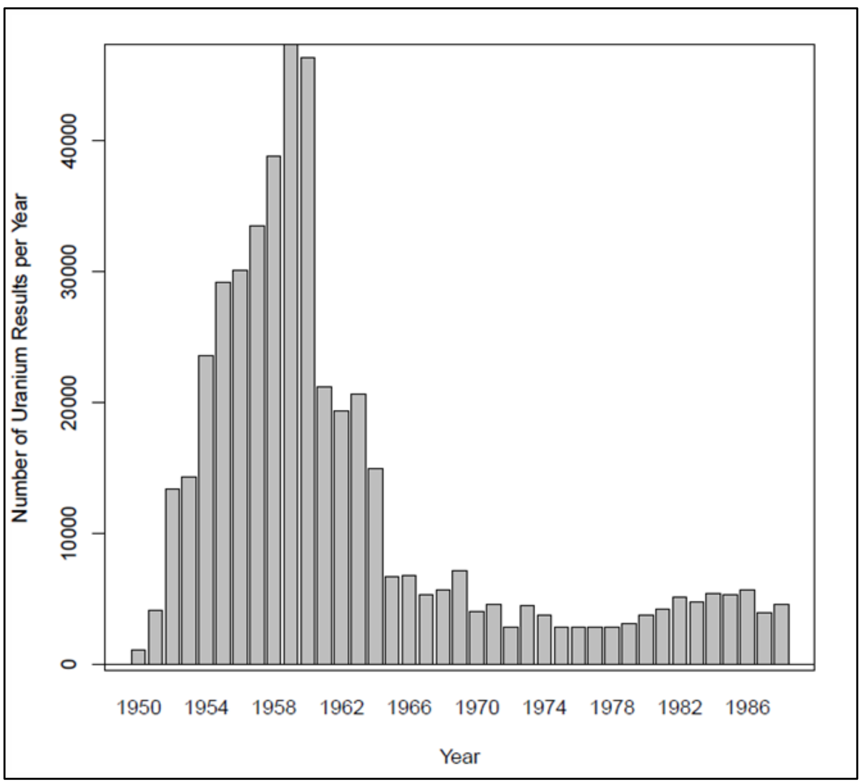


Figure 3-1. Number of samples submitted per year by all Y-12 workers.

---

[1] The statistical software R (RDCT 2008) was used for all calculations. The R codes are available as SRDB Ref ID: 151952. The data used in the analysis was time-weighted one-person-one-statistic (TWOPOS) as described in ORAUT-OTIB-0053 (ORAUT 2014).

[2] The complete dataset was extracted from the ORISE CER database and supplied in 10 separate comma-separated values (CSV) files.

[3] Five workers with extremely high urine results were excluded from the complete dataset because their data inflated the variance of the complete dataset to the extent that the claimant dataset would always agree with the complete dataset.

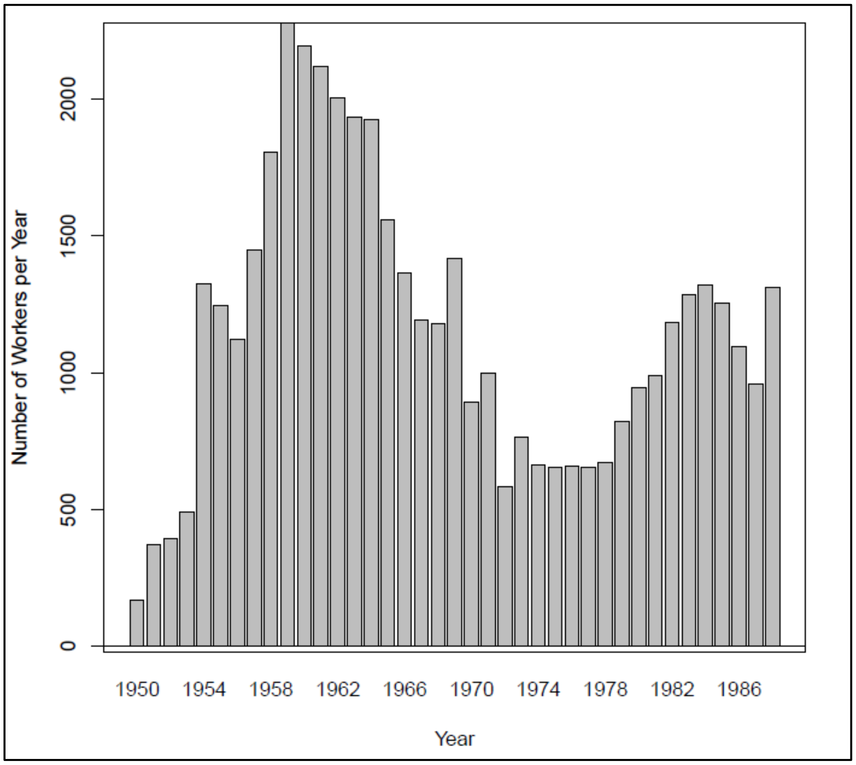


Figure 3-2. Number of Y-12 workers monitored in each year.

The second dataset[4] consists of all the uranium urine bioassay data for Y-12 workers who were monitored at any time during the period from 1950 to 1988 and who are named as energy employees in a claim under EEOICPA. For brevity, this dataset is referred to as the claimant dataset, and a member of the claimant dataset is referred to as a claimant. Note that this is not the same meaning for claimant as used in EEOICPA.[5]

The claimant dataset is a subset of the complete dataset; that is, all of the claimants are workers. A total of 1,585 claimants submitted 119,044 uranium urine samples. The year-by-year breakdowns of the number of submitted samples and number of monitored claimants are shown in Figures 3-3 and 3-4, respectively.

A summary of the numbers of samples and workers in the claimant and complete datasets is given in Table 3-1. It is worth noting that the ratio of claimants to all workers for a given year (the fourth column in Table 3-1) is approximately the same as the ratio of number of claimant-submitted samples to the total number of samples in that year (the last column in Table 3-1). This indicates that the claimants submitted samples at the same rate as the rest of the workers, which suggests that they were not monitored differently than other workers. It is also worth noting that the number of results in a year will equal the number of individuals after each individual's bioassay results for a given year are summarized into one result (i.e., the TWOPOS).

[4] The claimant data were extracted from the ORISE CER database and supplied as one CSV file that was separate and distinct from those for the complete dataset.
[5] Under EEOICPA, a claimant is the individual who filed a claim and is seeking compensation. The claimant can be the energy employee or a survivor of an energy employee.

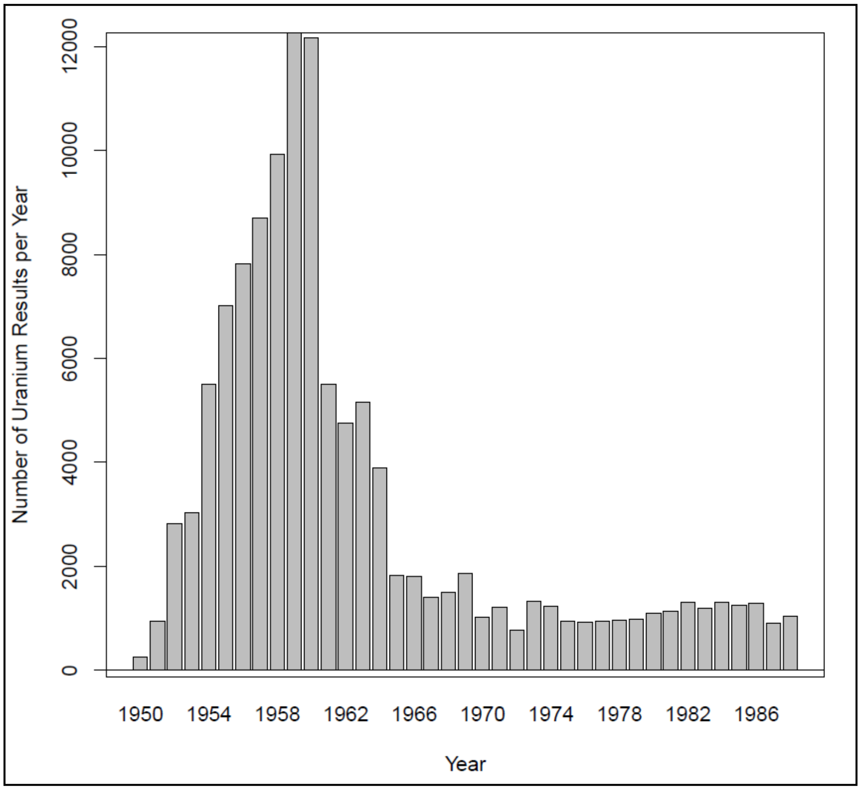


Figure 3-3. Number of samples submitted per year by Y-12 claimants.

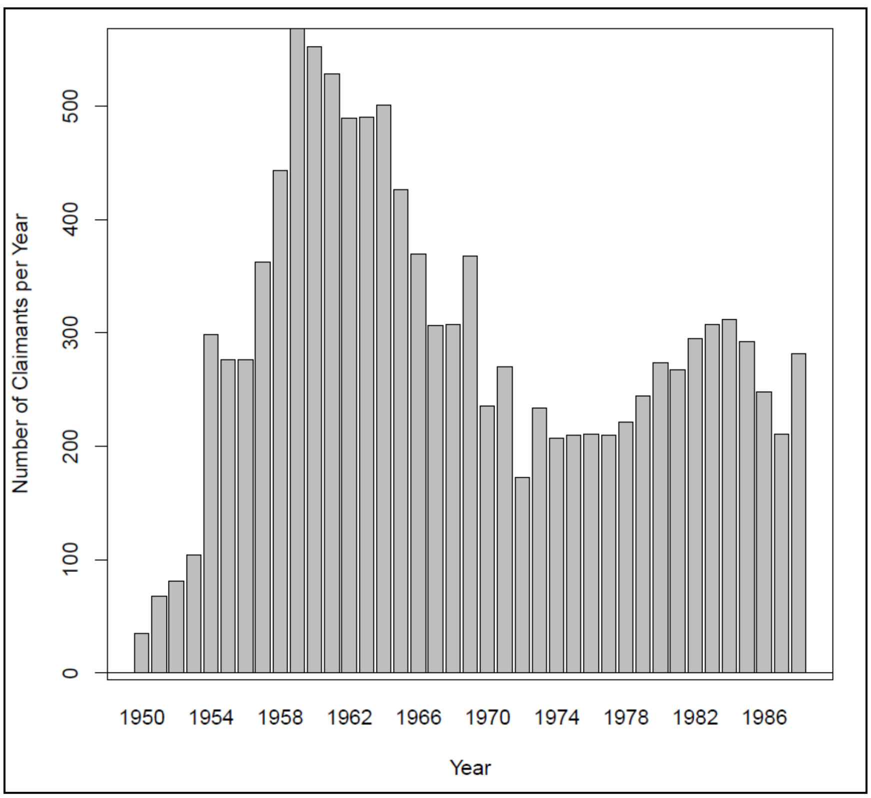


Figure 3-4. Number of Y-12 claimants monitored in each year.

Table 3-1. Summary of samples and workers in claimant and complete Y-12 datasets by year.

Year	Workers complete dataset	Workers claimant dataset	Workers ratio	Samples complete dataset	Samples claimant dataset	Samples ratio
1950	169	35	0.207	1,127	249	0.221
1951	370	68	0.184	4,175	954	0.229
1952	396	81	0.205	13,414	2,821	0.210
1953	493	104	0.211	14,312	3,025	0.211
1954	1,327	298	0.225	23,564	5,502	0.233
1955	1,244	276	0.222	29,160	7,021	0.241
1956	1,123	276	0.246	30,083	7,817	0.260
1957	1,448	362	0.250	33,551	8,693	0.259
1958	1,805	443	0.245	38,848	9,937	0.256
1959	2,277	568	0.249	47,339	12,260	0.259
1960	2,193	552	0.252	46,387	12,175	0.262
1961	2,120	528	0.249	21,214	5,497	0.259
1962	2,004	489	0.244	19,370	4,764	0.246
1963	1,933	490	0.253	20,709	5,160	0.249
1964	1,924	501	0.260	15,013	3,892	0.259
1965	1,559	426	0.273	6,734	1,829	0.272
1966	1,366	369	0.270	6,849	1,811	0.264
1967	1,191	306	0.257	5,369	1,399	0.261
1968	1,181	307	0.260	5,676	1,496	0.264
1969	1,417	368	0.260	7,161	1,862	0.260
1970	892	235	0.263	4,107	1,019	0.248
1971	998	270	0.271	4,631	1,207	0.261
1972	583	172	0.295	2,896	773	0.267
1973	763	234	0.307	4,520	1,328	0.294
1974	663	207	0.312	3,827	1,236	0.323
1975	654	210	0.321	2,896	953	0.329
1976	658	211	0.321	2,896	930	0.321
1977	653	210	0.322	2,877	938	0.326
1978	674	221	0.328	2,893	967	0.334
1979	823	244	0.296	3,141	985	0.314
1980	945	274	0.290	3,798	1,105	0.291
1981	992	267	0.269	4,213	1,131	0.268
1982	1,184	295	0.249	5,175	1,309	0.253
1983	1,285	307	0.239	4,786	1,203	0.251
1984	1,320	312	0.236	5,404	1,302	0.241
1985	1,255	292	0.233	5,356	1,257	0.235
1986	1,096	248	0.226	5,672	1,292	0.228
1987	957	211	0.220	3,958	899	0.227
1988	1,312	282	0.215	4,648	1,046	0.225

4.0 MODELING OF DATA

Using the methodology in ORAUT-RPRT-0053, *Analysis of Stratified Coworker Datasets* (ORAUT 2014), the complete dataset for 1952 is presented in Figure 4-1 in the form of a lognormal probability plot. The year 1952 was chosen as an example because it is one of the years that has relatively good agreement between the complete and claimant datasets. The data are fit with a straight line and the slope *m* and intercept *b* that were calculated from the fit are given on the plot. The geometric mean (GM) and geometric standard deviation (GSD) of the data are:

$$GM = e^{b} = e^{3.216} = 24.93 \text{ dpm/d} \qquad (4\text{-}1)$$

$$GSD = e^{m} = e^{0.6092} = 1.839 \tag{4-2}$$

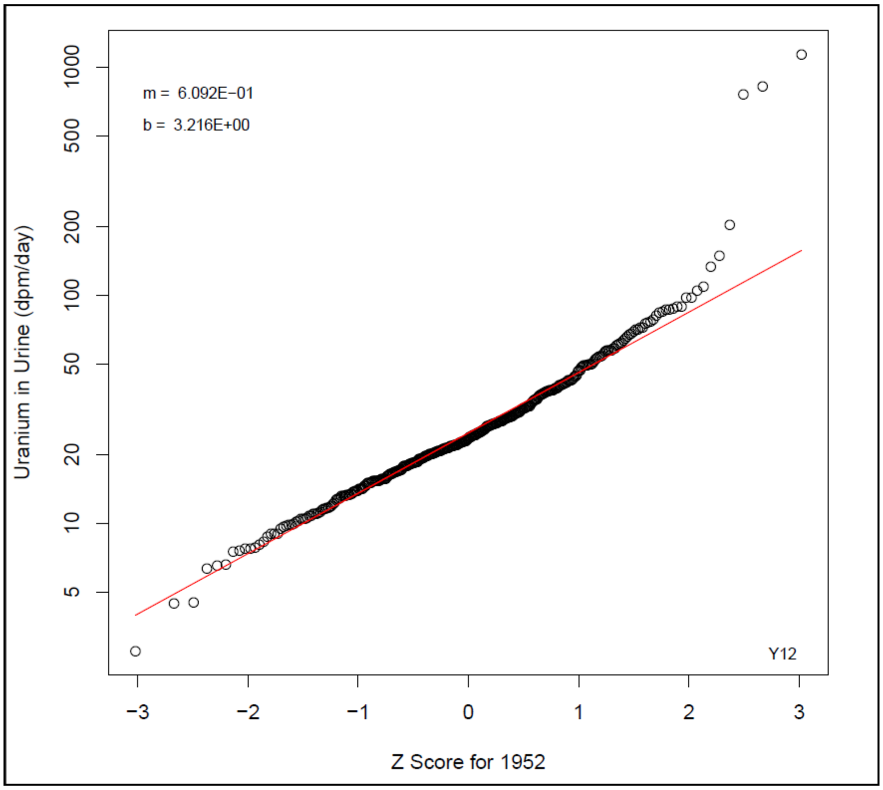


Figure 4-1. Lognormal probability plot of the complete Y-12 dataset for 1952.

The statistics of interest for the coworker model are the 50th and 84th percentiles. The 50th percentile of the data is equal to the GM, and the 84th percentile is equal to the product of the GM and the GSD:

- 50th percentile = 24.93 dpm/d
- 84th percentile = 45.85 dpm/d

These are parametric percentiles, which means that they derive from the parameters of the lognormal model that is fit to the data rather than the data themselves. The validity of the parametric percentiles depends in part on how well a lognormal probability model fits the data. Ultimately, the 50th and 84th percentiles are modeled by an internal dosimetrist and uranium intake rates are calculated for each set of data. The intake rate that is typically calculated from the 50th percentile data is assigned for the period of interest, and the ratio of the 84th-percentile intake rate to the 50th-percentile intake rate gives the GSD of the intake.

The lognormal probability plot for the 1952 claimant dataset is given in Figure 4-2. The relevant statistics from this plot are:

$$GM = e^{3.278} = 26.52 \text{ dpm/d} \tag{4-3}$$

$$GSD = e^{0.6279} = 1.874 \tag{4-4}$$

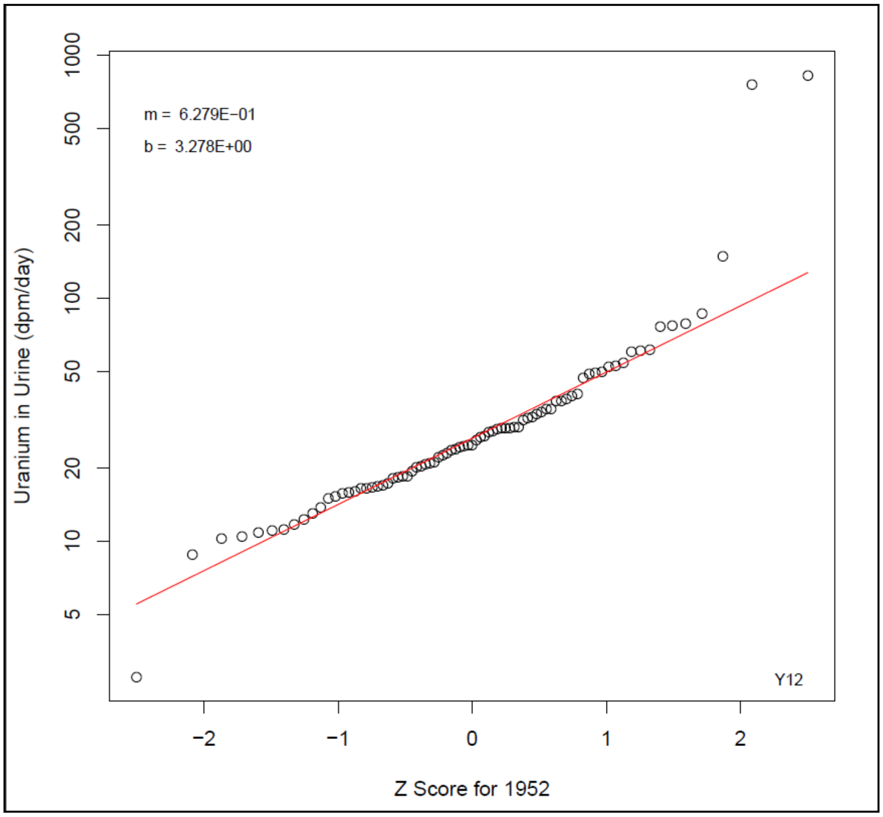


Figure 4-2. Lognormal probability plot of the claimant Y-12 dataset for 1952.

The 50th and 84th percentiles are then:

- 50th percentile = 26.52 dpm/d
- 84th percentile = 49.69 dpm/d

The annual lognormal probability plots for complete and claimant datasets for 1950 to 1988 are given in Attachments A and B, respectively, and the results of the fits are given in Table 4-1. Graphical comparisons of the annual 50th and 84th percentiles for the complete and claimant datasets from Table 4-1 are shown in Figures 4-3 and 4-4. In general, there is excellent agreement between the percentiles that derive from the complete dataset and those from the claimant dataset.

In this specific case (i.e., for Y-12 uranium urine bioassay from 1950 to 1988) the sample consisting of the claimant dataset gives estimates of slope and intercept that appear to agree quite well with the slope and intercept of the complete dataset. This raises the question of whether the claimant dataset can be expected to give good estimates of the slope and intercept for all sites, radionuclides, and years.

To answer this question, consider the situation in which neither the complete dataset nor the claimant dataset are available in a database. The urine bioassay data for all 7,537 workers could be entered into a database and then used to calculate the slope and intercept for each year for the complete dataset. This approach requires the most effort, but it would give the best available answer.

Another approach that requires less effort is to pick a number of the workers as a sample and calculate the slopes and intercepts from the sample. The ideal result would be that the slopes and intercepts of the sample would be the same as the slopes and intercepts of the complete dataset.

Table 4-1. Summary of the slopes and intercepts of a fit to a lognormal probability model and the corresponding 50th and 84th percentiles for Y-12. The corresponding plots are in Attachments A and B.

Year	Complete slope	Complete intercept	Claimant slope	Claimant intercept	Complete 50th percentile	Complete 84th percentile	Claimant 50th percentile	Claimant 84th percentile
1950	1.457	2.276	1.164	1.972	9.73	41.80	7.18	23.00
1951	0.848	2.380	0.871	2.347	10.81	25.22	10.45	24.98
1952	0.609	3.216	0.628	3.278	24.93	45.85	26.52	49.69
1953	0.940	3.191	0.867	3.347	24.31	62.25	28.43	67.63
1954	1.267	2.561	1.220	2.602	12.94	45.93	13.49	45.71
1955	1.047	2.159	0.936	2.329	8.66	24.66	10.26	26.17
1956	0.915	2.302	0.851	2.403	10.00	24.97	11.06	25.90
1957	1.067	2.043	1.033	2.088	7.72	22.41	8.07	22.67
1958	1.077	2.381	0.986	2.471	10.82	31.74	11.83	31.74
1959	1.108	2.361	1.042	2.470	10.60	32.10	11.82	33.53
1960	1.149	1.907	1.153	1.992	6.73	21.24	7.33	23.21
1961	1.524	1.320	1.471	1.281	3.74	17.18	3.60	15.68
1962	1.419	1.879	1.412	1.874	6.55	27.07	6.51	26.73
1963	1.455	1.465	1.510	1.460	4.33	18.54	4.31	19.50
1964	1.415	0.857	1.330	0.914	2.36	9.70	2.49	9.43
1965	1.687	1.095	1.615	1.212	2.99	16.15	3.36	16.90
1966	1.610	1.350	1.793	1.177	3.86	19.28	3.24	19.49
1967	1.342	1.632	1.281	1.714	5.12	19.58	5.55	19.99
1968	1.176	1.830	1.078	1.896	6.23	20.21	6.66	19.56
1969	1.253	1.562	1.254	1.540	4.77	16.70	4.66	16.35
1970	0.951	2.277	0.976	2.267	9.75	25.22	9.65	25.62
1971	1.032	2.151	0.957	2.170	8.60	24.14	8.76	22.81
1972	0.747	2.728	0.682	2.707	15.30	32.31	14.98	29.62
1973	0.942	2.526	1.034	2.510	12.51	32.09	12.31	34.62
1974	0.871	2.313	0.884	2.255	10.10	24.13	9.54	23.09
1975	0.741	2.510	0.720	2.520	12.30	25.81	12.43	25.53
1976	0.851	2.411	0.822	2.445	11.14	26.10	11.53	26.23
1977	0.790	2.511	0.763	2.571	12.32	27.13	13.08	28.05
1978	1.085	2.034	1.091	1.998	7.65	22.64	7.37	21.96
1979	1.172	1.781	1.103	1.739	5.94	19.16	5.69	17.15
1980	1.286	1.687	1.338	1.665	5.40	19.55	5.28	20.14
1981	1.144	1.622	1.089	1.687	5.06	15.90	5.40	16.05
1982	1.116	1.695	1.173	1.632	5.45	16.63	5.12	16.54
1983	1.286	1.412	1.146	1.598	4.10	14.84	4.95	15.56
1984	1.316	1.394	1.315	1.388	4.03	15.04	4.00	14.92
1985	1.483	0.796	1.564	0.763	2.22	9.76	2.14	10.24
1986	1.459	0.802	1.456	0.882	2.23	9.60	2.42	10.36
1987	1.470	0.861	1.421	1.004	2.37	10.28	2.73	11.30
1988	1.527	0.787	1.585	0.732	2.20	10.12	2.08	10.15

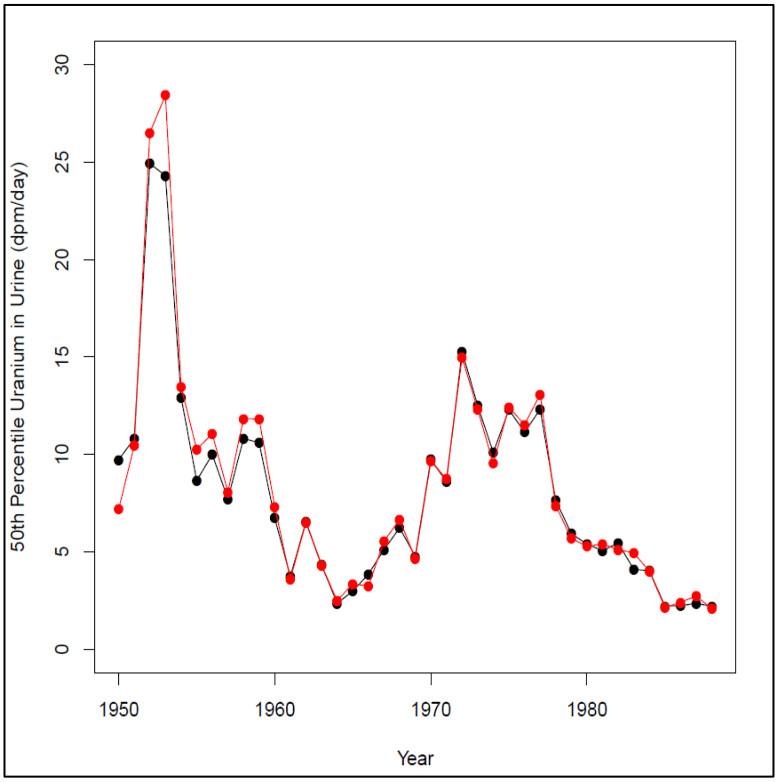


Figure 4-3. Comparison of 50th percentiles for the complete (black) and claimant (red) Y-12 datasets.

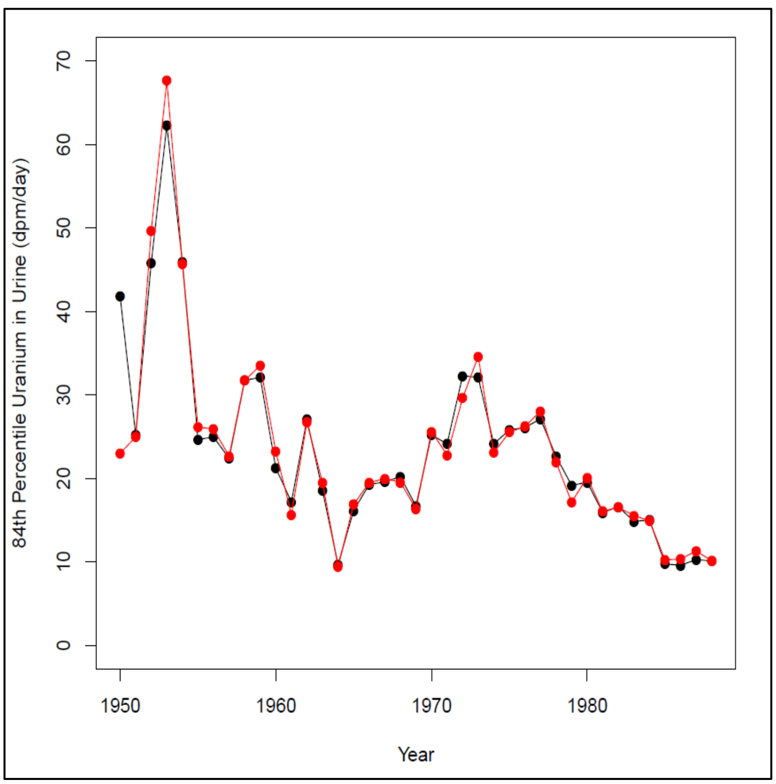


Figure 4-4. Comparison of 84th percentiles for the complete (black) and claimant (red) Y-12 datasets.

Given that the slopes and intercepts of the sample are unlikely to be exactly equal to the slopes and intercepts of the complete dataset, it is desirable that:

- On average, the slopes and intercepts of the sample are equal to the slopes and intercepts from the complete dataset, and

- The sampling error is acceptably small.

To state the information under the first bullet another way, the slopes and intercepts of the sample should be an unbiased estimate of the slopes and intercepts of the complete dataset. Unbiased estimates of the slopes and intercepts can be obtained from properly collected random samples, and the likely size of the sampling error decreases as the size of this random sample increases. Therefore, to generalize the conclusion that the claimant dataset is useful for estimating the slopes and intercepts of the complete dataset, it must be shown that the claimant dataset can be treated as if it were a random sample.

5.0 RANDOM SAMPLES

The claimant datasets are best described as samples of convenience; that is, they were not constructed through a designed random sampling of the complete dataset. Nevertheless, this section shows that, in general, claimant datasets can be used as if they were true random samples from the complete dataset. Verifying that the claimant datasets behave like true random samples is important because coworker models that are based on random samples will, on average, provide the same coworker models that would derive from the complete dataset. Random sampling provides a statistical basis for stating that the sample is representative of the complete dataset and eliminates selection bias. Therefore, if the claimant dataset is large enough and has properties very much like those of a truly random sample from the complete dataset then, in relation to the estimates of the lognormal parameters, meaningful statements can be made about the sampling errors that are associated with the percentiles without having access to the complete dataset.

5.1 THEORY AND METHODS

There are n = 7,537 monitored workers in the complete dataset, each of which has a mostly unique set of bioassay data. The slopes and intercepts of the complete dataset are the true slopes and intercepts, and they are the parameters to be estimated from a sample of the complete dataset (the claimant dataset is such a sample). For example, the slope and intercept of the complete dataset for 1952 are (0.609, 3.216). The claimant dataset of k = 1,585 workers (some of whom do not have results in 1952) has a slope and intercept of (0.628, 3.278) for 1952. If the complete dataset was not available, the slope and intercept of the claimant dataset for 1952 would be the logical choice to be the estimate of the complete dataset slope and intercept for 1952. The natural question at this point is how good an estimate is the slope and intercept of the claimant data; that is, how close is it to the slope and intercept of the complete dataset?

One way to answer this question begins with an examination of the distribution of annual slopes and intercepts that can be calculated from all possible samples of k = 1,585 claimants of the n = 7,537 monitored workers. Equation 5-1 calculates the number of possible random samples. There are

$$C = \frac{7{,}537!}{1{,}585!(7{,}537-1{,}585)!} = 4.542 \times 10^{1681} \qquad (5\text{-}1)$$

equally probable ways of selecting 1,585 different workers from a population of 7,537 workers. Each of the C combinations of the 1,585 workers can potentially have a different collection of bioassay results with its own slope and intercept for each year. Note that only one of these C ways (the claimant dataset) was actually observed (realized). In theory, the slope and intercept for each of the *C* datasets for each year could be calculated, and the ranges where most (95% for example) of the slopes and intercepts occurred could be established. For a given year, the estimated slope and intercept would be considered to be statistically the same as the true slope and intercept if the estimated slope and intercept fall in the range where 95% of the results would be observed.[6]

While theoretically possible, the calculations above are not feasible because of the immense number[7] of samples of size *k* = 1,585 that can be drawn from the complete dataset. However, the 95% range can be estimated using the following simulation, which is a type of bootstrap analysis (Efron and Tibshirani 1998; Davison and Hinkley 1997):

1. Assume that there are *n* workers in the complete dataset and k claimants in the claimant dataset, where *n* > *k* and the claimant dataset is a subset of the complete dataset.

2. Calculate the slope and intercept of the lognormal fit to the bioassay results for the claimant dataset for each year. Do the same for the complete dataset and plot these two points.

3. Randomly draw *k* = 1,585 workers without replacement[8] from the complete dataset. Fit a lognormal model to the bioassay results for the *k* workers for each year.

4. Repeat step 3 *m* times (e.g., *m* = 1,000 times). Note that each time the experiment is run, a new set of *k* bootstrap workers is drawn and new slopes and intercepts are calculated for each year.

5. Plot the *m* points for each year at their slope *x* and intercept *y* coordinates.

6. Calculate a 95% joint confidence ellipse of all *m* slopes and intercepts for the bootstrap workers in each year.

For example, the results of the bootstrap analysis for *k* = 1,585 and *m* = 1,000 for 1952 are shown in Figure 5-1, where:

- The solid red point is the slope and intercept (0.609, 3.216) of the complete dataset.

- The solid blue point is the slope and intercept (0.628, 3.278) of the claimant dataset.

- The open pink points are the slopes and intercepts of the fits to the *m* random sets of 1,585 workers taken from the complete dataset.

- The ellipse is the 95% joint normal confidence ellipse.

- The numbers in the corners are the full scale of the respective axes in terms of positive and negative percentages of the slope or intercept of the complete dataset.

The desired result is that the slope and intercept of the claimant data are close to the slope and intercept of the complete data. One way to gauge how close the two points are is to see if the slope

---

[6] The number of combinations, *n* choose *k* or *n*C*k*, was calculated with Maxima 5.37.3 (SourceForge 2016). Note that the order of the workers does not count and a particular worker only appears once in each sample of 1,585 workers.
[7] For comparison, there are approximately 1 × 10^{80} atoms in the observable universe.
[8] “Without replacement” means that no workers appear more than once in any given sample of k workers.

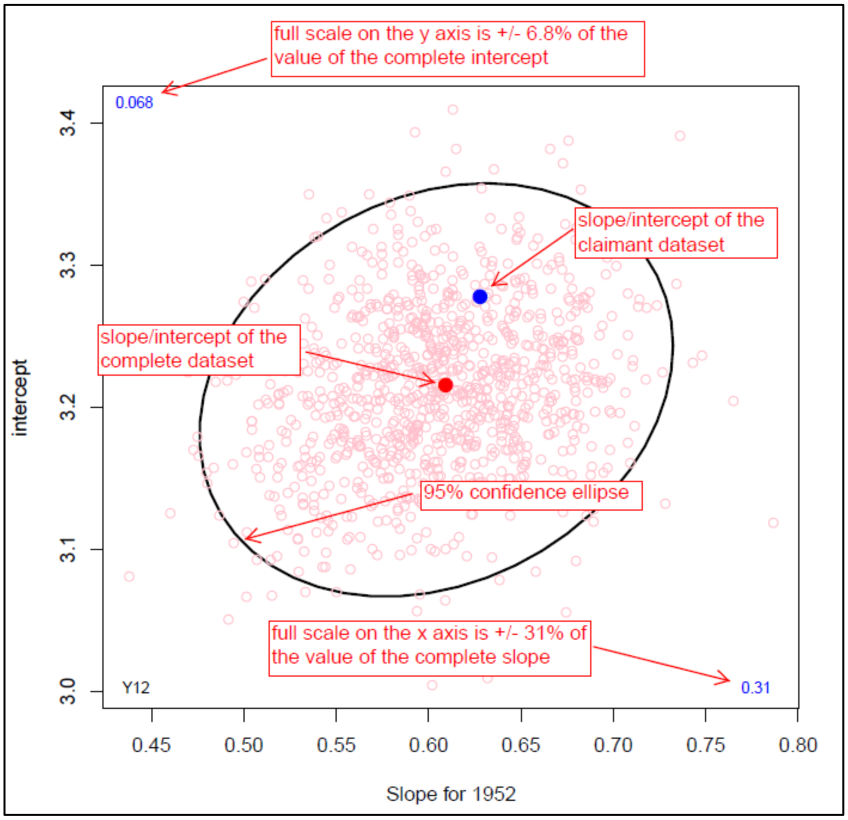


Figure 5-1. Bootstrap analysis of slope and intercept for 1952 complete Y-12 dataset with a sample of 1,585 workers.

and intercept of the claimant data falls in the 95% confidence ellipse of the bootstrap samples. When they do, it is concluded that the slope and intercept of the claimant data is statistically no different from the slope and intercept of a true random sample of 1,585 workers from the complete dataset. As can be seen in Figure 5-1, there was excellent agreement for 1952. If there is a strong tendency for the slopes and intercepts of the claimant data to fall in the 95% confidence ellipse each year, this would be evidence to support the assumption that the claimant dataset behaves as if it were a random sample from the complete dataset. In addition, it is important to note that two sets of slopes and intercepts can be statistically different but practically the same. For example, Table 5-1 lists a summary of the percentiles for the claimant and complete datasets for 1952.

Table 5-1. Summary of the percentiles for the claimant and complete Y-12 datasets for 1952.

Dataset	50th percentile	84th percentile
Claimant	26.52	49.69
Complete	24.93	45.85

Whether or not the percentiles of the claimant dataset for 1952 are different statistically, in consideration of how these percentiles are to be used by an internal dosimetrist, there is clearly no difference of practical significance.

5.2 Y-12 RESULTS

Based on the discussion in the previous section, the general strategy was to draw 1,585 unique workers at random[9] from the 7,537 workers in the complete dataset and determine the slope and

[9] Sample without replacement.

intercept of the fit. The results of the m = 1,000 iterations[10] are presented in scatter plots like the one shown in Figure 5-1 for 1952. The results of the fits for 1950 to 1988 are summarized in Table 4-1, and the plots are given in Attachment C. The slopes and intercepts of the claimant datasets fall in the 95% confidence ellipse for 38 of the 39 years, with only the slopes and intercepts for 1950 being outside of the 95% confidence ellipse.

6.0 MOUND PLUTONIUM URINE BIOASSAY DATA AND RESULTS

For the Mound Laboratory a single plutonium urine bioassay dataset, PURECON, was provided for 1960 to 1990 along with a list that identified claimants in the dataset. This list was used to extract claimants from the complete dataset to construct the claimant dataset.

The PURECON database was created in 1991 by the University of Lowell Research Foundation. This work was undertaken to meet the requirements of DOE Order 5480.11, “Radiation Protection for Occupational Workers” (DOE 1988), which became effective on January 1, 1989. Mound staff recognized the need to develop more reliable internal dosimetry recordkeeping for plutonium; this resulted in the construction of a computer database from the original raw data. Subsequent quality assurance and repair work on the PURECON database has been performed (and documented) by the MJW Corporation during its work on the Pre-1989 Dose Assessment Project (MJW 2002). PURECON is considered to be a complete dataset.

There are 2,070 workers in the complete dataset who submitted a total of 53,338 plutonium urine samples. Urine samples that were influenced by chelation therapy were not included in the analysis. The year-by-year breakdowns of the total number of submitted samples and total number of monitored workers are shown in Figures 6-1 and 6-2, respectively. The claimant dataset is a subset of the complete dataset; that is, all of the claimants are workers. A total of 301 claimants submitted 12,088 plutonium urine samples. The year-by-year breakdowns of the number of submitted samples and number of monitored claimants are shown in Figures 6-3 and 6-4, respectively. Note that the number of results in a year will equal the number of individuals monitored in that year after each individual’s bioassay results are summarized into one result (i.e., the TWOPOS).

A summary of the number of samples and workers in the claimant and complete datasets is given in Table 6-1. A summary of the slopes and intercepts of the fits to the complete and claimant data is given in Table 6-2. Plots of the 50th and 84th percentile data in Table 6-2 are given in Figures 6-5 and 6-6, respectively. The lognormal probability plots for the complete and claimant datasets are given in Attachments D and E, respectively.

Following the general procedure established for the Y-12 data, k = 301 unique workers were drawn at random without replacement from the n = 2,070 workers in the complete dataset and the slope and intercept of the fit were determined. The results of the m = 2,000 iterations of this process are presented in scatter plots for 1960 to 1990 in Attachment F. The slopes and intercepts of the claimant datasets fall inside the 95% confidence ellipse for 29 out of 31 years. The slopes and intercepts for 1982 and 1990 fall outside of their respective 95% confidence ellipses.

---

[10] The simulations for the three sites were run for lengths of time that were approximately equal. Y-12 was the largest dataset and ran the fewest number of iterations, m = 1000. Anything over ~500 iterations is considered adequate for the purposes of this discussion.

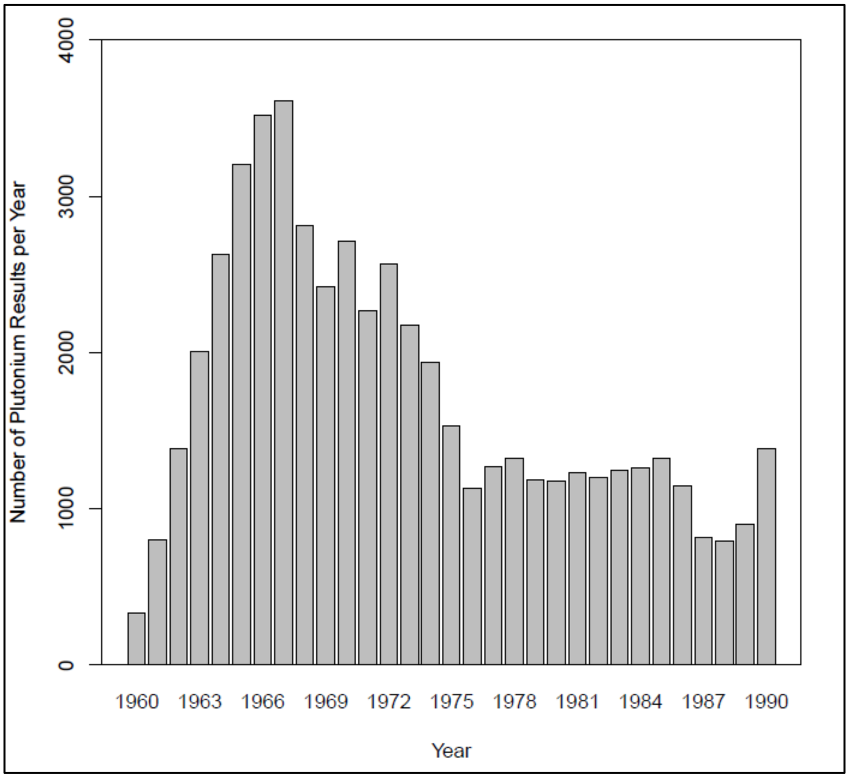


Figure 6-1. Number of samples submitted per year by all Mound workers.

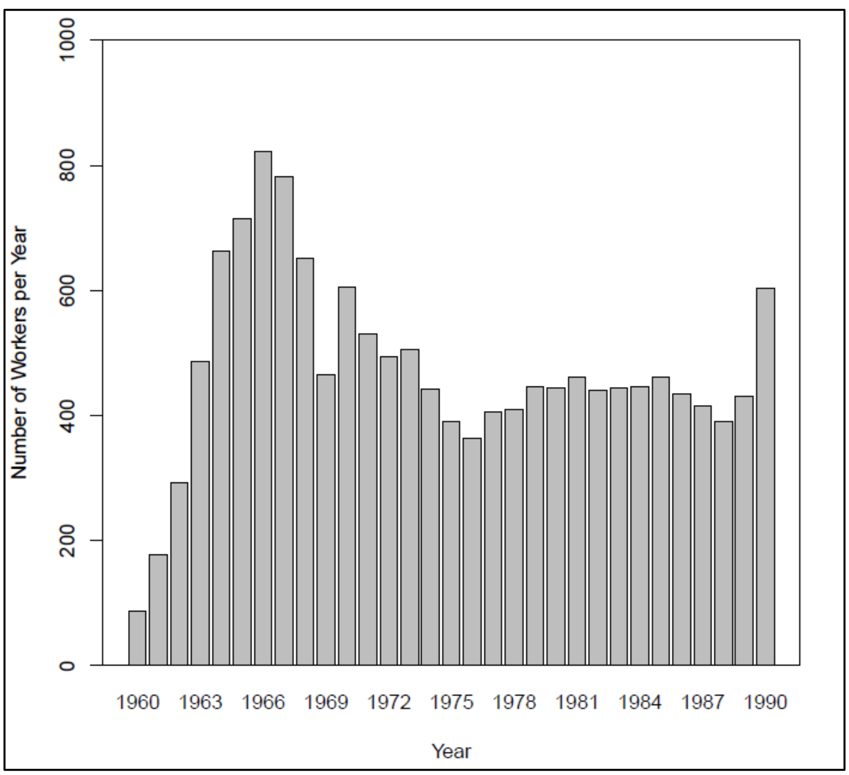


Figure 6-2. Number of Mound workers monitored in each year.

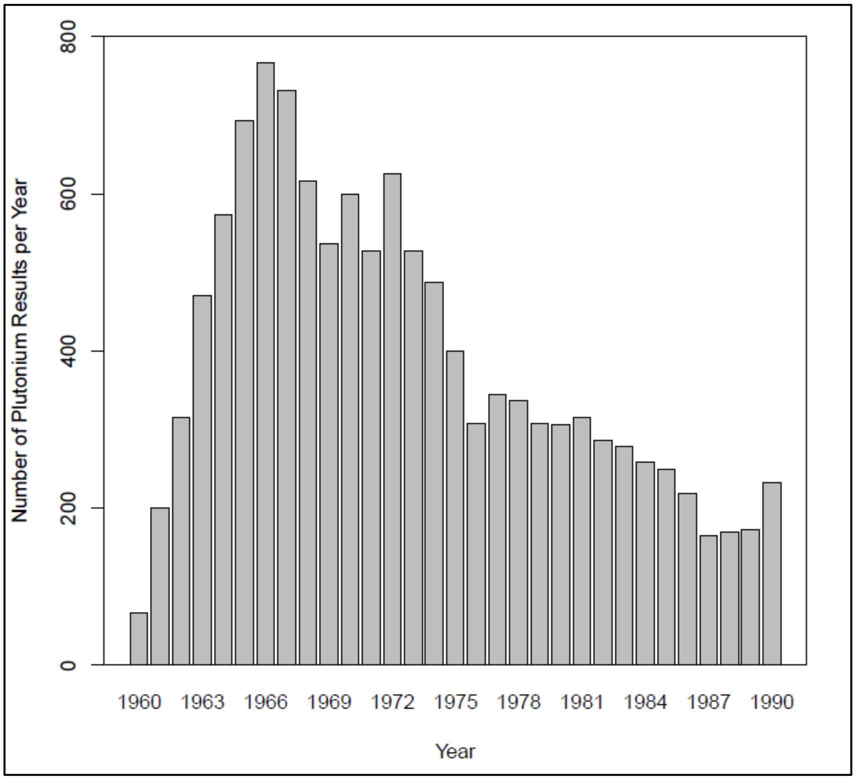


Figure 6-3. Number of samples submitted per year by all Mound claimants.

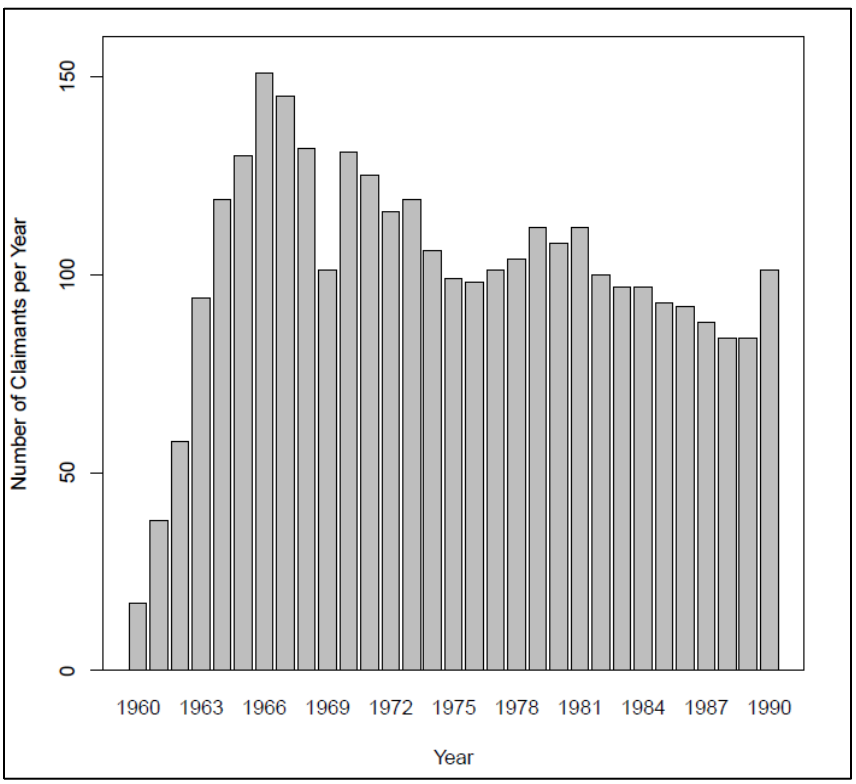


Figure 6-4. Number of Mound claimants monitored in each year.

Table 6-1. Summary of samples and workers in claimant and complete Mound datasets by year.

	Workers			Samples		
Year	**Complete dataset**	**Claimant dataset**	**Ratio**	**Complete dataset**	**Claimant dataset**	**Ratio**
1960	88	17	0.193	336	66	0.196
1961	178	38	0.213	802	200	0.249
1962	292	58	0.199	1385	315	0.227
1963	487	94	0.193	2009	471	0.234
1964	662	119	0.180	2625	574	0.219
1965	715	130	0.182	3203	693	0.216
1966	822	151	0.184	3520	767	0.218
1967	781	145	0.186	3611	732	0.203
1968	652	132	0.202	2815	616	0.219
1969	466	101	0.217	2425	536	0.221
1970	605	131	0.217	2716	599	0.221
1971	531	125	0.235	2264	527	0.233
1972	495	116	0.234	2568	626	0.244
1973	505	119	0.236	2177	528	0.243
1974	442	106	0.240	1940	488	0.252
1975	391	99	0.253	1533	400	0.261
1976	363	98	0.270	1134	308	0.272
1977	405	101	0.249	1270	344	0.271
1978	409	104	0.254	1321	337	0.255
1979	447	112	0.251	1188	308	0.259
1980	445	108	0.243	1181	306	0.259
1981	461	112	0.243	1235	315	0.255
1982	440	100	0.227	1198	286	0.239
1983	444	97	0.218	1247	279	0.224
1984	446	97	0.217	1262	258	0.204
1985	461	93	0.202	1325	249	0.188
1986	435	92	0.211	1149	219	0.191
1987	415	88	0.212	818	165	0.202
1988	391	84	0.215	793	170	0.214
1989	430	84	0.195	899	173	0.192
1990	603	101	0.167	1389	233	0.168

Table 6-2. Summary of the slopes and intercepts of a fit to a lognormal probability model and the corresponding 50th and 84th percentiles for Mound. The corresponding plots are in Attachments D and E.

Year	Complete slope	Complete intercept	Claimant slope	Claimant intercept	Complete 50th percentile	Complete 84th percentile	Claimant 50th percentile	Claimant 84th percentile
1960	1.0285	-1.8960	1.1559	-1.7109	0.1502	0.4200	0.1807	0.5741
1961	1.3832	-2.6173	1.2341	-2.5006	0.0730	0.2911	0.0820	0.2818
1962	1.2627	-2.4775	1.1975	-2.3977	0.0840	0.2968	0.0909	0.3011
1963	1.4106	-2.7808	1.2576	-2.5725	0.0620	0.2541	0.0763	0.2685
1964	0.8885	-2.1258	0.9705	-2.0416	0.1193	0.2902	0.1298	0.3426
1965	1.3581	-2.5238	1.4105	-2.4247	0.0802	0.3117	0.0885	0.3627
1966	0.9219	-2.6932	0.9611	-2.5615	0.0677	0.1701	0.0772	0.2018
1967	1.1224	-2.7545	1.0280	-2.5635	0.0636	0.1955	0.0770	0.2153
1968	1.2522	-3.0720	1.3147	-3.0879	0.0463	0.1621	0.0456	0.1698
1969	1.6109	-3.4352	1.5358	-3.3294	0.0322	0.1613	0.0358	0.1664
1970	1.4390	-3.7412	1.3322	-3.5402	0.0237	0.1000	0.0290	0.1099
1971	1.3972	-3.6546	1.3815	-3.6198	0.0259	0.1046	0.0268	0.1066
1972	1.4298	-3.5429	1.4422	-3.5319	0.0289	0.1209	0.0293	0.1237
1973	1.3044	-3.3533	1.0626	-3.1753	0.0350	0.1289	0.0418	0.1209
1974	1.1206	-2.9539	0.9513	-2.8416	0.0521	0.1599	0.0583	0.1510
1975	1.6222	-3.7457	1.3743	-3.6731	0.0236	0.1196	0.0254	0.1004
1976	1.5412	-4.2120	1.5360	-4.2282	0.0148	0.0692	0.0146	0.0677
1977	1.7134	-4.6201	1.7084	-4.3528	0.0099	0.0547	0.0129	0.0710
1978	2.0784	-5.4187	1.7476	-5.1082	0.0044	0.0354	0.0060	0.0347
1979	1.7643	-4.7977	1.8308	-4.7066	0.0082	0.0482	0.0090	0.0564
1980	1.3337	-4.0778	1.3808	-4.0150	0.0169	0.0643	0.0180	0.0718
1981	1.4640	-3.9524	1.4195	-3.7631	0.0192	0.0830	0.0232	0.0960
1982	1.4665	-4.2334	1.2198	-3.7440	0.0145	0.0629	0.0237	0.0801
1983	1.7624	-4.5486	1.6137	-4.1738	0.0106	0.0617	0.0154	0.0773
1984	1.6793	-4.8136	1.5284	-4.3897	0.0081	0.0435	0.0124	0.0572
1985	2.5635	-6.3277	2.5470	-5.7437	0.0018	0.0232	0.0032	0.0409
1986	2.0940	-5.5520	2.1252	-5.2856	0.0039	0.0315	0.0051	0.0424
1987	1.7788	-5.4406	1.8982	-5.1192	0.0043	0.0257	0.0060	0.0399
1988	2.1273	-5.6891	2.1432	-5.1145	0.0034	0.0284	0.0060	0.0512
1989	2.5266	-6.5060	2.5131	-5.8045	0.0015	0.0187	0.0030	0.0372
1990	2.6762	-7.1818	2.4367	-6.1129	0.0008	0.0110	0.0022	0.0253

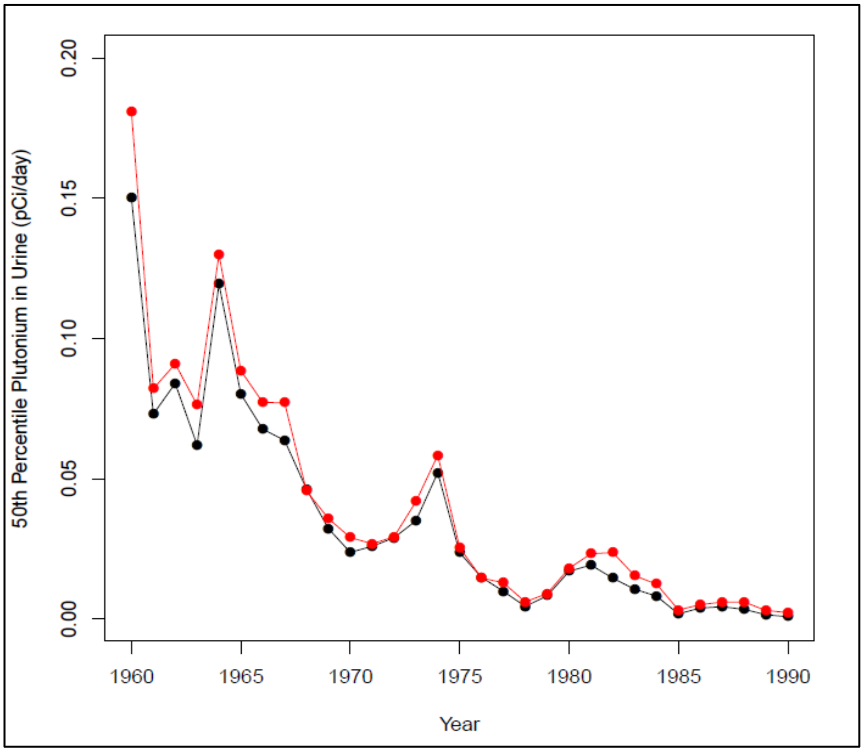


Figure 6-5. Comparison of 50th percentiles for the complete (black) and claimant (red) Mound datasets.

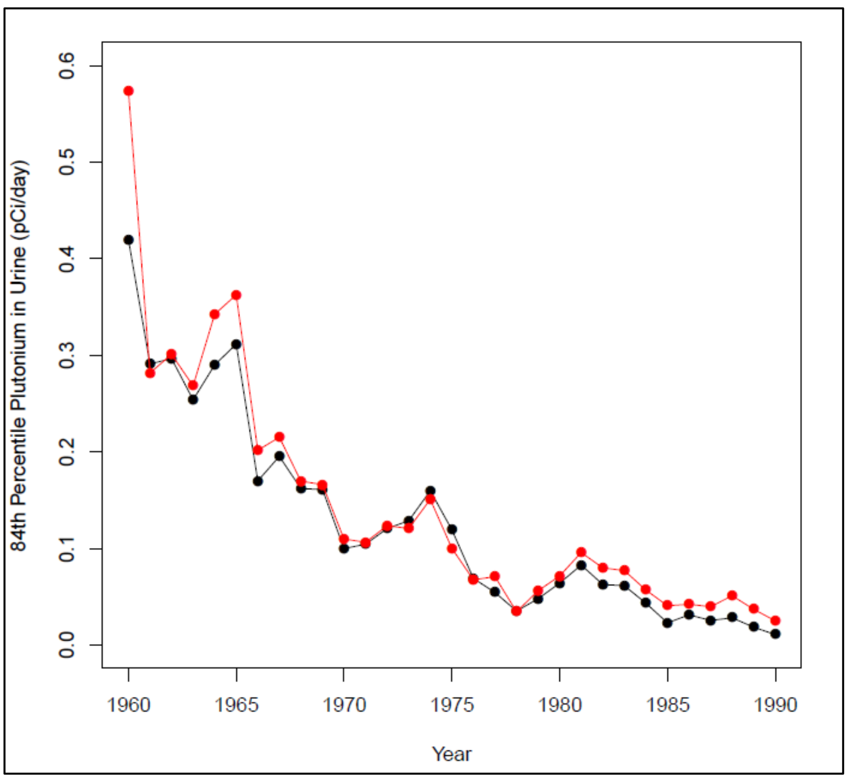


Figure 6-6. Comparison of 84th percentiles for the complete (black) and claimant (red) Mound datasets.

7.0 SAVANNAH RIVER SITE TRITIUM DOSE DATA AND RESULTS

The ideal type of data with which to develop a coworker model is dose data as opposed to bioassay data because the relatively complicated step in which the bioassay data are converted to dose can be skipped. This is the case with external coworker models where the relevant data are reported in terms of dose. However, the datasets for most internal dose coworker models (such as Y-12 and Mound) consist of urine bioassay data. The required effort to convert all of the uranium and plutonium urine bioassay data to doses for each individual is prohibitive, so the coworker modeling is performed on the bioassay data. One case where it is feasible[11] to convert urine bioassay data to individual doses is for tritium urine bioassay data. Therefore, the SRS analysis used tritium dose data rather than tritium urine bioassay data. These doses were calculated from the tritium urine bioassay dataset that was extracted from the SRS Health Protection Radiation Exposure Database (HPRED), which is considered complete for 1991 onward.

The tritium dose dataset for this analysis is for 1991 to 2000. There are 10,712 workers in the complete dose dataset and 920 claimants in the claimant dose dataset. The original urine bioassay dataset from HPRED contained 521,798 individual bioassay results. It is important to note that while an individual might have hundreds of tritium urine samples in a given year, there is only one annual dose (i.e., dose is inherently TWOPOS).

The year-by-year breakdowns of the total number of monitored workers and claimants are shown in Figures 7-1 and 7-2, respectively. A summary of the data in these plots is given in Table 7-1. A summary of the slopes and intercepts of the fits to the complete and claimant dose data is given in Table 7-2, and plots of these data are given in Figures 7-3 and 7-4. The lognormal probability plots for the complete and claimant datasets are given in Attachments G and H, respectively.

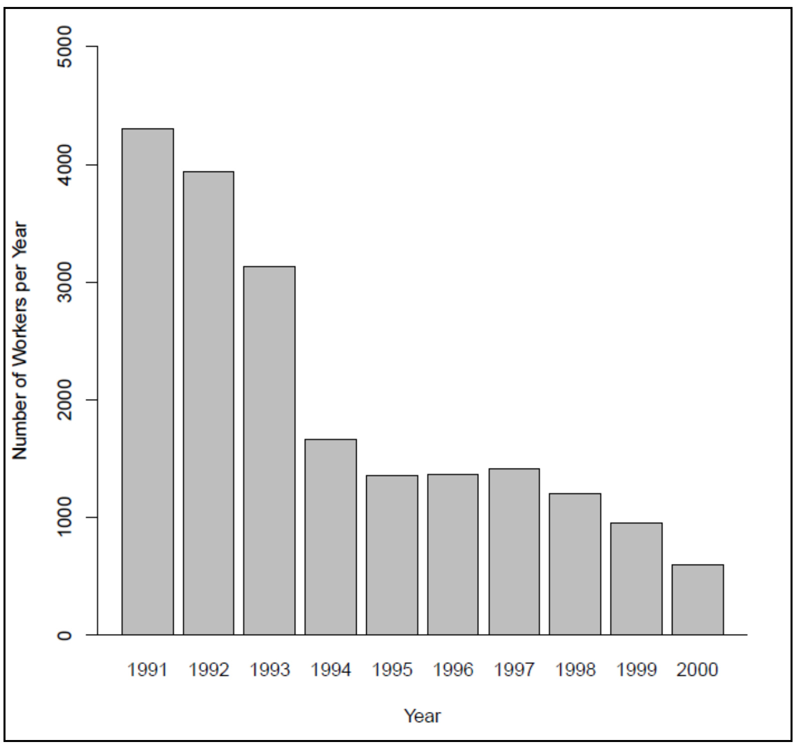


Figure 7-1. Number of SRS workers monitored in each year.

[11] The methodology of ORAUT-OTIB-0011, Technical Information Bulletin: Tritium Calculated and Missed Dose Estimates (ORAUT 2004) can be implemented in a computer program that is capable of quickly calculating dose from tritium urine bioassay datasets without manual intervention.

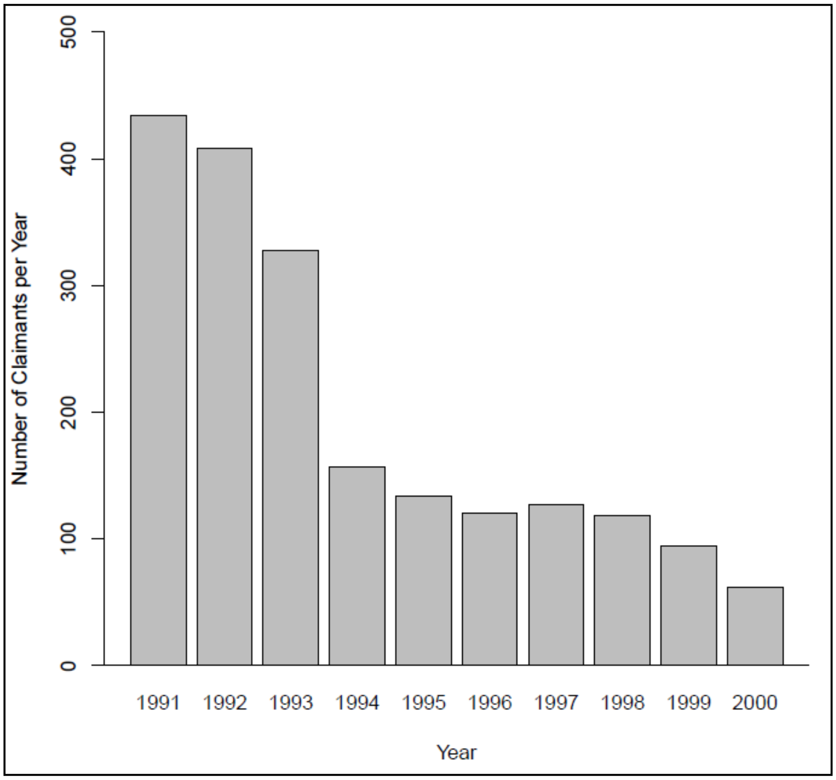


Figure 7-2. Number of SRS claimants monitored in each year.

Table 7-1. Summary of workers in claimant and complete SRS datasets by year.

Year	Workers in complete dataset	Workers in claimant dataset	Ratio
1991	4,303	434	0.1009
1992	3,939	408	0.1036
1993	3,130	328	0.1048
1994	1,667	157	0.0942
1995	1,355	134	0.0989
1996	1,367	120	0.0878
1997	1,414	127	0.0898
1998	1,200	118	0.0983
1999	953	94	0.0986
2000	599	62	0.1035

Once again following the general procedure established for the Y-12 data, k = 920 unique workers were drawn at random without replacement from the n = 10,712 workers in the complete dataset and the slope and intercept of the fit were determined. The results of the m = 3,000 iterations of this process are presented in scatter plots for 1991 to 2000 in Attachment I. The slopes and intercepts in 9 out of the 10 claimant datasets are within the 95% confidence ellipses. The slopes and intercepts for 1991 fall outside of the 95% confidence ellipses.

Table 7-2. Summary of the slopes and intercepts of a fit to a lognormal probability model and the corresponding 50th and 84th percentiles for SRS. The corresponding plots are in Attachments G and H.

Year	Complete slope	Complete intercept	Claimant slope	Claimant intercept	Complete 50th percentile	Complete 84th percentile	Claimant 50th percentile	Claimant 84th percentile
1991	0.8138	1.3796	0.8173	1.4962	3.9734	8.9655	4.4647	10.1093
1992	0.6726	1.1651	0.6662	1.2299	3.2061	6.2816	3.4210	6.6604
1993	0.6485	1.0756	0.6784	1.0937	2.9318	5.6075	2.9853	5.8831
1994	0.6614	1.2011	0.6982	1.3208	3.3237	6.4397	3.7462	7.5302
1995	0.5975	1.1877	0.6244	1.2702	3.2796	5.9612	3.5616	6.6498
1996	0.5439	1.2287	0.5819	1.2969	3.4167	5.8857	3.6578	6.5454
1997	0.5006	1.2346	0.4889	1.2547	3.4370	5.6702	3.5069	5.7180
1998	0.4768	1.2311	0.4142	1.2397	3.4250	5.5173	3.4547	5.2275
1999	0.4988	1.1111	0.4749	1.1390	3.0376	5.0020	3.1238	5.0224
2000	0.4313	1.1685	0.4540	1.0929	3.2172	4.9521	2.9830	4.6970

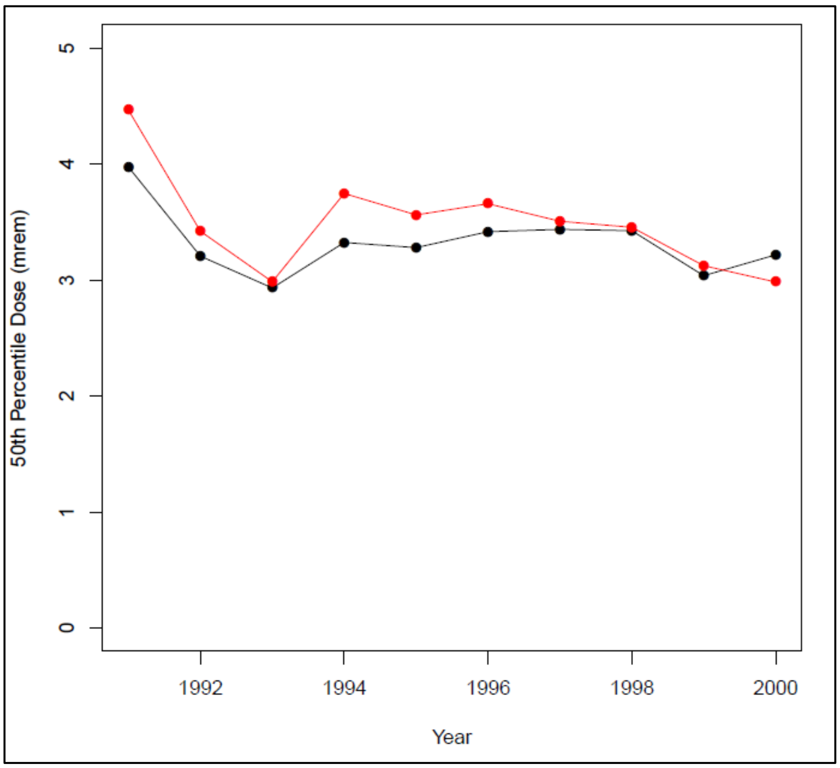


Figure 7-3. Comparison of 50th percentiles for the complete (black) and claimant (red) SRS datasets.

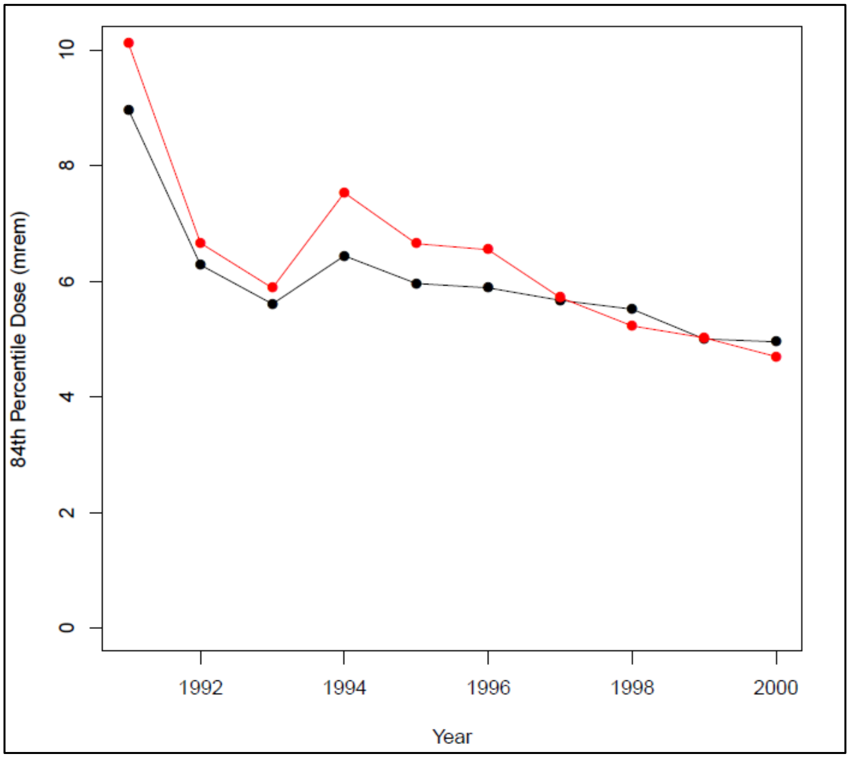


Figure 7-4. Comparison of 84th percentiles for the complete (black) and claimant (red) SRS datasets.

8.0 SUMMARY AND CONCLUSIONS

In the dose reconstruction process, coworker models are used to assign doses to workers who were not monitored at certain times during their employment but probably should have been. Coworker models of the 50th- and 84th-percentile bioassay results for specific periods are typically developed from essentially complete datasets of worker monitoring data. However, there are sites where complete worker monitoring datasets are not available for developing coworker models. Because of the considerable cost and effort necessary to create a complete dataset, there is interest in developing coworker models from a sample of the complete dataset. There is no technical reason this cannot be carried out provided that the sample is selected by a random process and is sufficiently large. Samples that are collected in this way provide unbiased estimates of the percentiles of the complete dataset and permit estimation of the uncertainty in the percentile estimates.

A sample of the complete dataset that is readily obtainable is the bioassay data of the workers who have a claim under EEOICPA that names them as the energy employees (i.e., the claimants for a site who make up the claimant dataset). However, the claimant datasets were not intentionally selected by an appropriate sampling process, so it is not clear that they would provide unbiased estimates of the percentiles of the complete dataset. It would be ideal to show that the claimant datasets behave essentially like a random sample in relation to the parameters that are entered into coworker models. The approach taken here to show this was:

1. Select sites with well-defined claimant and complete datasets.
2. Examine the distribution of slopes and intercepts from randomly selected datasets the same size as the claimant dataset that are drawn from the complete dataset.
3. Use this distribution to determine how likely it would be to observe slopes and intercepts like those from the claimant dataset.

4. If the slopes and intercepts from the claimant dataset are indistinguishable from those from the random samples, then the claimant dataset can be used as a random sample of the complete dataset and can be used to develop coworker models.

At the three sites a total of 4 years out of 80 (5%) had coworker model parameters outside of the respective 95% confidence ellipses (Y-12 in 1950, Mound in 1982 and 1990, and SRS in 1991). This result is in excellent agreement with what would be expected if the NOCTS datasets were random draws from the complete datasets (i.e., in the long run, we would expect to see 5% of the parameter estimates outside of the 95% confidence ellipse). This proof-of-principle technique was used to show that the claimant datasets from Y-12, Mound, and SRS can, in general, be used as if they were random samples for the purpose of estimating the slopes and intercepts of lognormal fits to the data. Further, this exercise can be used as a technical justification for applying this assumption to other sites for which complete datasets are not available.

9.0 ATTRIBUTIONS AND ANNOTATIONS

All information requiring identification was addressed via references integrated into the reference section of this document.

REFERENCES

Efron, B., and R. J. Tibshirani, 1998, *An Introduction to the Bootstrap*, CRC Press, Boca Raton, Florida.

Davison, A. C., and D. V. Hinkley, 1997, *Bootstrap Methods and their Application*, Cambridge University Press, New York, New York, October.

DOE (U.S. Department of Energy), 1988, *Radiation Protection for Occupational Workers*, Order 5480.11, Office of Environment, Safety and Health, Washington D.C., December 21. [SRDB Ref ID: 8073]

ORAUT (Oak Ridge Associated Universities Team), 2004, *Technical Information Bulletin: Tritium Calculated and Missed Dose Estimates*, ORAUT-OTIB-0011, Rev. 00, Oak Ridge, Tennessee, June 29.

ORAUT (Oak Ridge Associated Universities Team), 2005, *Analysis of Coworker Bioassay Data for Internal Dose Assignment*, ORAUT-OTIB-0019, Rev. 01, Oak Ridge, Tennessee, October 7.

ORAUT (Oak Ridge Associated Universities Team), 2012, *Y-12 National Security Complex – Occupational Internal Dose*, ORAUT-TKBS-0014-5, Rev. 03, Oak Ridge, Tennessee, March 12.

ORAUT (Oak Ridge Associated Universities Team), 2014, *Analysis of Stratified Coworker Datasets*, ORAUT-RPRT-0053, Rev. 02, Oak Ridge, Tennessee, October 8.

MJW (MJW Corporation), 2002, *Pre-1989 Dose Assessment Project, Phase I Final Report, Volume 1 (Non-SUD Version)*, Williamsville, New York, April. [SRDB Ref ID: 8745]

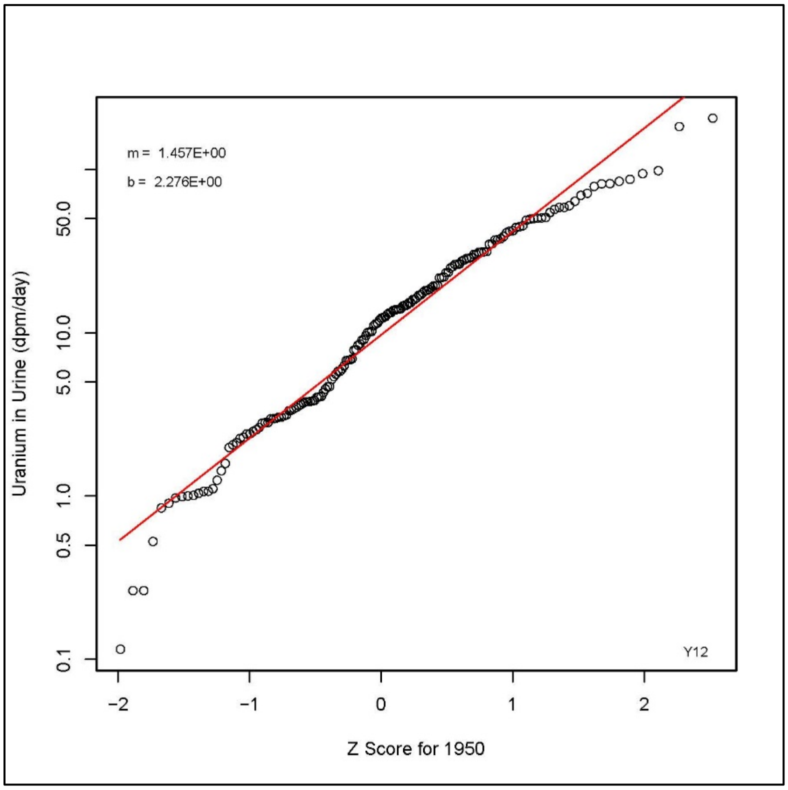
RDCT (R Development Core Team), 2008, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

SourceForge, 2016, *Maxima Manual*, Mountain View, California.

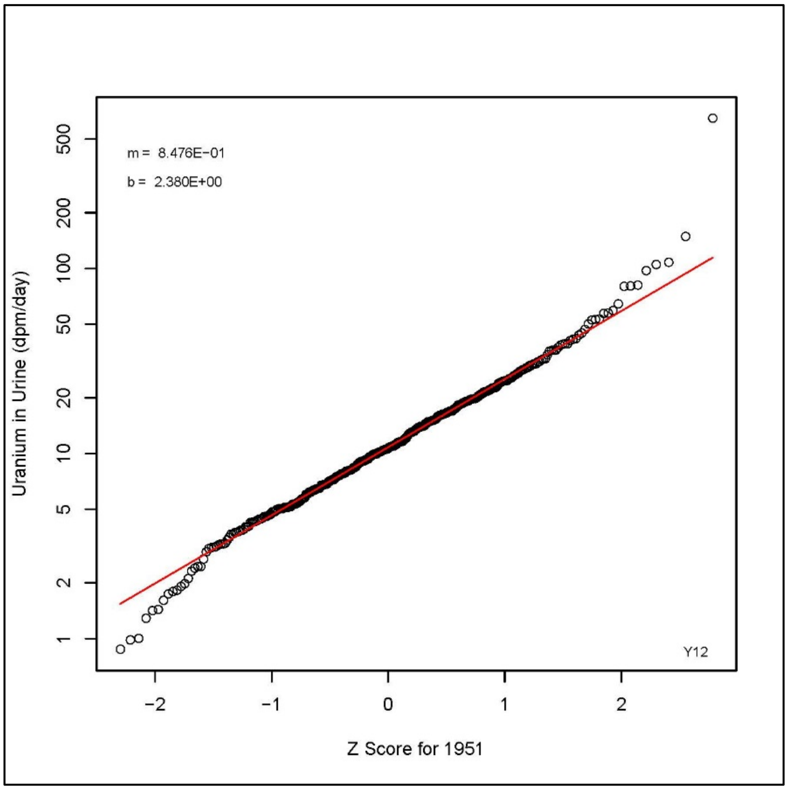
Watkins, J. P., J. L. Reagan, D. L. Cragle, E. L. Frome, C. M. West, D .J. Crawford-Brown, and W. G. Tankersley, 1993, *Data Collection, Validation, and Description for the Oak Ridge Nuclear Facilities Mortality Study*, ORISE 93/J-42, Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee, October. [SRDB Ref ID: 10133]

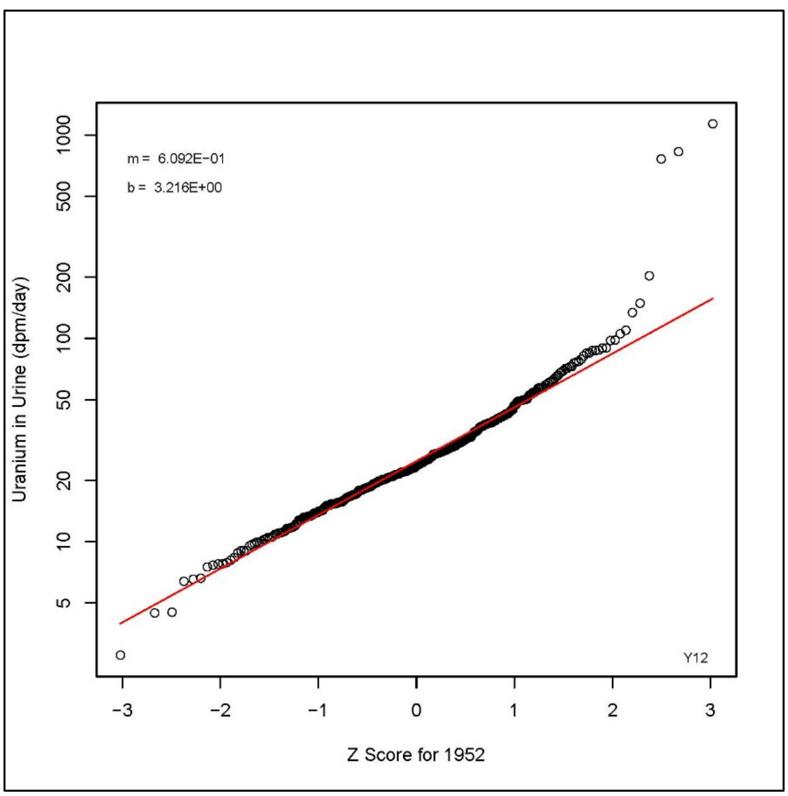
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET

This attachment presents lognormal probability plots for the complete Y-12 uranium urine bioassay dataset for 1950 to 1988. The slope and intercept for the line of best fit to the data are given on the plot.

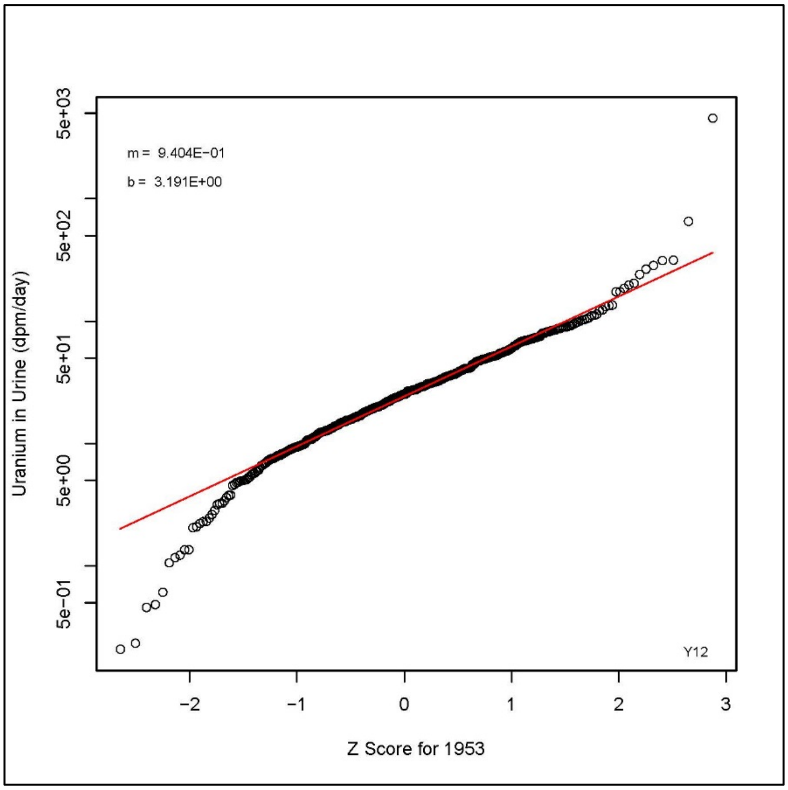


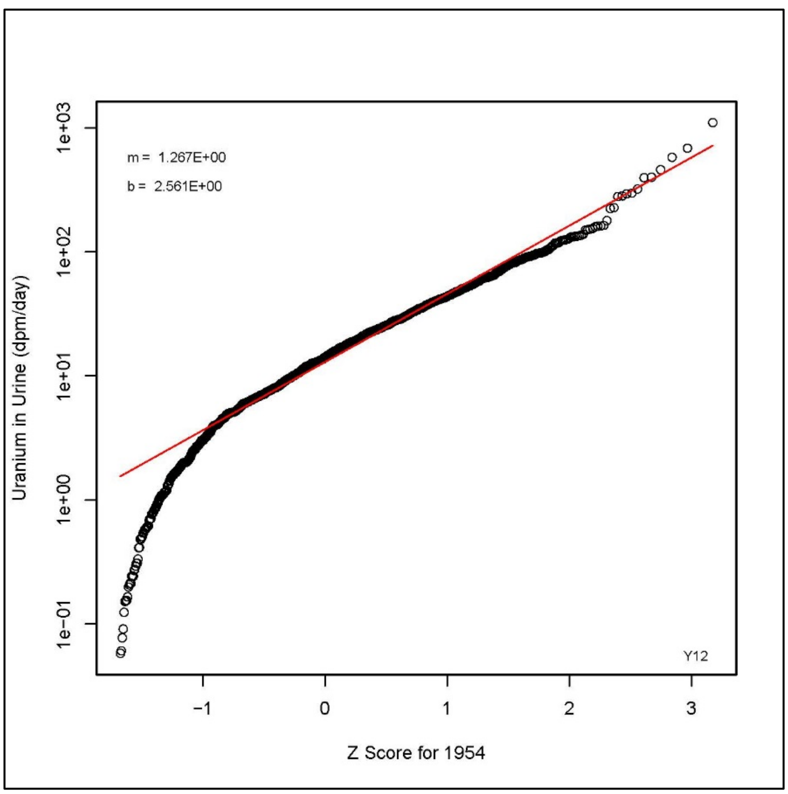
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



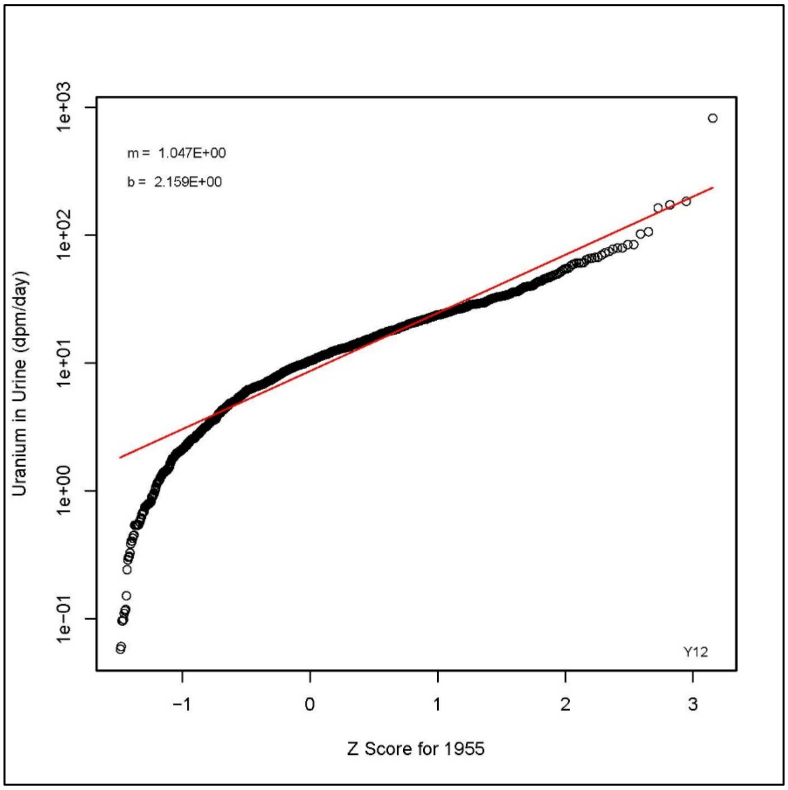


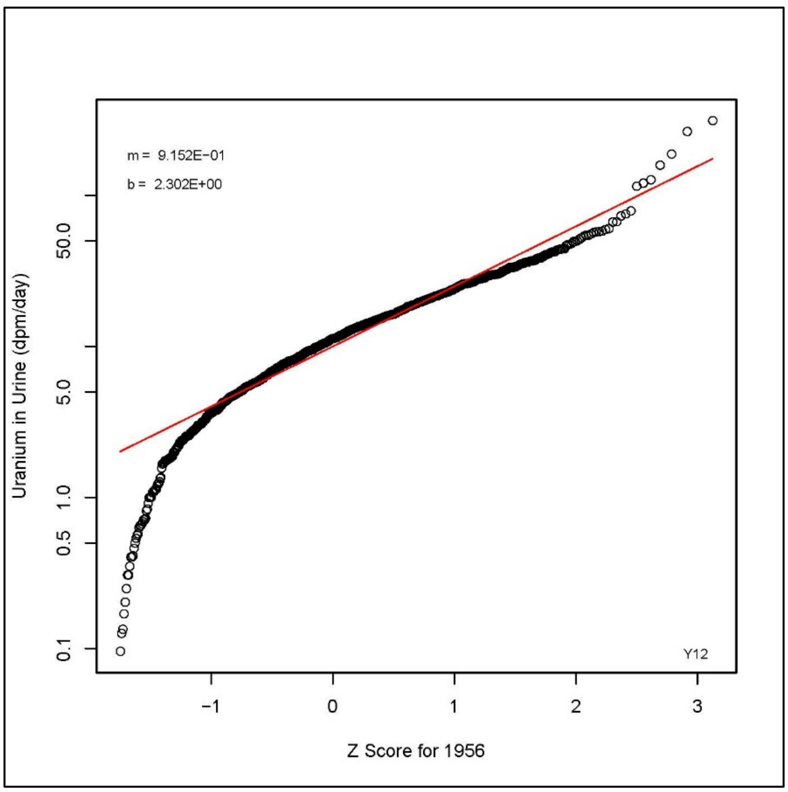
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



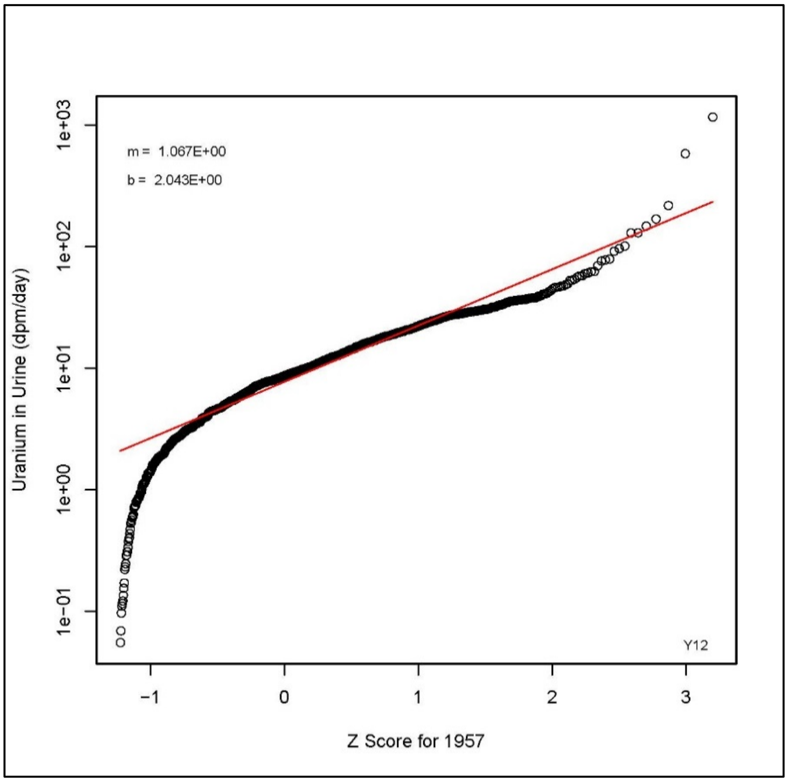


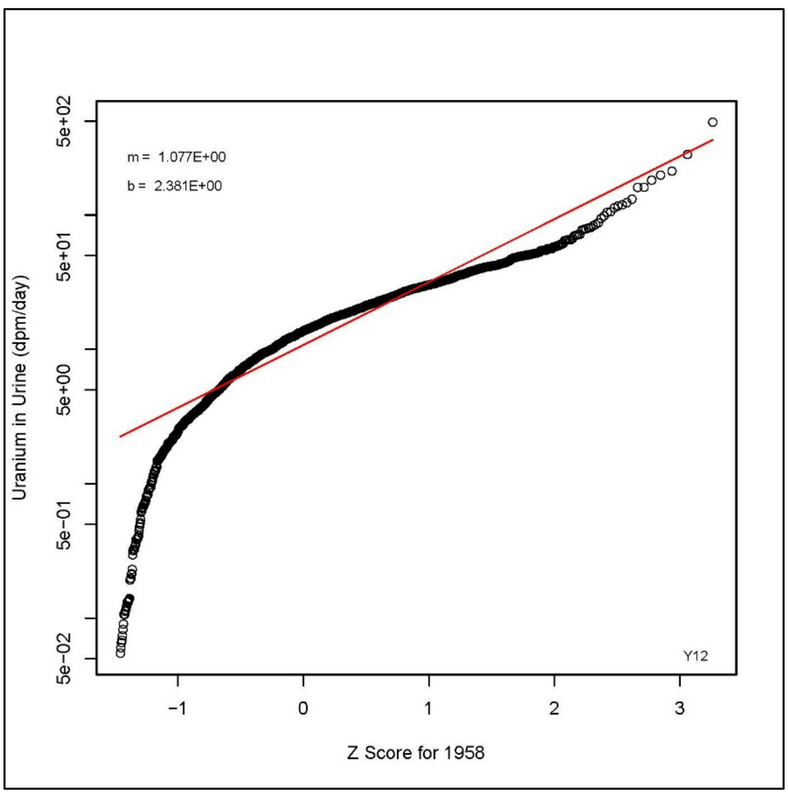
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



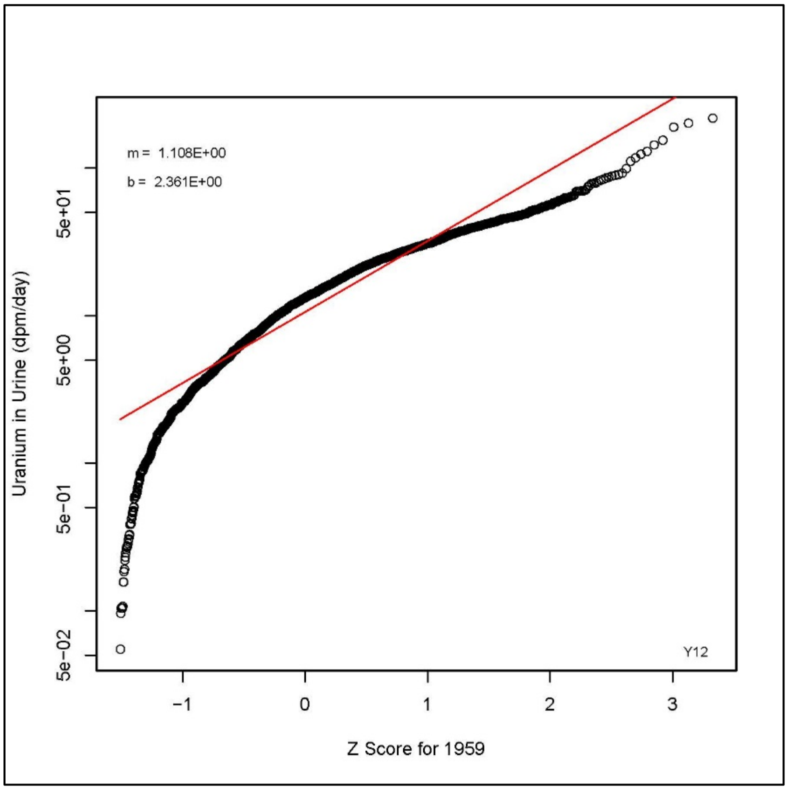


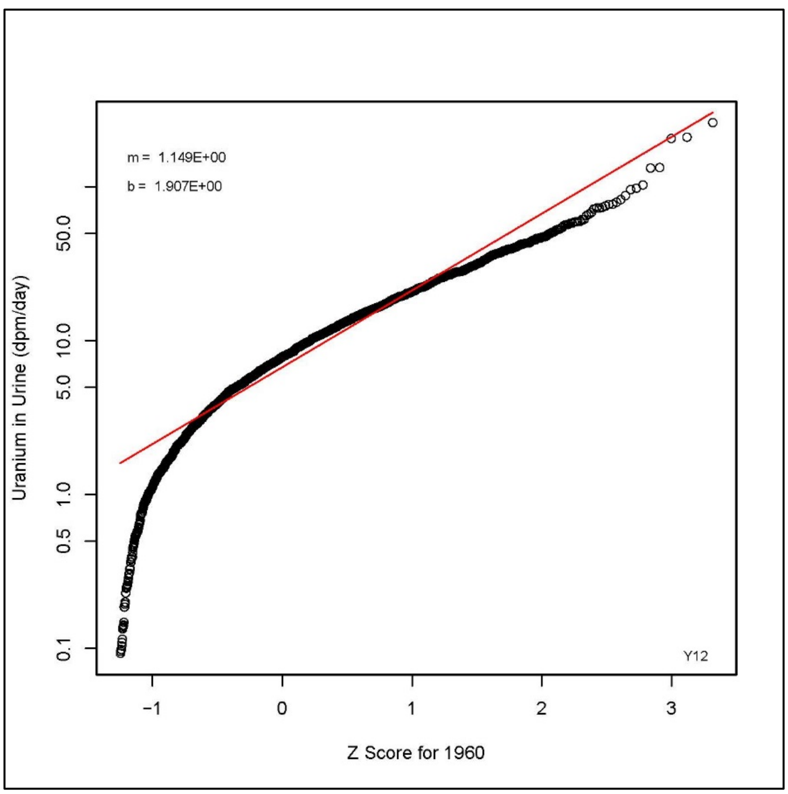
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



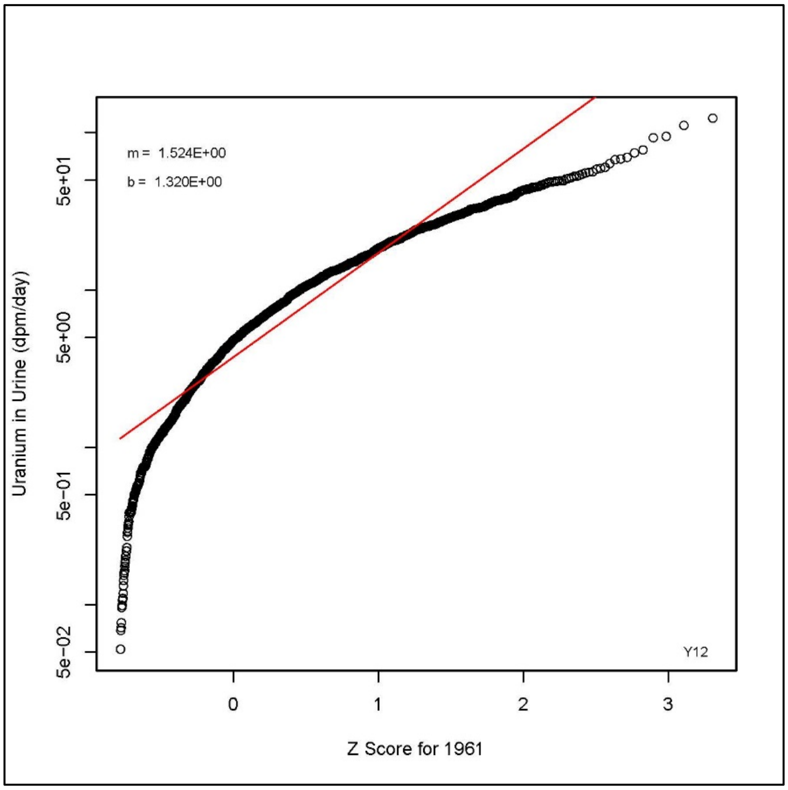


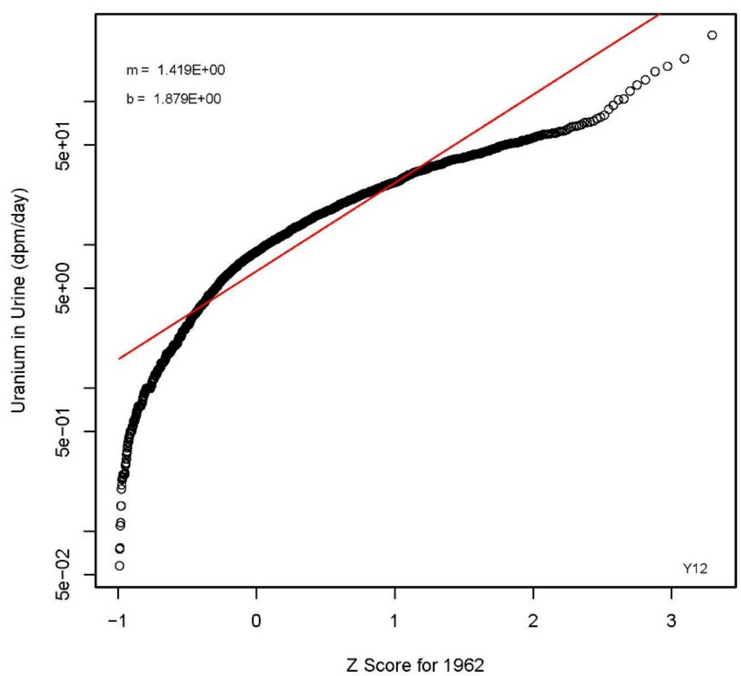
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



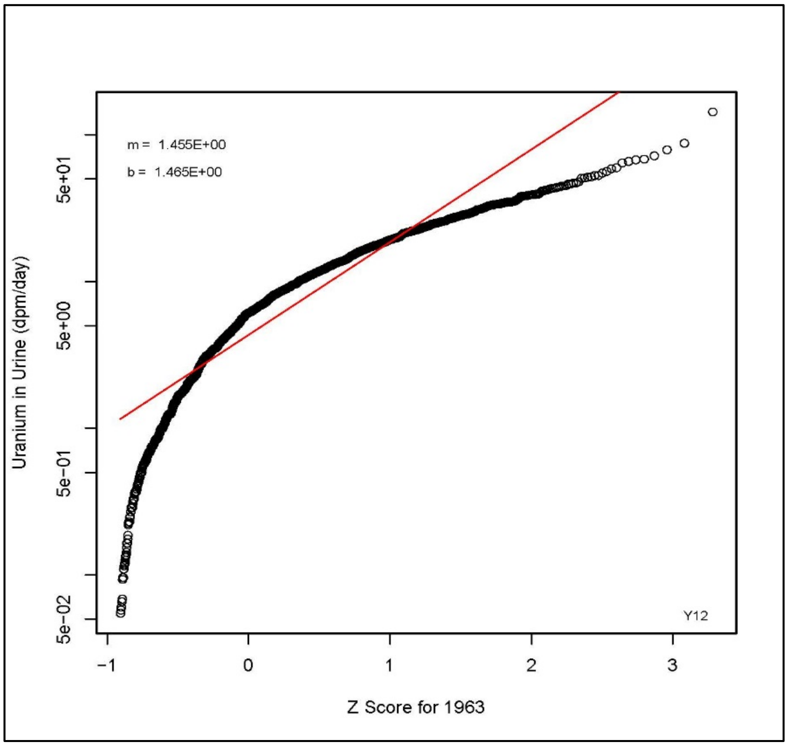


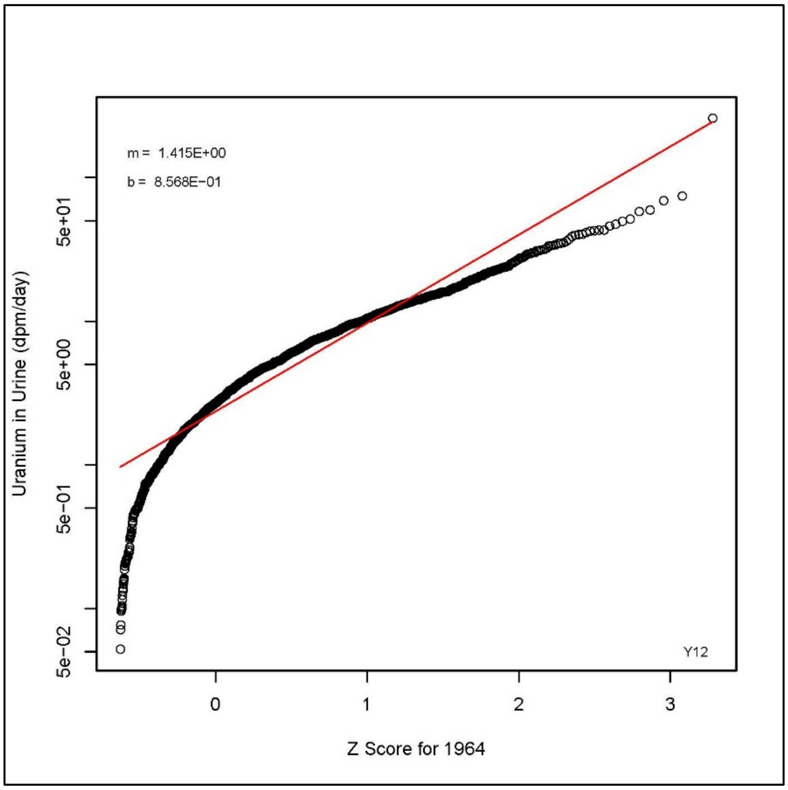
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



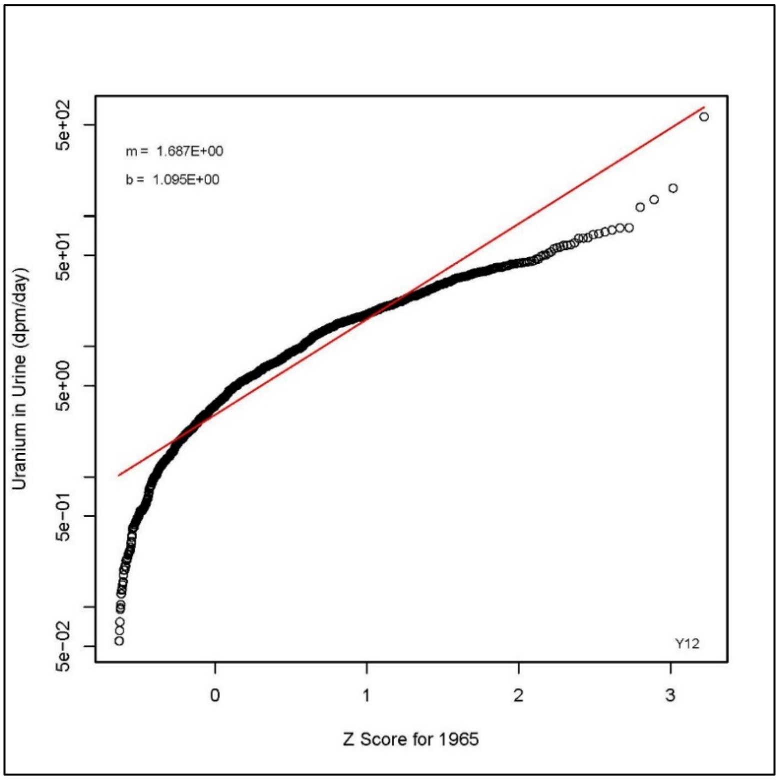


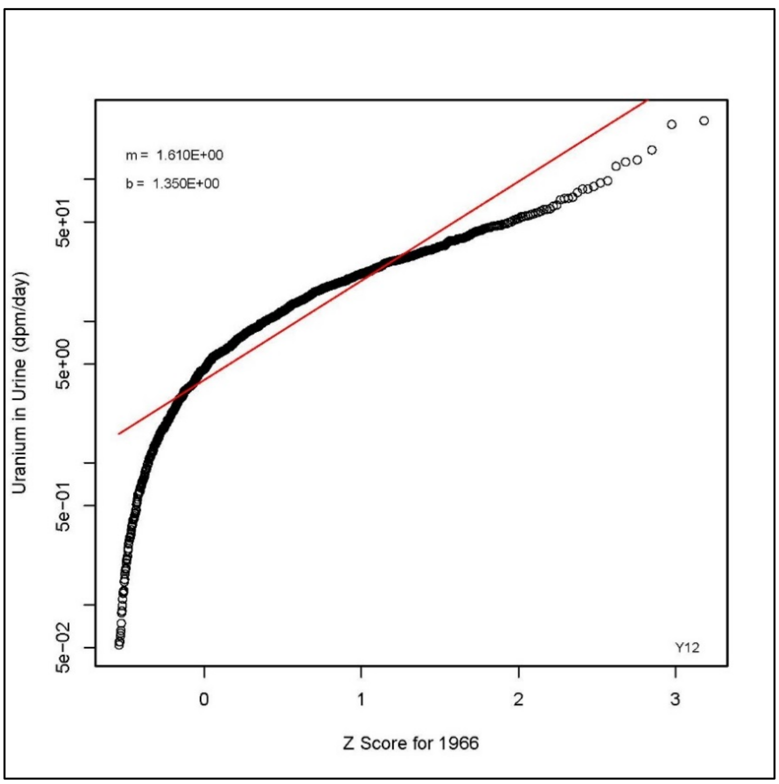
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



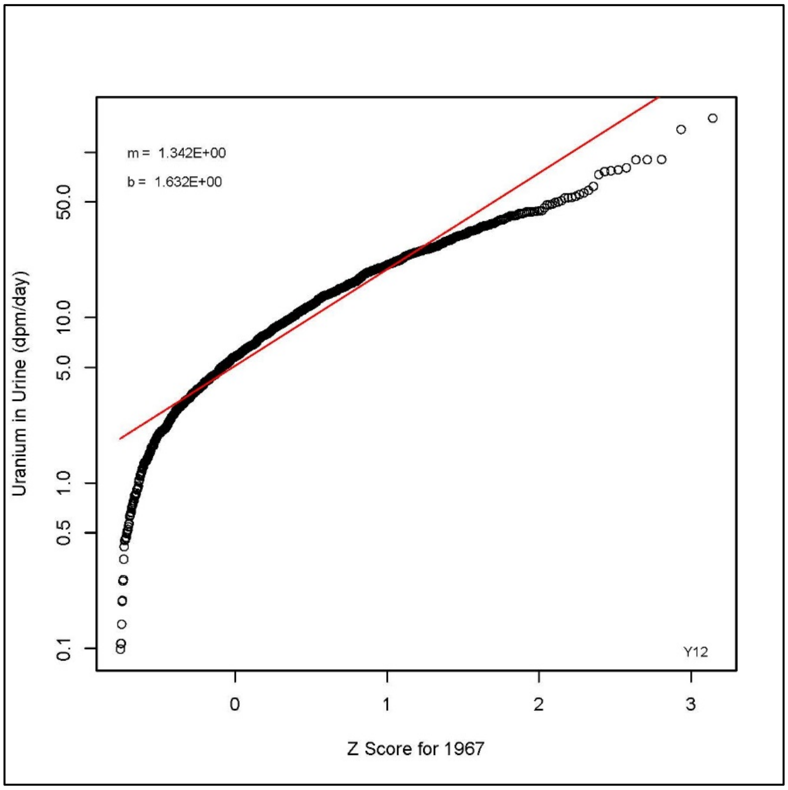


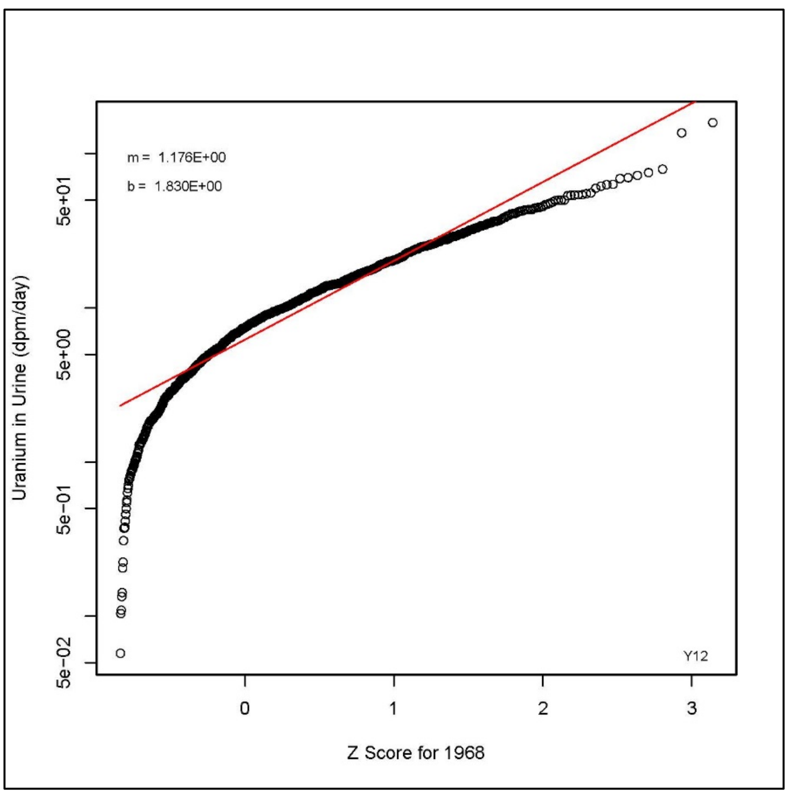
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



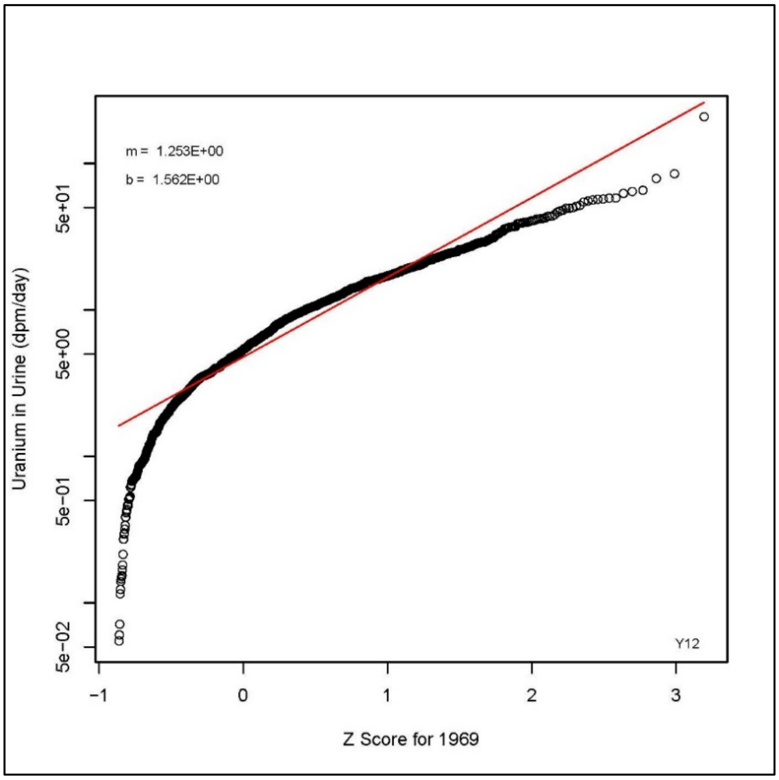


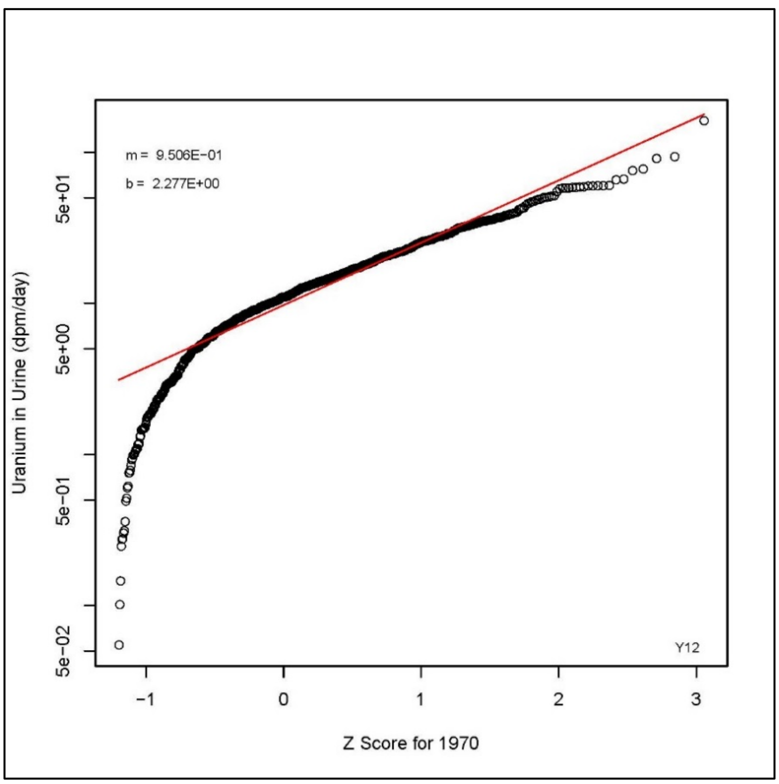
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



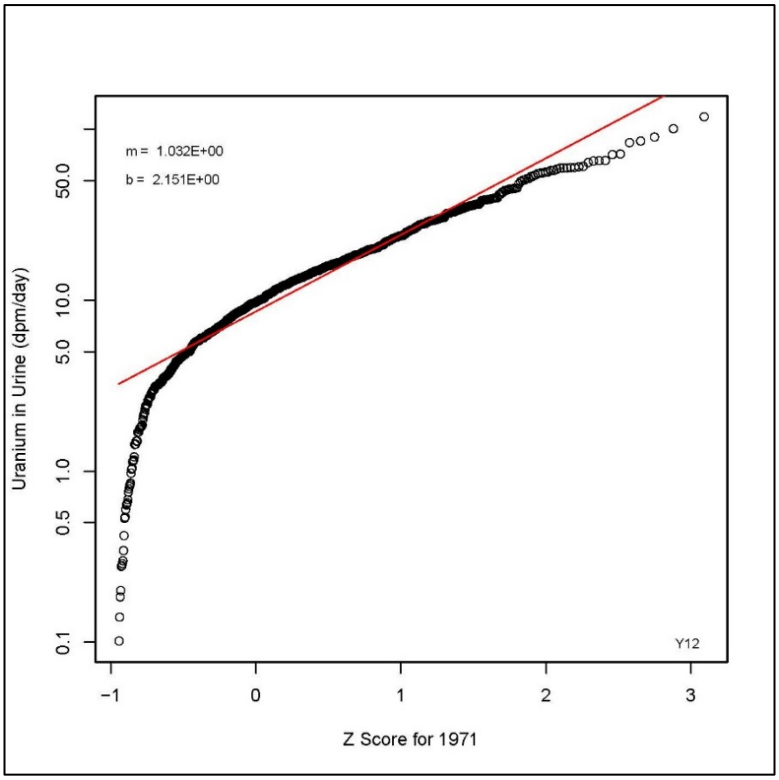


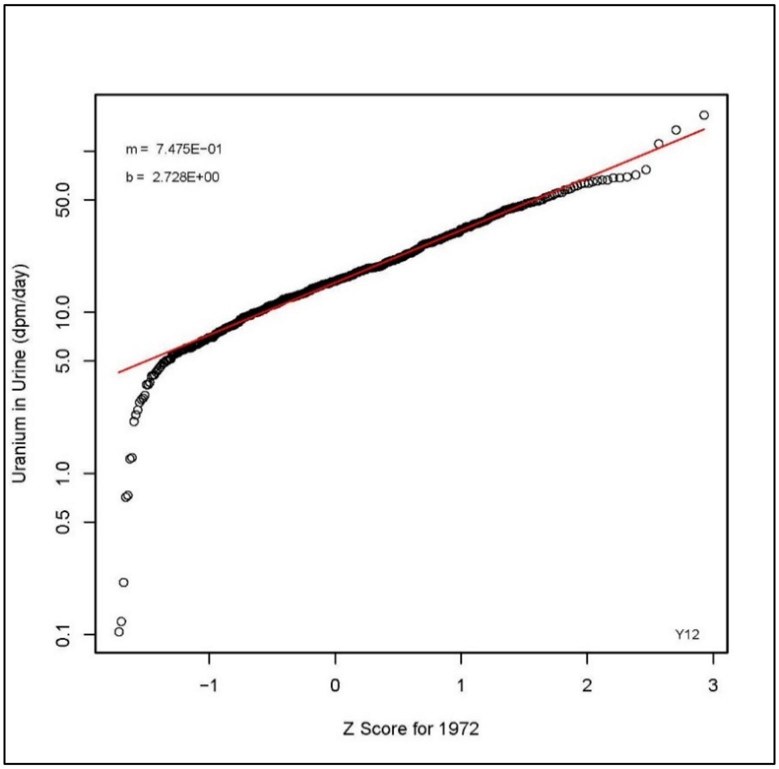
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



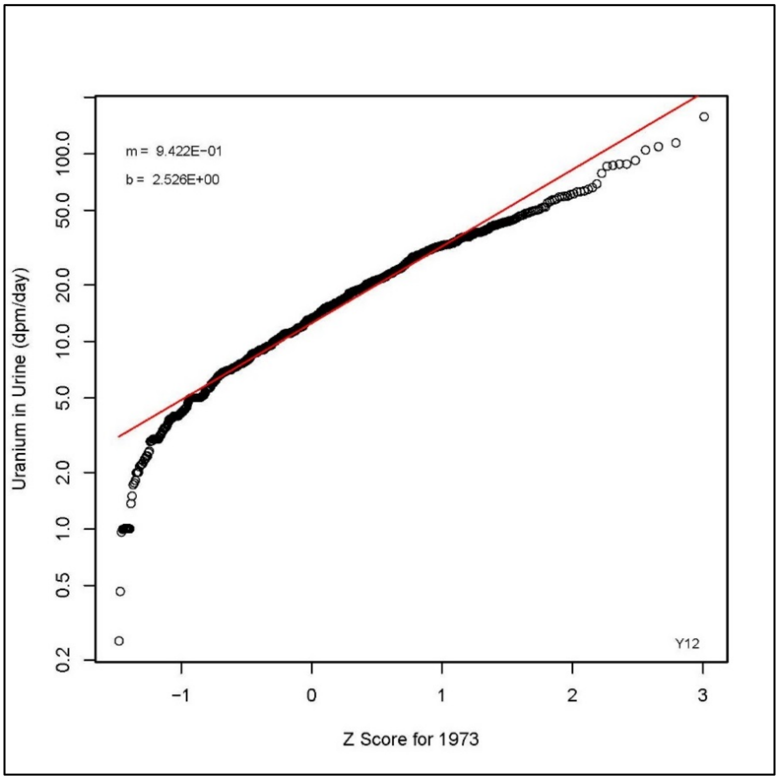


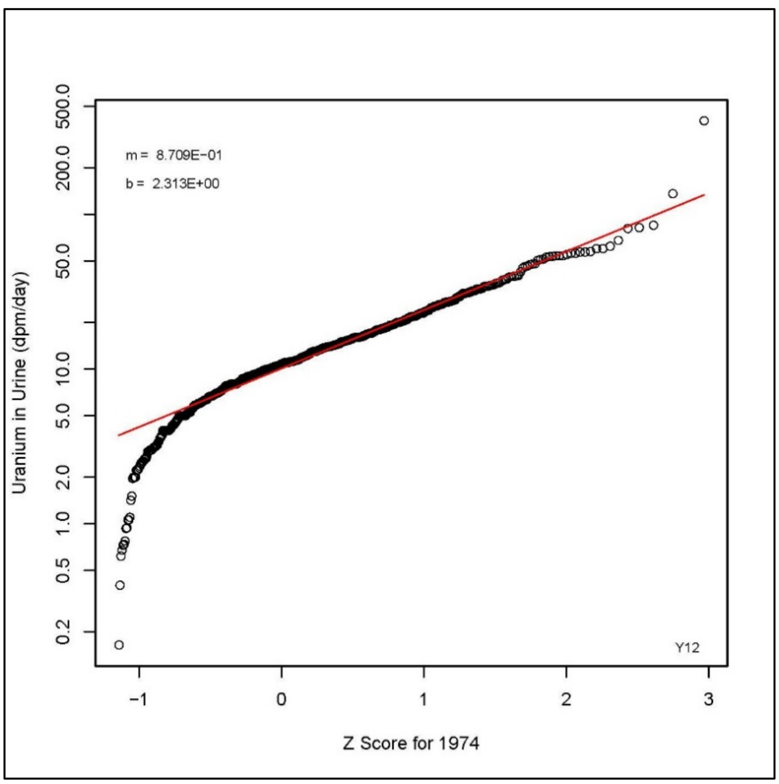
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



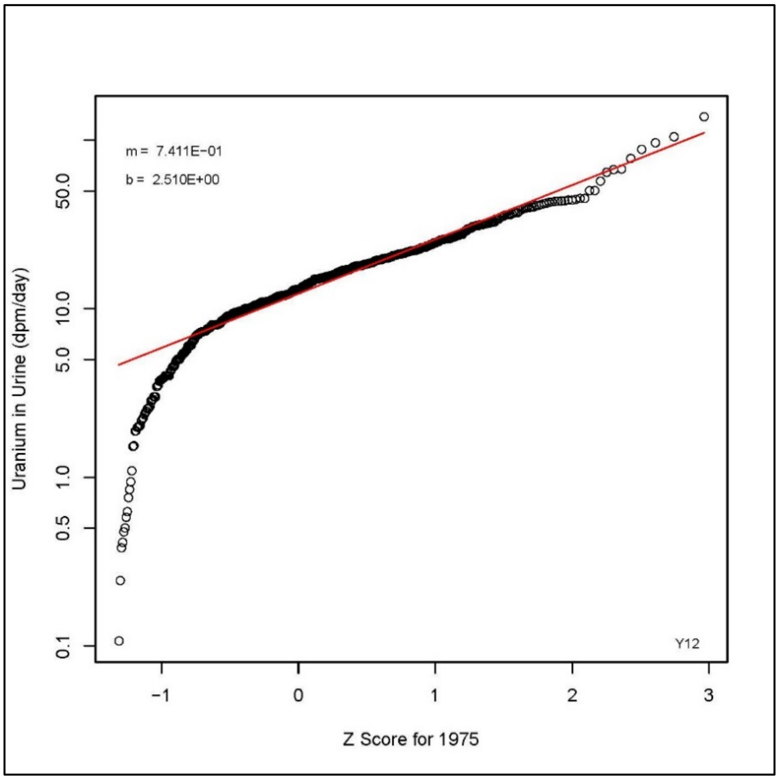


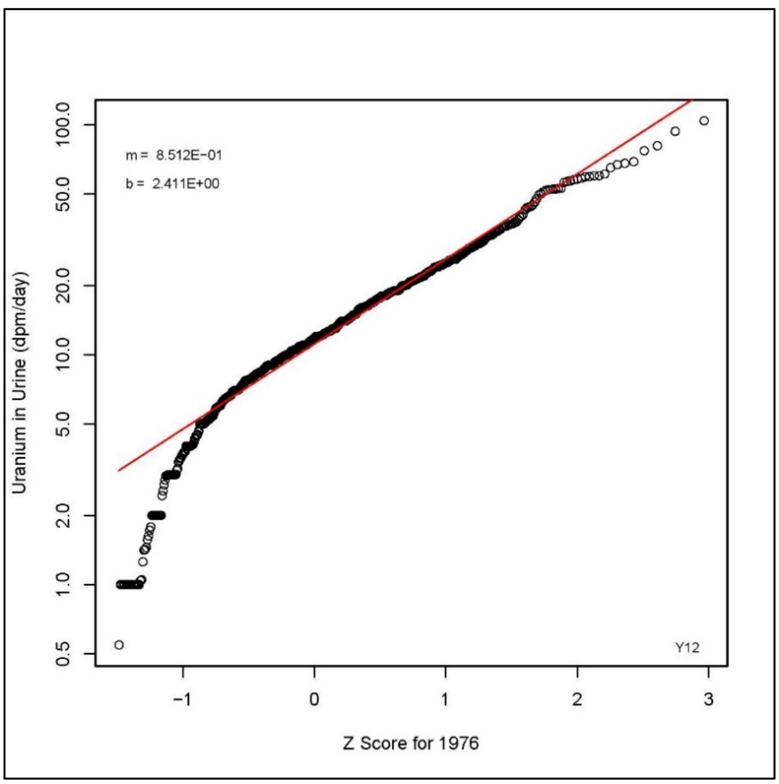
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



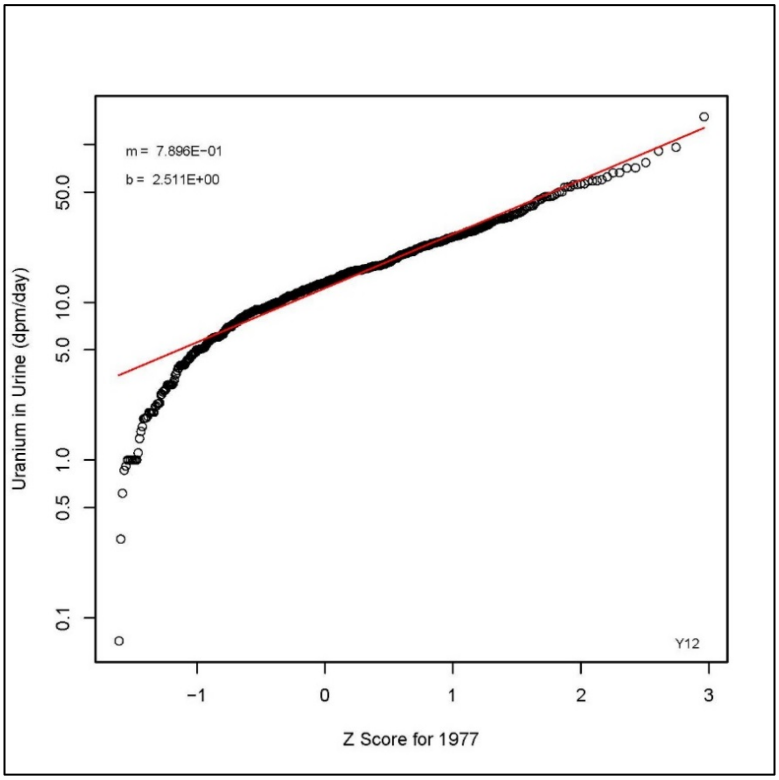


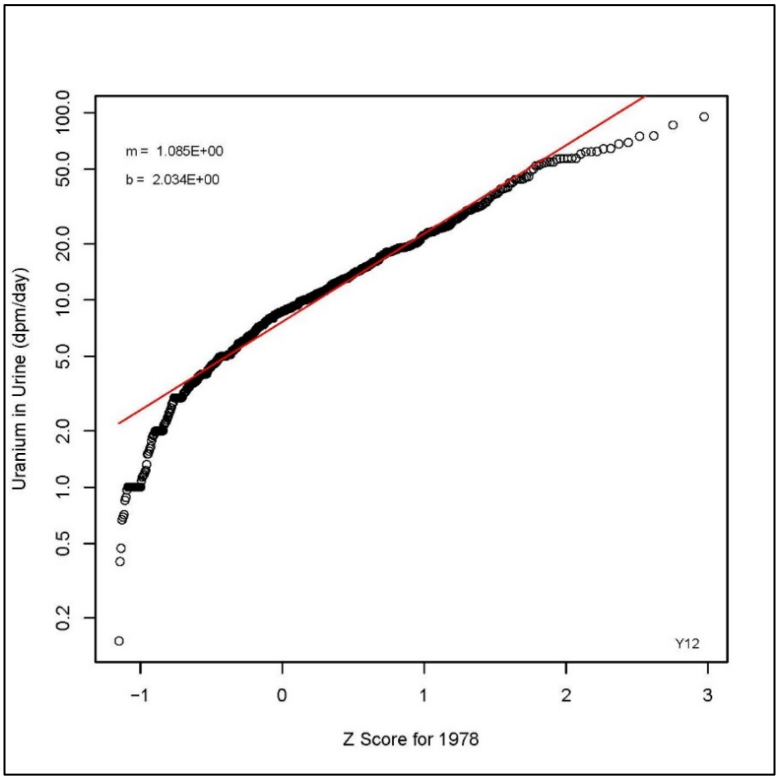
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



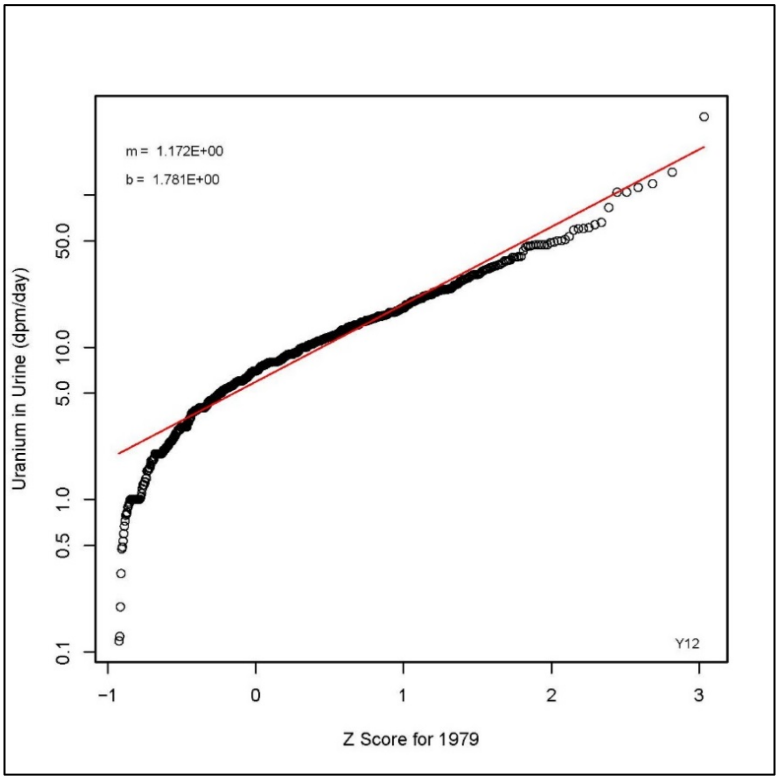


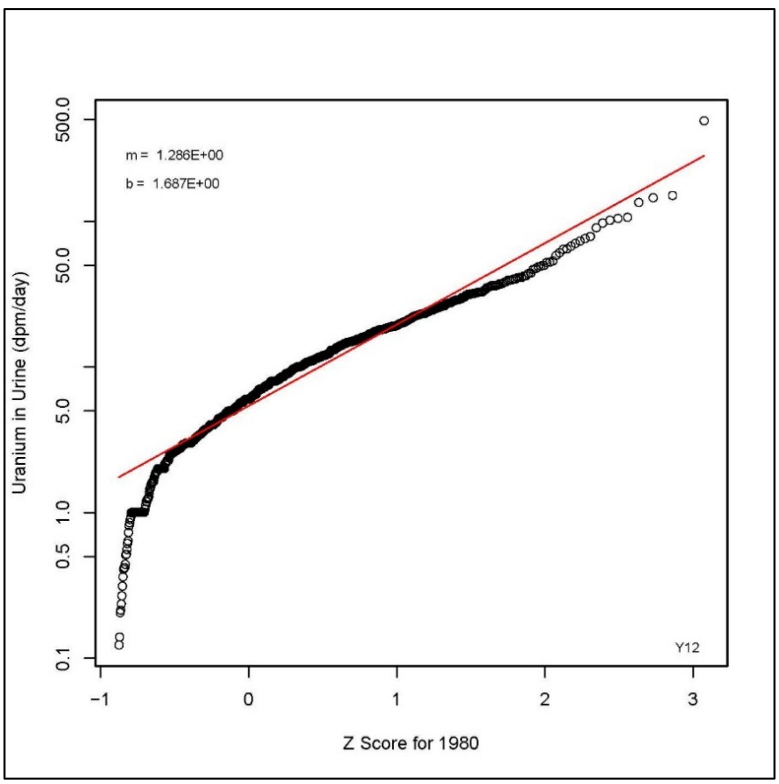
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



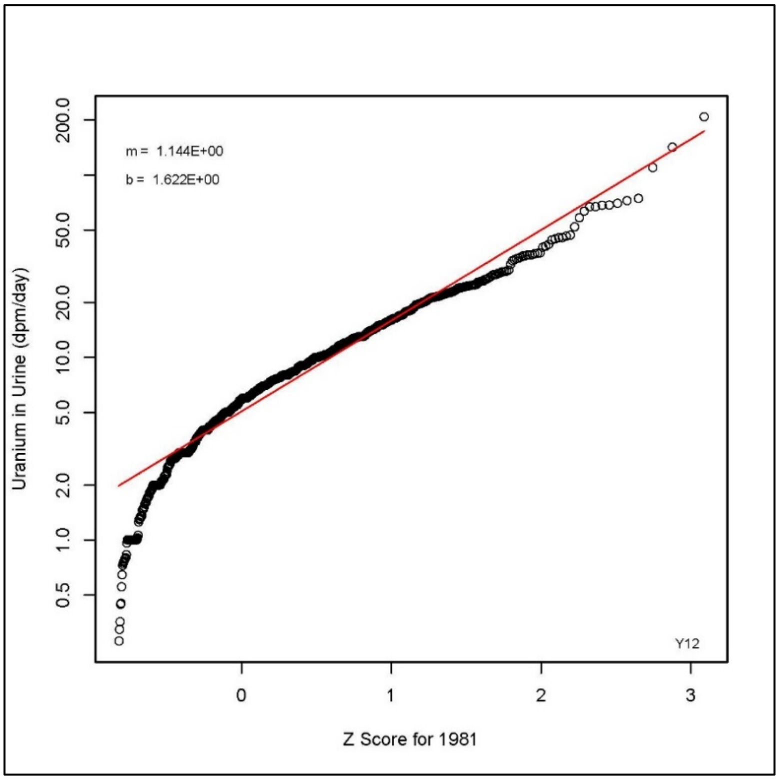


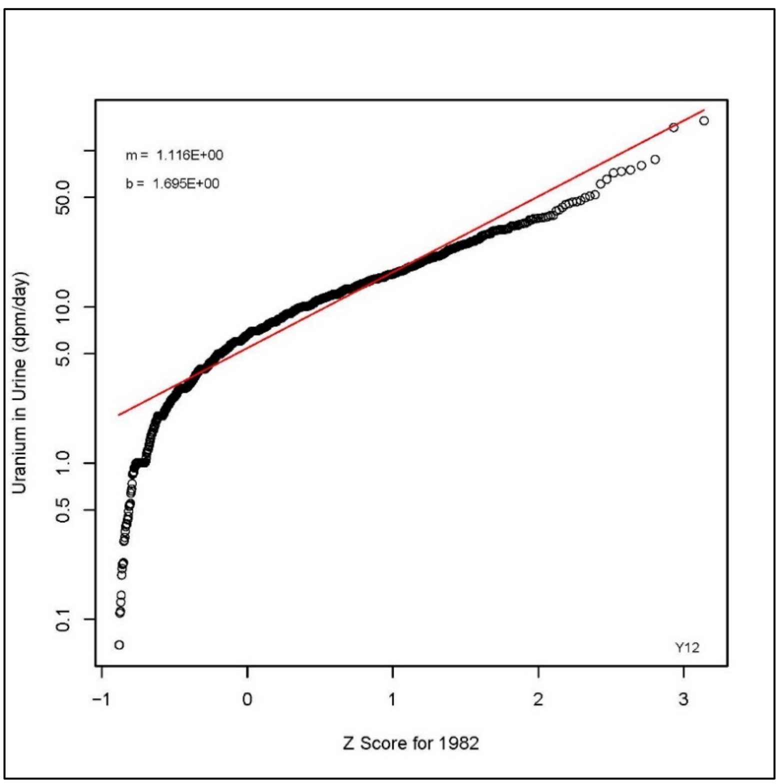
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



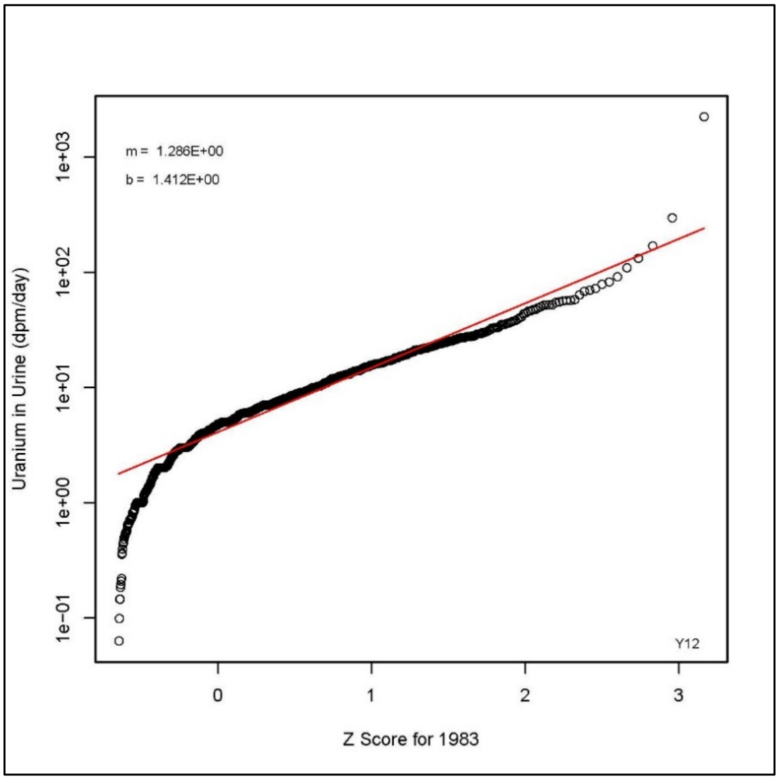


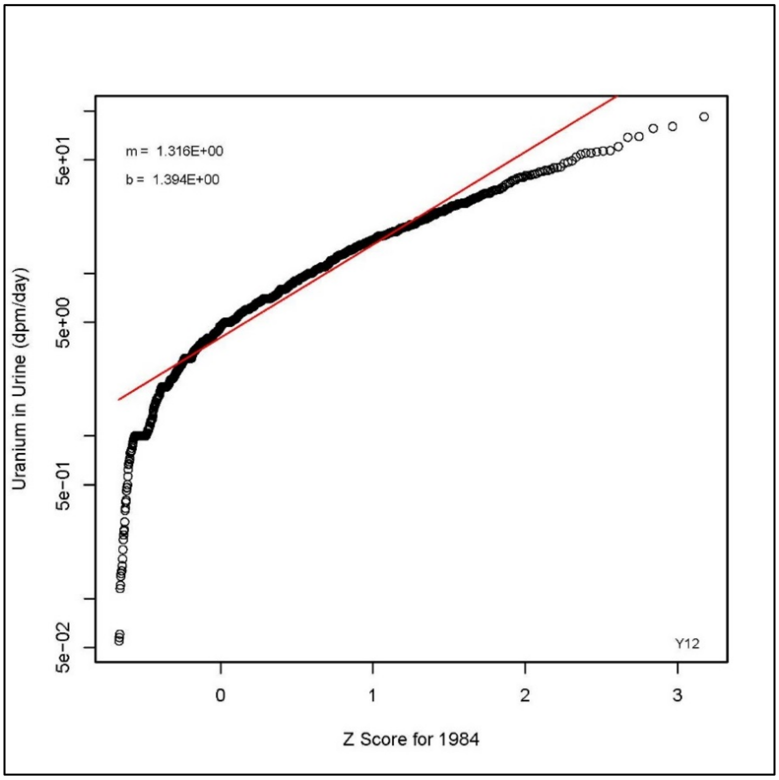
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



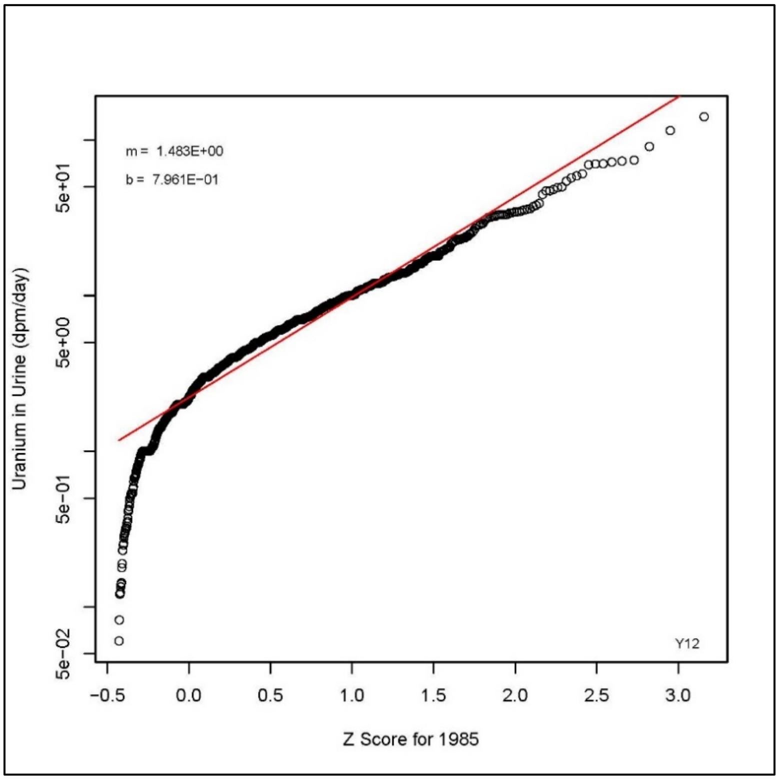


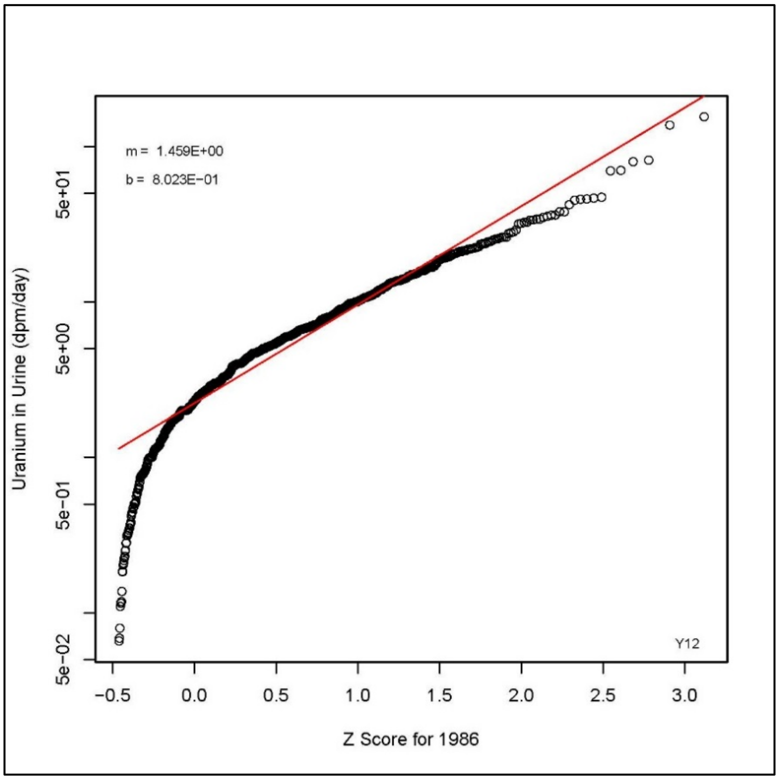
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



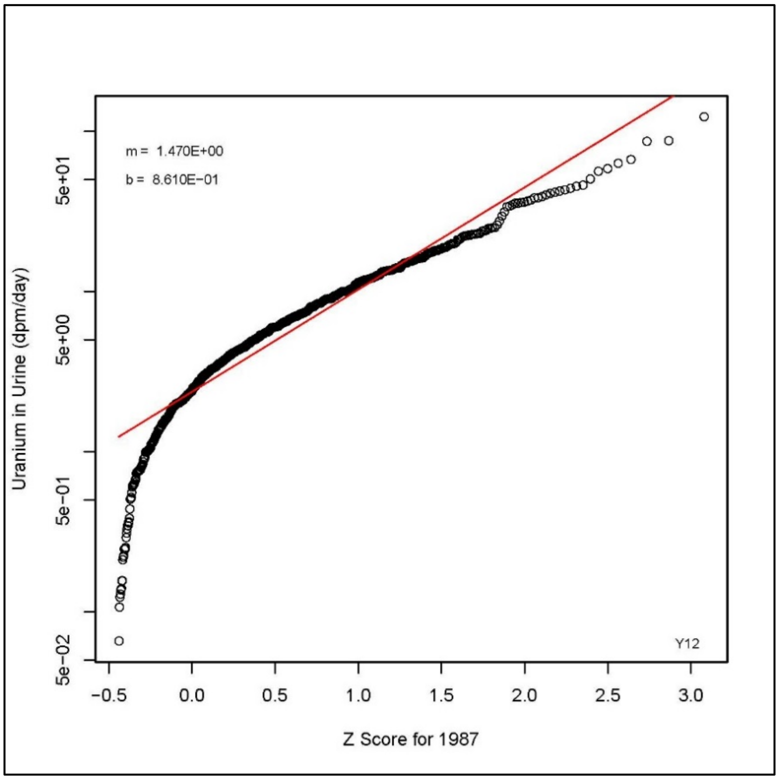


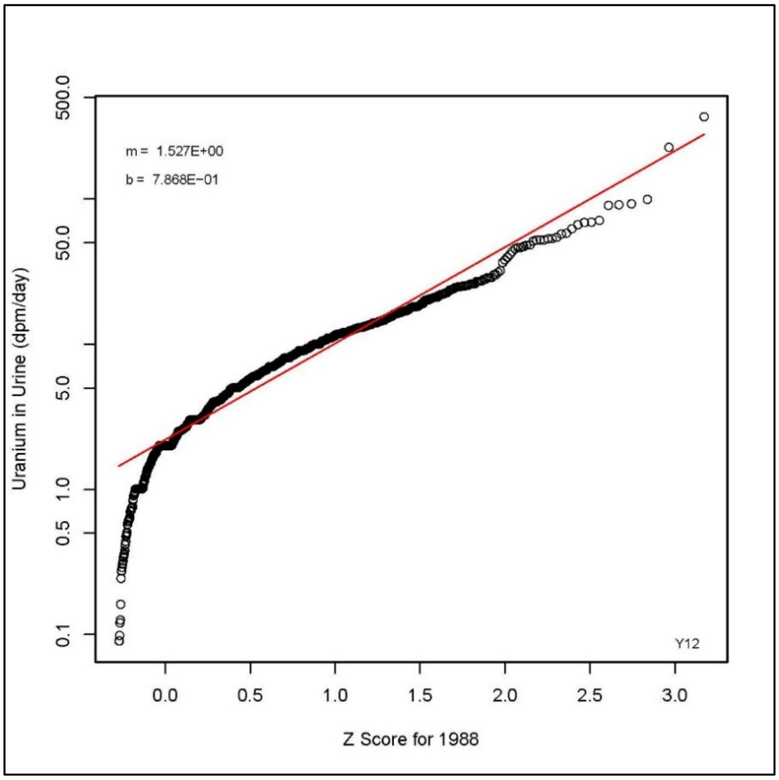
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)





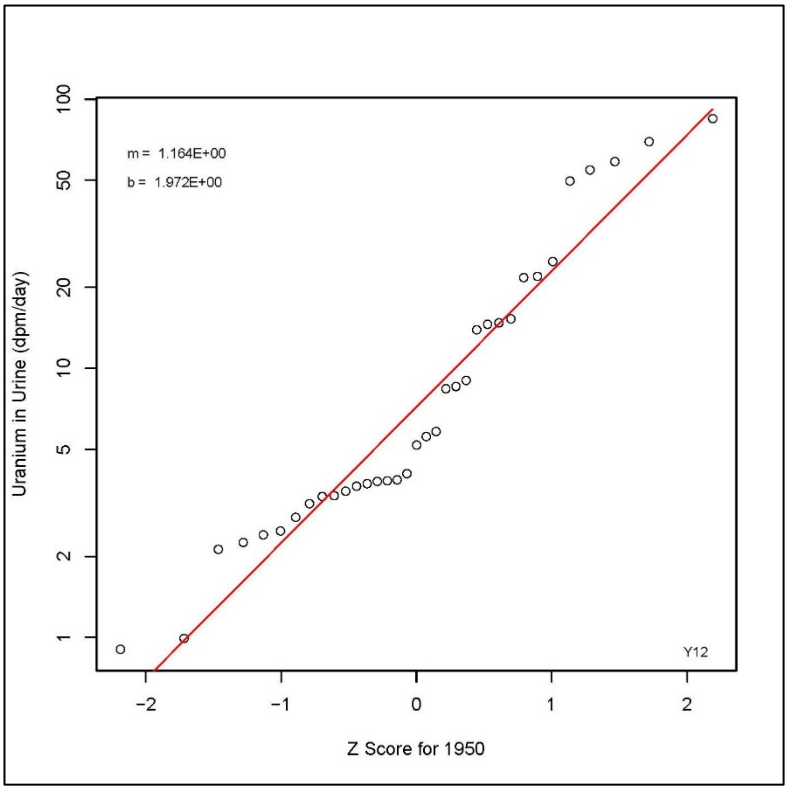
ATTACHMENT A
LOGNORMAL PROBABILITY PLOTS, COMPLETE Y-12 DATASET (continued)



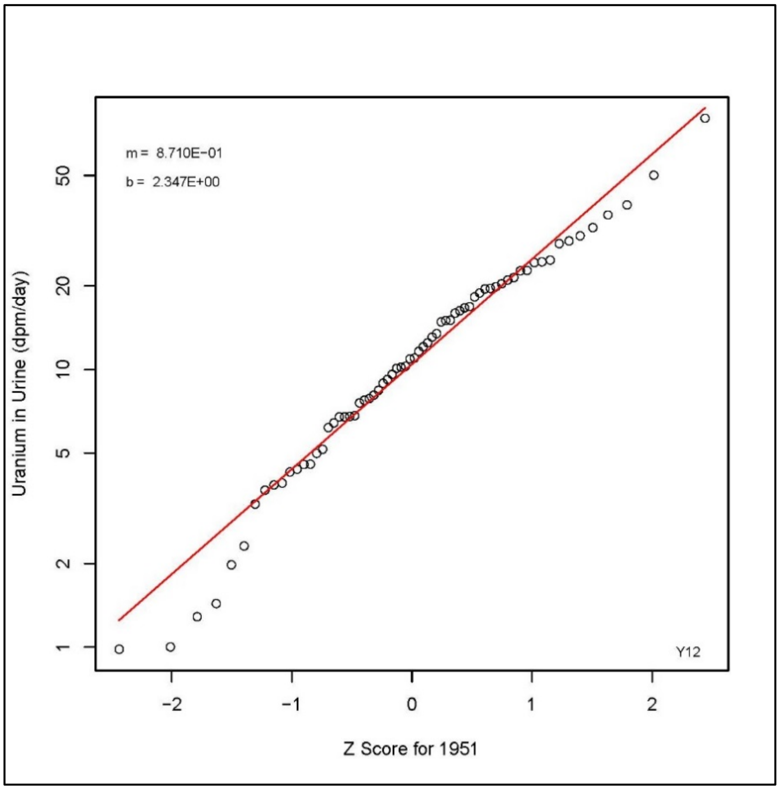


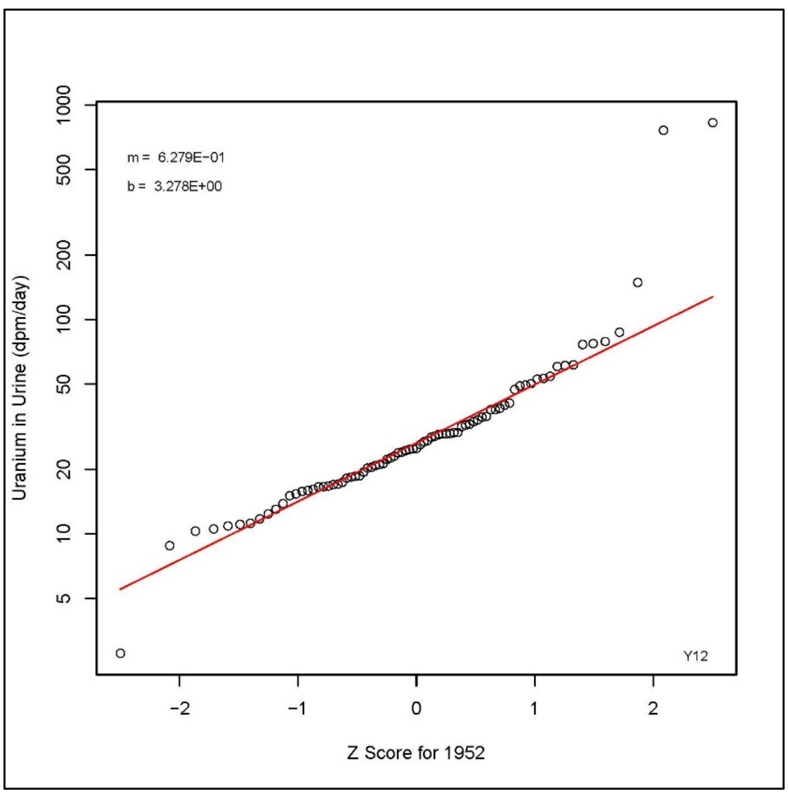
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET

This attachment presents lognormal probability plots for the claimant Y-12 uranium urine bioassay dataset for 1950 to 1988. The slope and intercept for the line of best fit to the data are given on the plot.

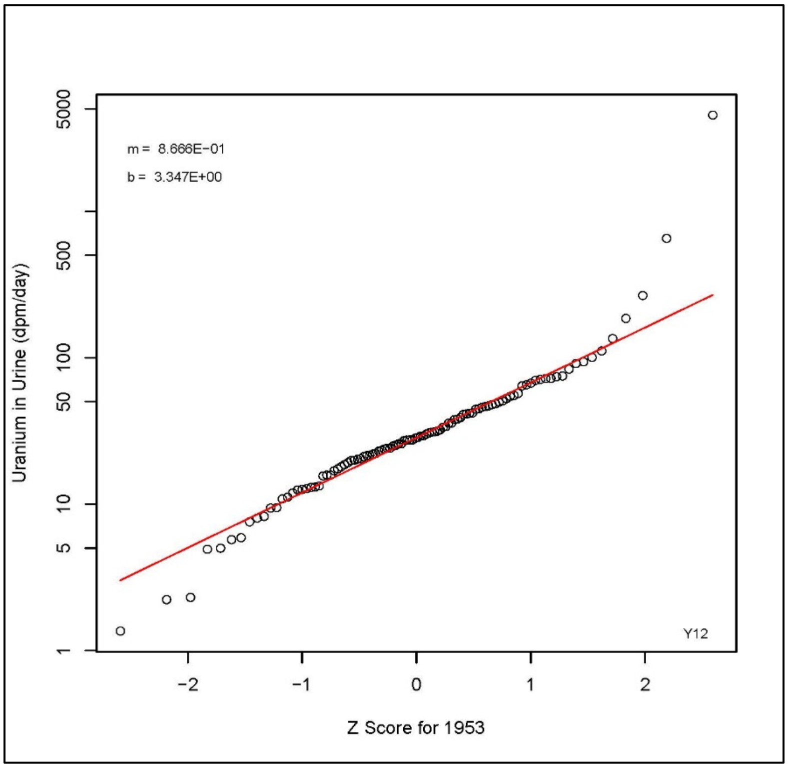


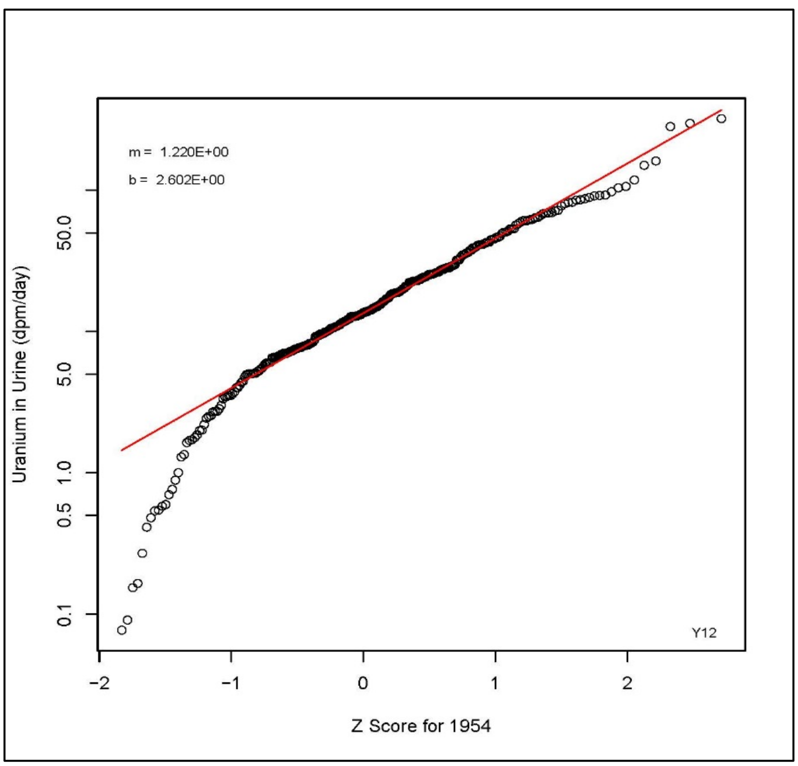
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



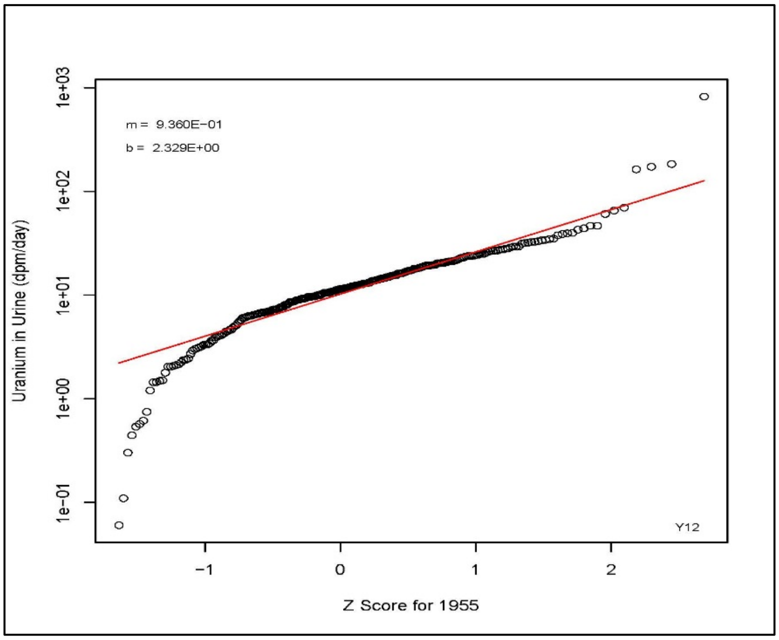


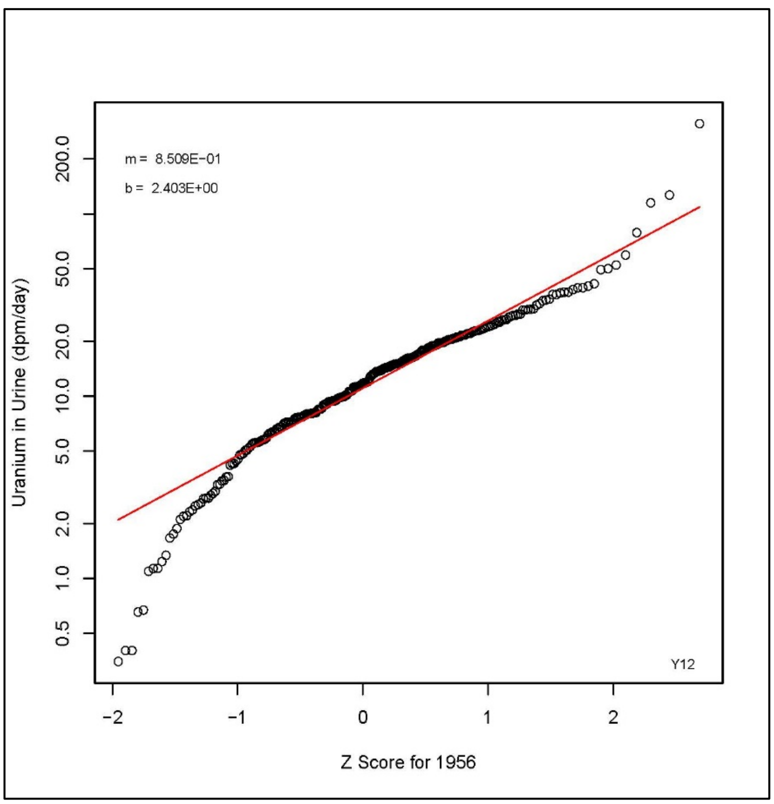
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



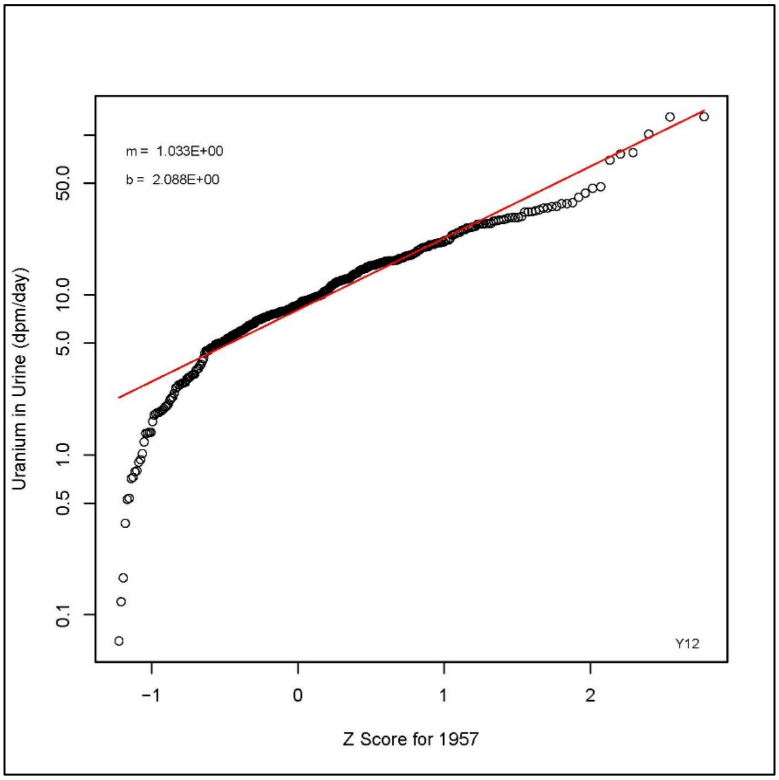


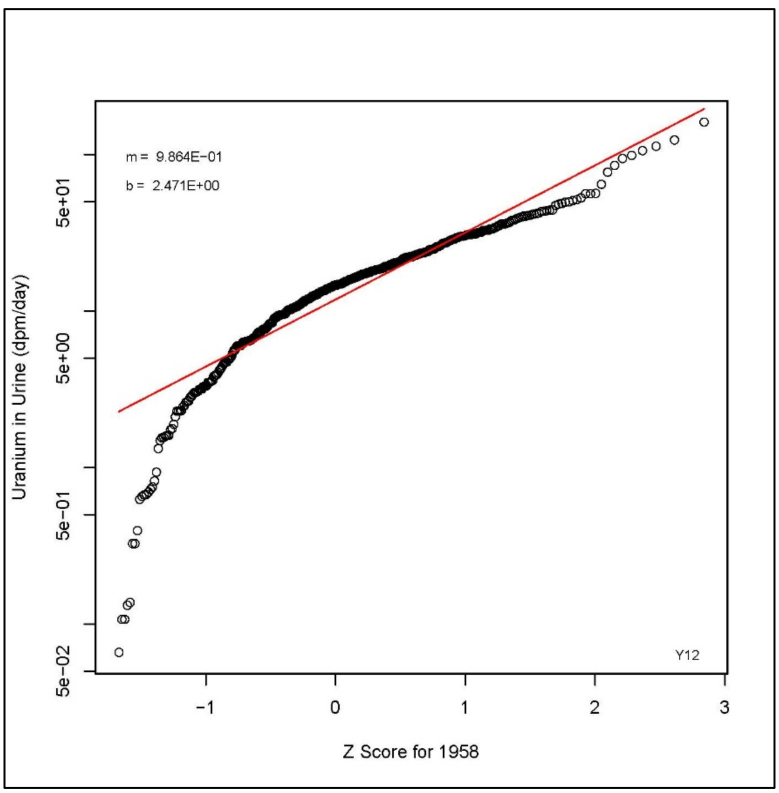
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



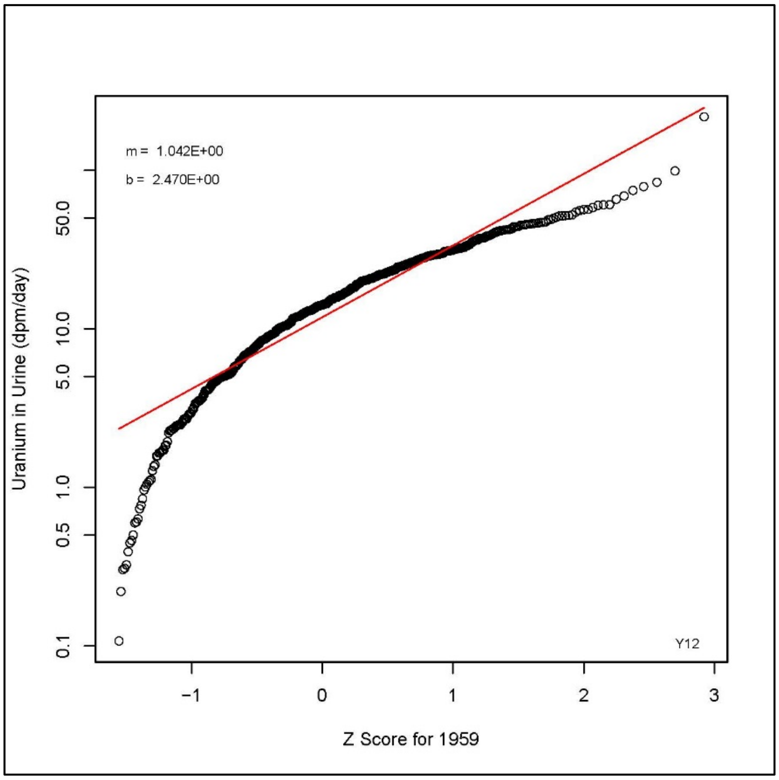


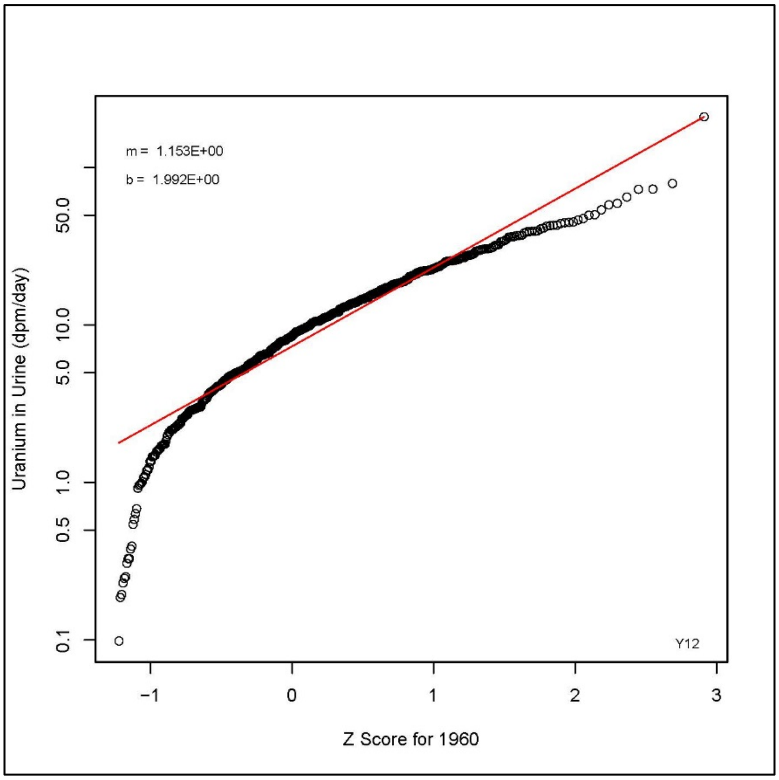
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



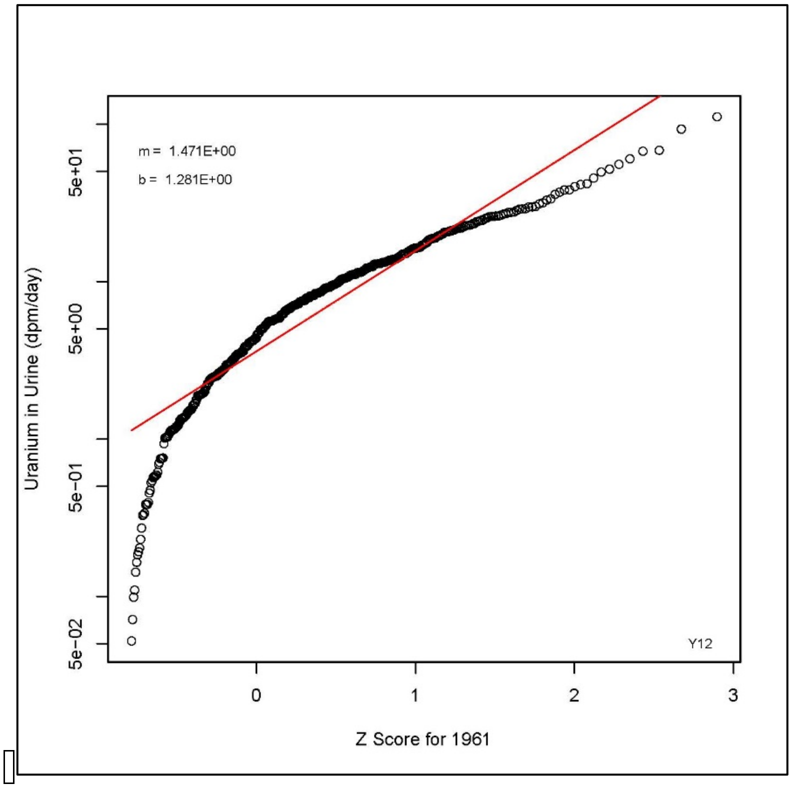


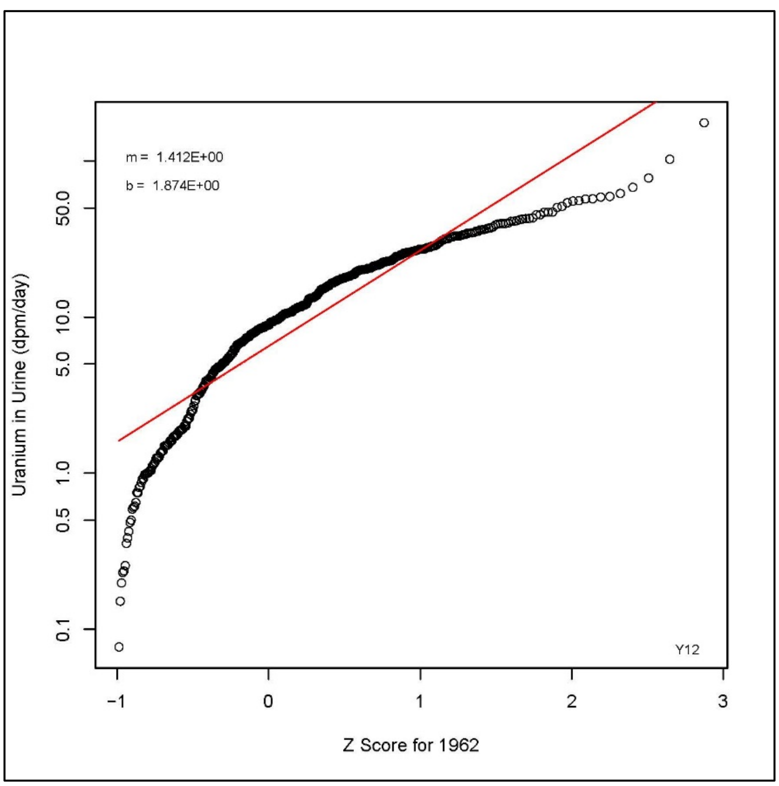
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



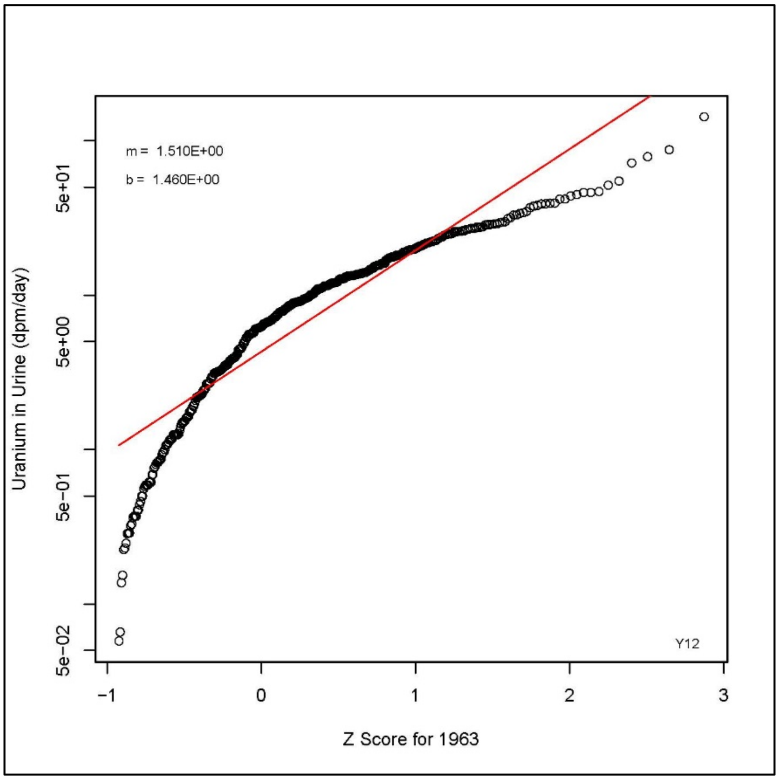


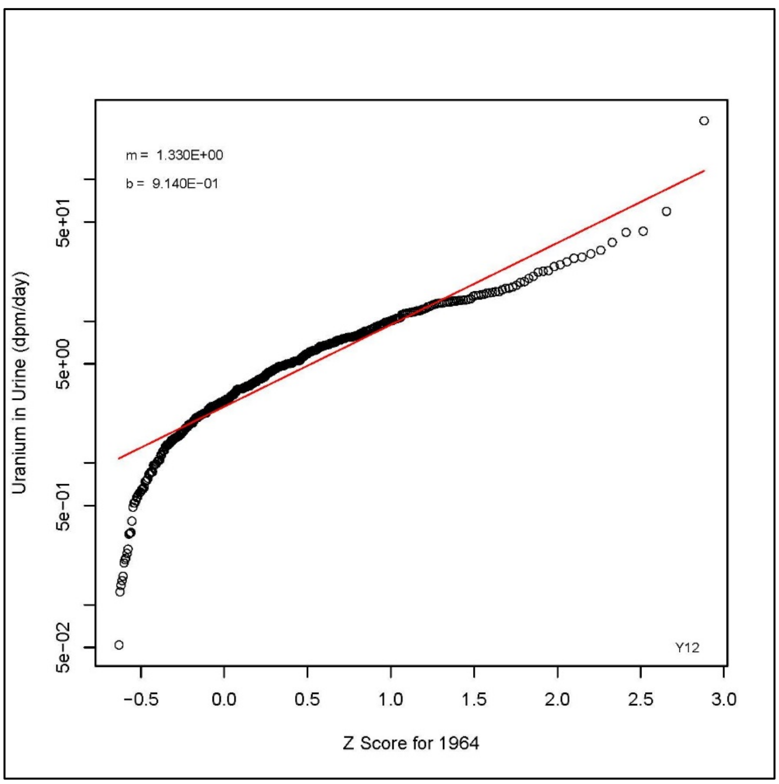
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



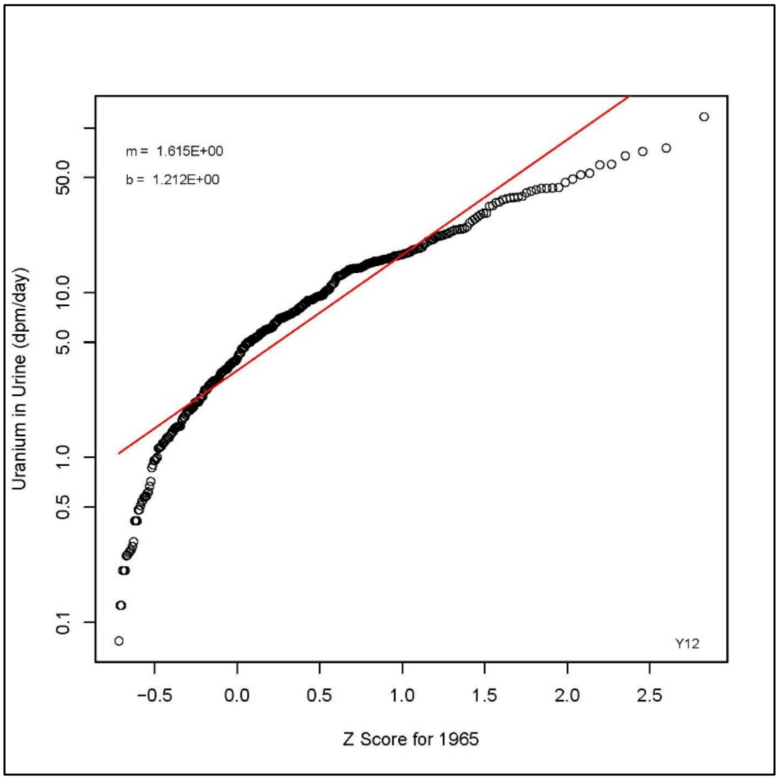


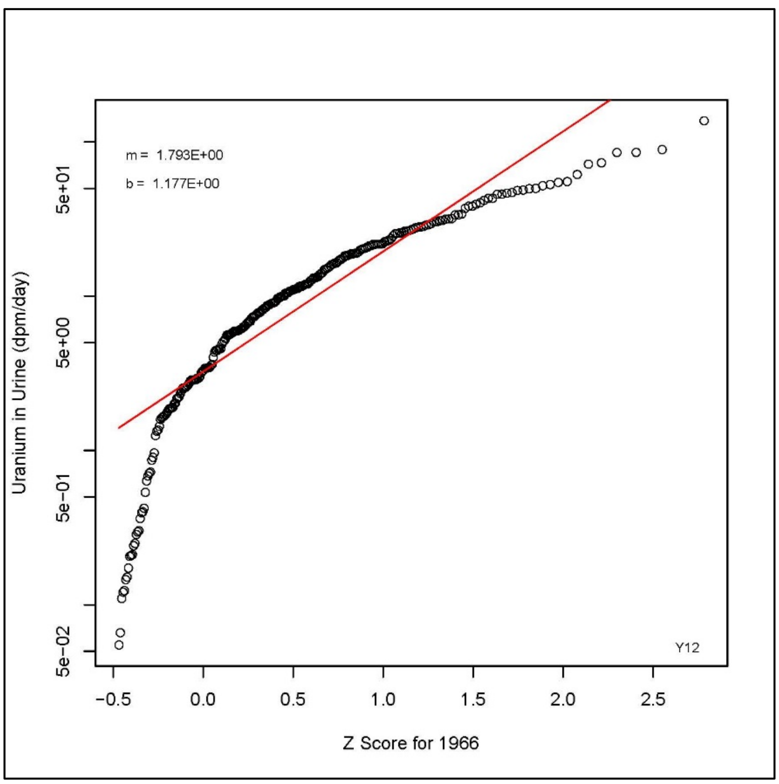
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



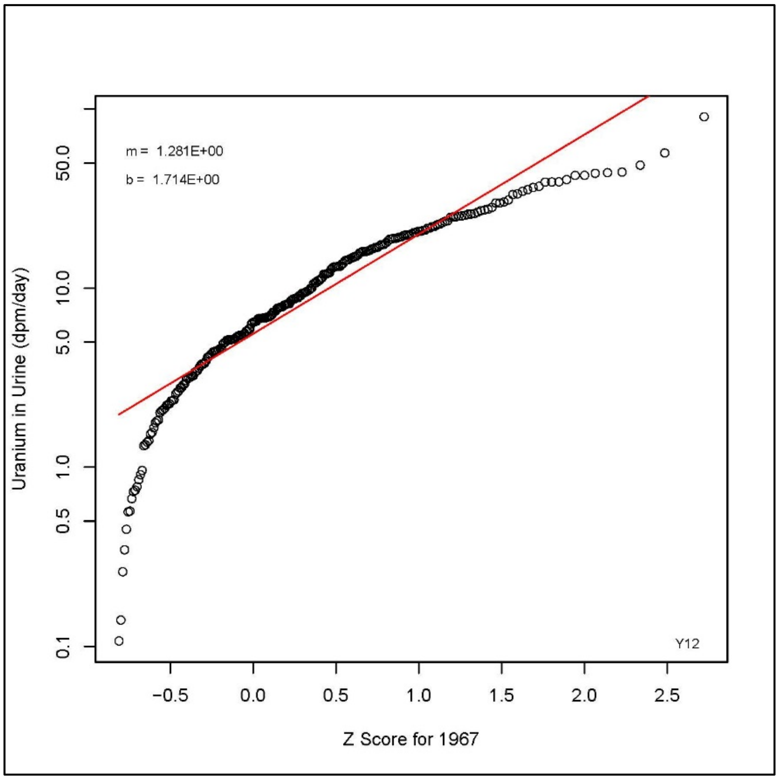


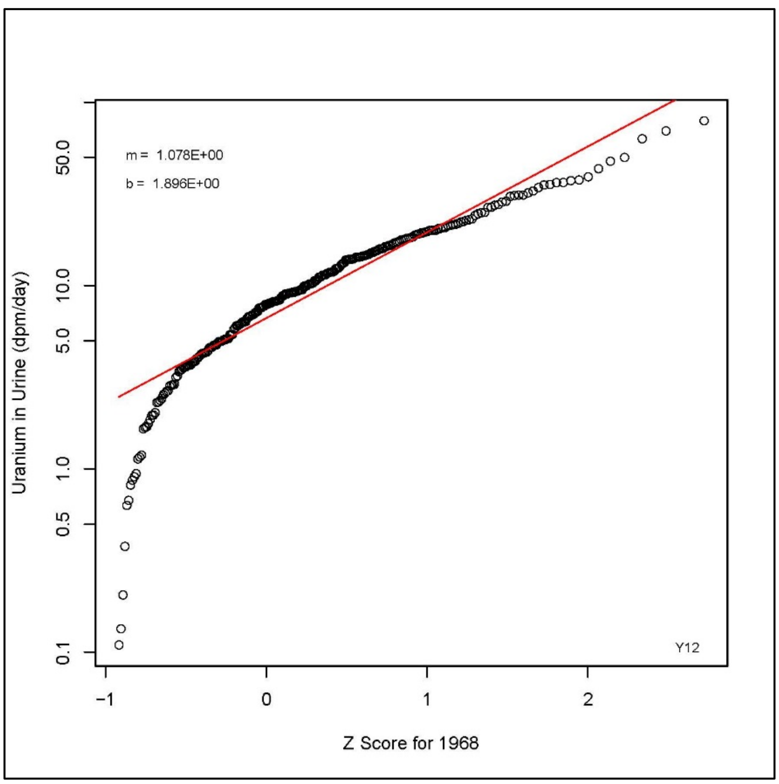
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



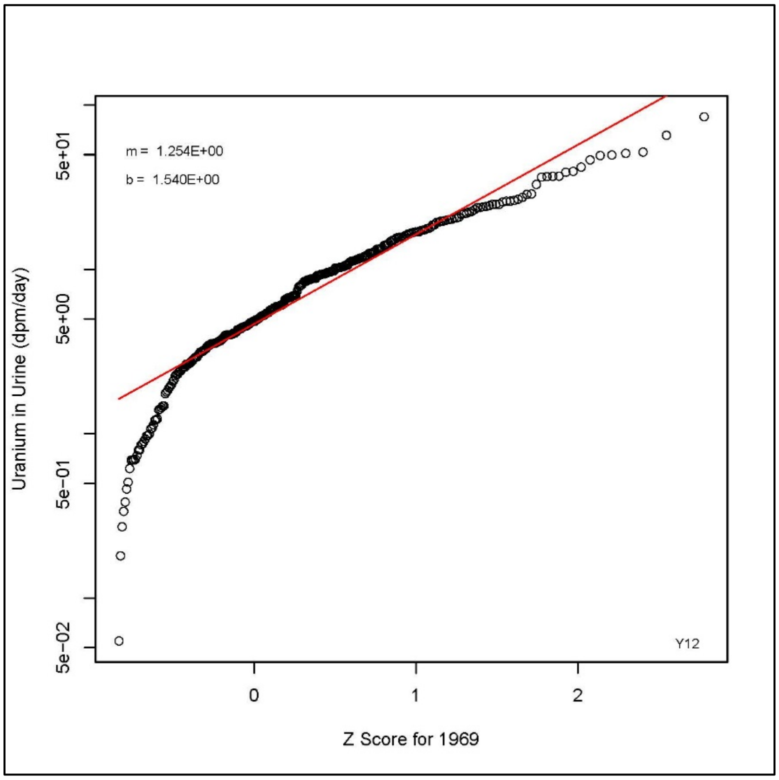


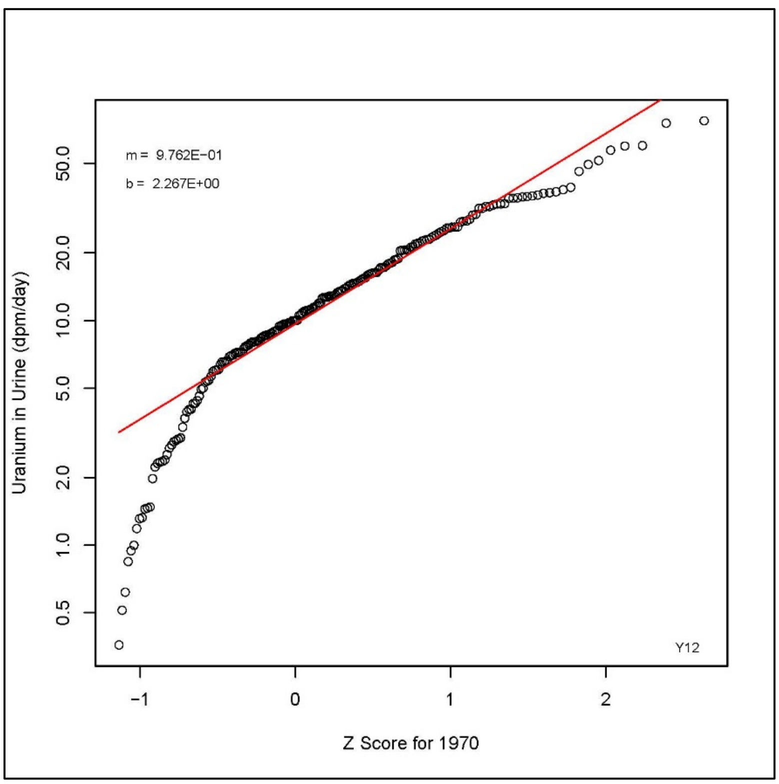
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



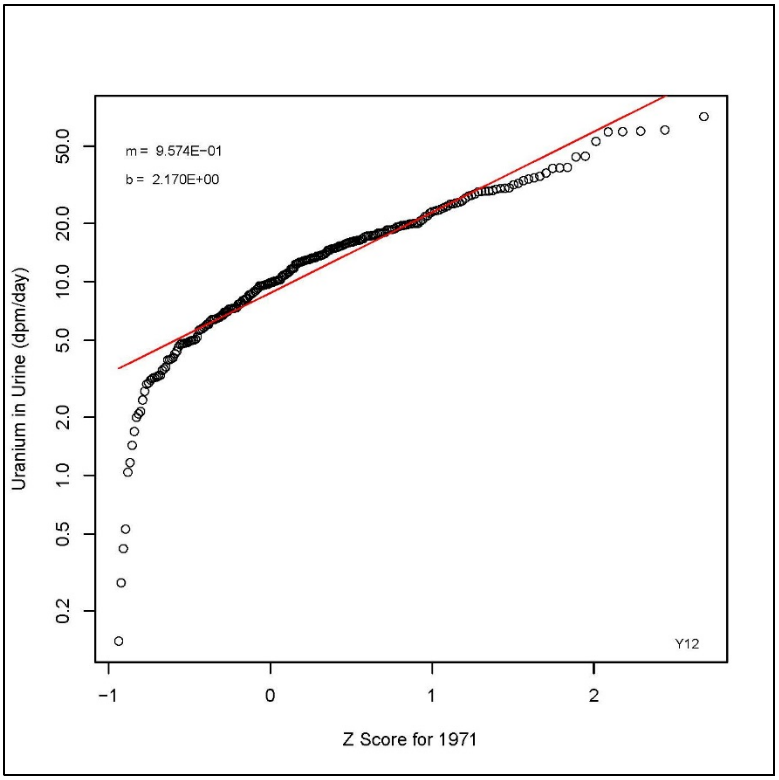


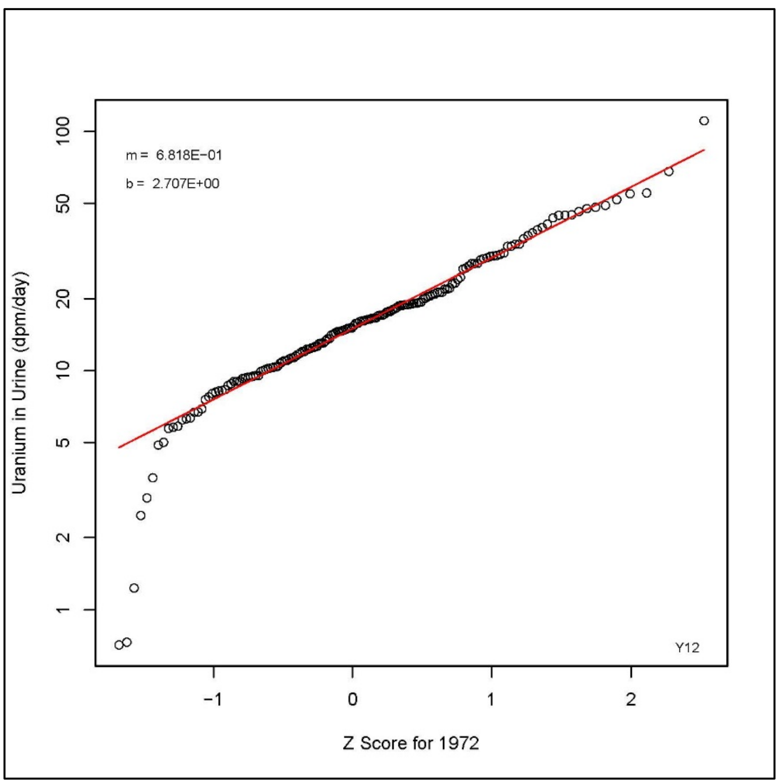
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



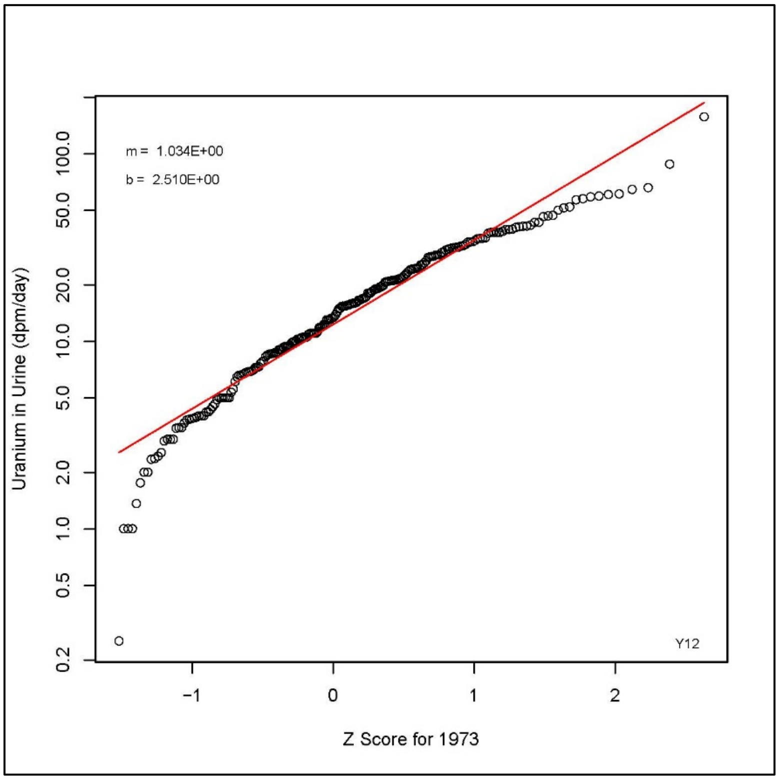


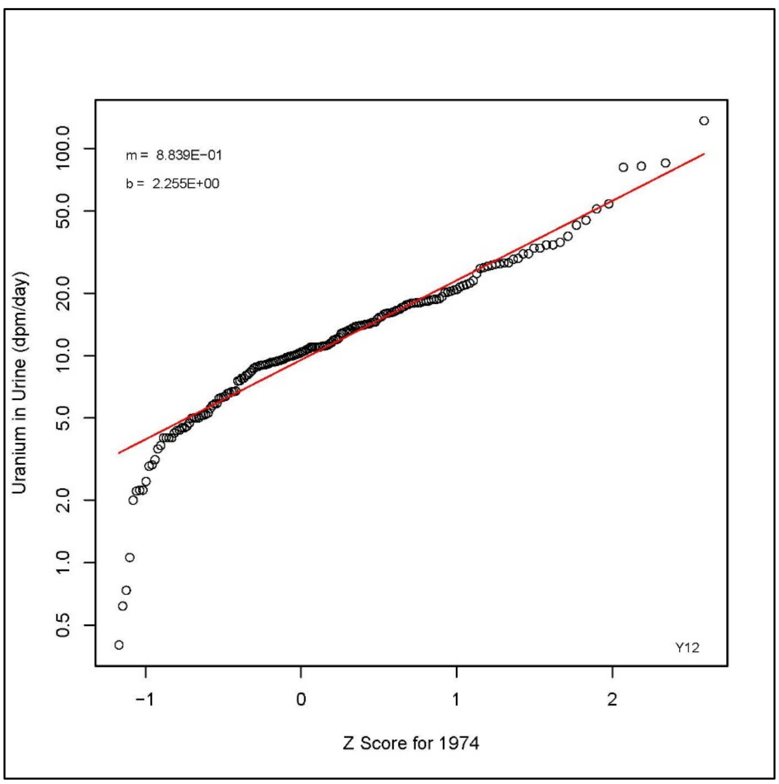
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



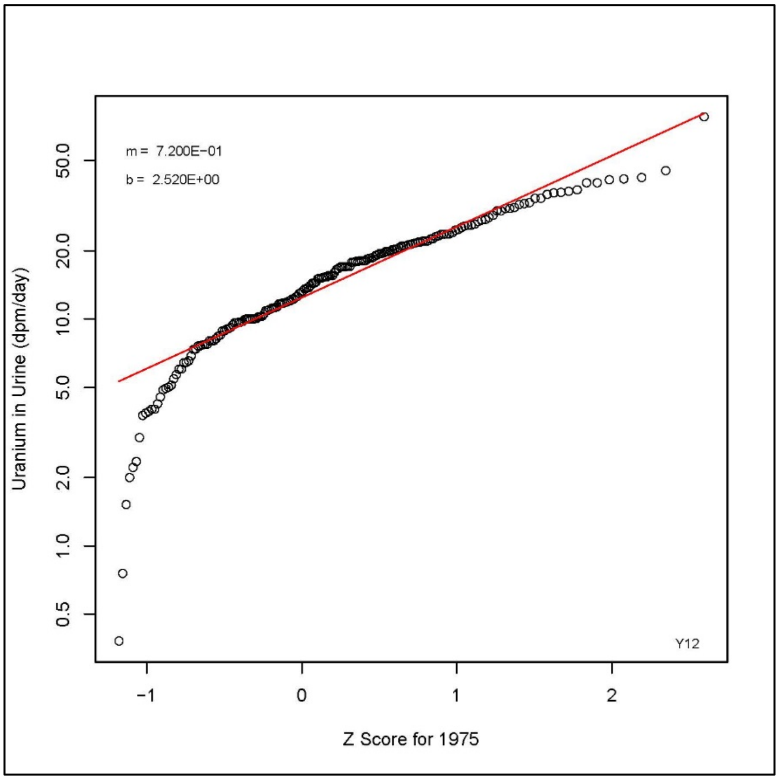


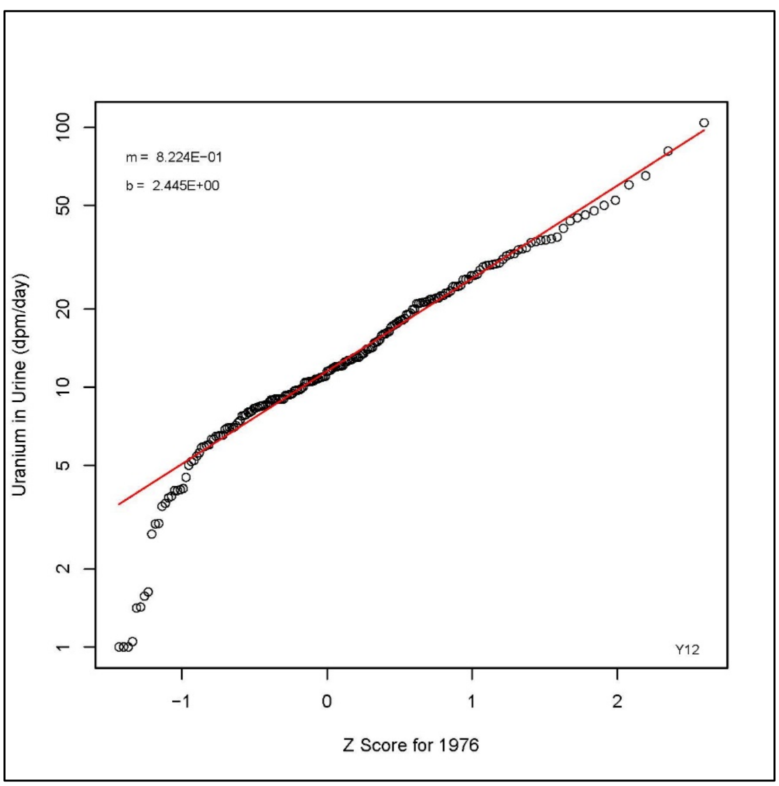
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



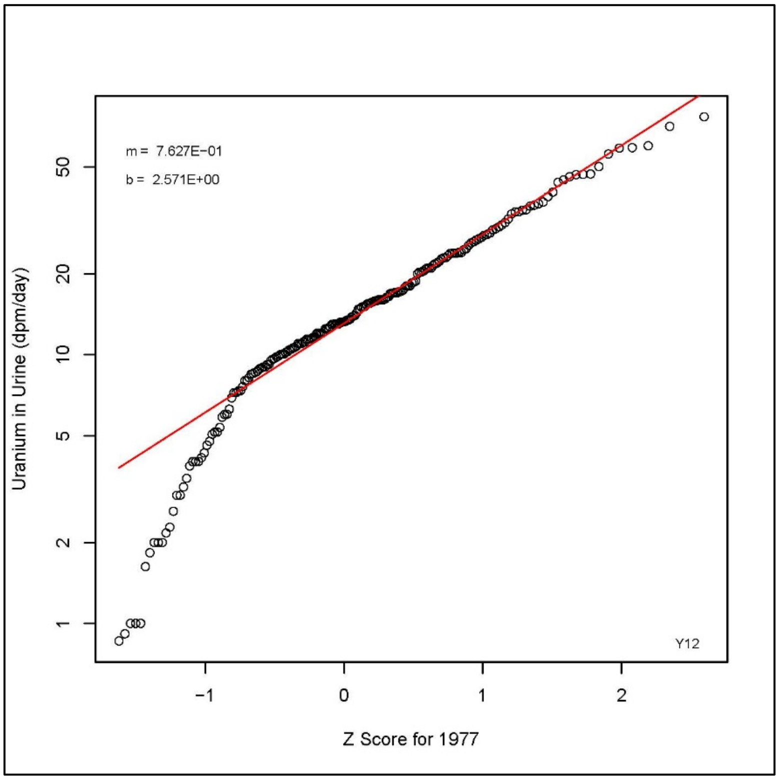


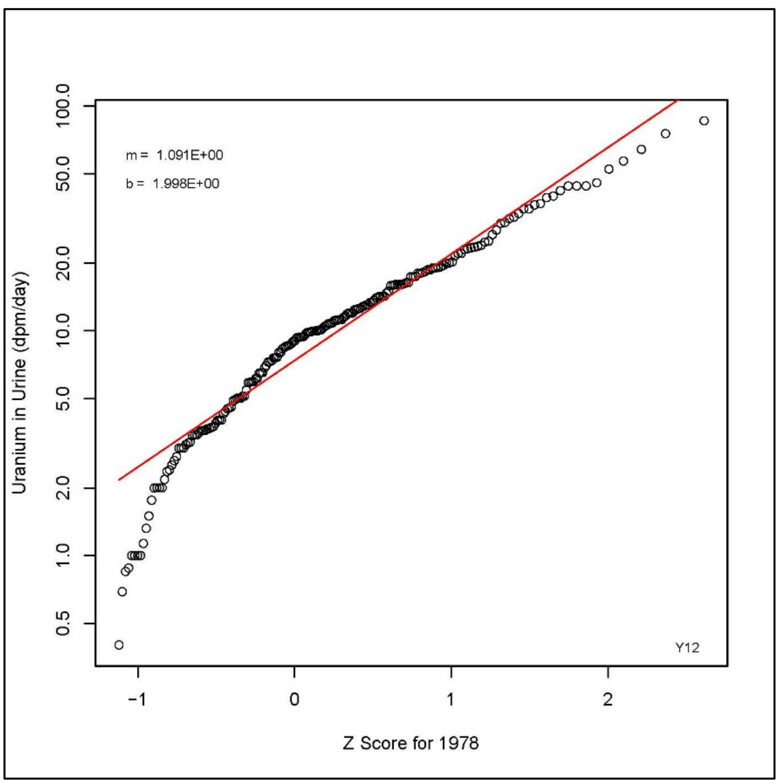
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



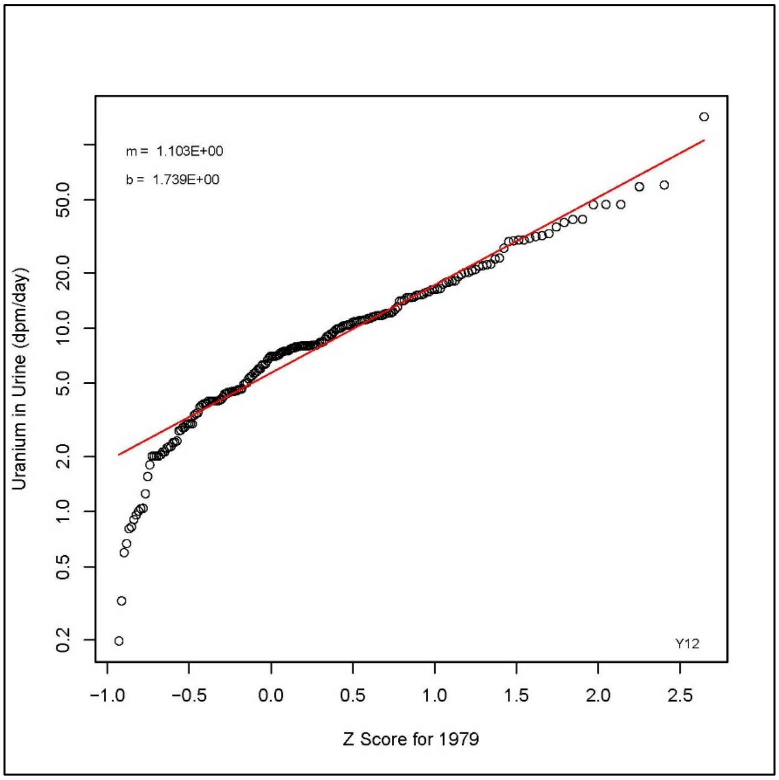


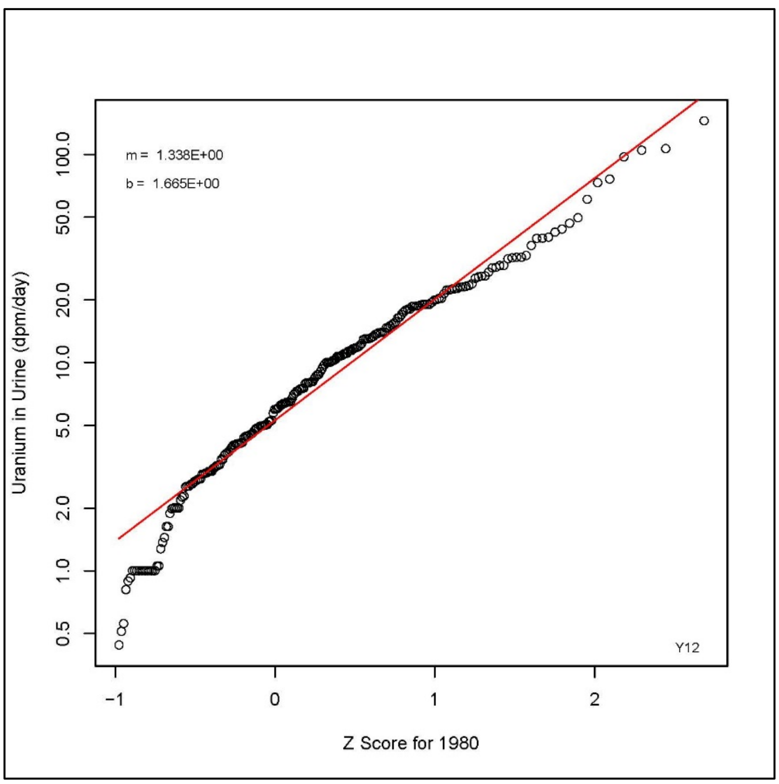
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



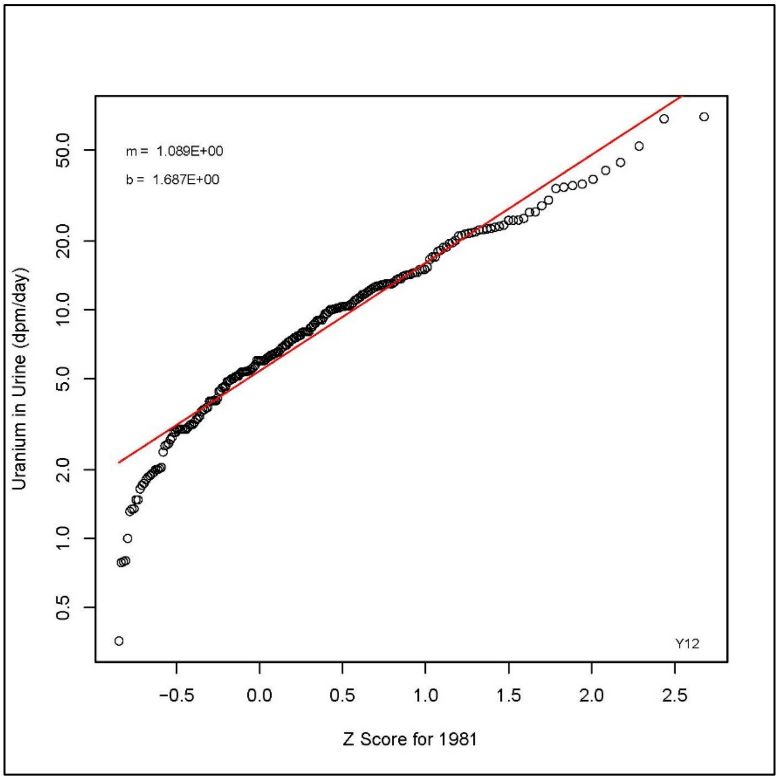


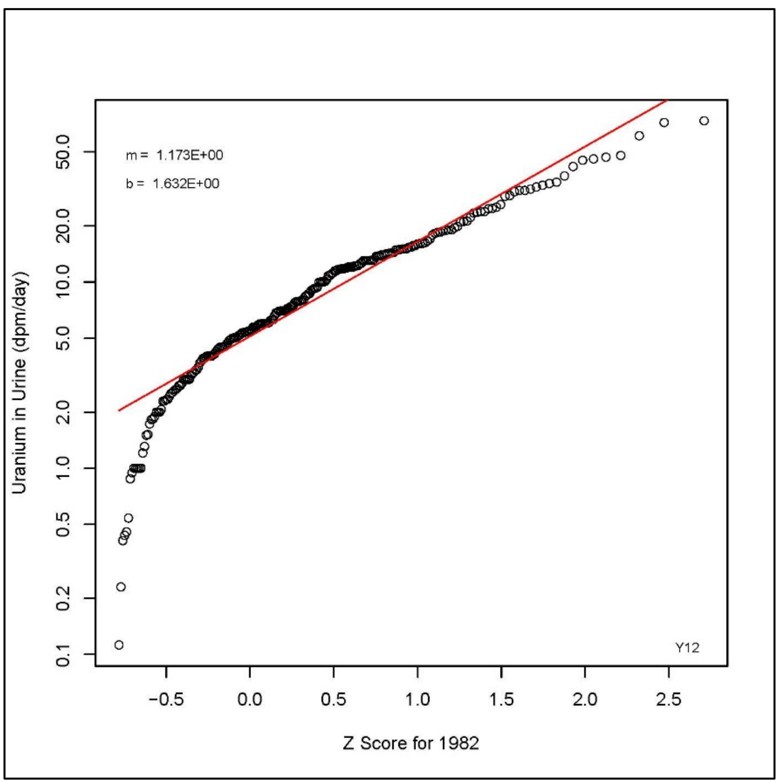
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



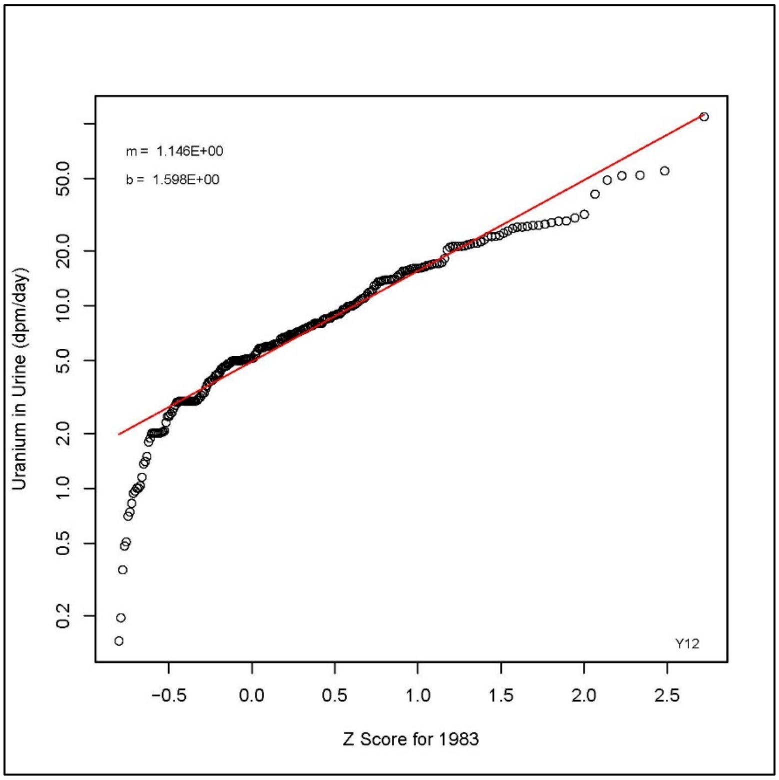


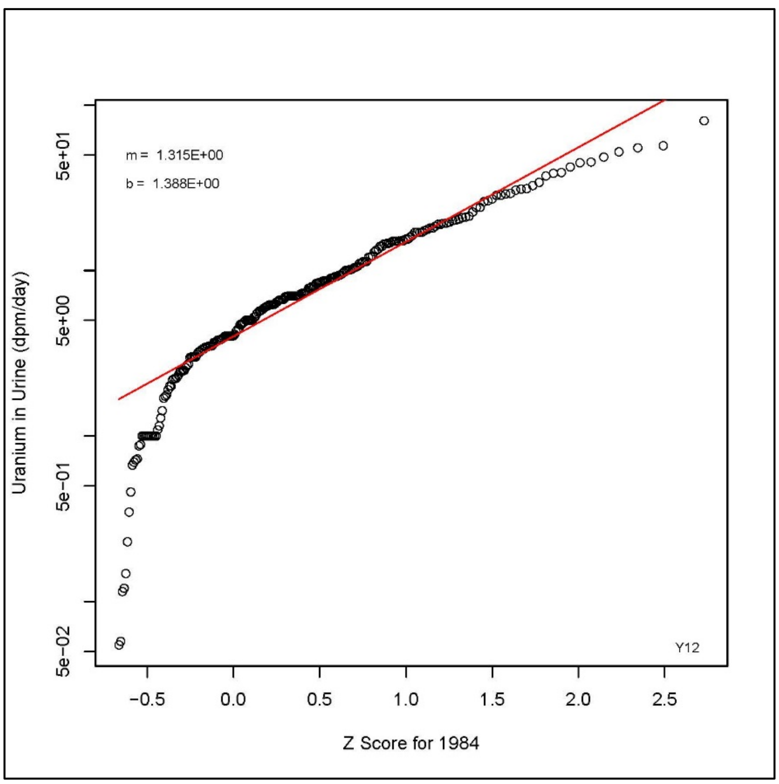
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



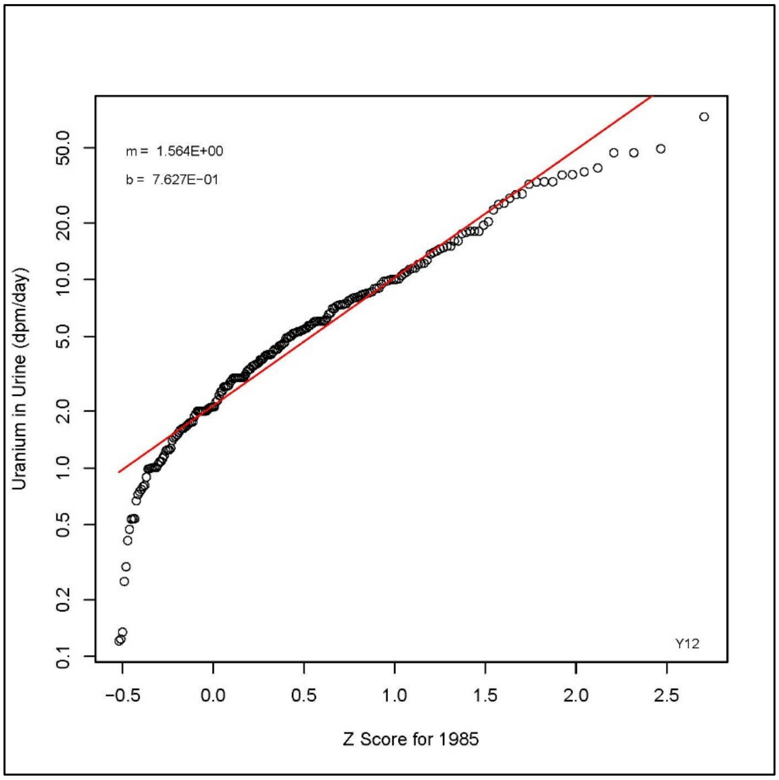


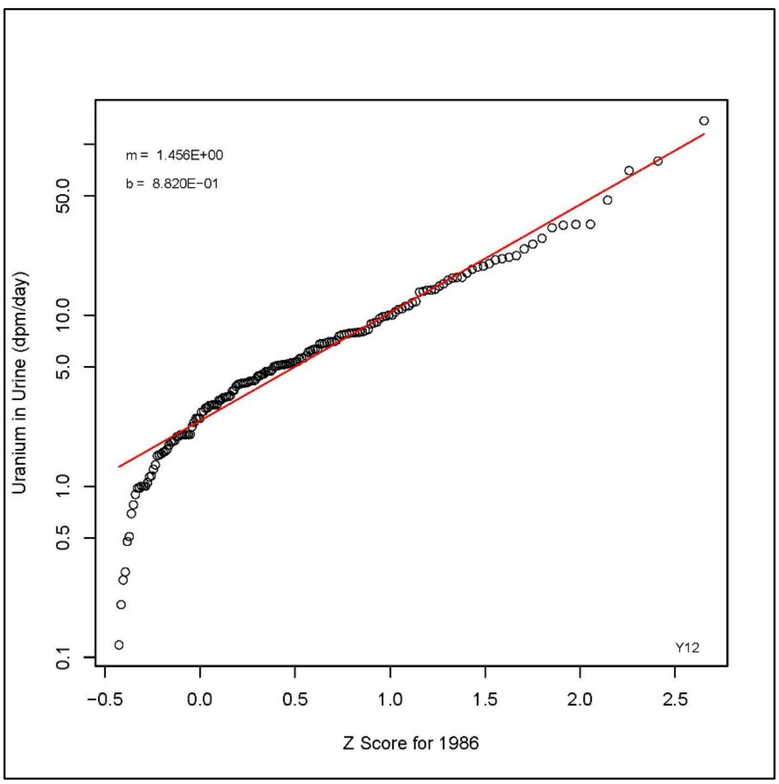
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



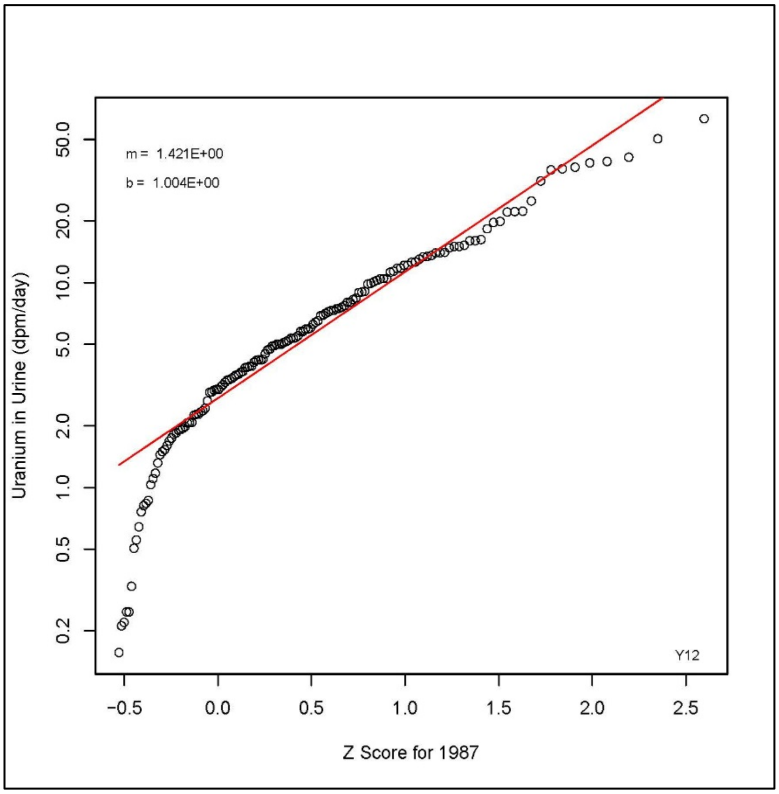


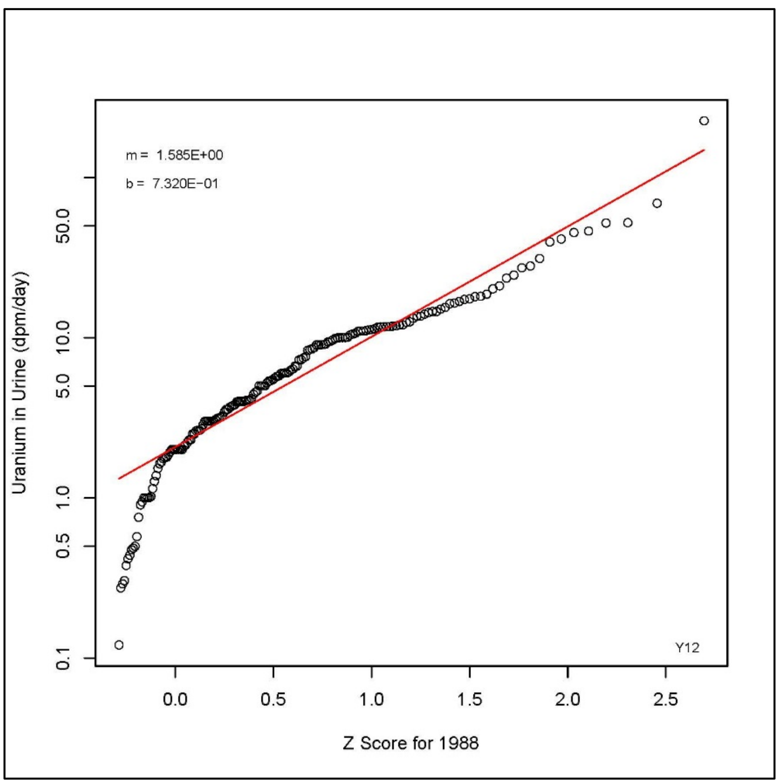
ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)





ATTACHMENT B
LOGNORMAL PROBABILITY PLOTS, CLAIMANT Y-12 DATASET (continued)



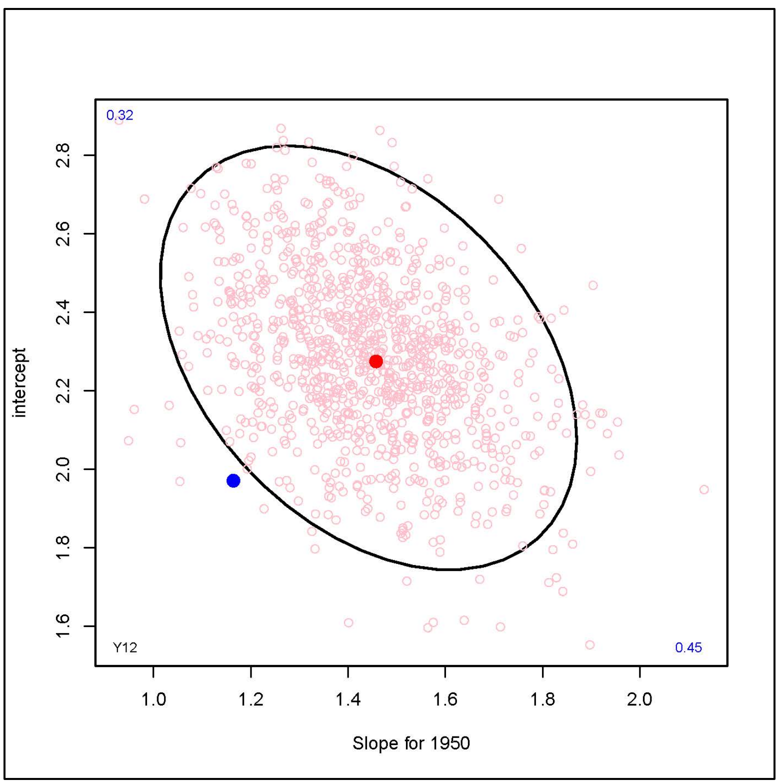


ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS

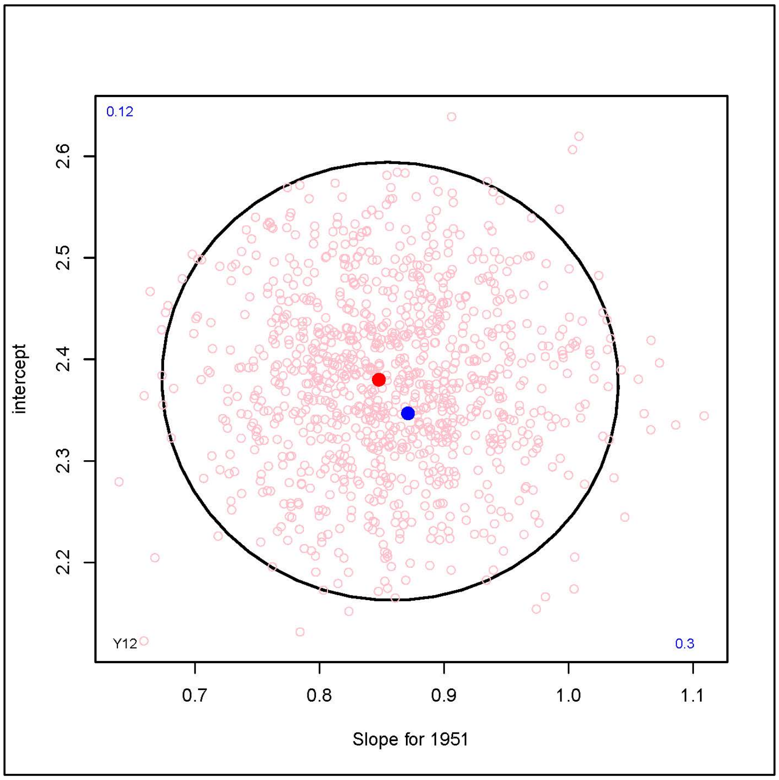
This attachment presents lognormal probability plots for a sample of 1585 unique workers that was drawn at random without replacement from the 7,537 workers in the complete dataset. The slope and intercept of the plot for each year from 1950 to 1988 was calculated based on this data. This process was repeated *m* = 1,000 times to generate the 95% confidence ellipses. In each plot:

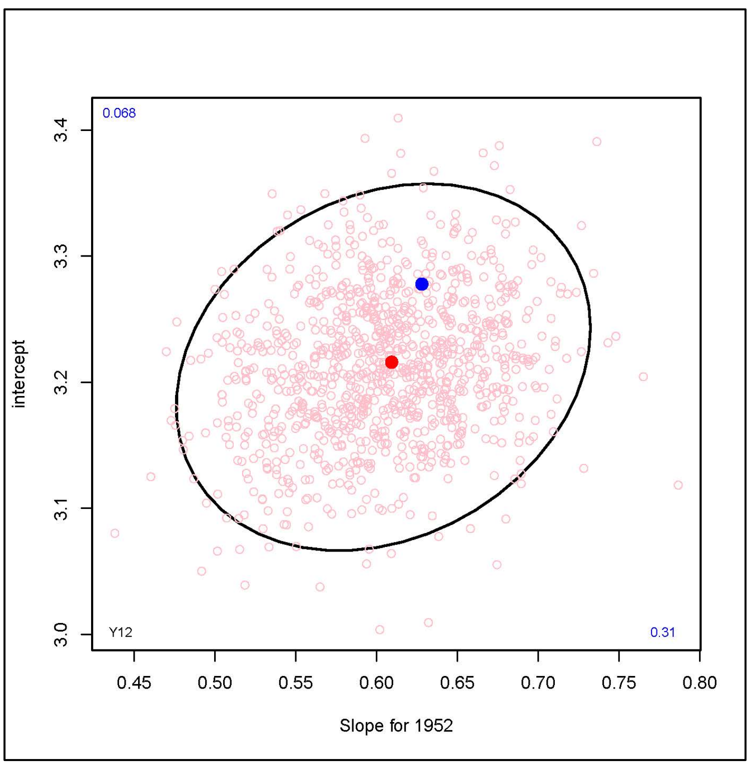
- The solid red point is the slope and intercept of the complete dataset.
- The solid blue point is the slope and intercept of the claimant dataset.
- The open pink points are the slopes and intercepts of the fits to the *m* random sets of 1585 workers taken from the complete dataset.
- The ellipse is the 95% joint normal confidence ellipse.
- The numbers in the corners are the full scale of the respective axes in terms of +/- the slope or intercept of the complete dataset.

These plots are presented to support the idea that the claimant dataset is basically no different from a random draw from the complete dataset.

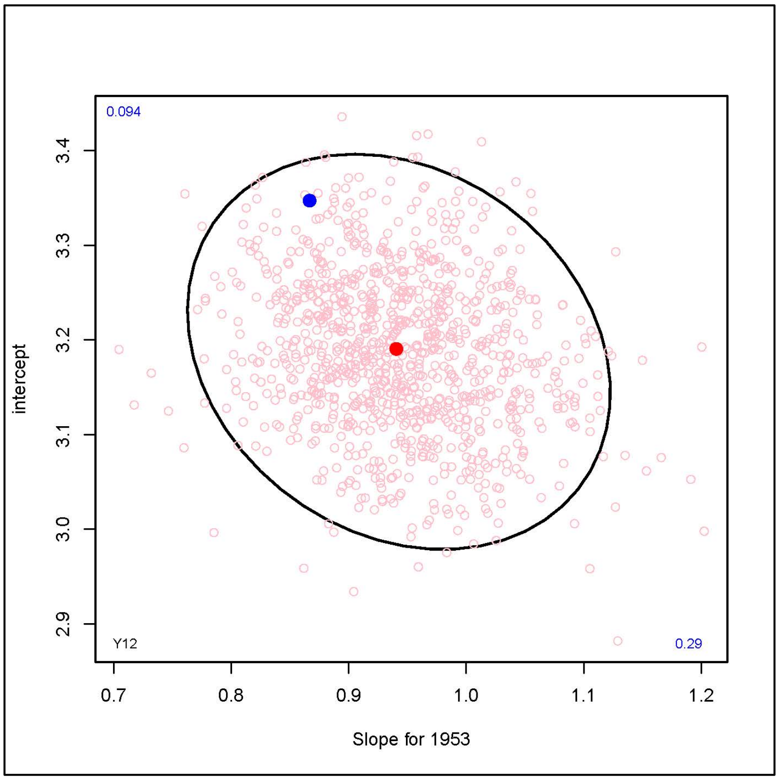


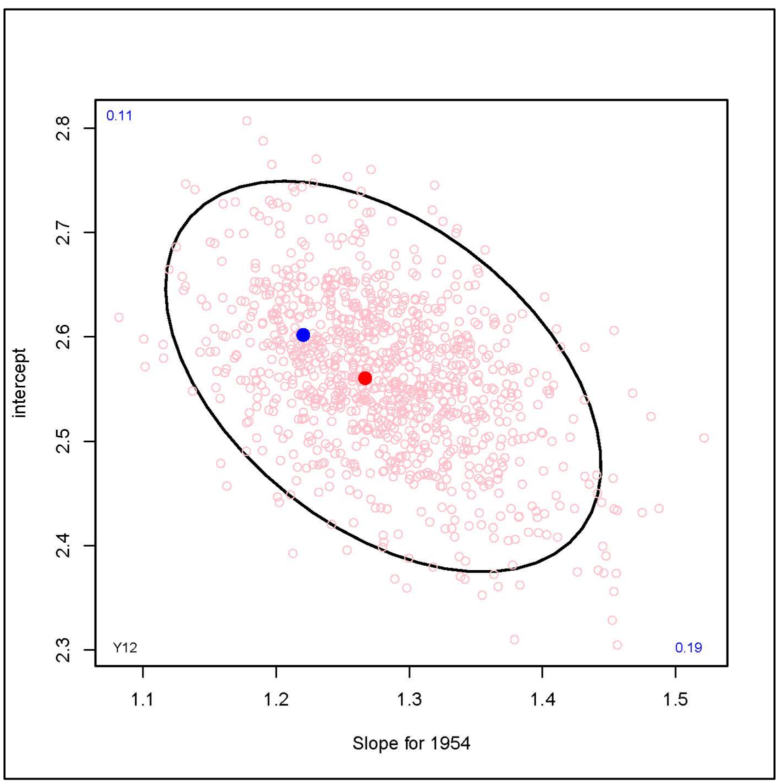
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



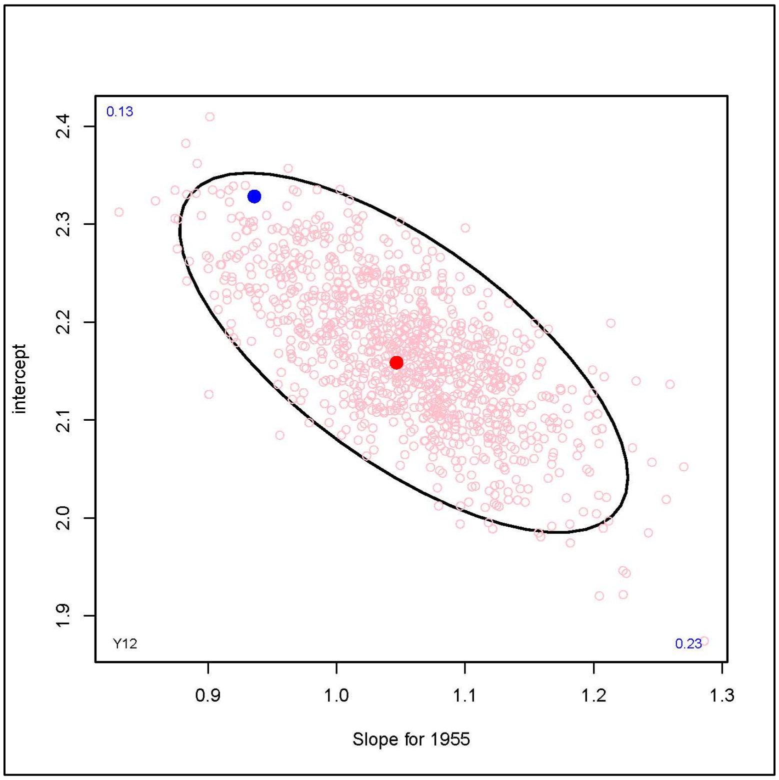


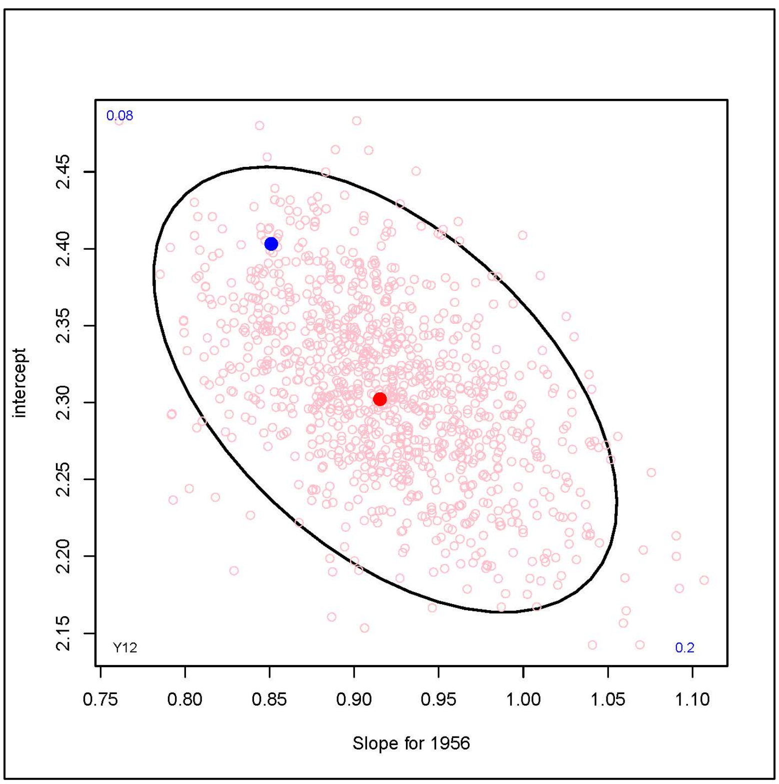
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



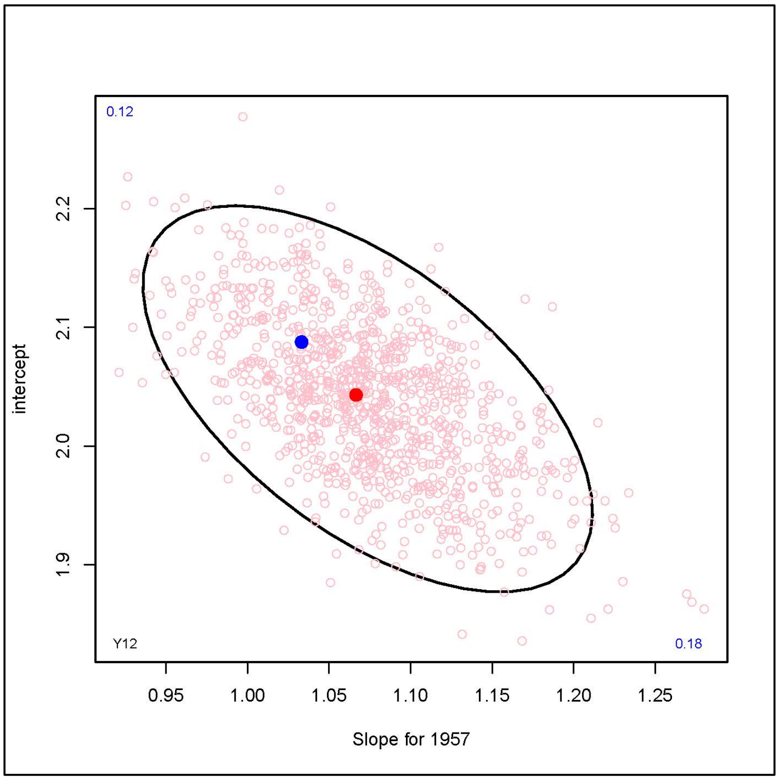


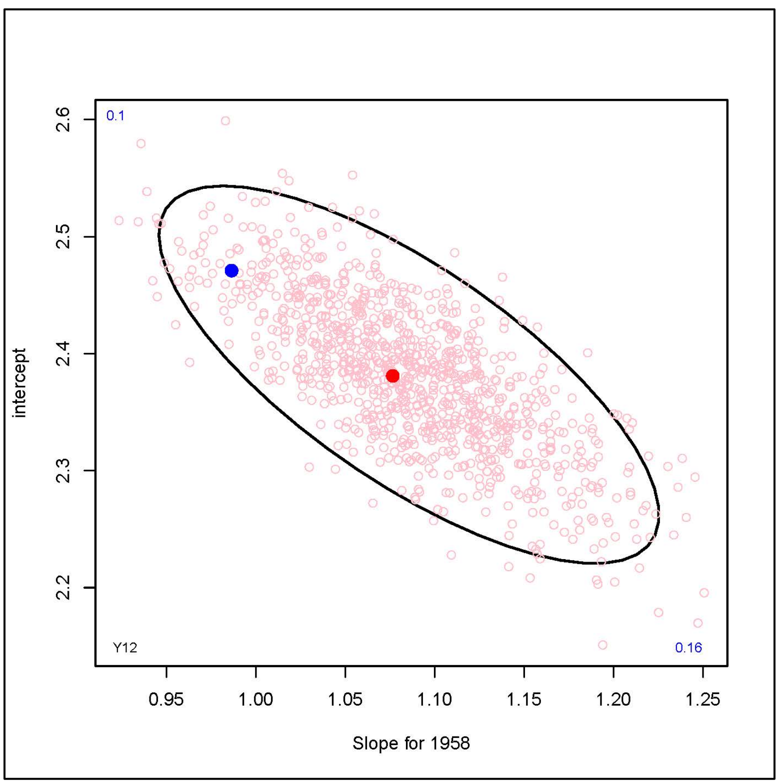
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



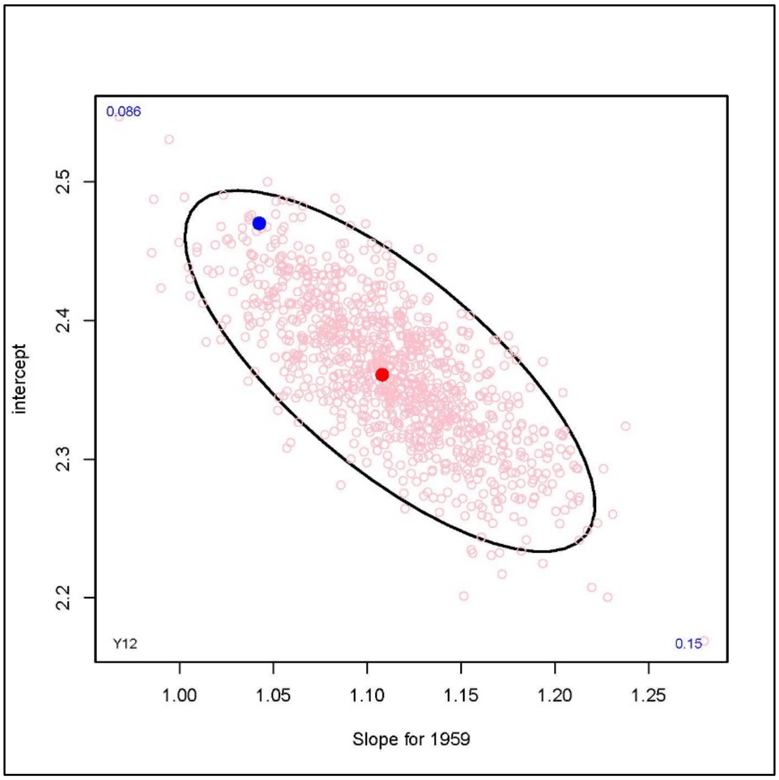


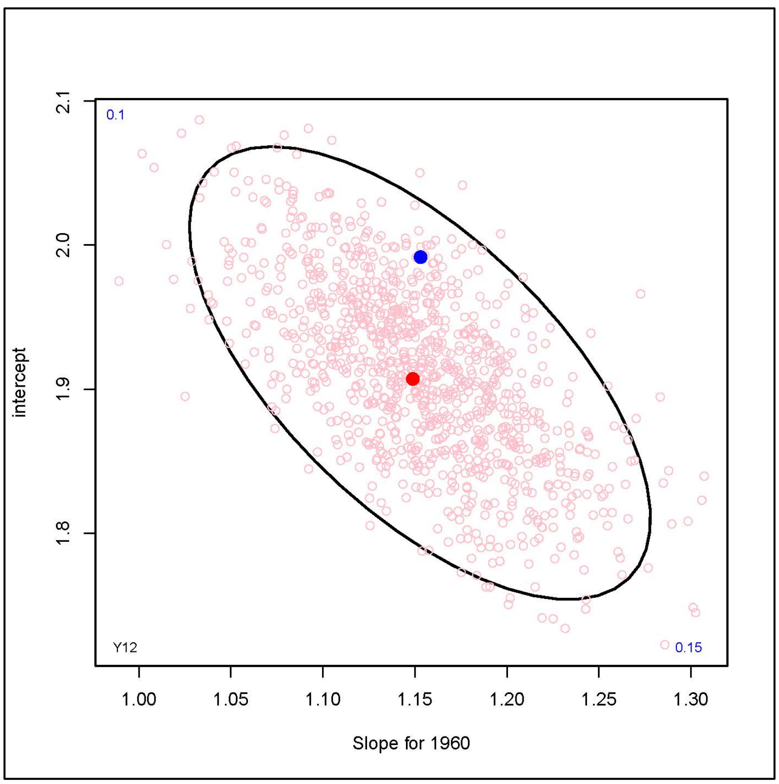
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



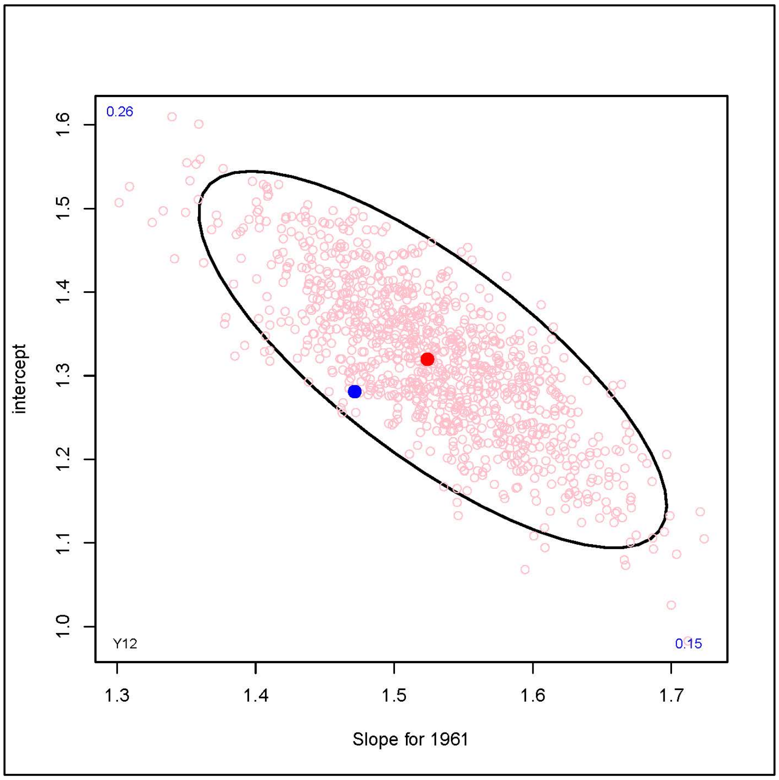


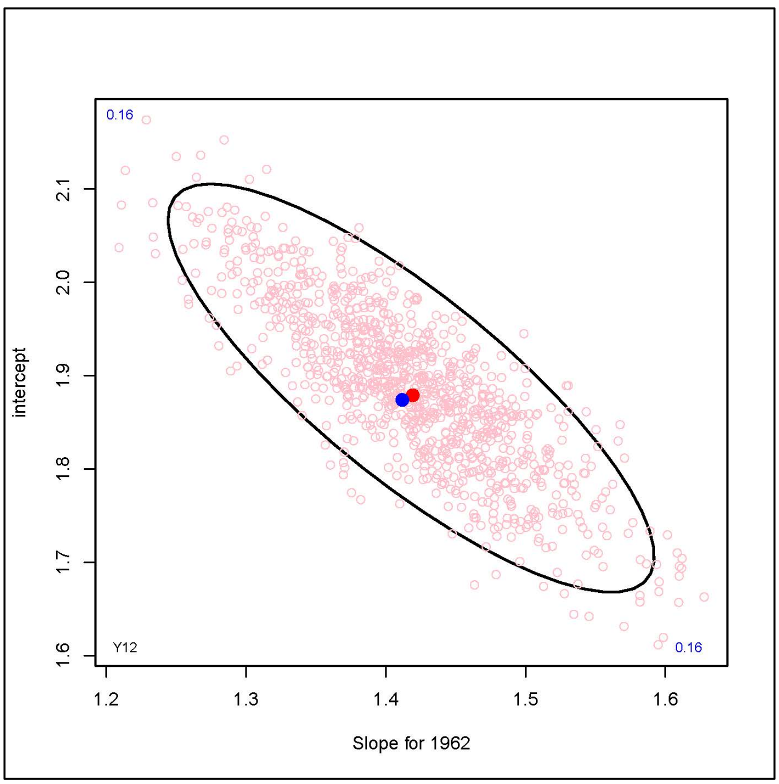
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



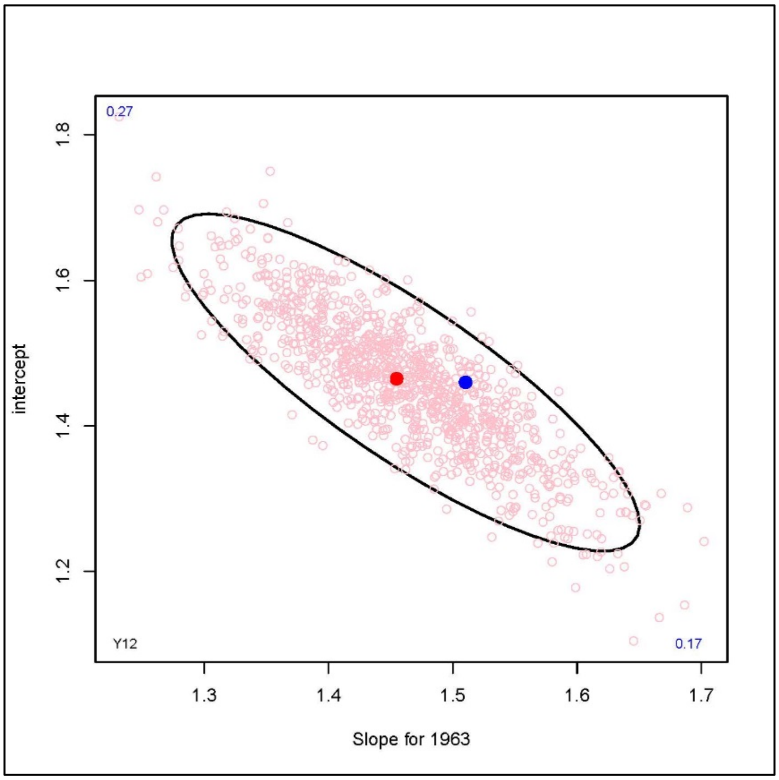


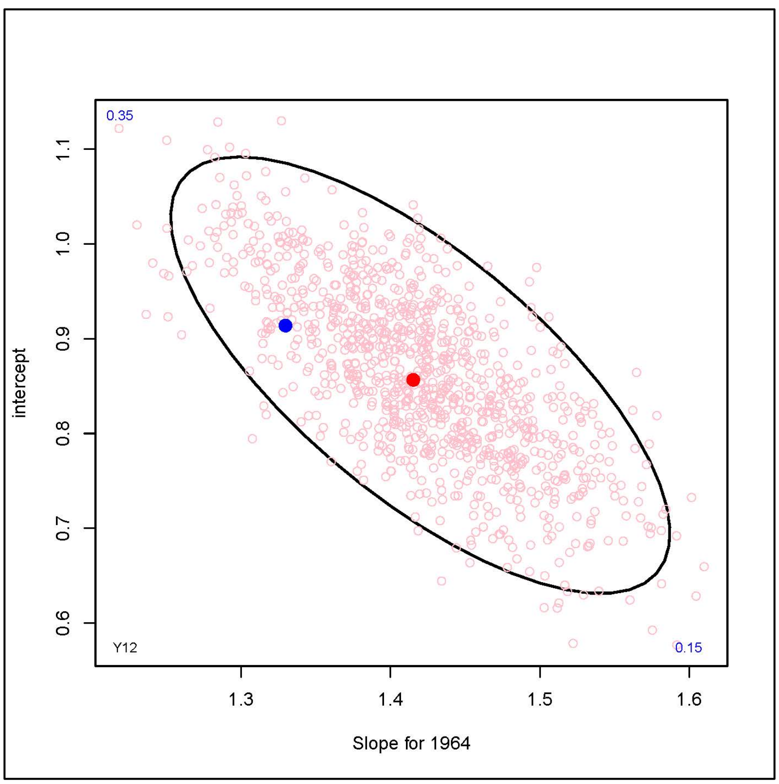
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



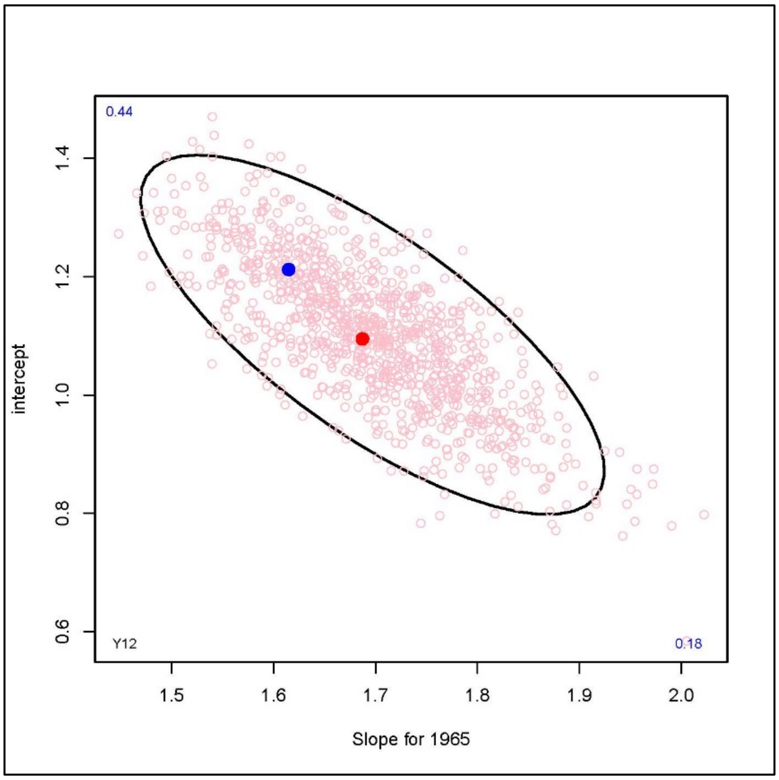


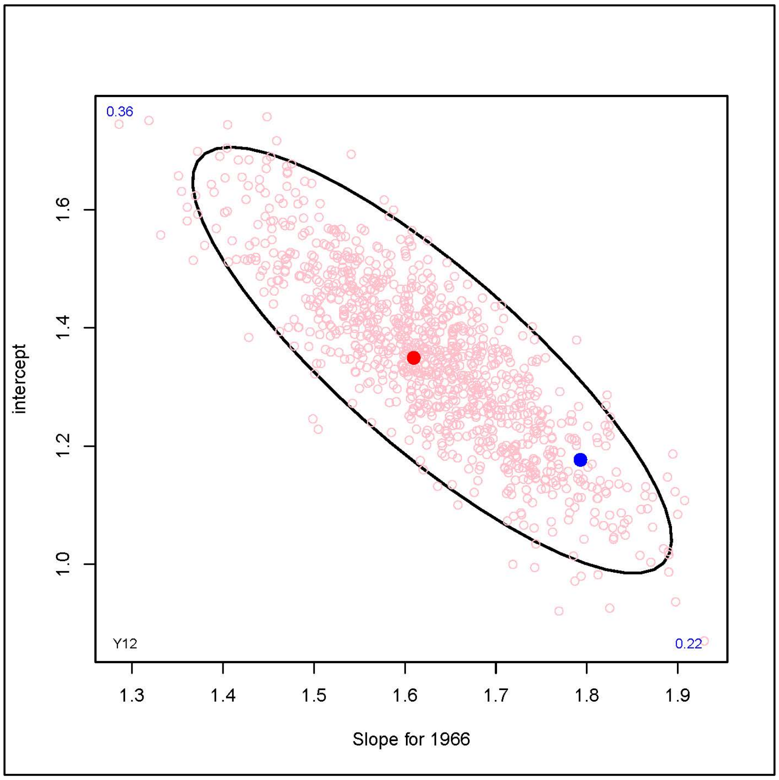
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



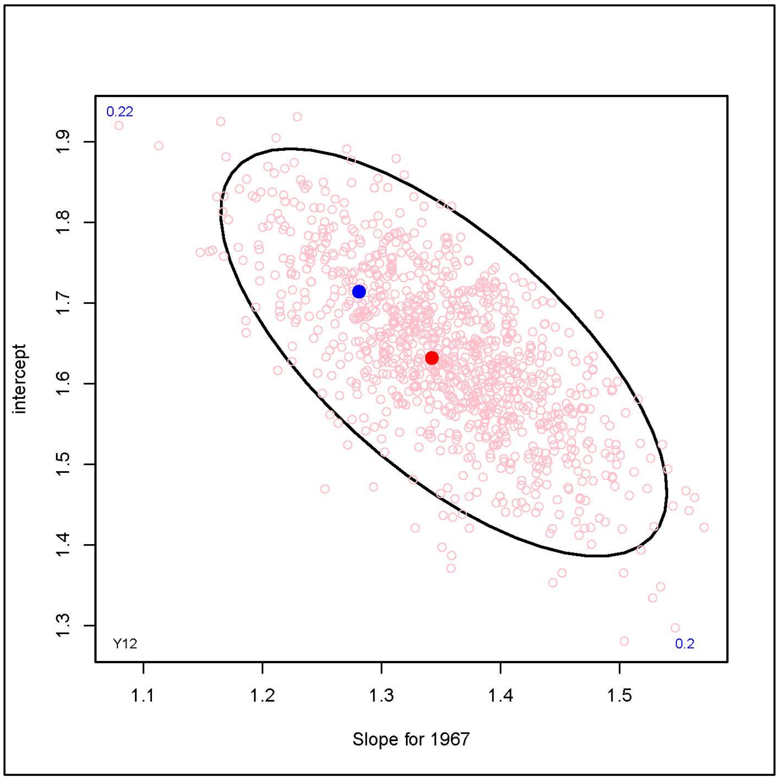


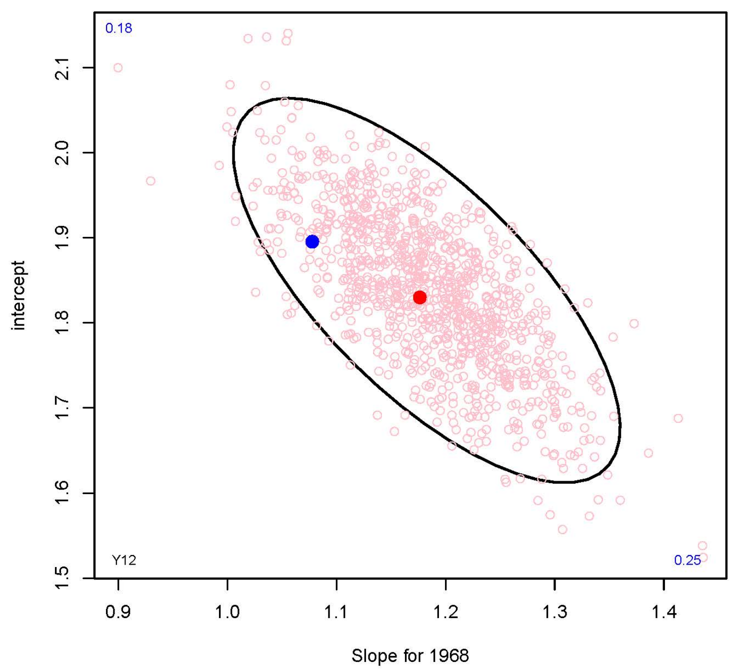
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



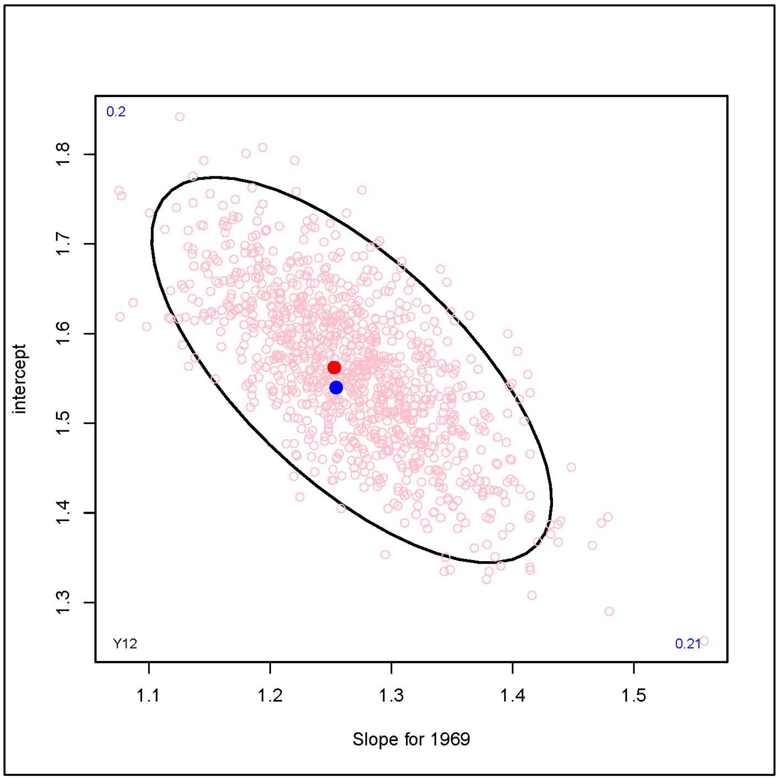


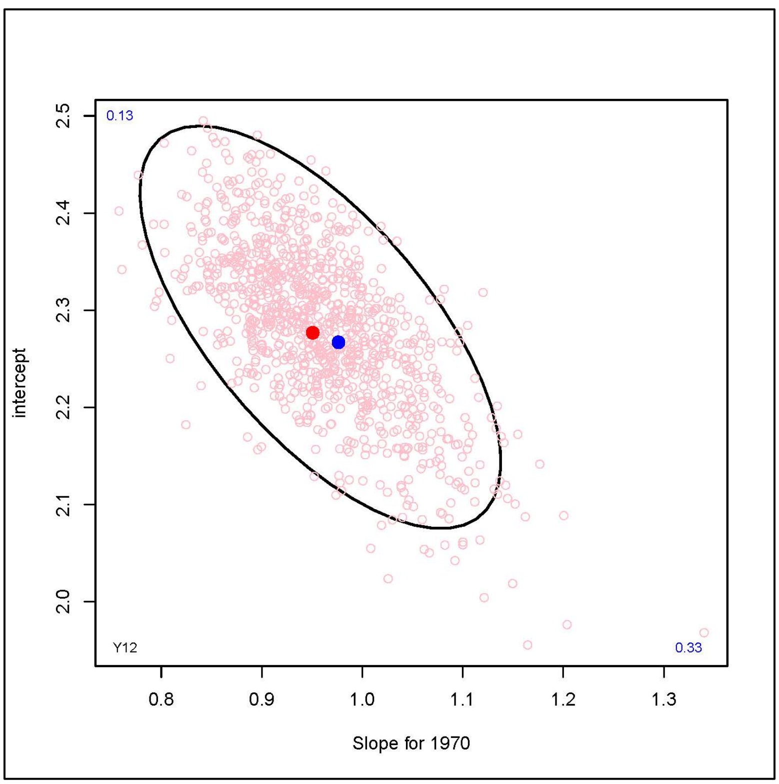
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



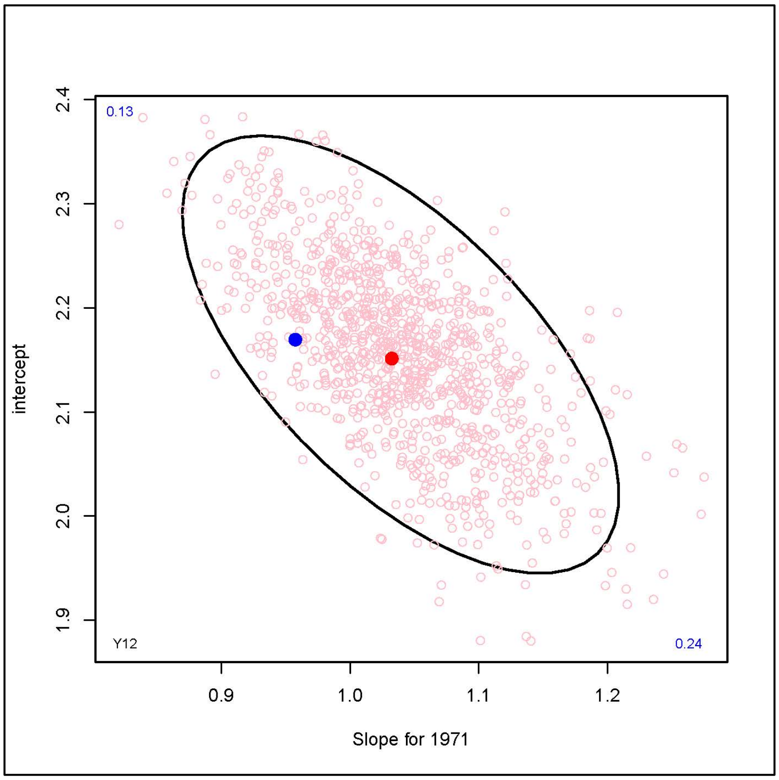


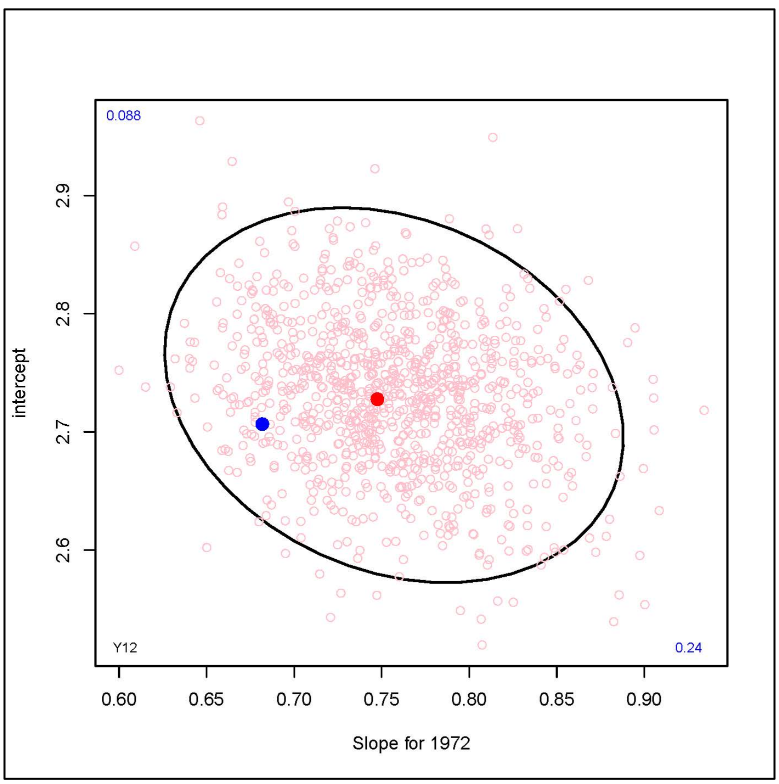
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



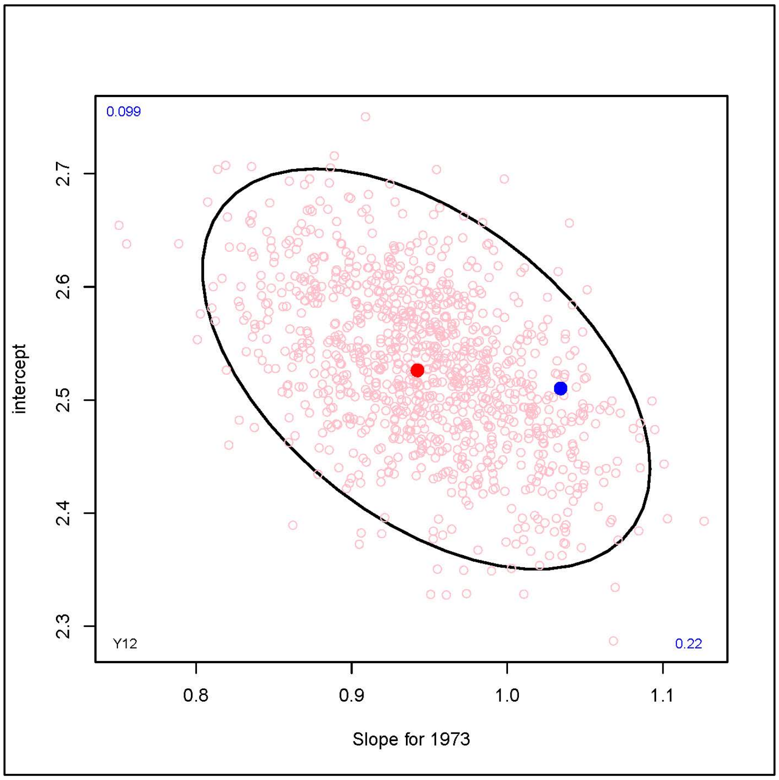


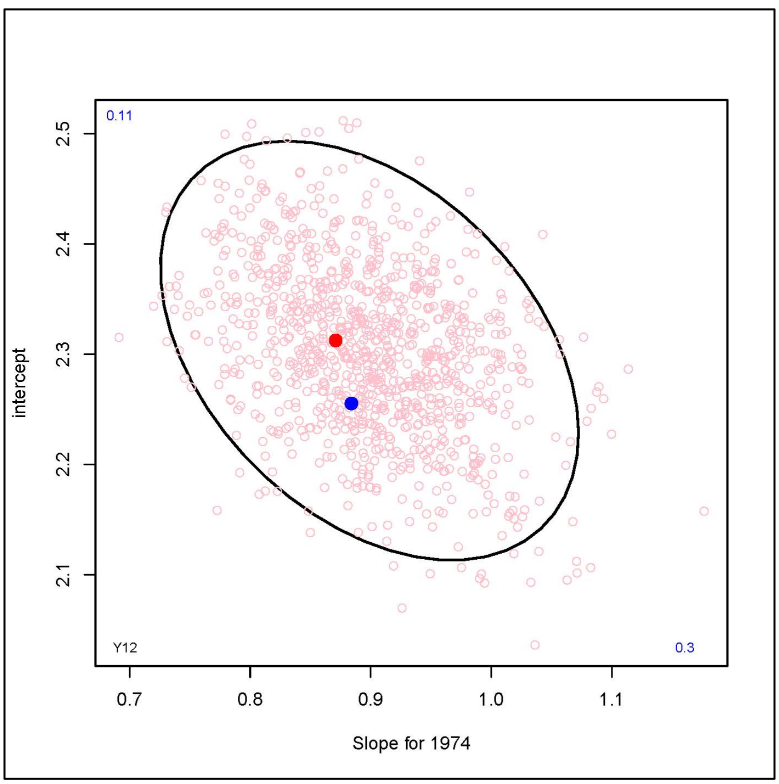
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



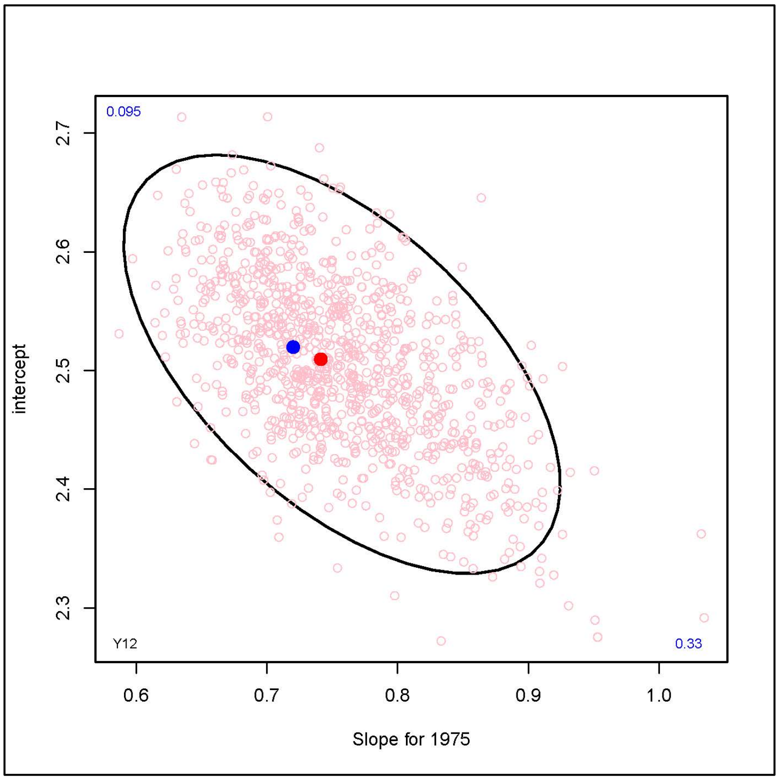


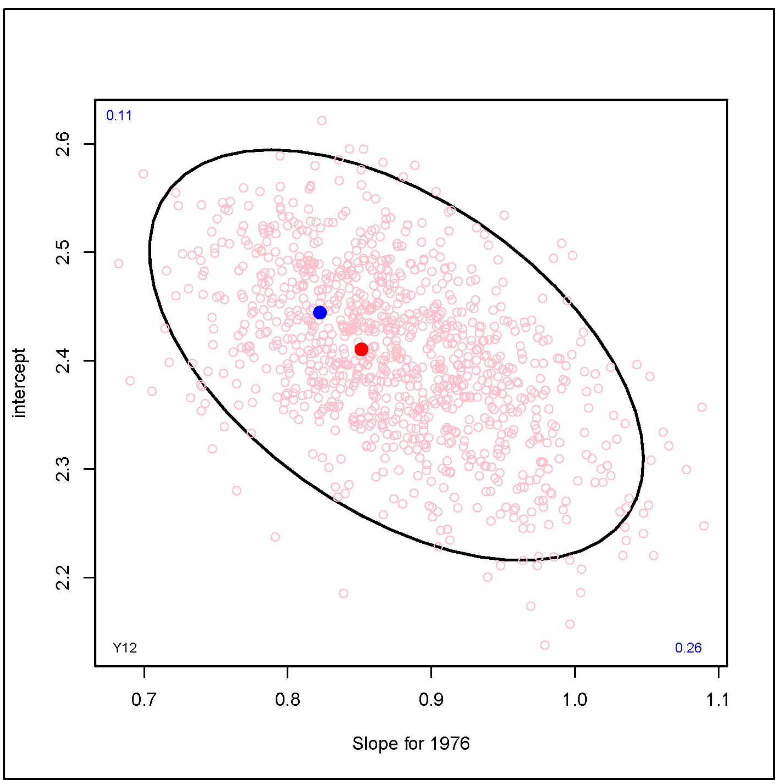
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



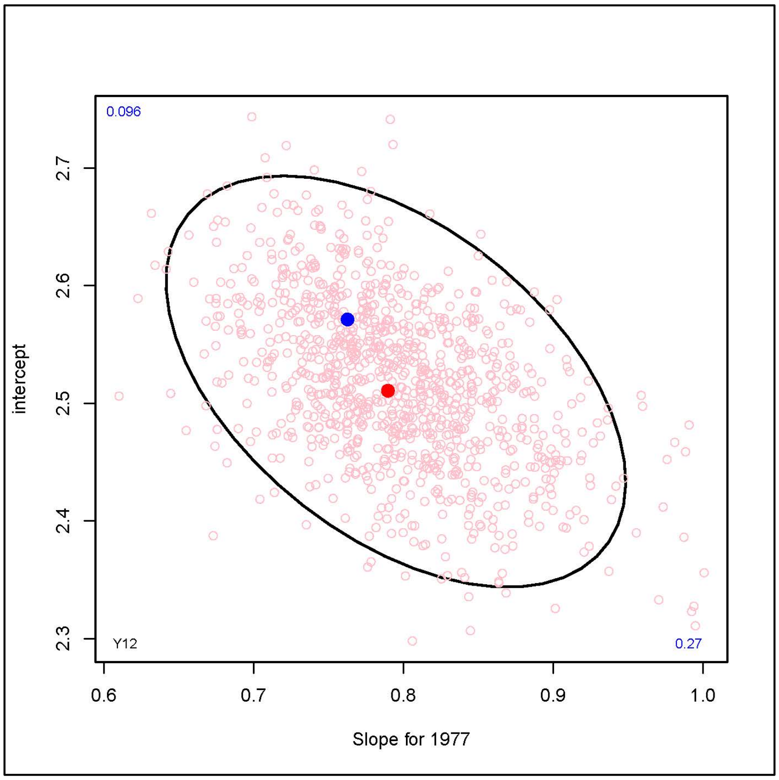


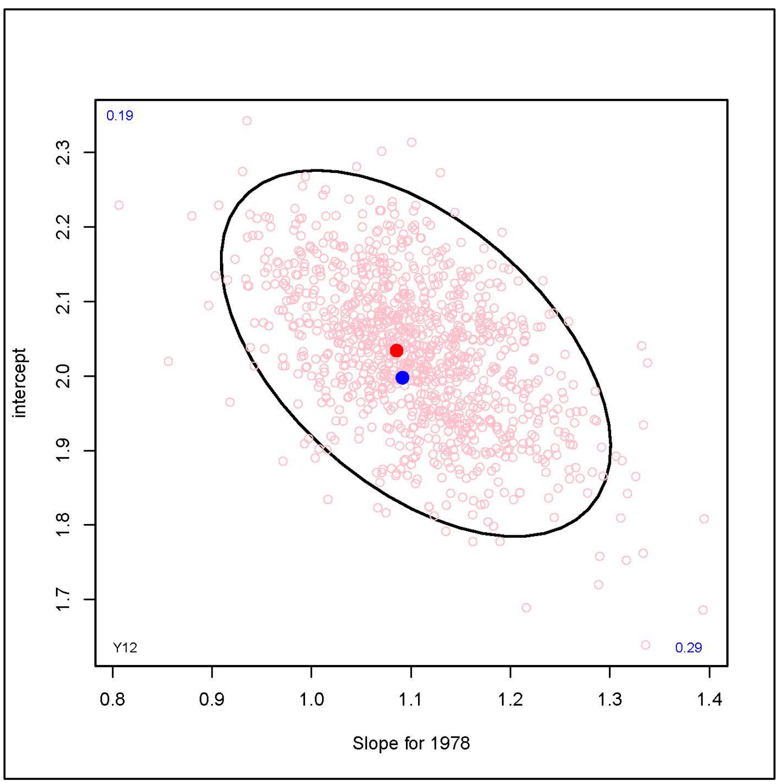
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



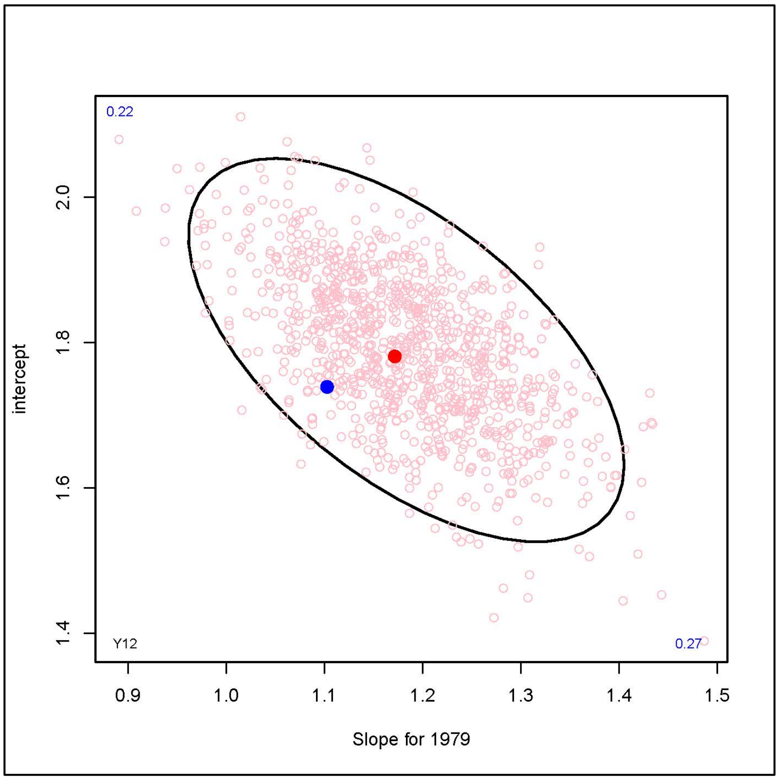


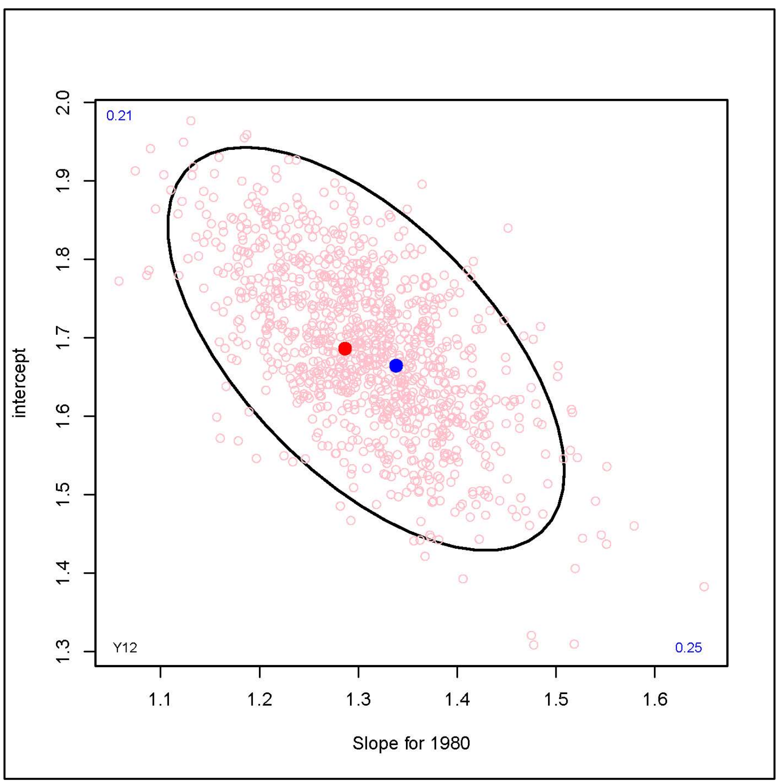
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



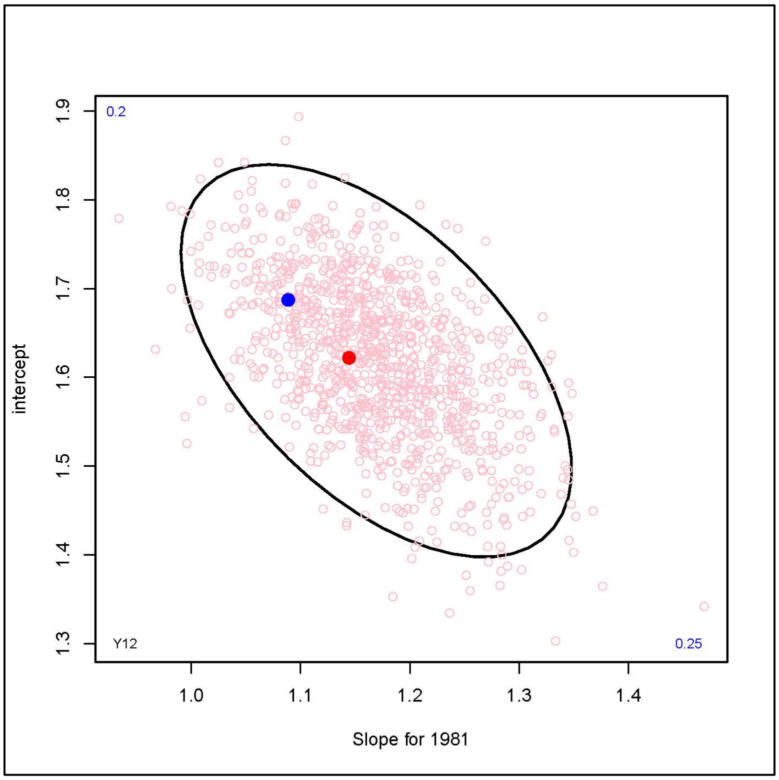


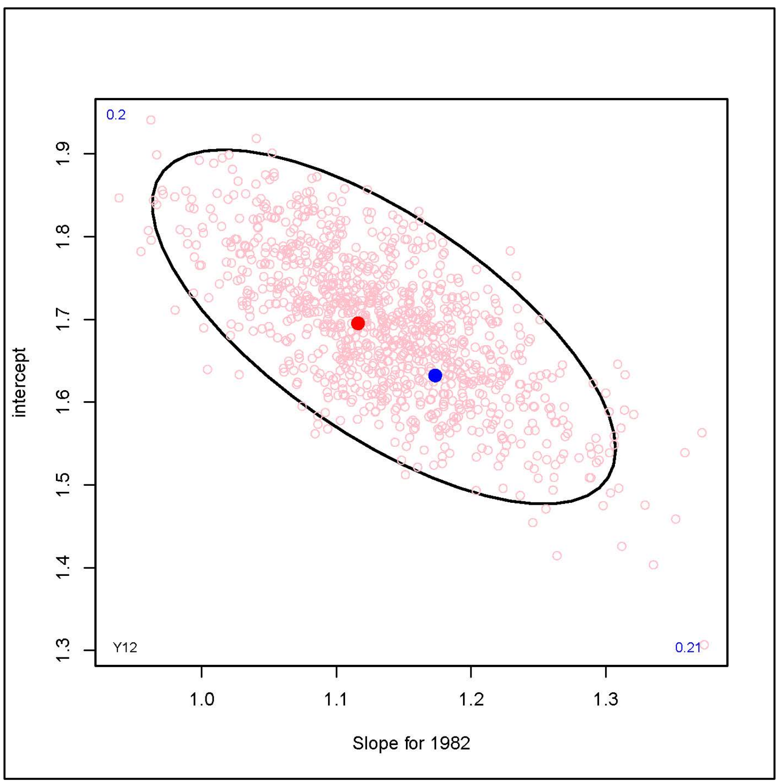
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



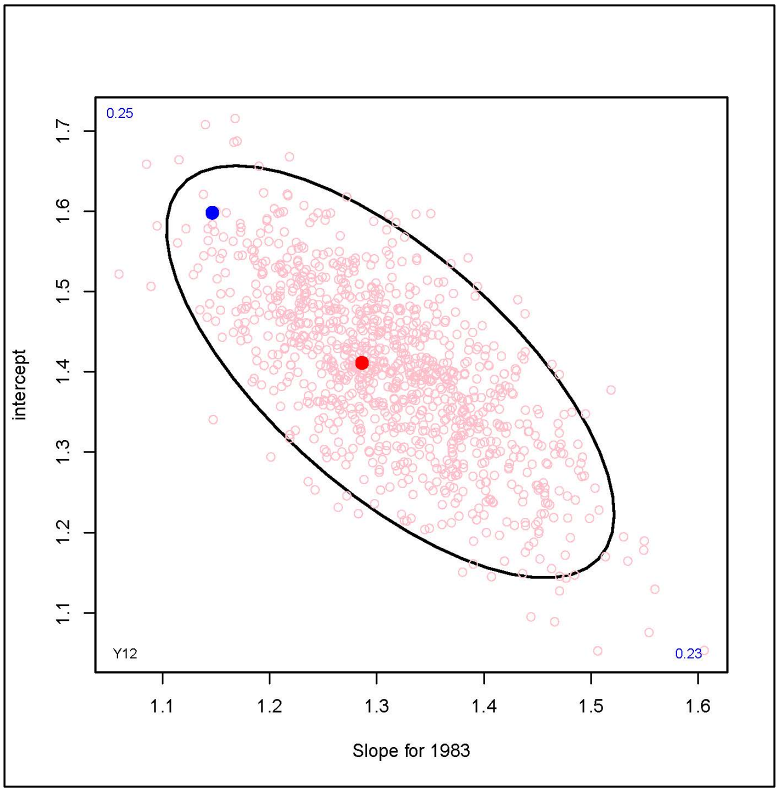


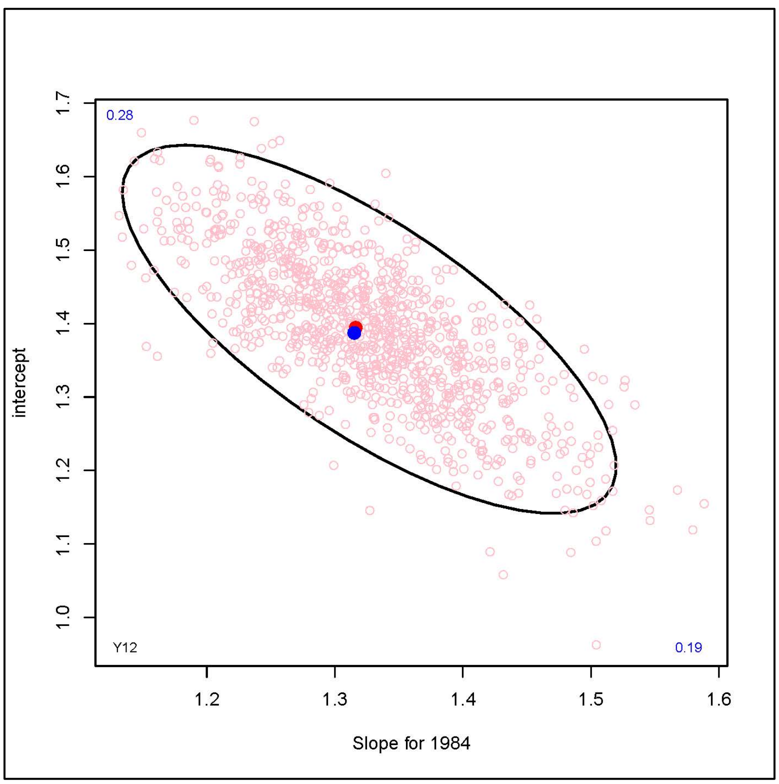
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



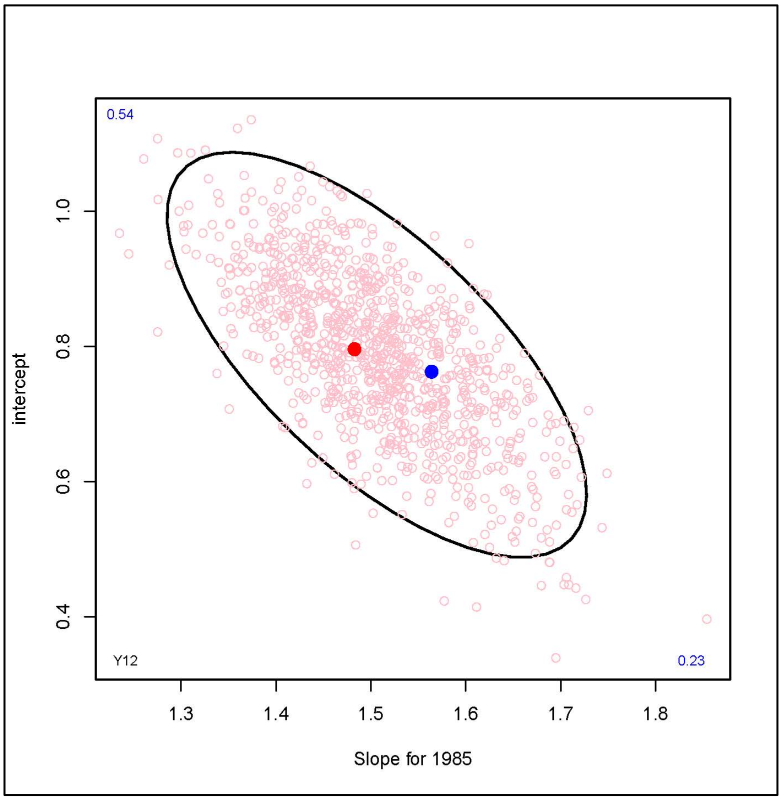


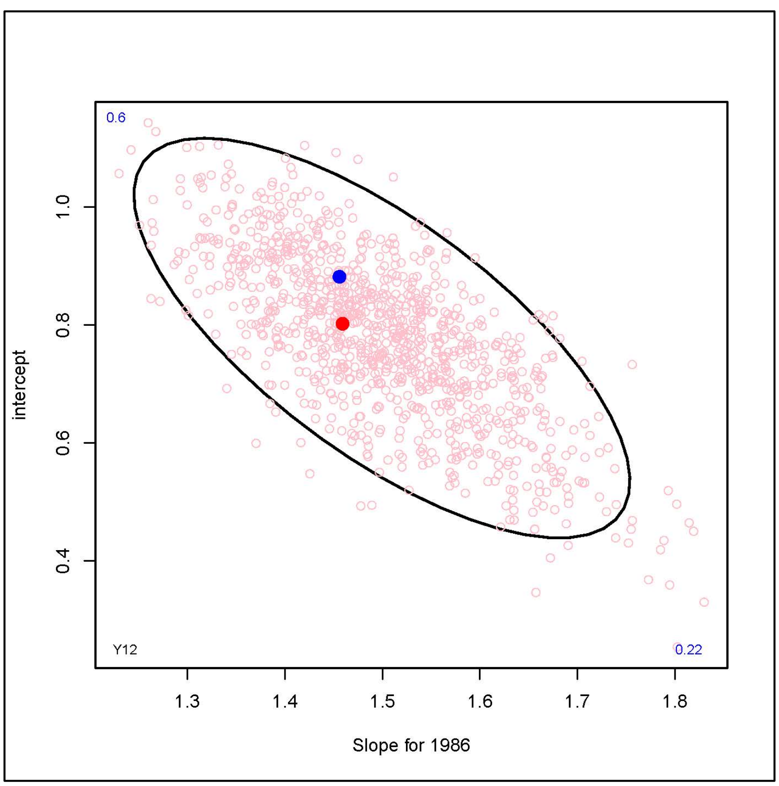
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



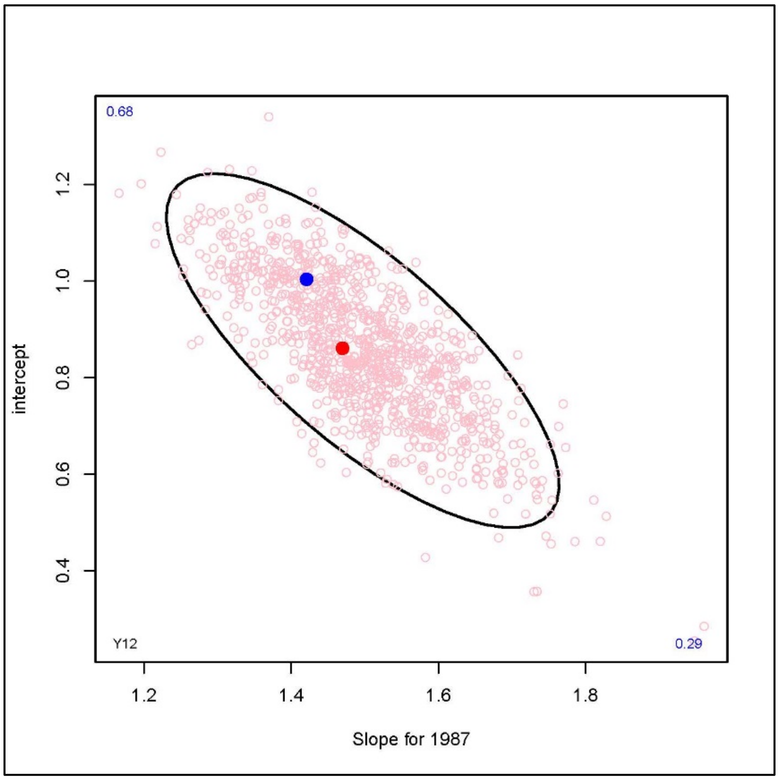


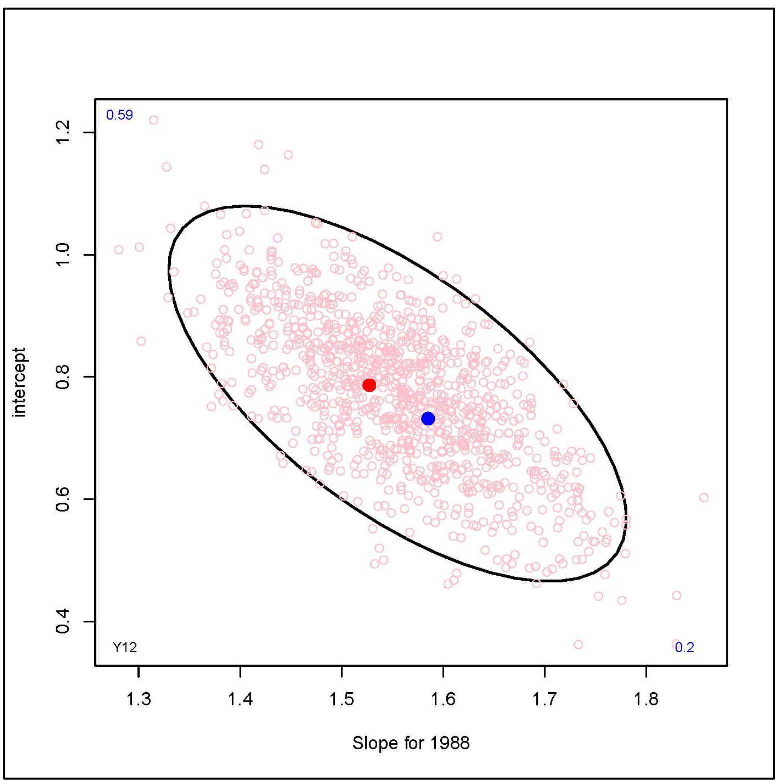
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)





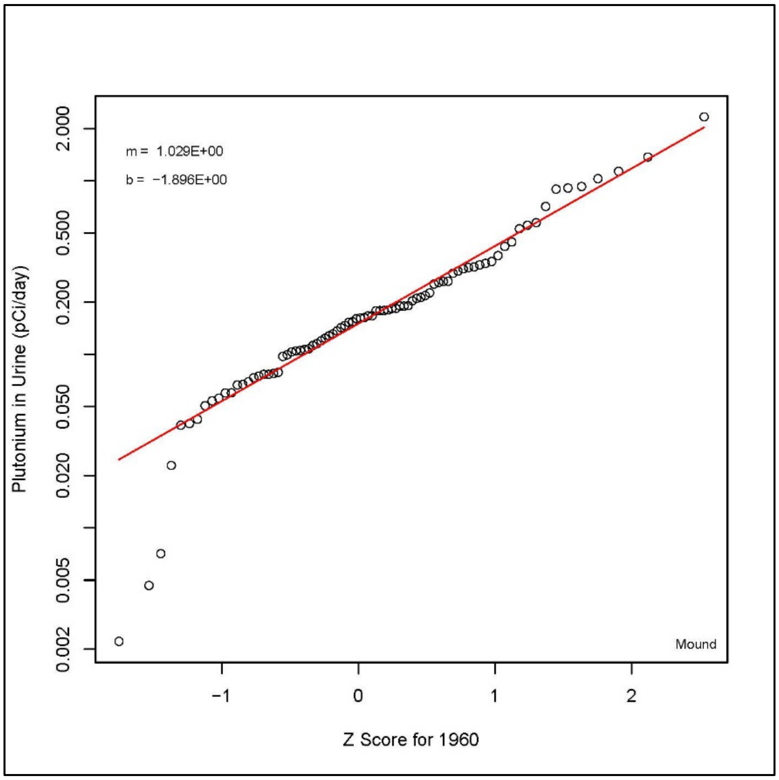
ATTACHMENT C
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE Y-12 DATASET, 1585 WORKERS (continued)



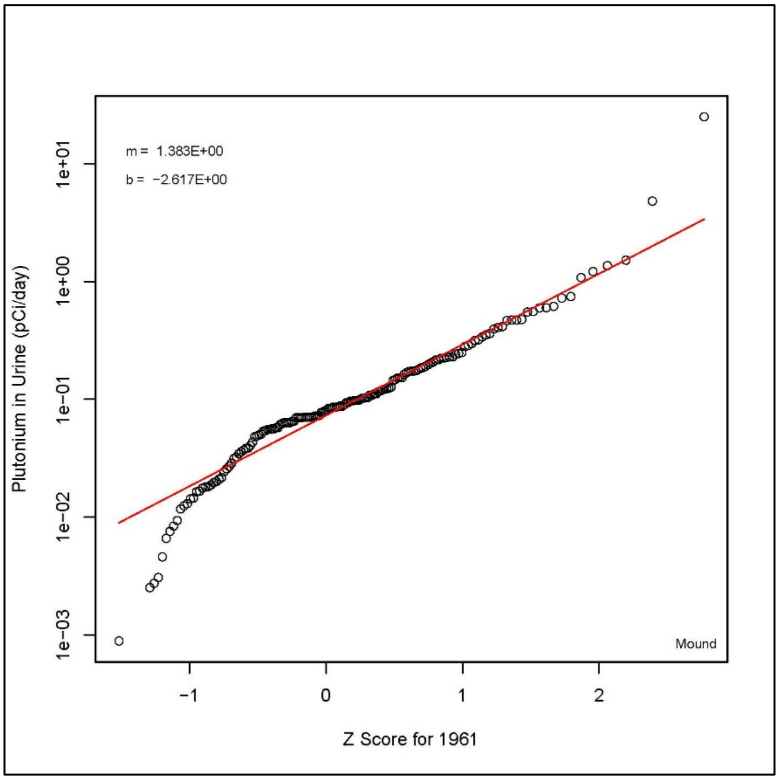


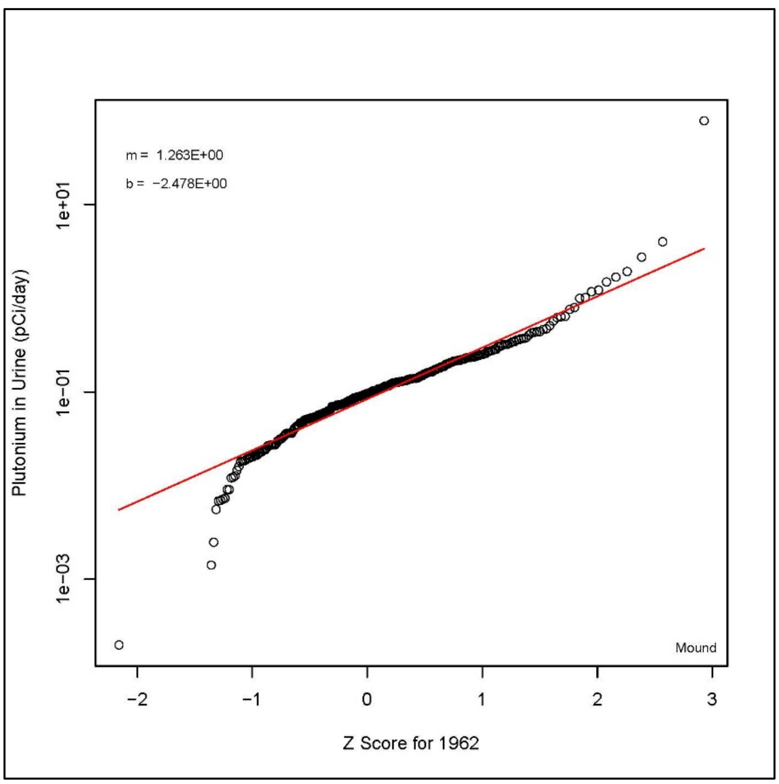
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET

This attachment presents lognormal probability plots for the complete Mound plutonium urine bioassay dataset for 1960 to 1990. The slope and intercept for the line of best fit to the data are given on the plot.

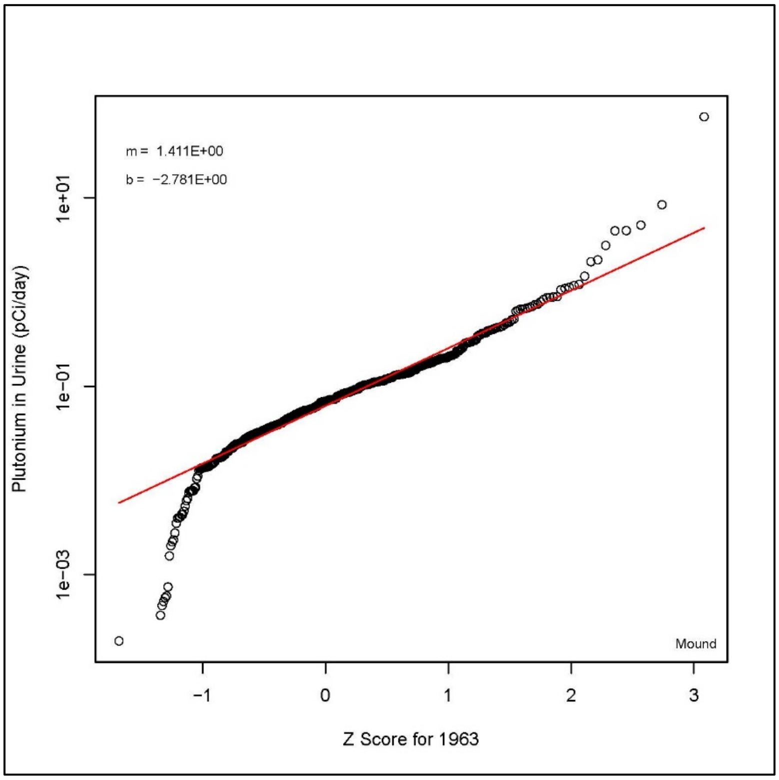


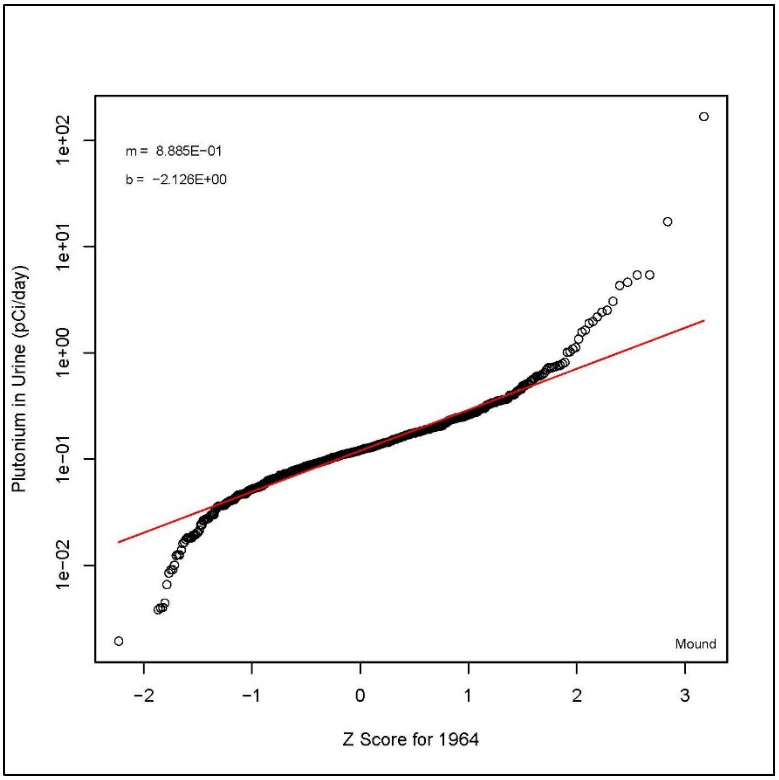
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



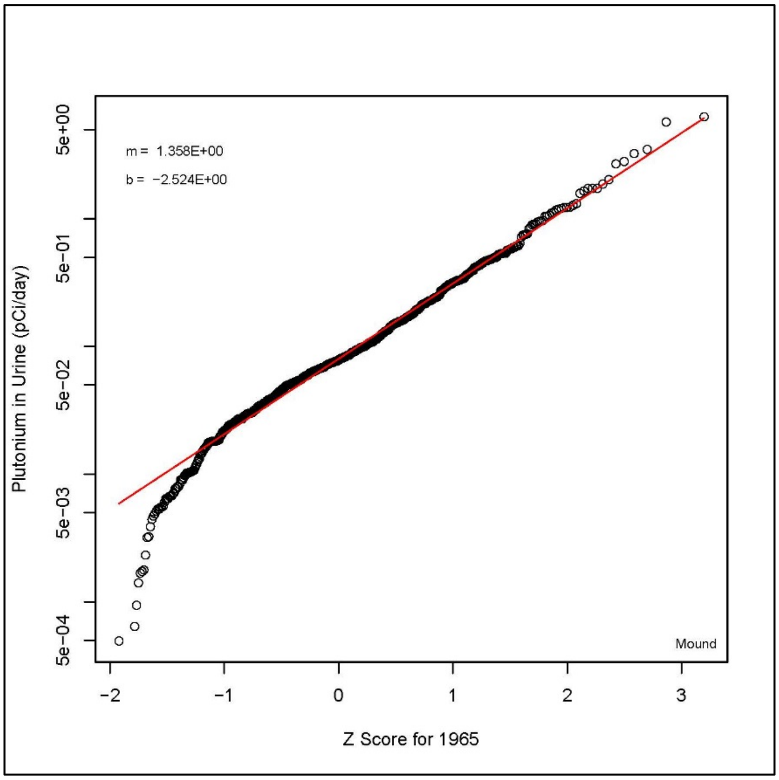


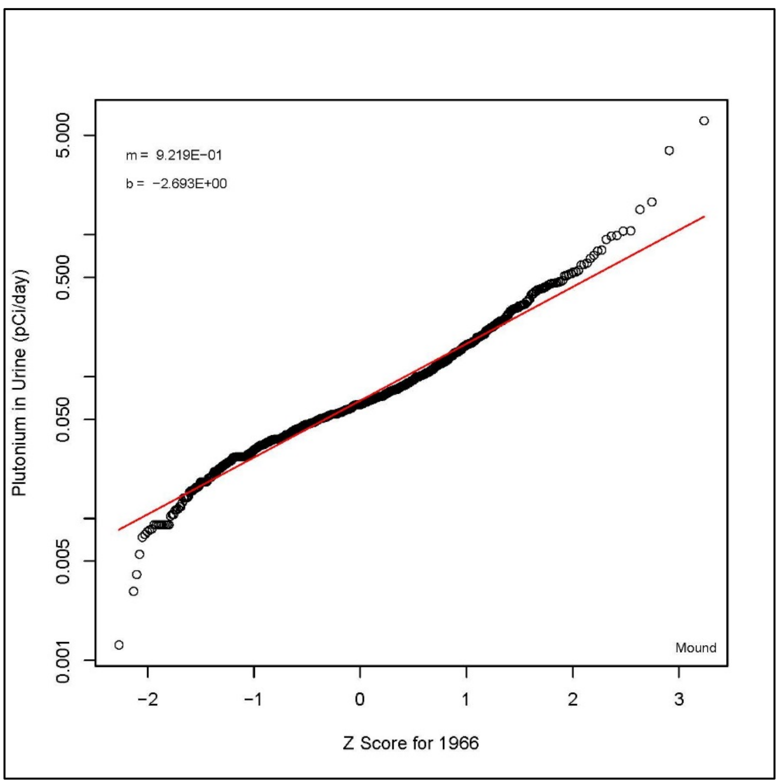
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



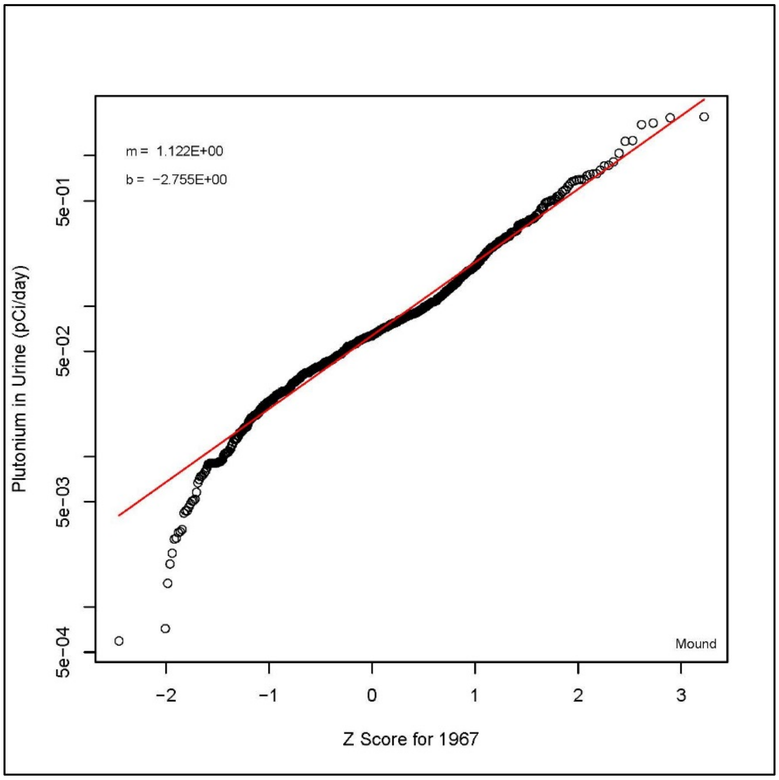


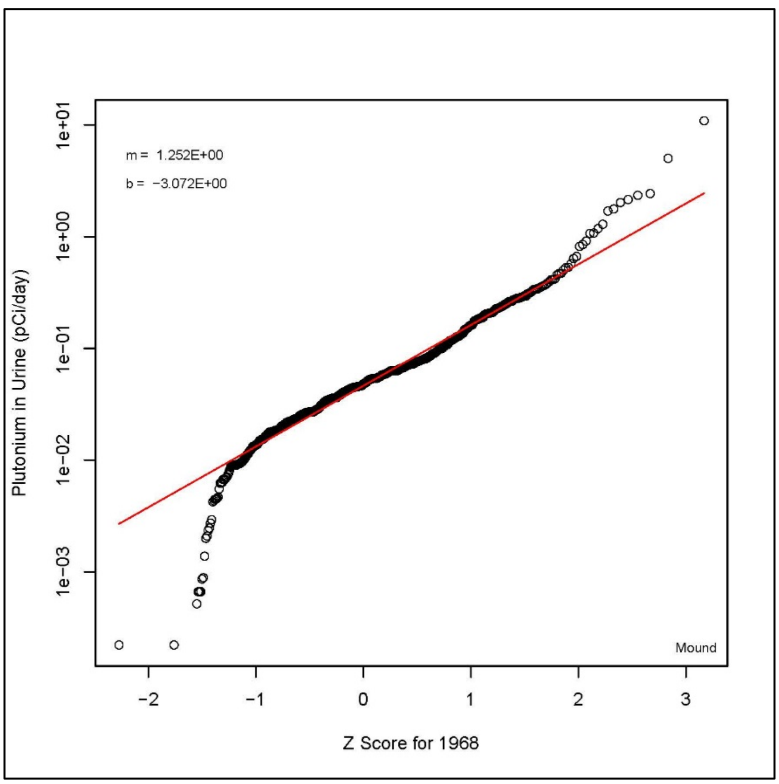
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



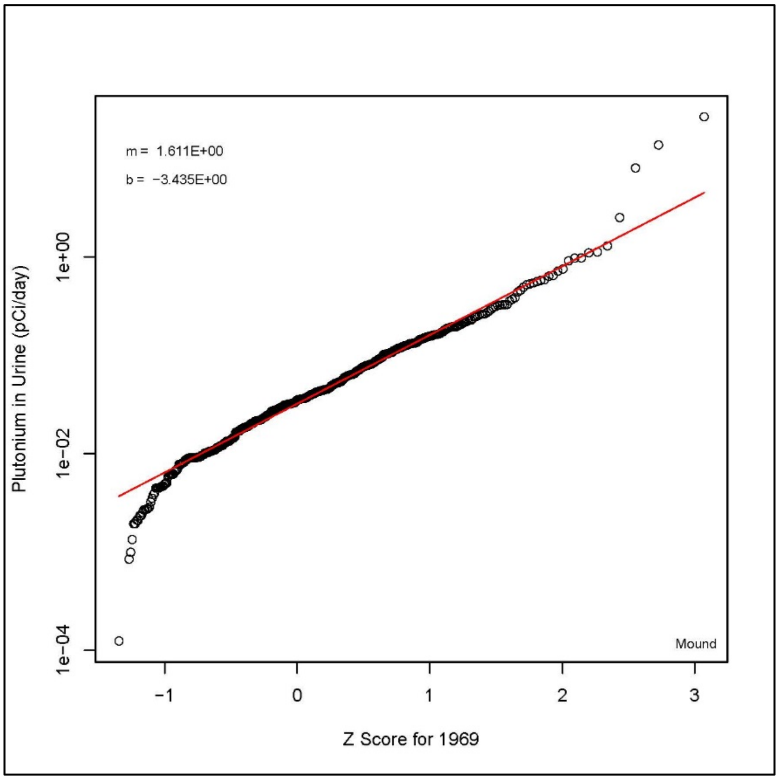


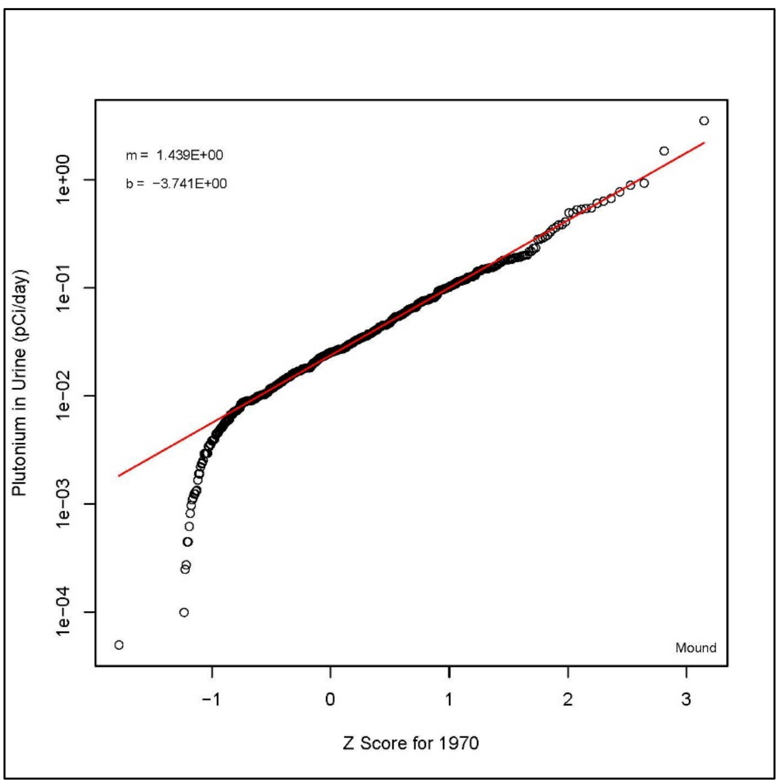
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



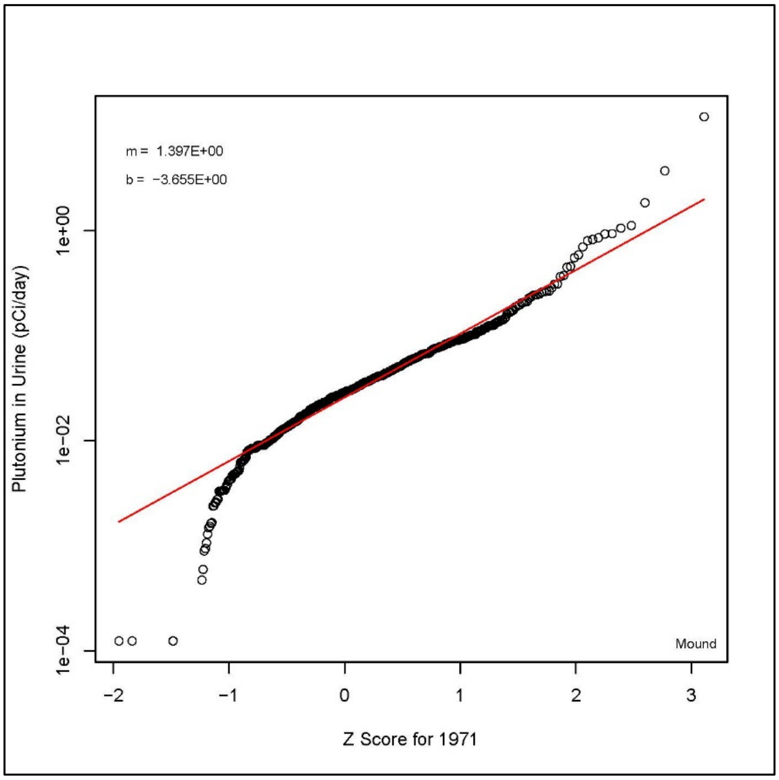


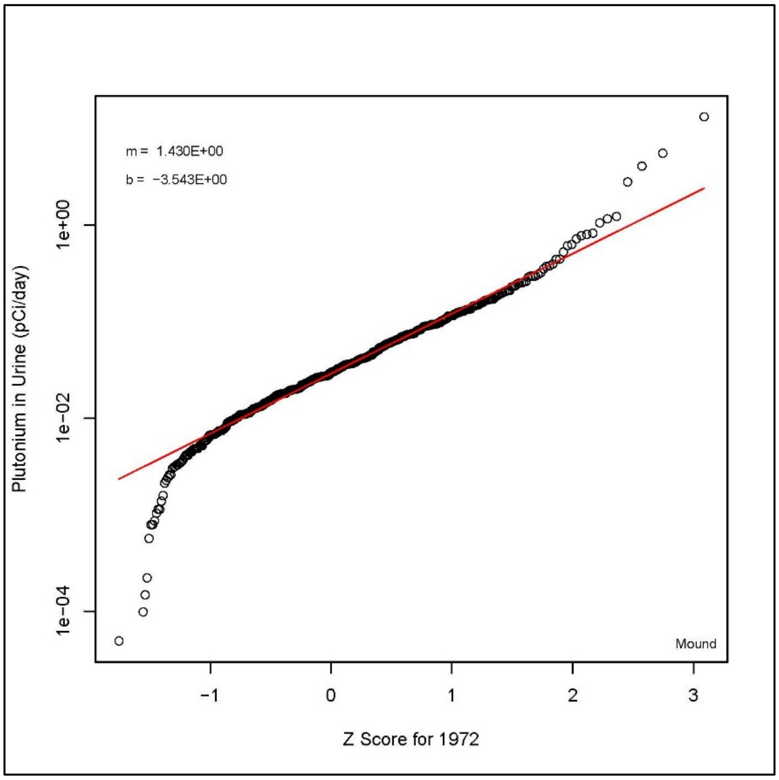
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



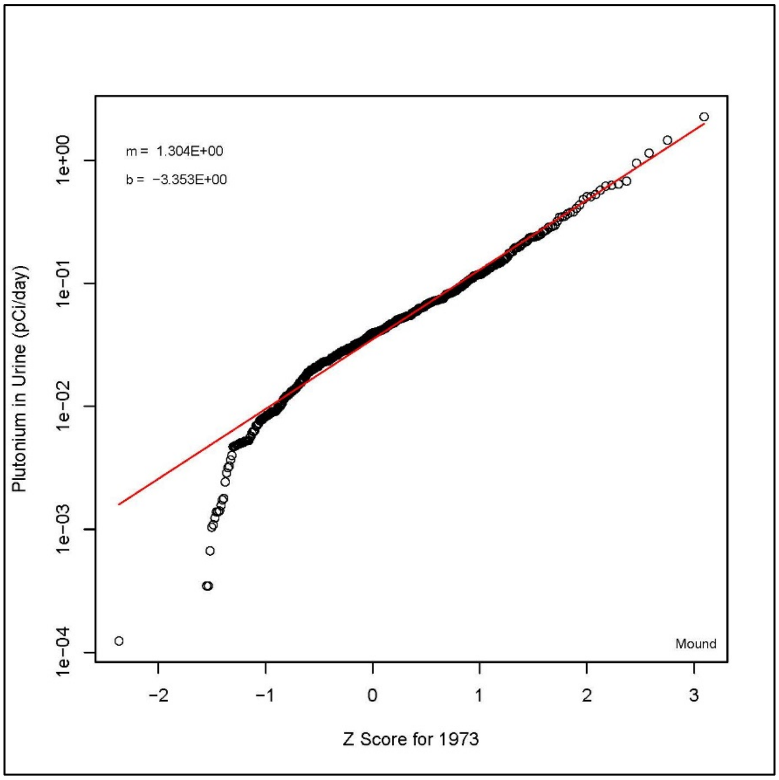


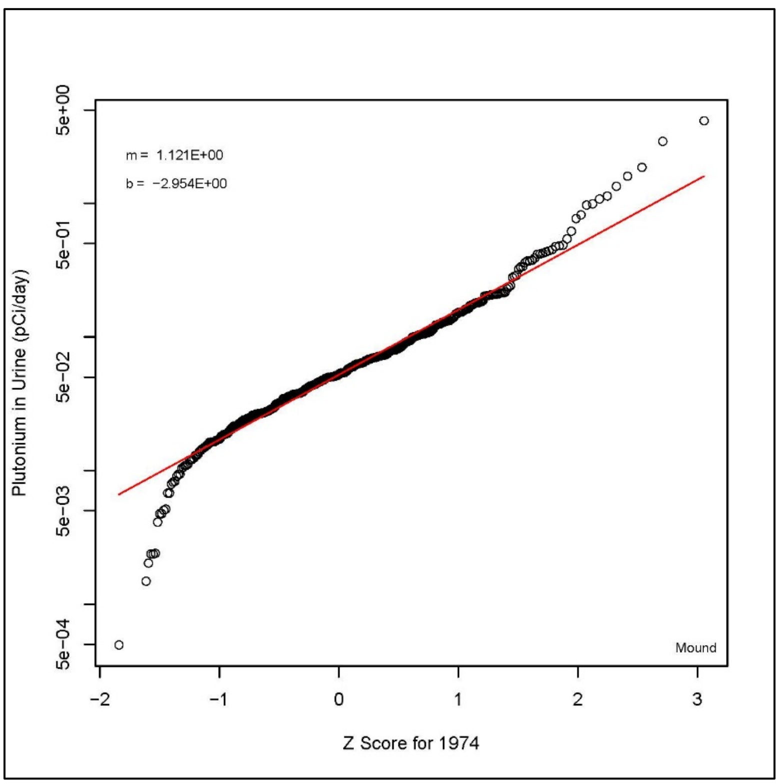
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



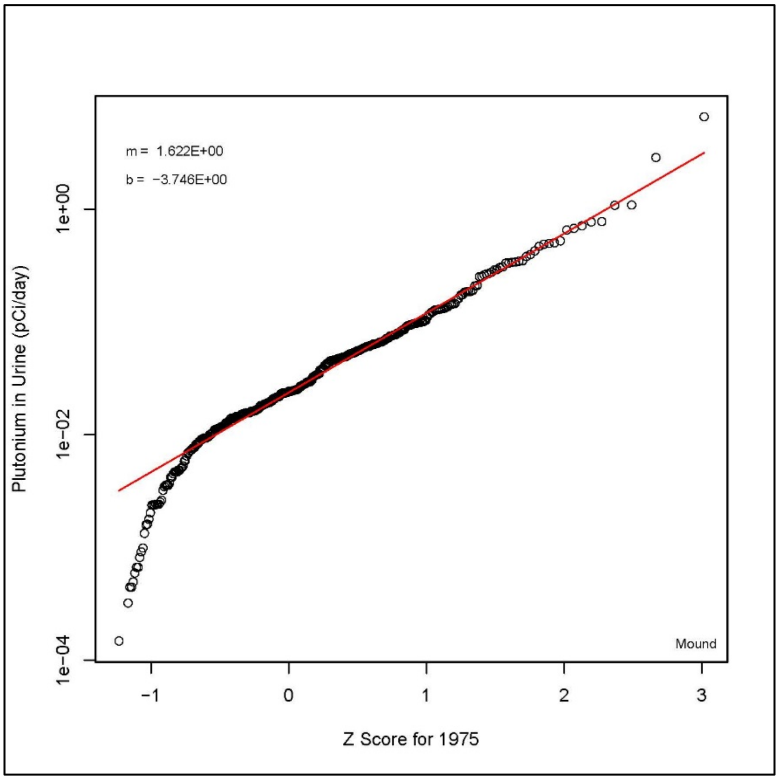


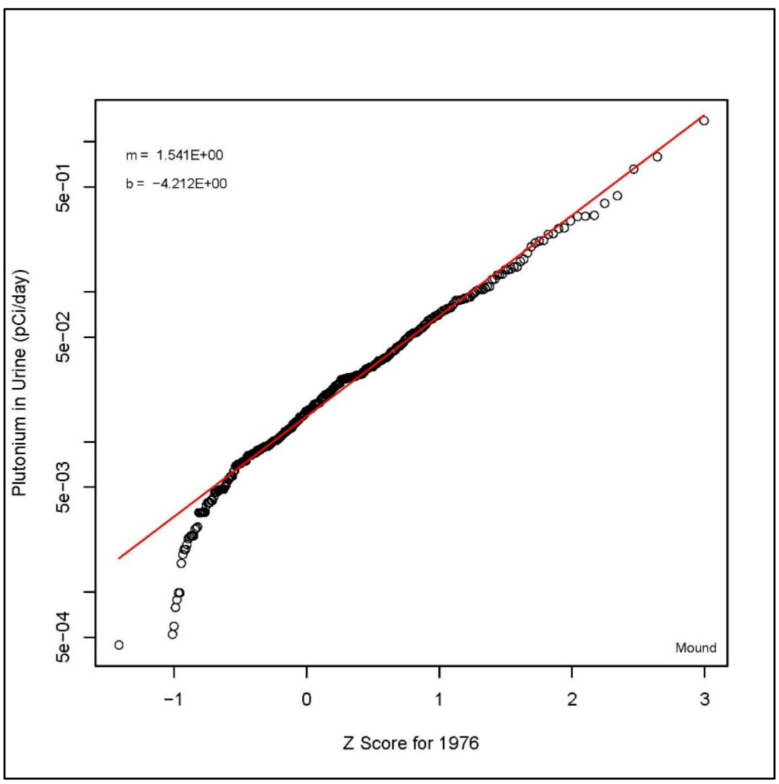
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



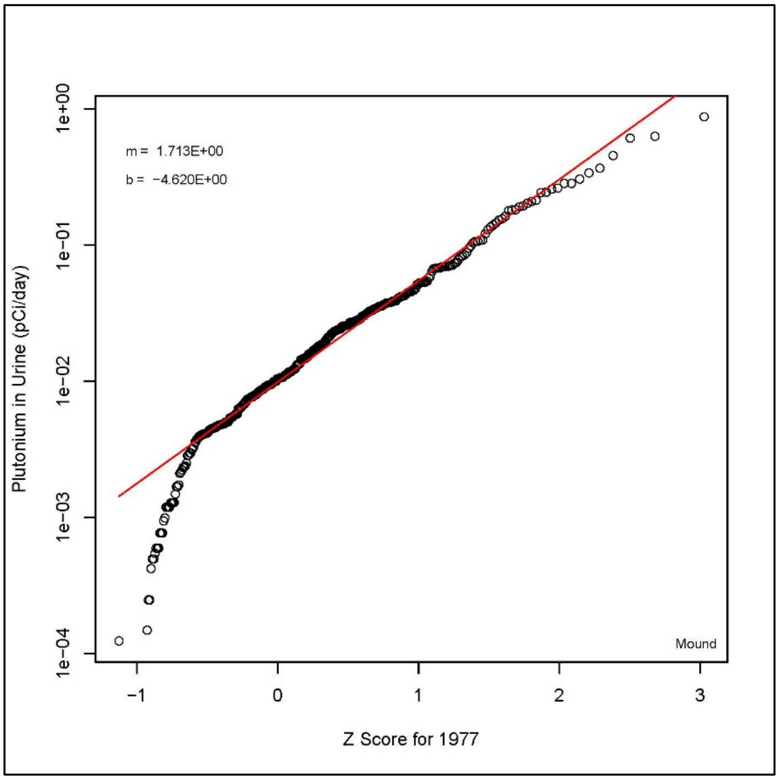


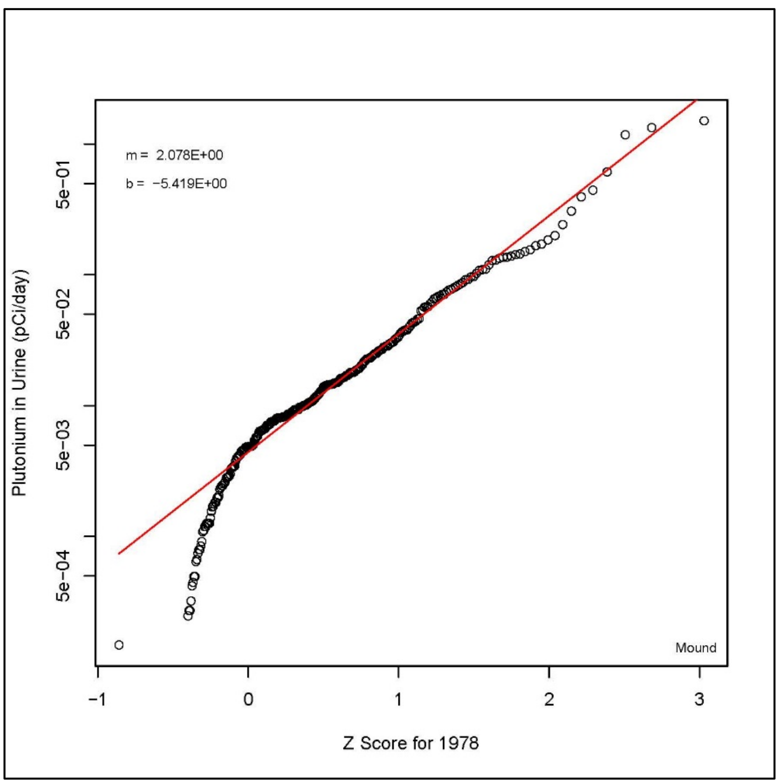
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



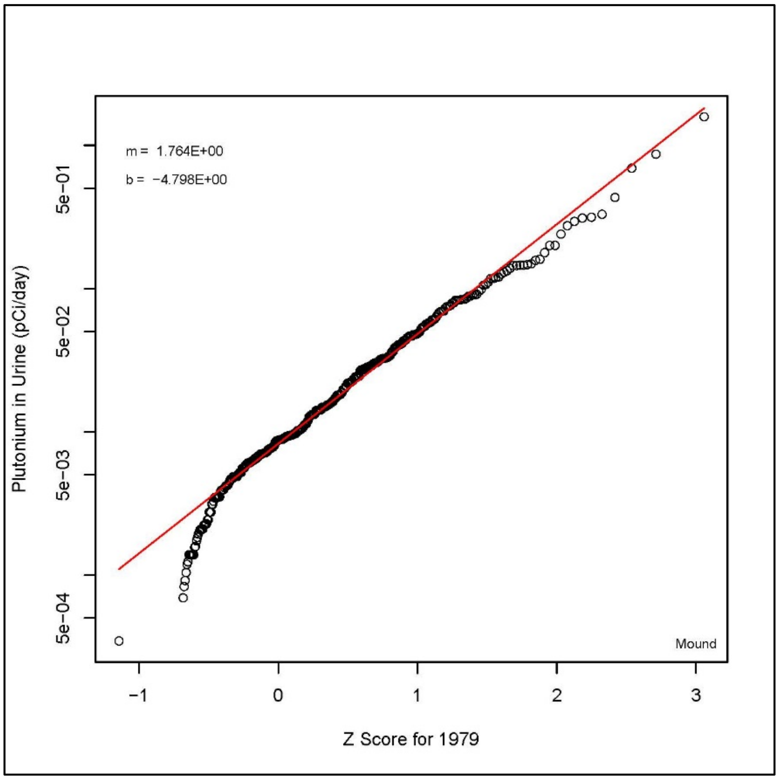


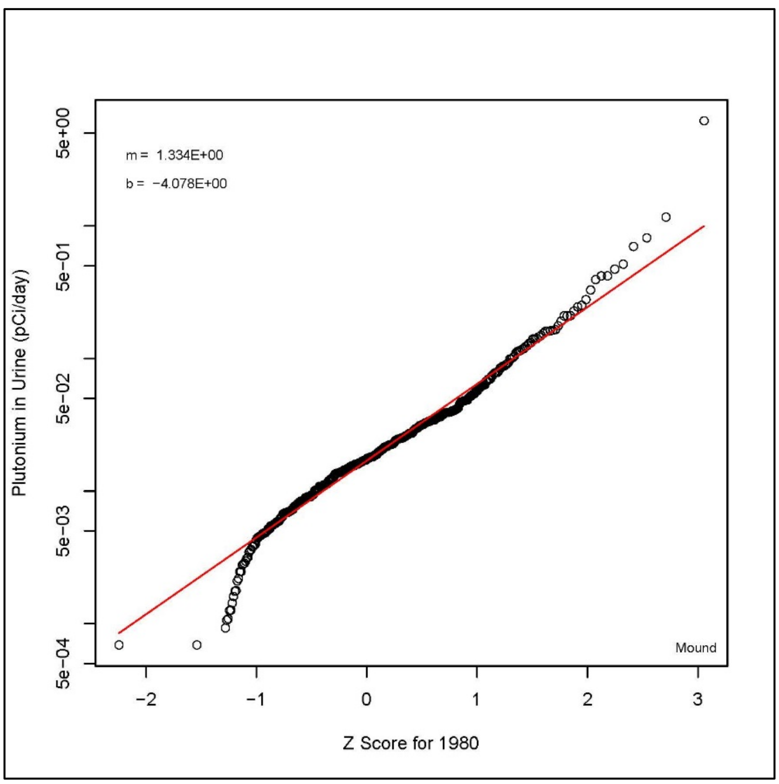
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



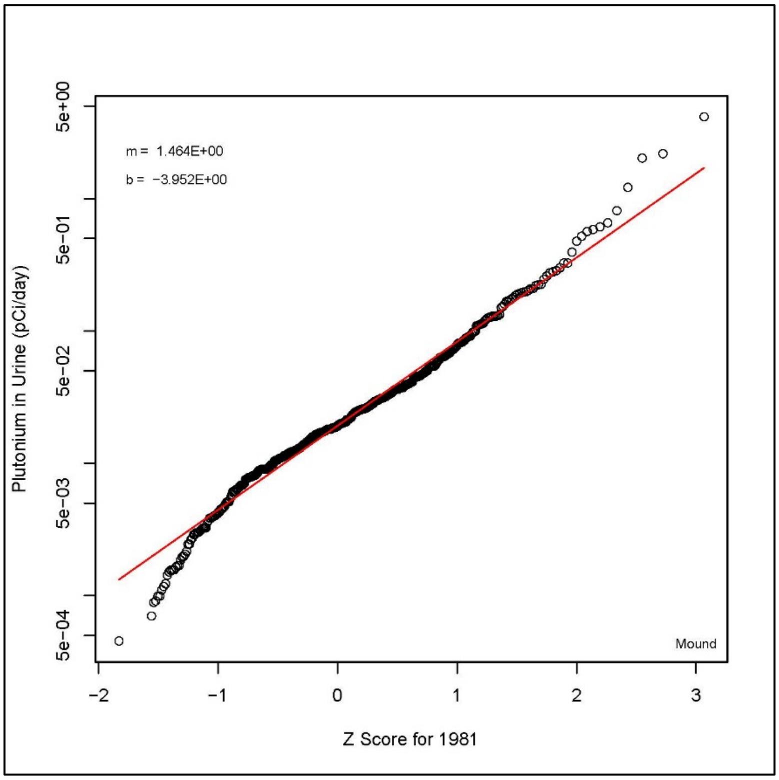


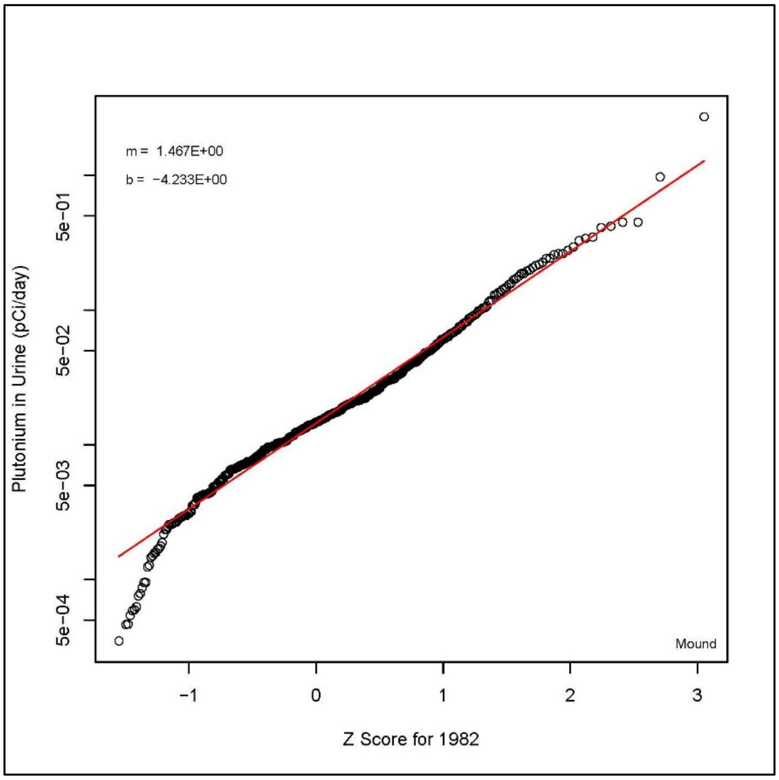
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



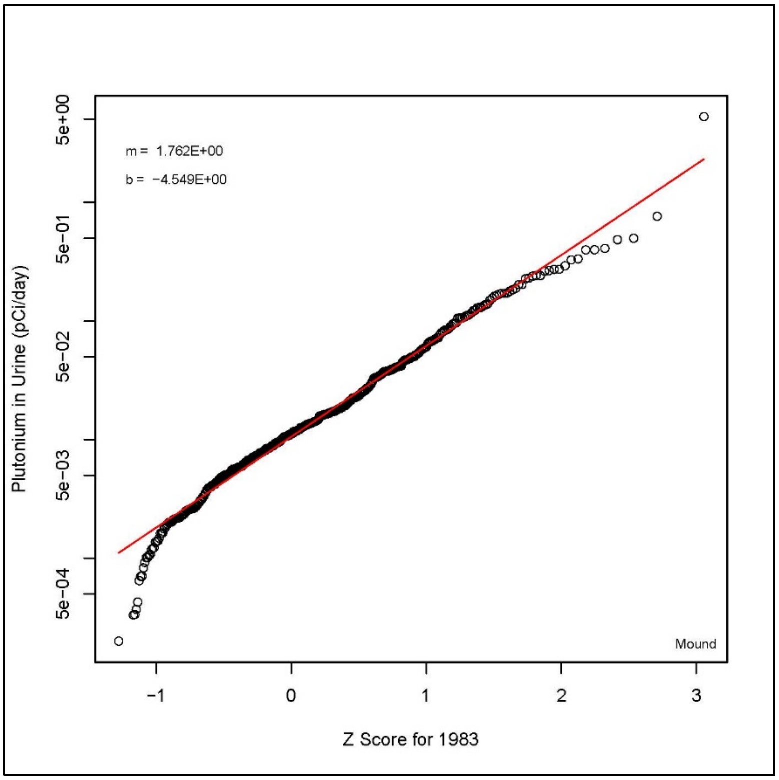


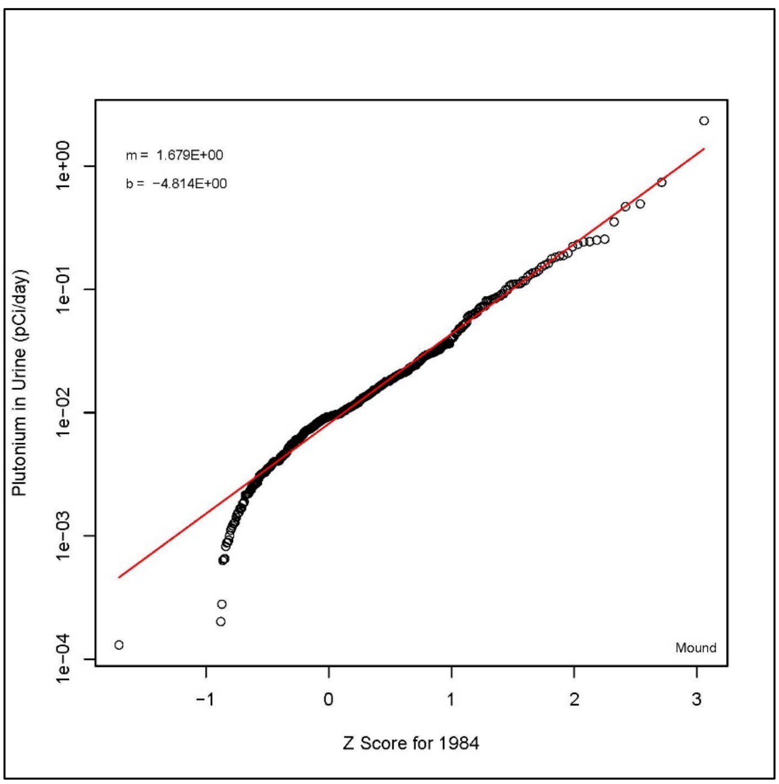
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



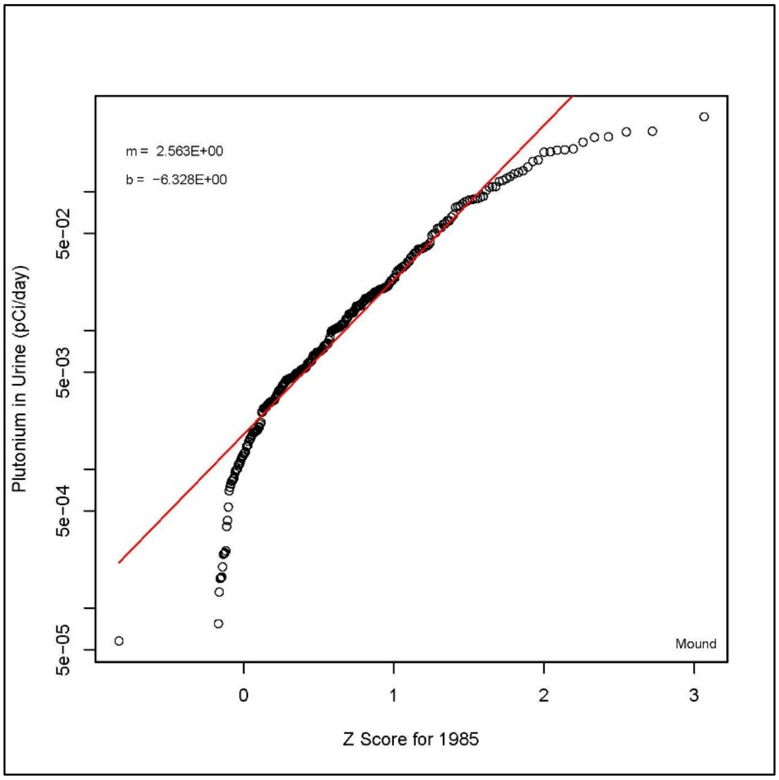


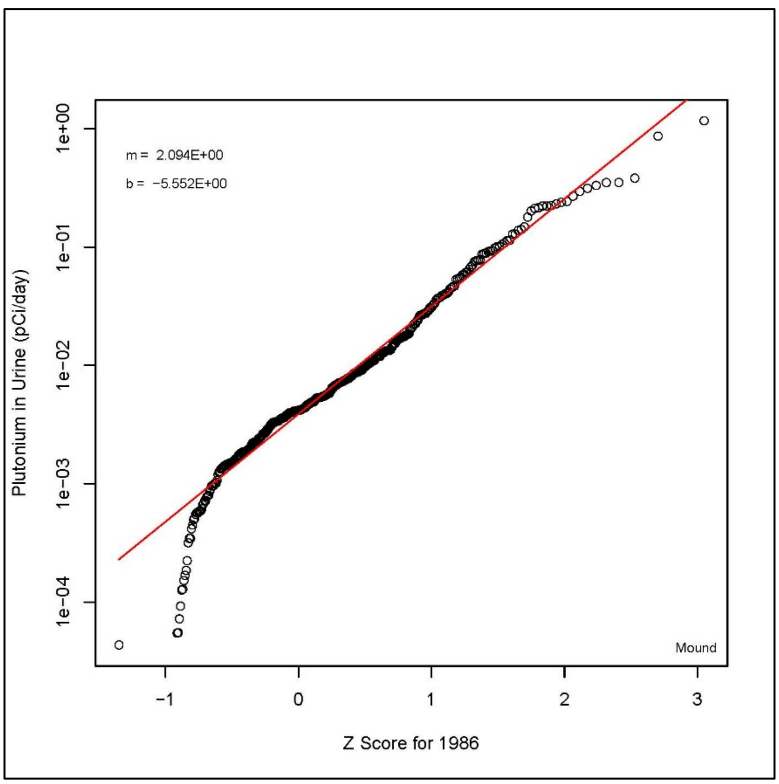
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



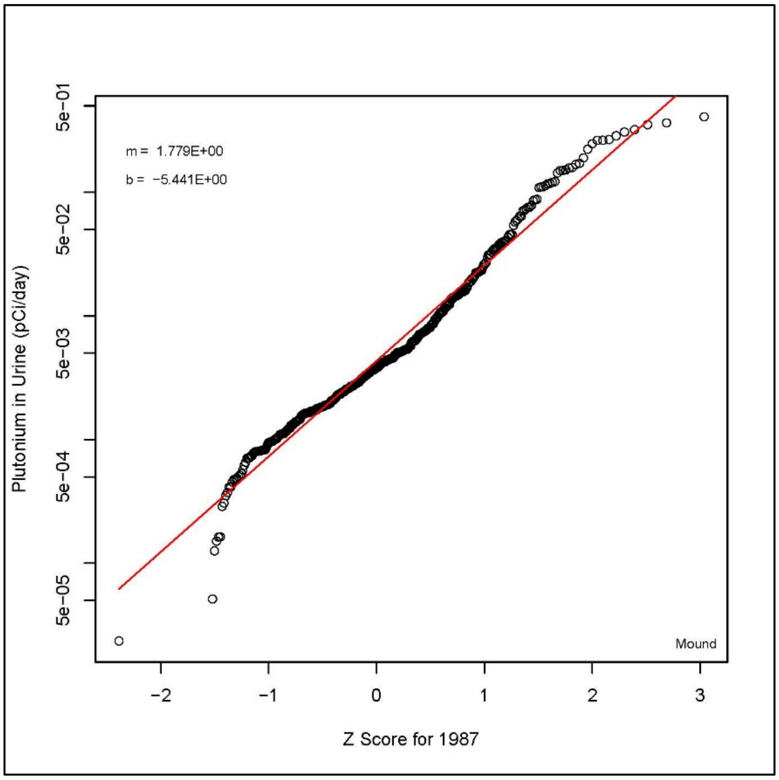


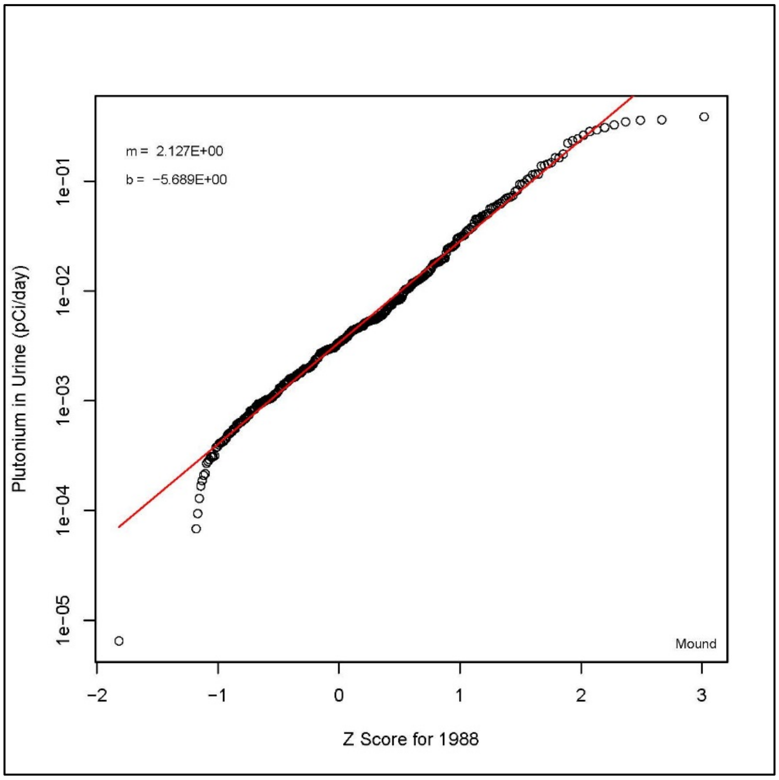
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



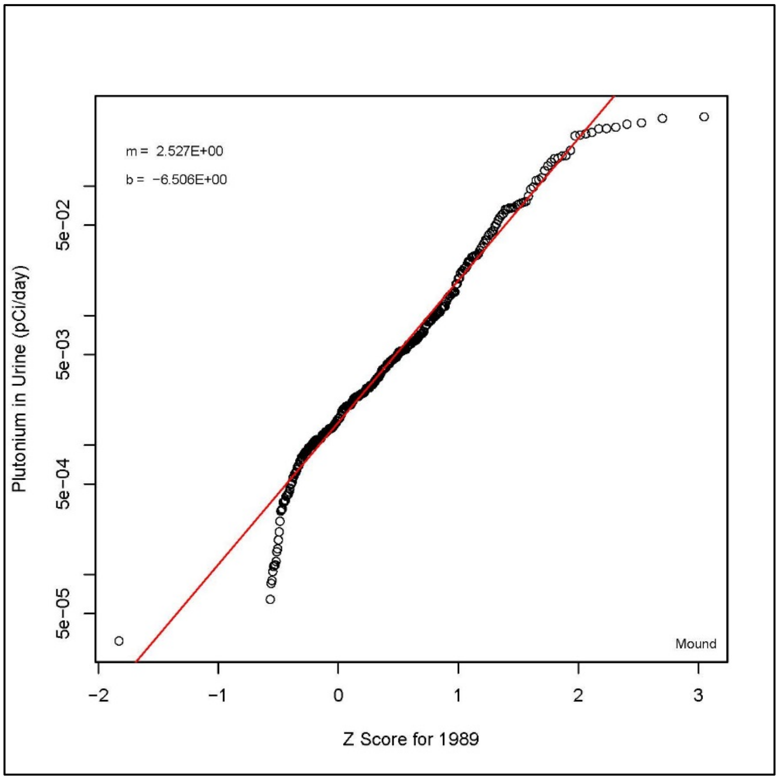


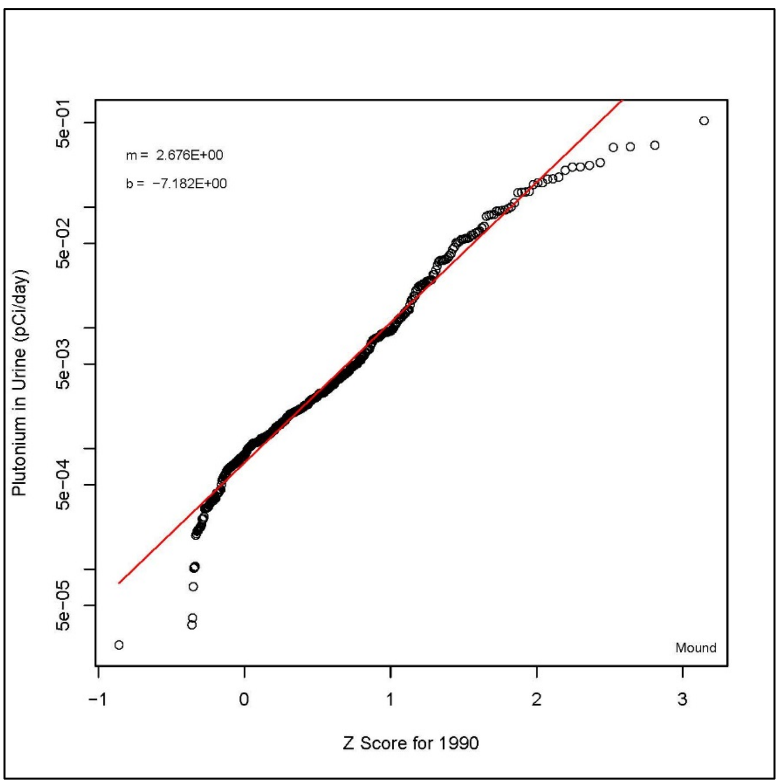
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)





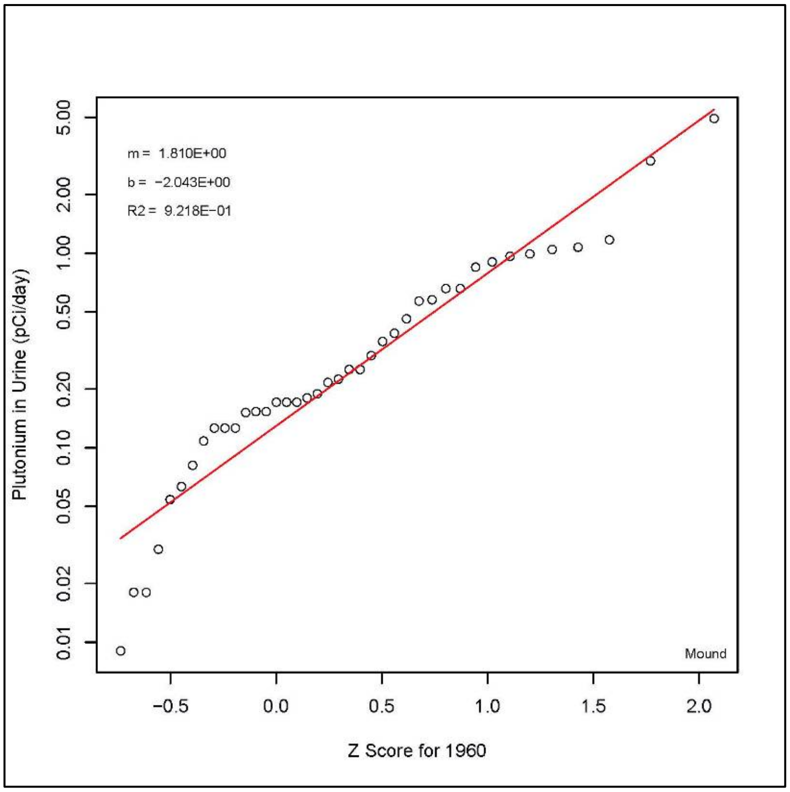
ATTACHMENT D
LOGNORMAL PROBABILITY PLOTS, COMPLETE MOUND DATASET (continued)



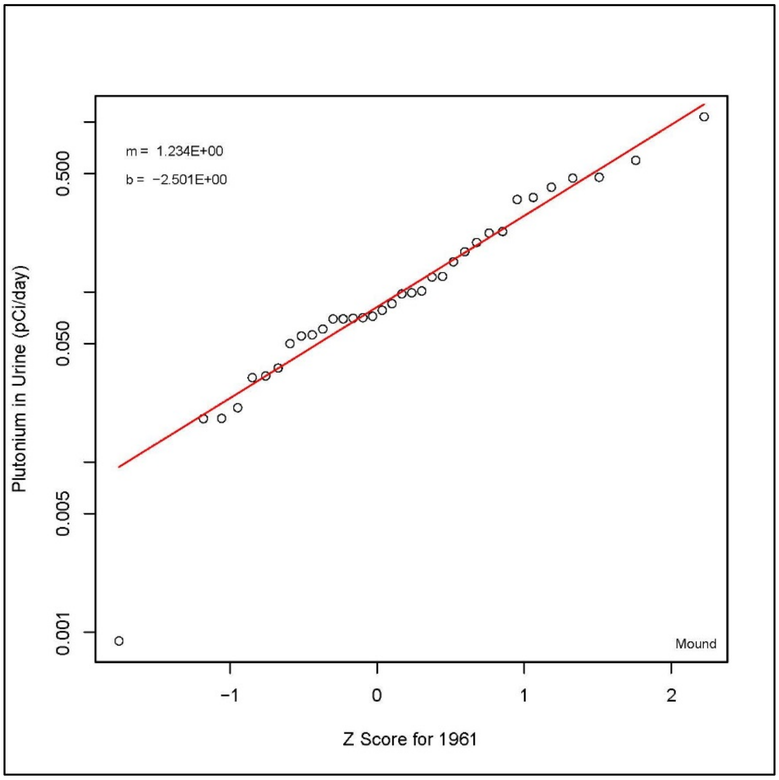


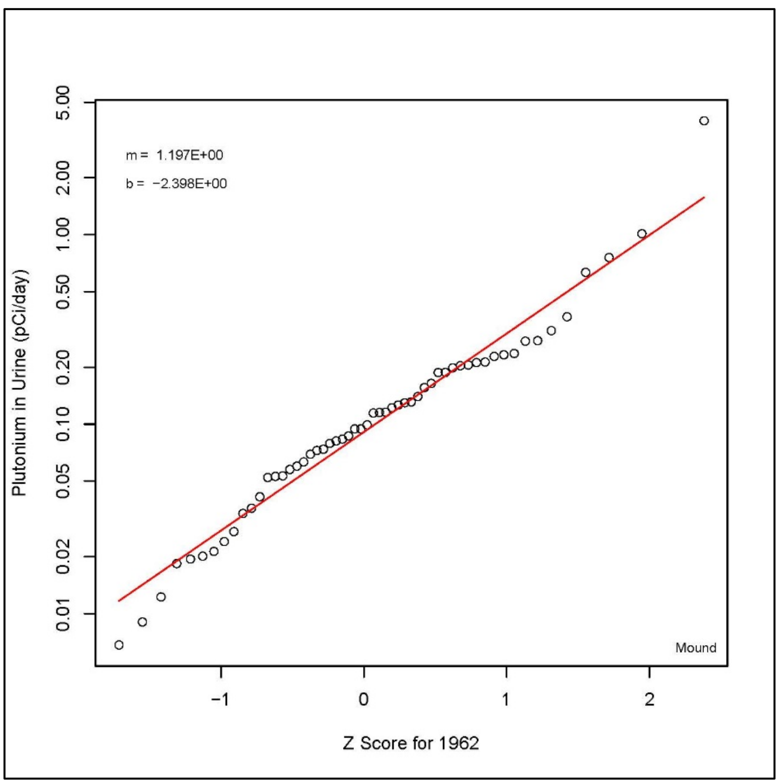
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET

This attachment presents lognormal probability plots for the claimant Mound plutonium urine bioassay dataset for 1960 to 1990. The slope and intercept for the line of best fit to the data are given on the plot.

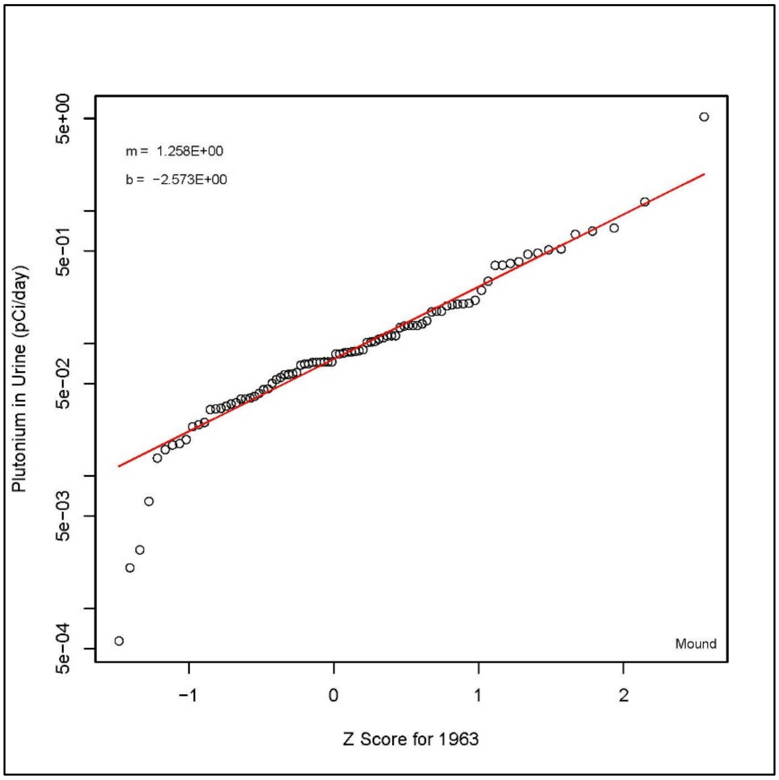


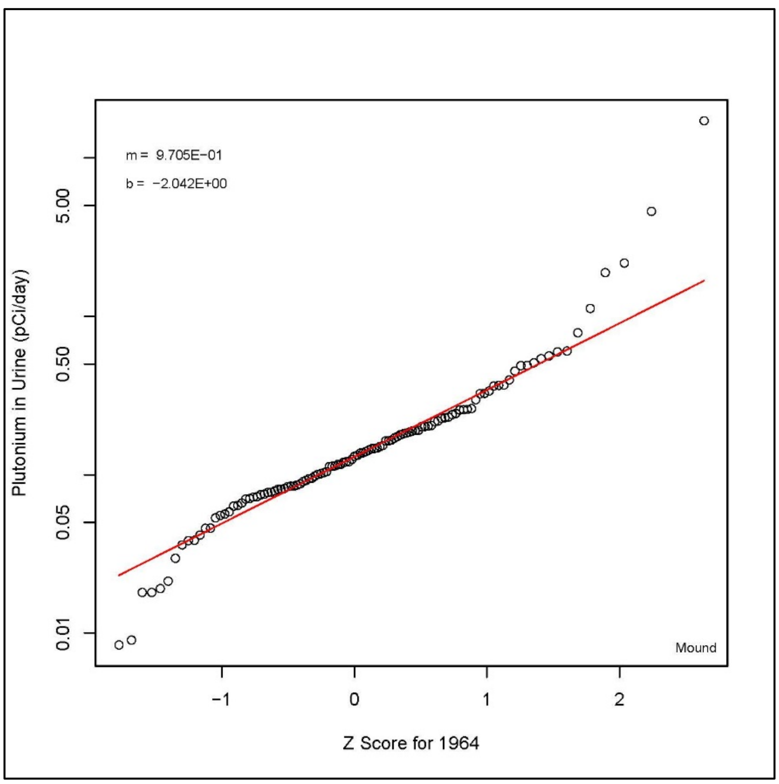
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



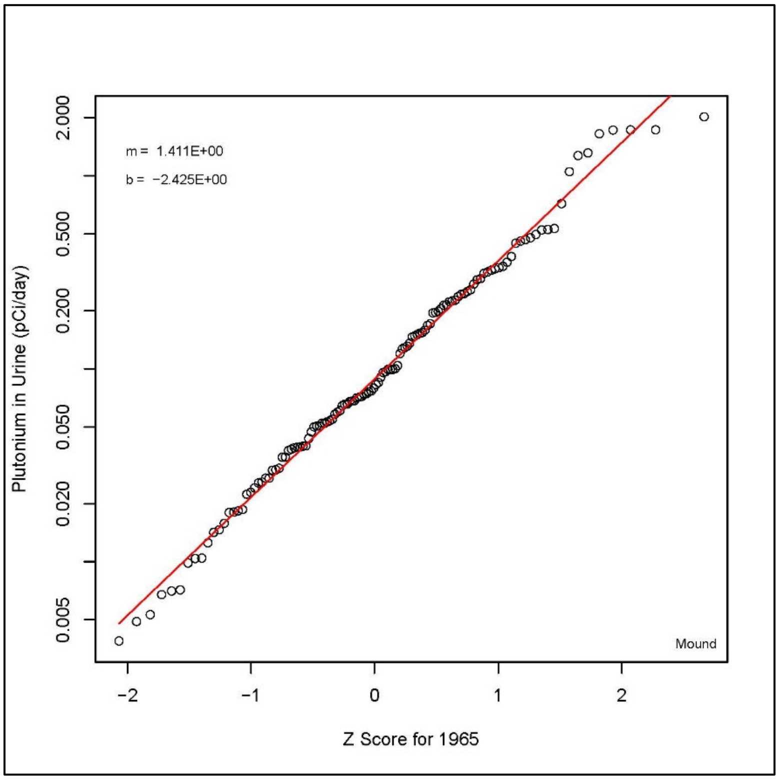


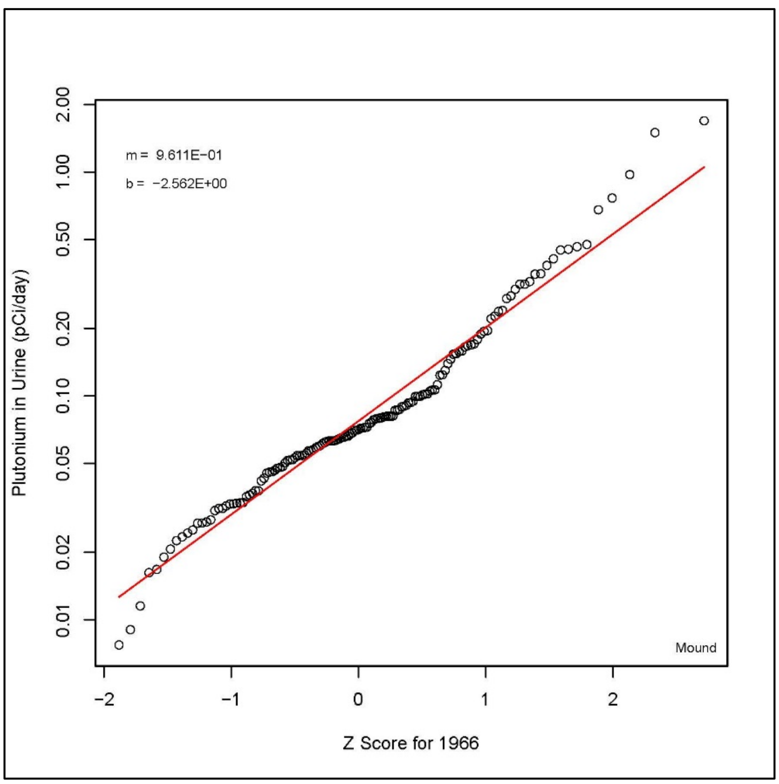
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



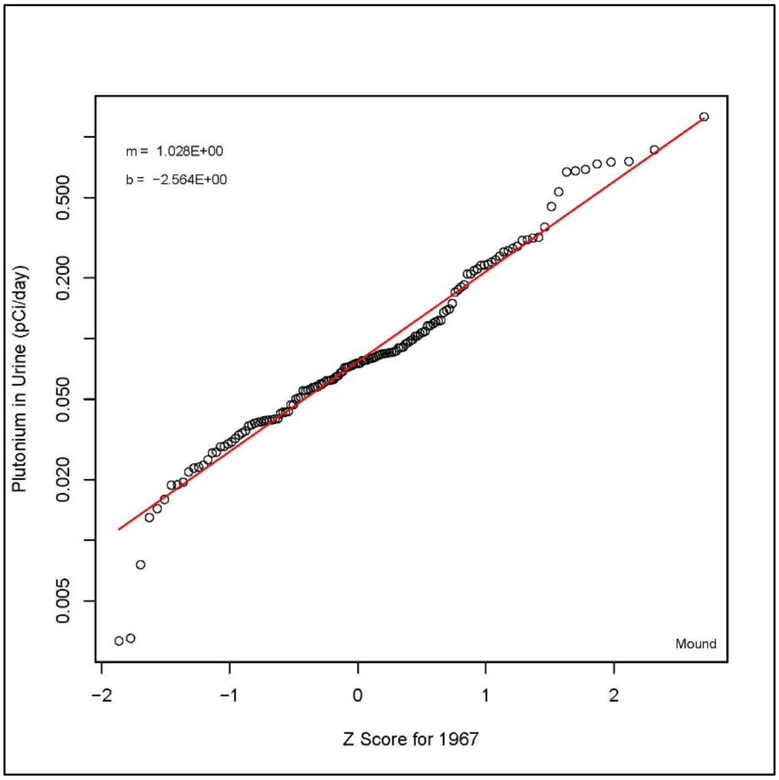


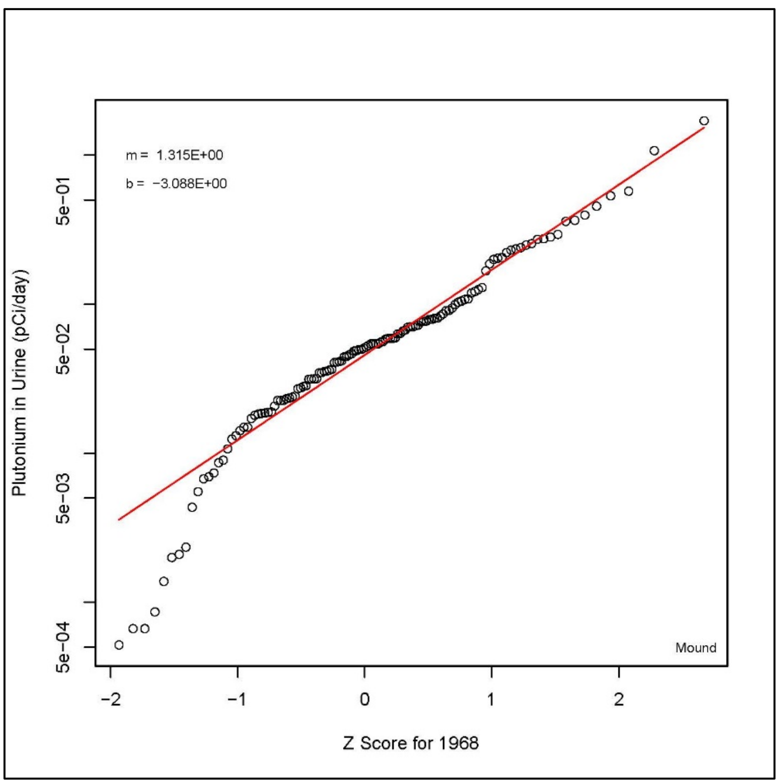
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



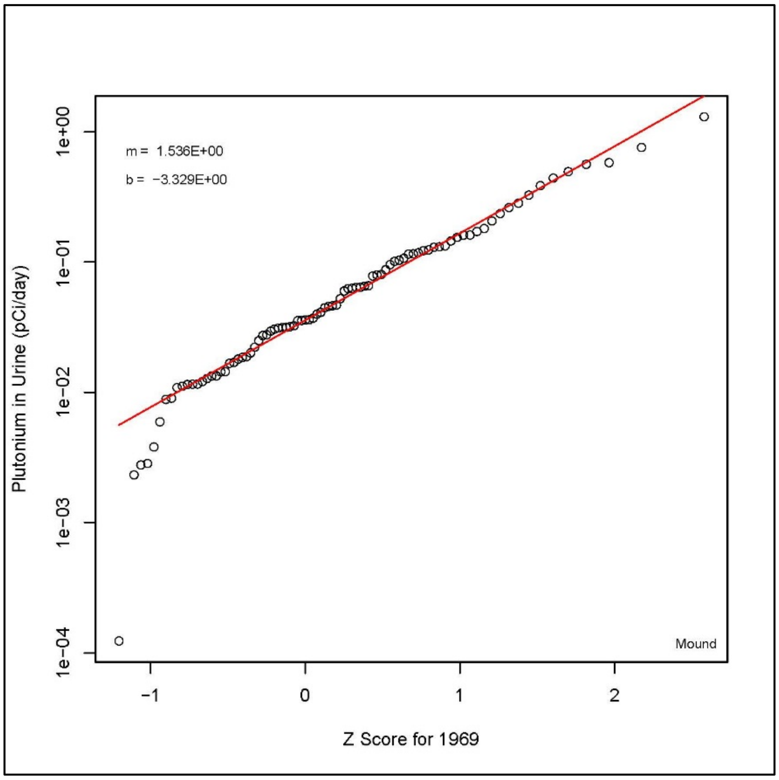


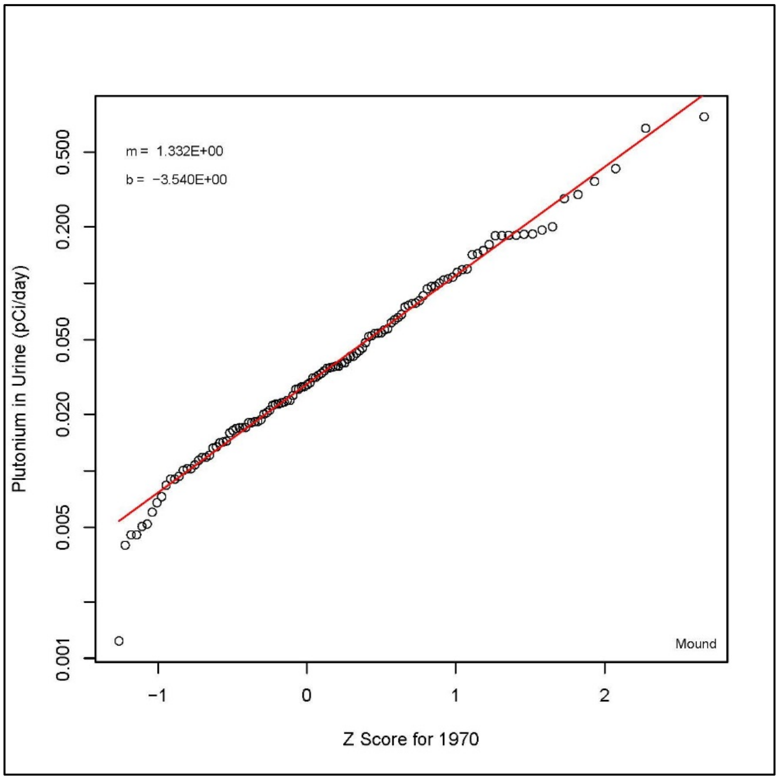
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



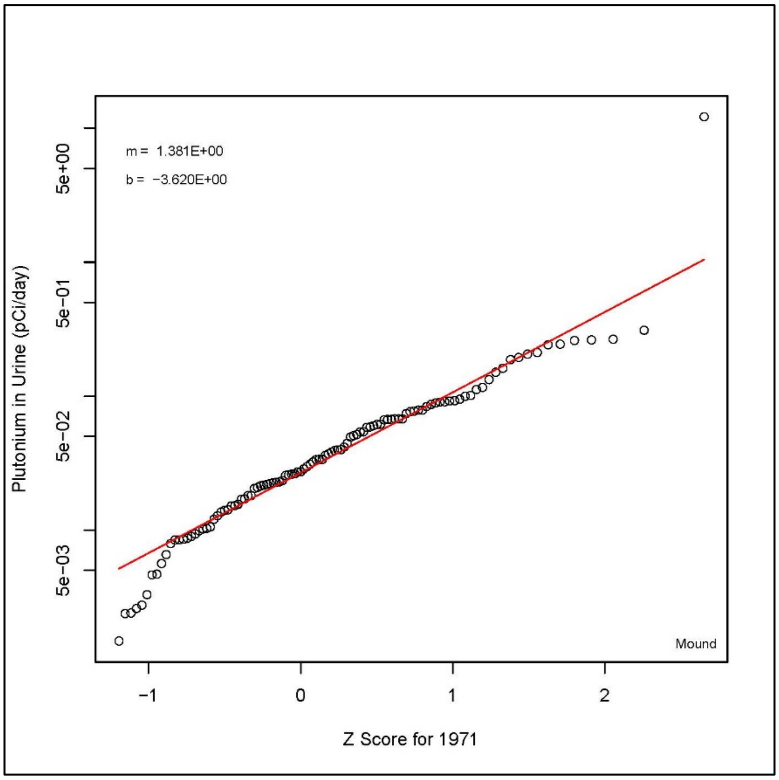


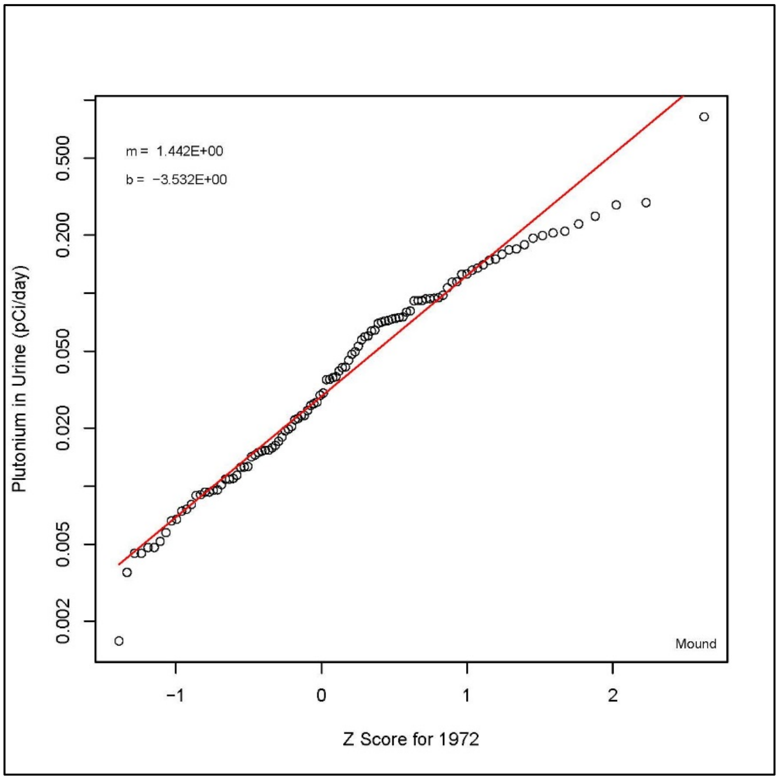
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



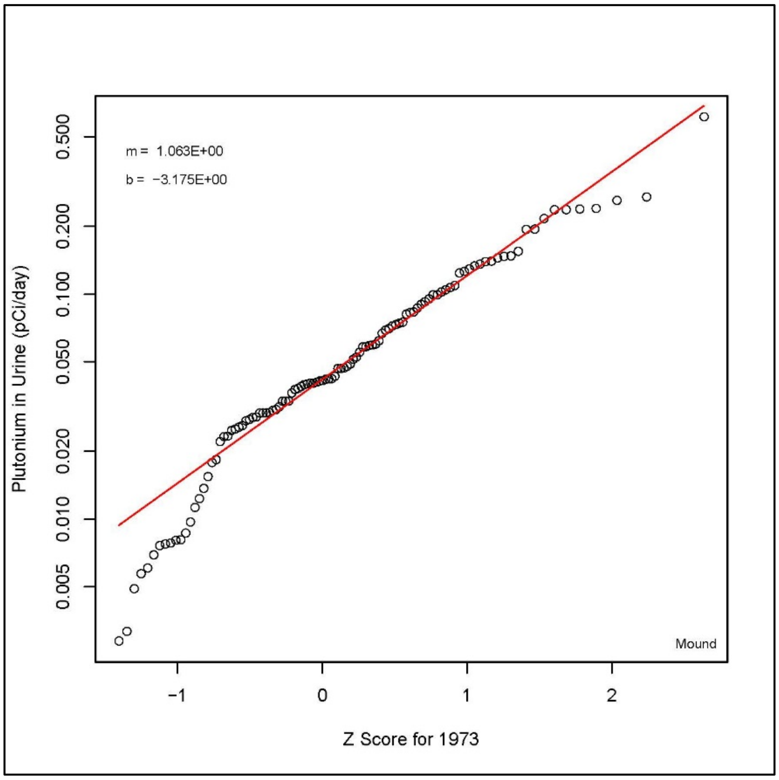


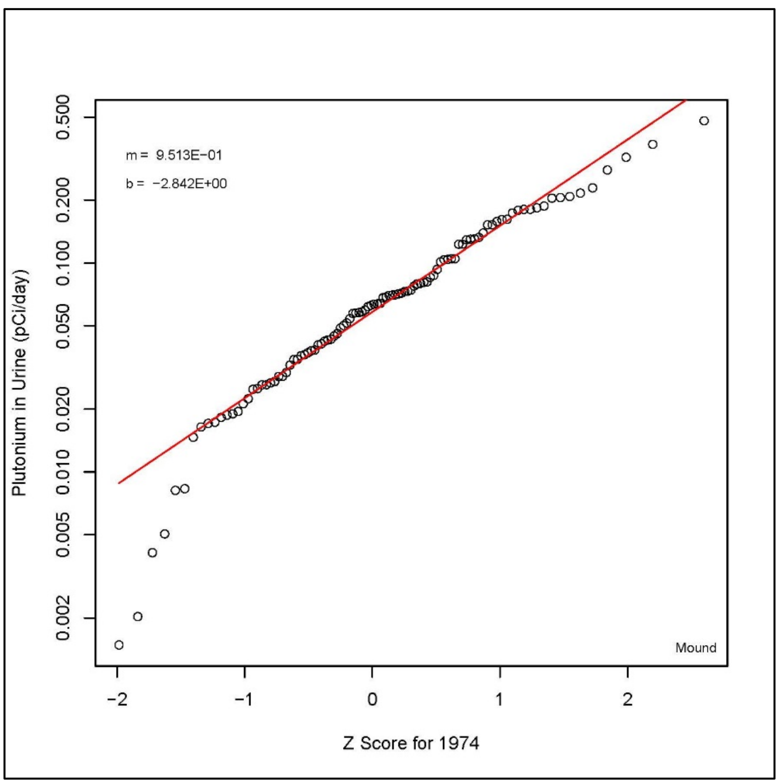
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



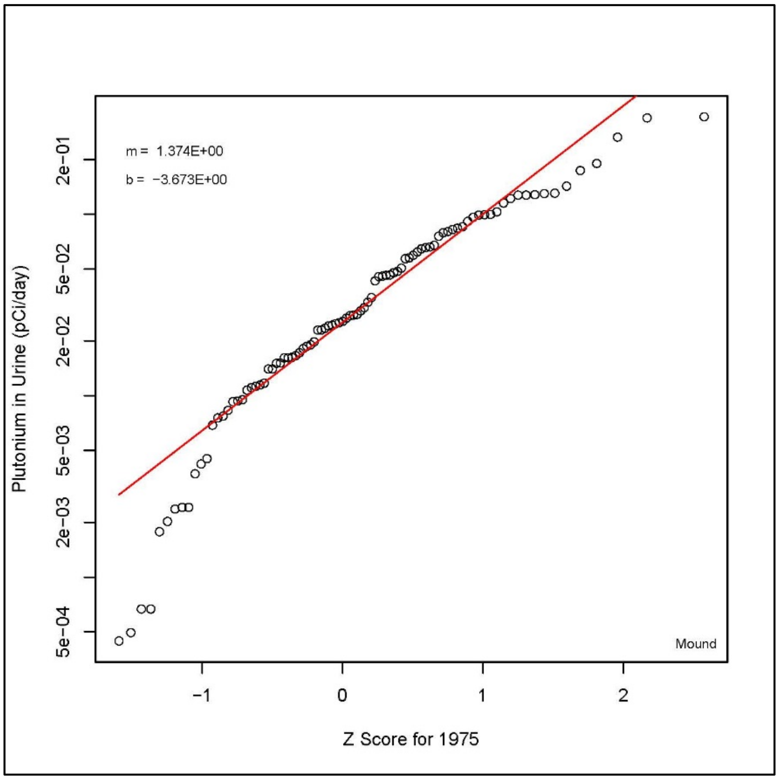


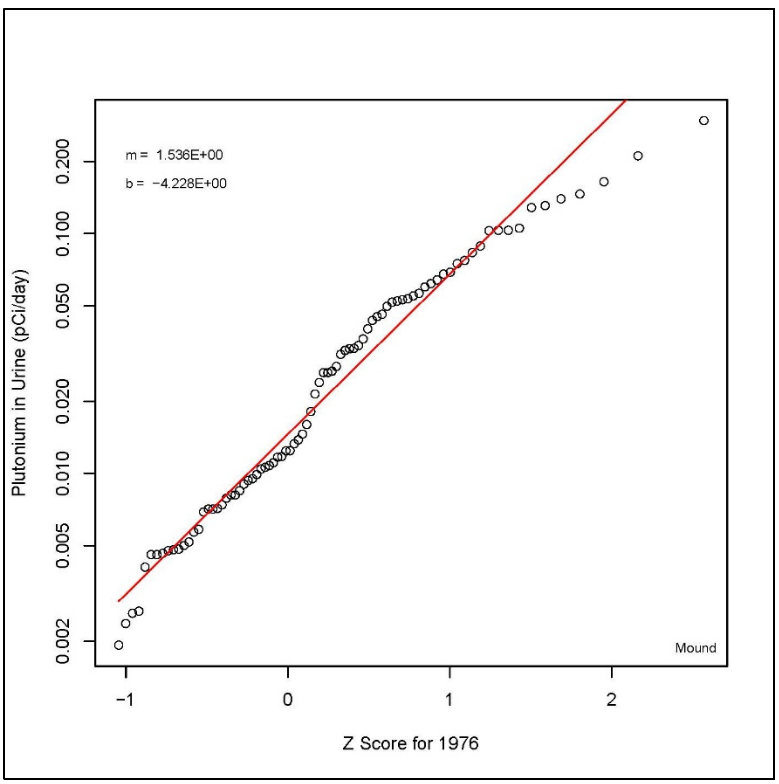
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



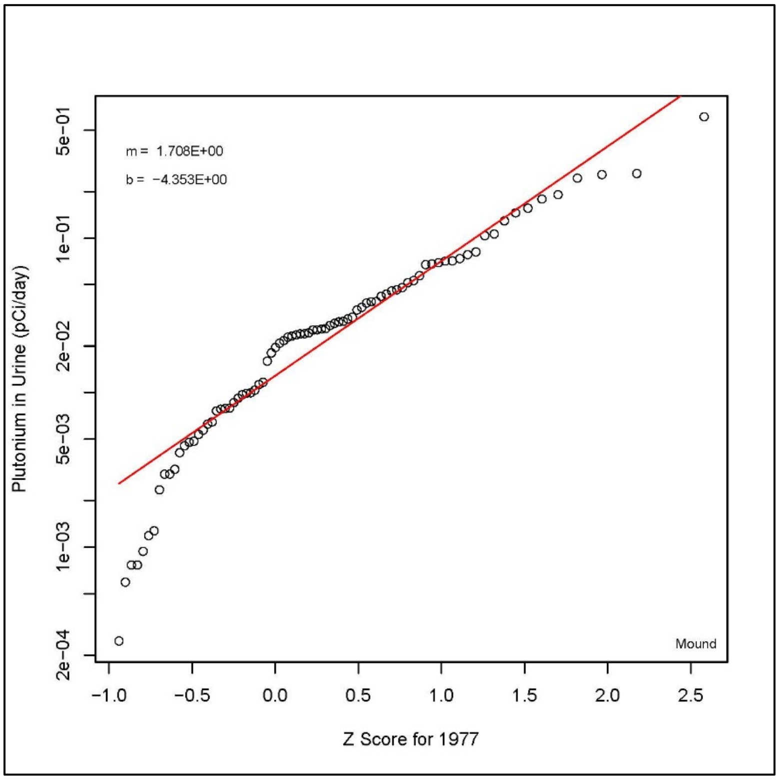


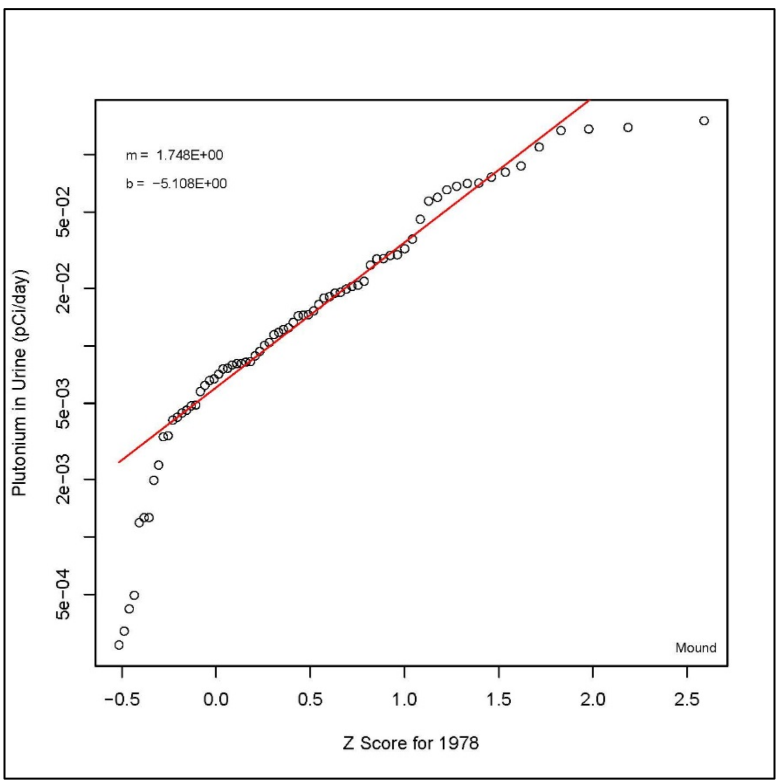
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



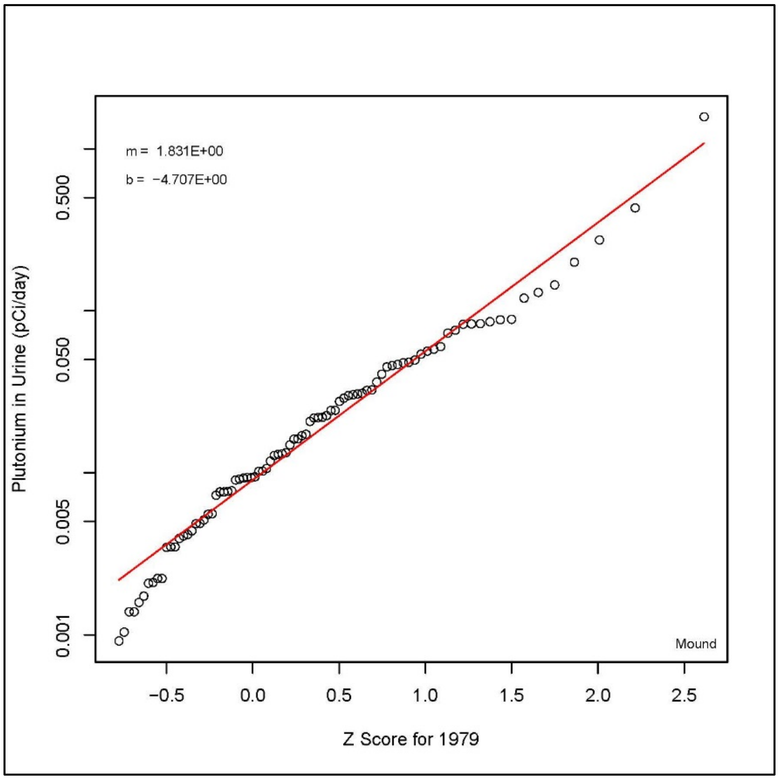


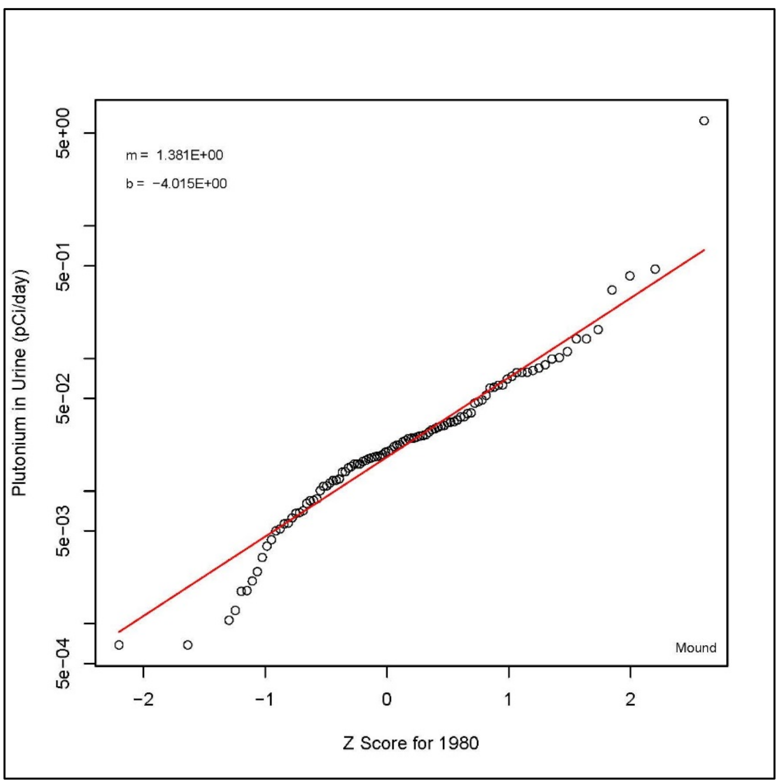
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



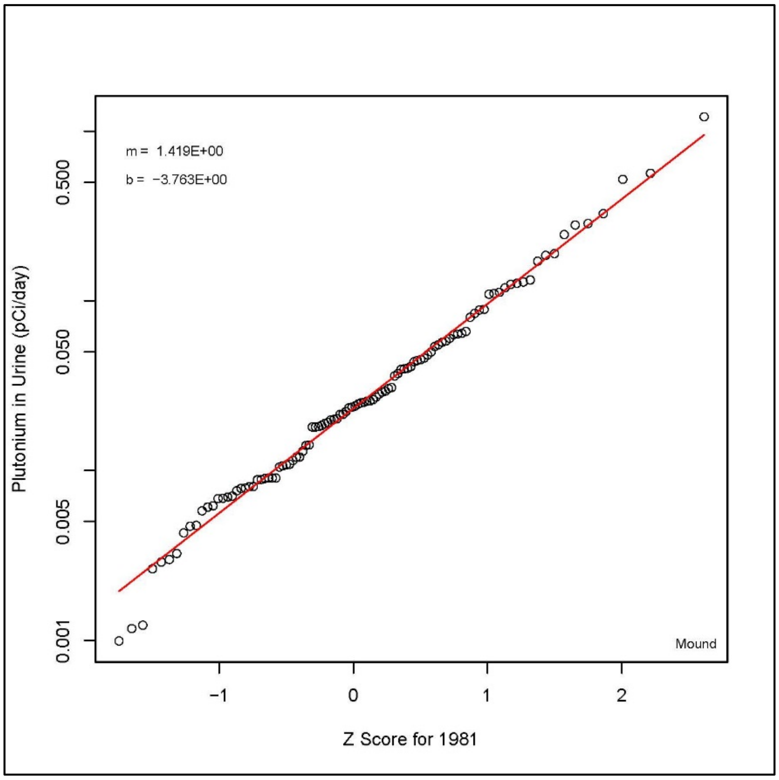


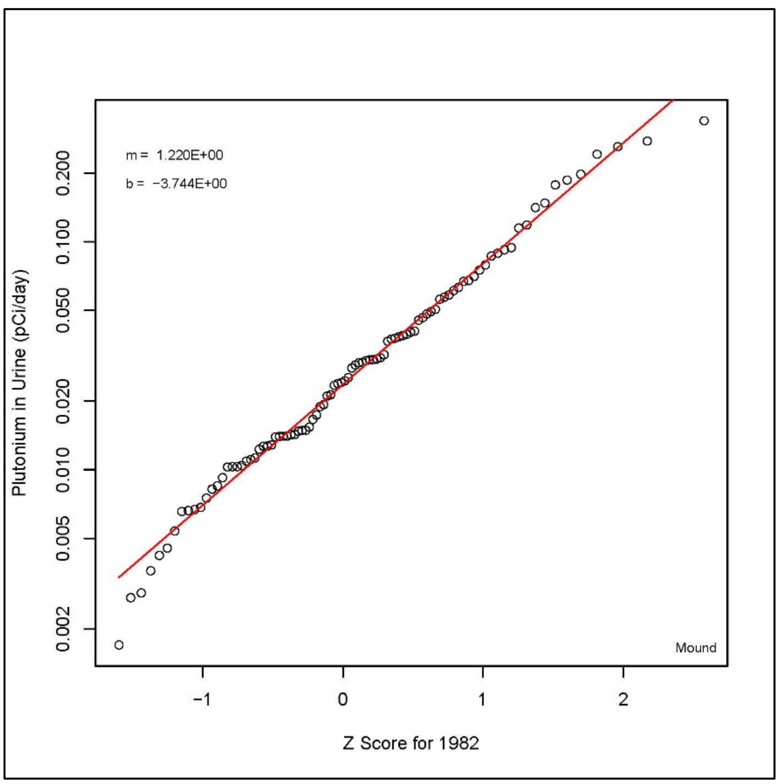
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



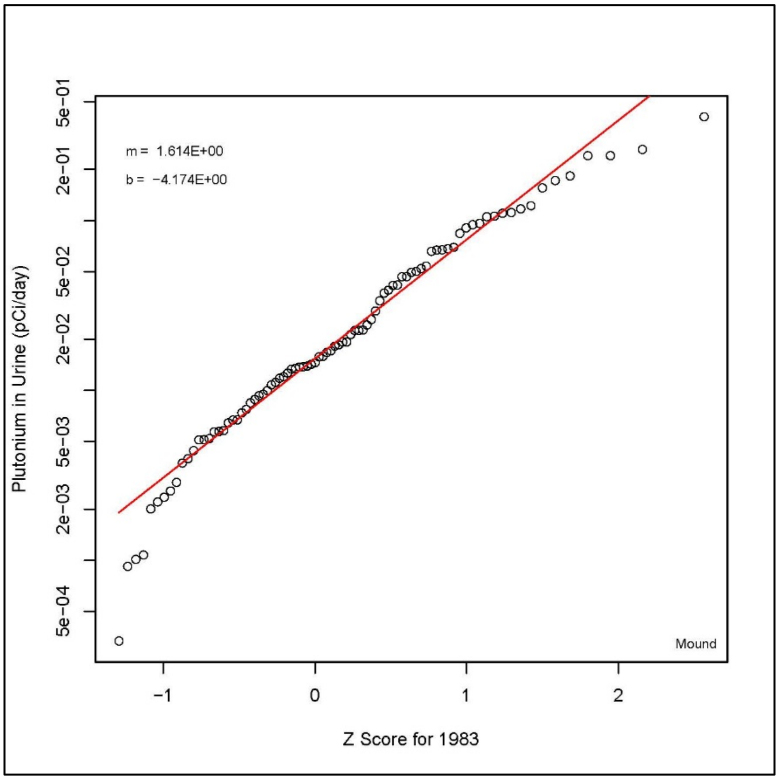


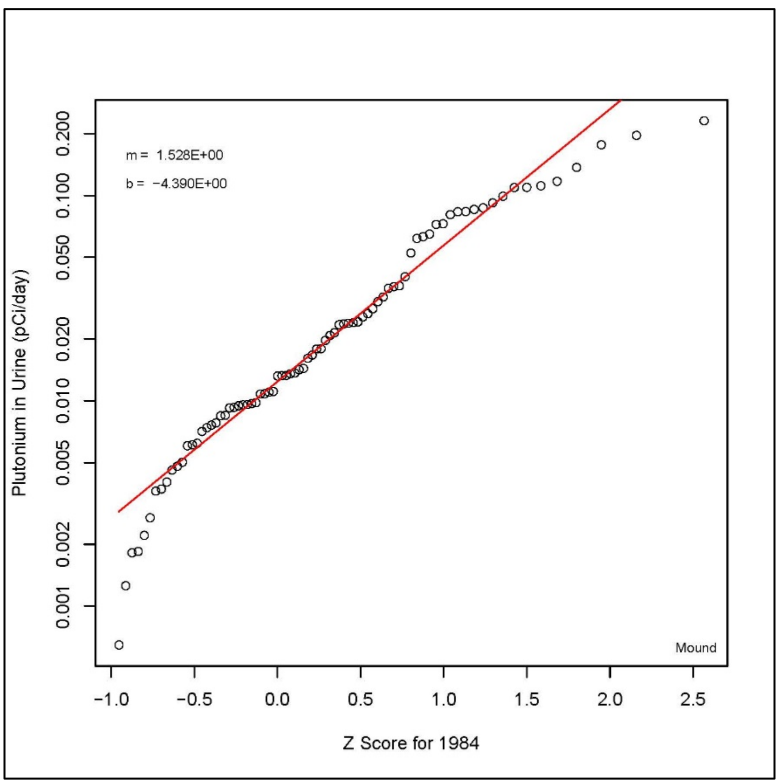
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



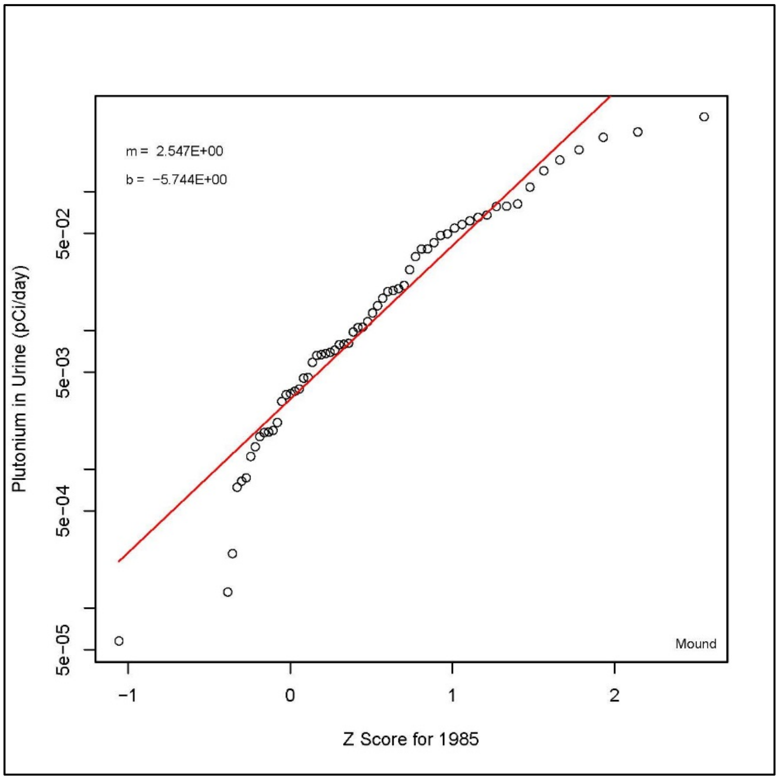


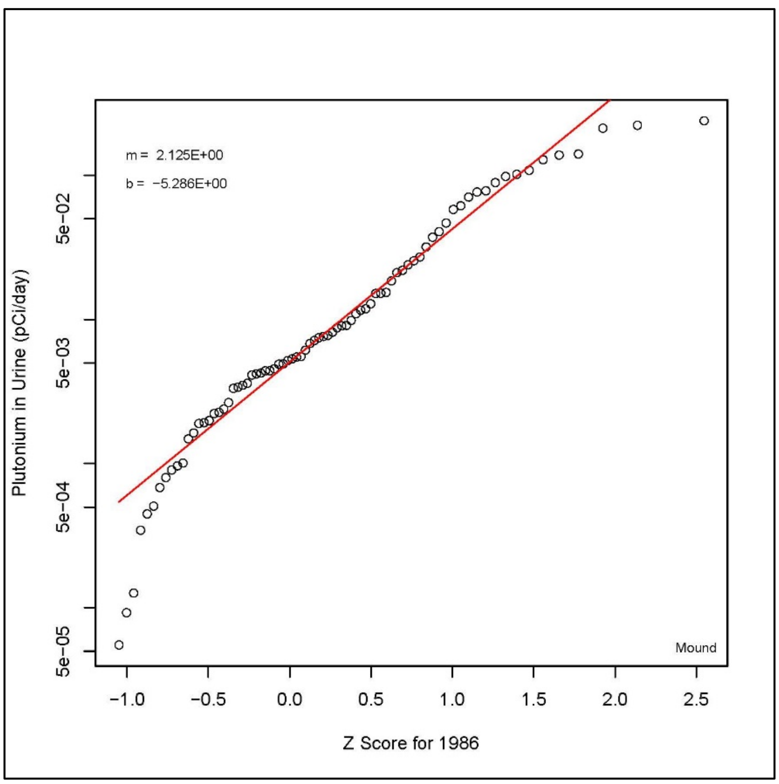
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



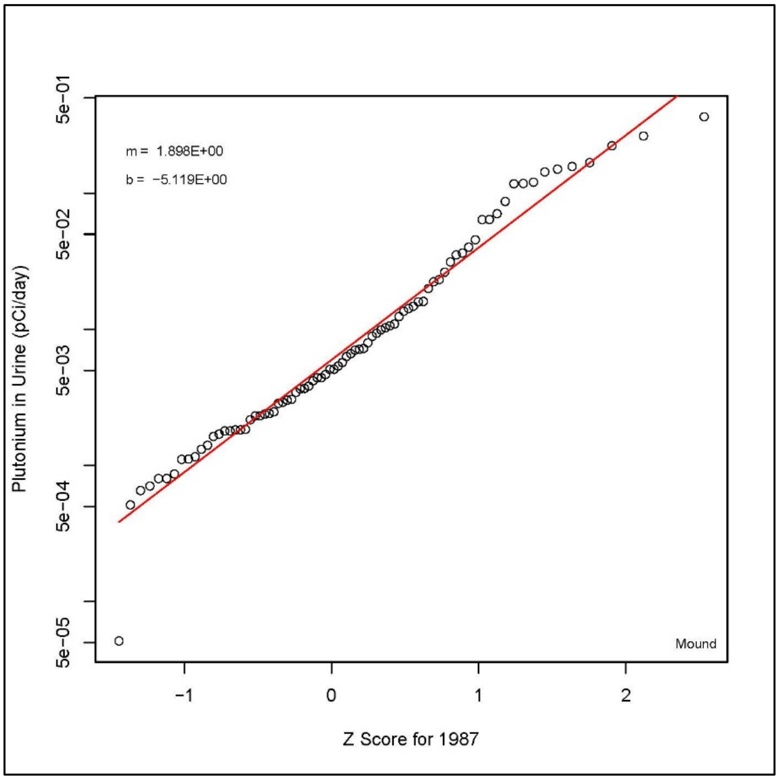


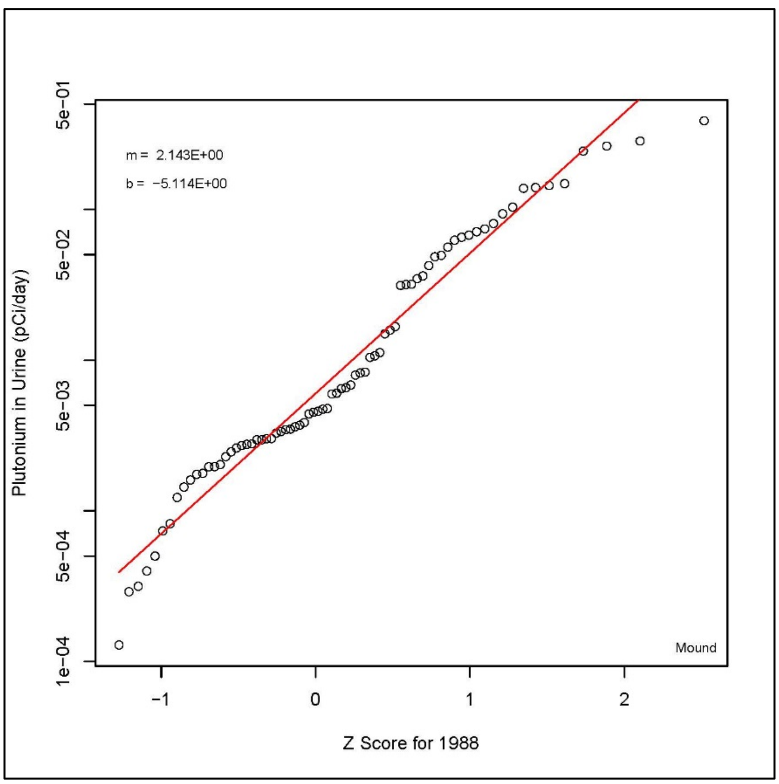
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



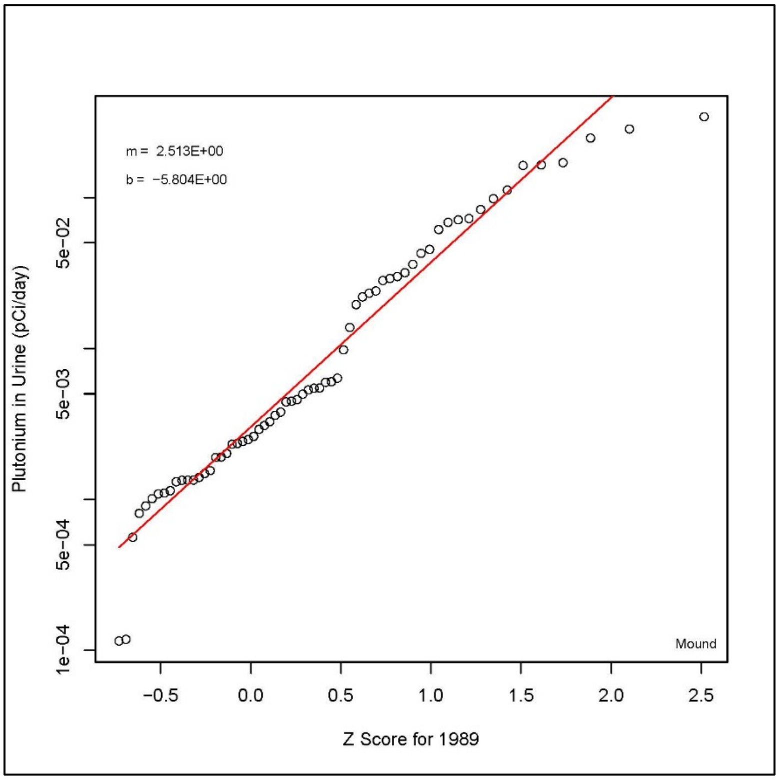


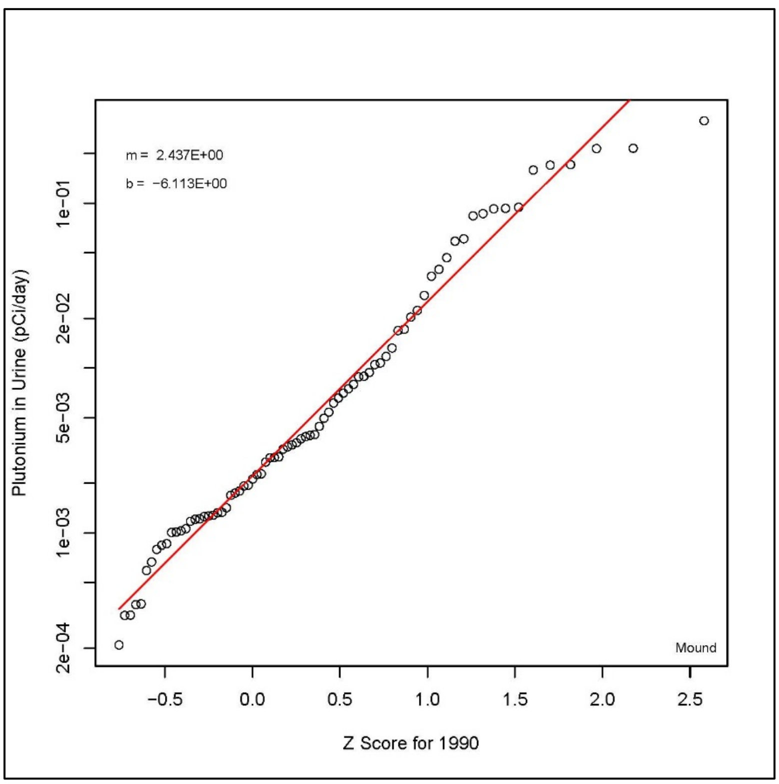
ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)





ATTACHMENT E
LOGNORMAL PROBABILITY PLOTS, CLAIMANT MOUND DATASET (continued)



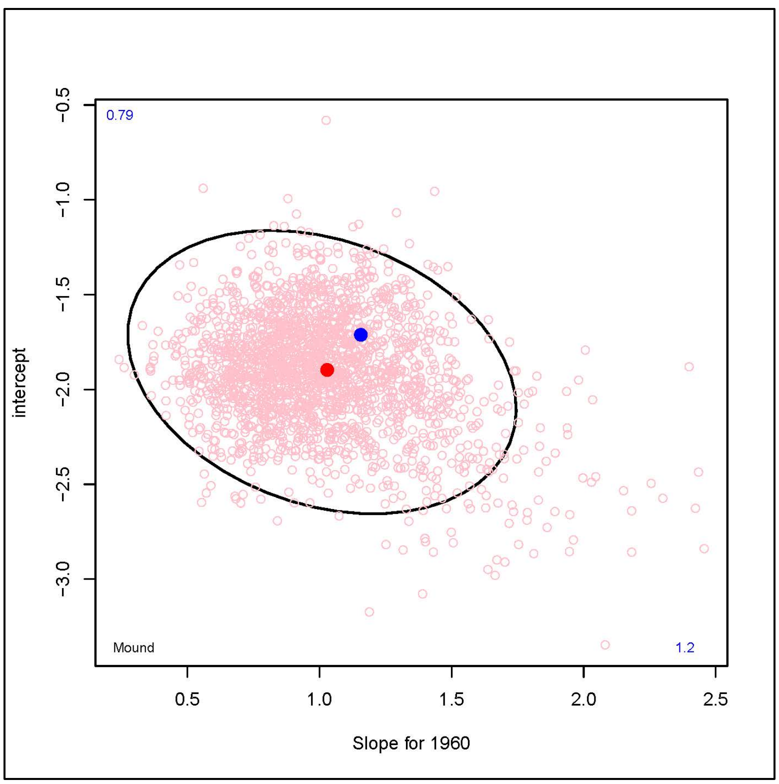


ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS

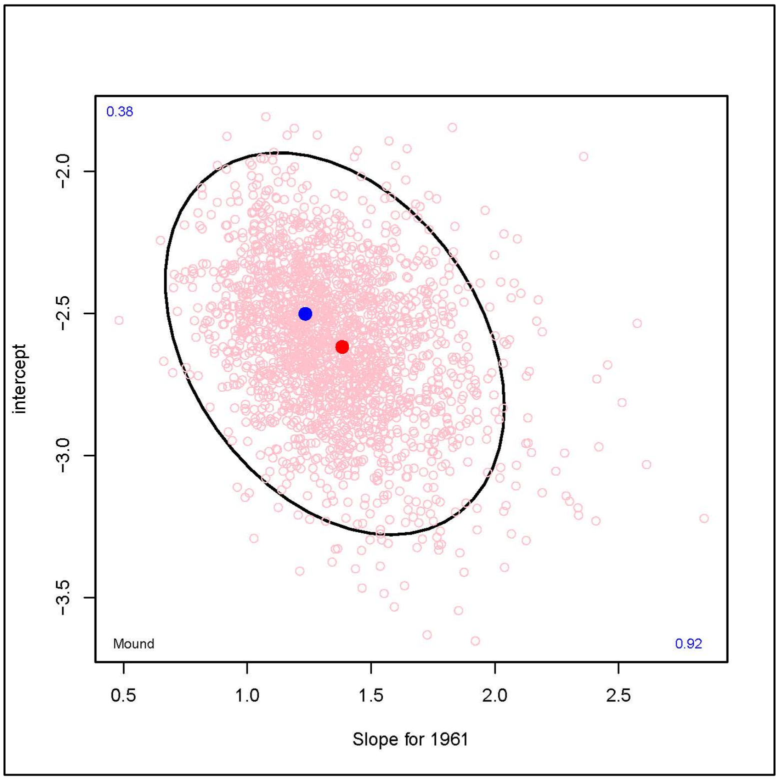
This attachment presents lognormal probability plots for a sample of 301 unique workers that was drawn at random without replacement from the 2,070 workers in the complete dataset. The slope and intercept of the plot for each year from 1960 to 1990 was calculated based on this data. This process was repeated m = 2,000 times to generate the 95% confidence ellipses. In each plot:

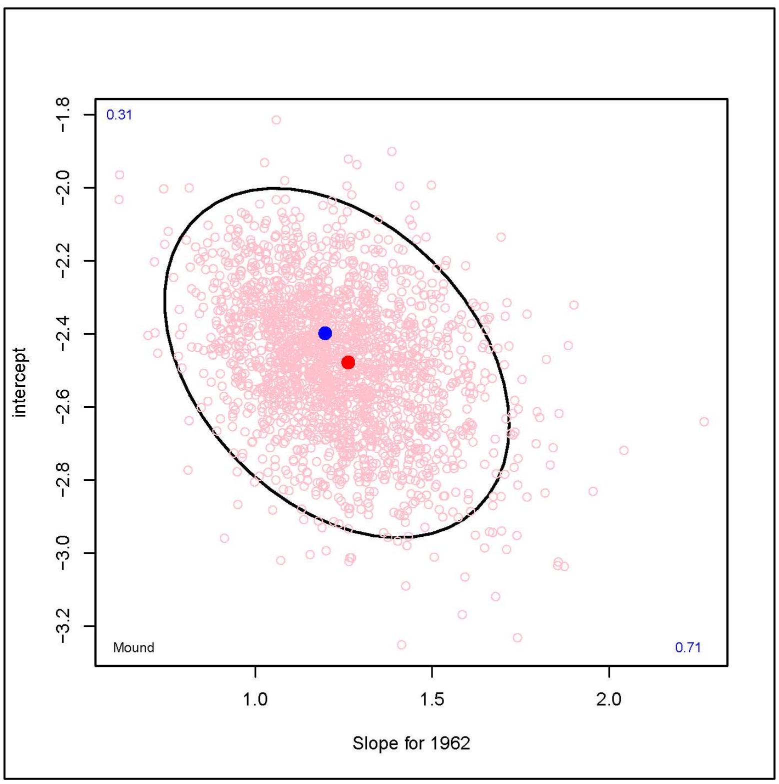
- The solid red point is the slope and intercept of the complete dataset.
- The solid blue point is the slope and intercept of the claimant dataset.
- The open pink points are the slopes and intercepts of the fits to the m random sets of 301 workers taken from the complete dataset.
- The ellipse is the 95% joint normal confidence ellipse.
- The numbers in the corners are the full scale of the respective axes in terms of +/- the slope or intercept of the complete dataset.

These plots are presented to support the idea that the claimant dataset is basically no different from a random draw from the complete dataset.

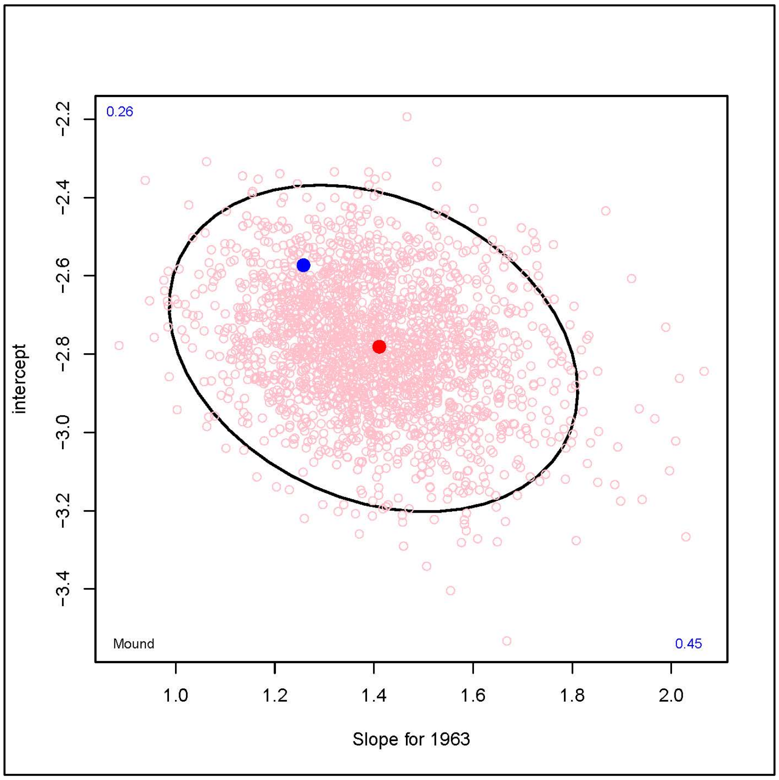


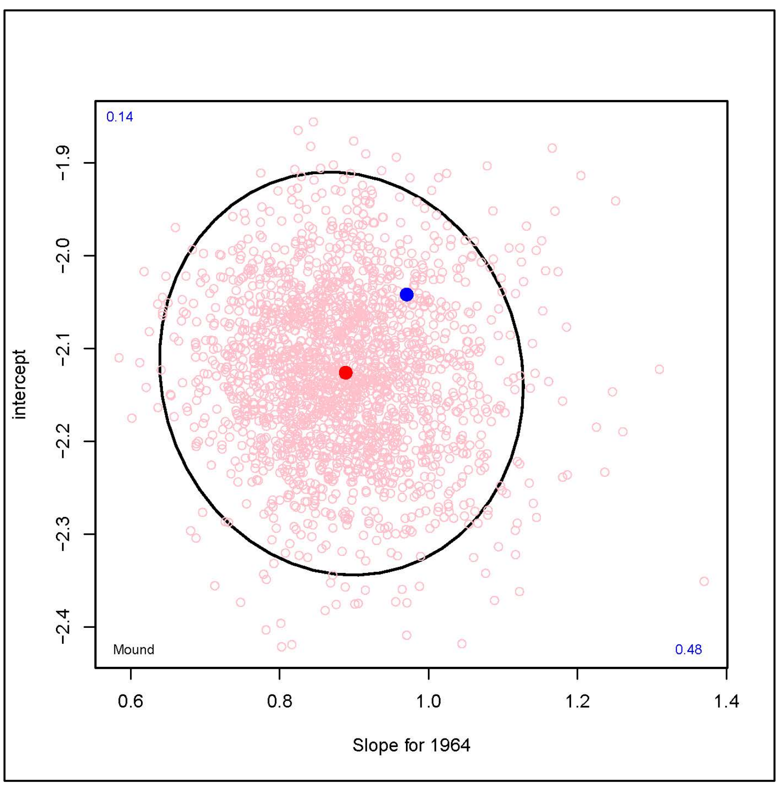
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



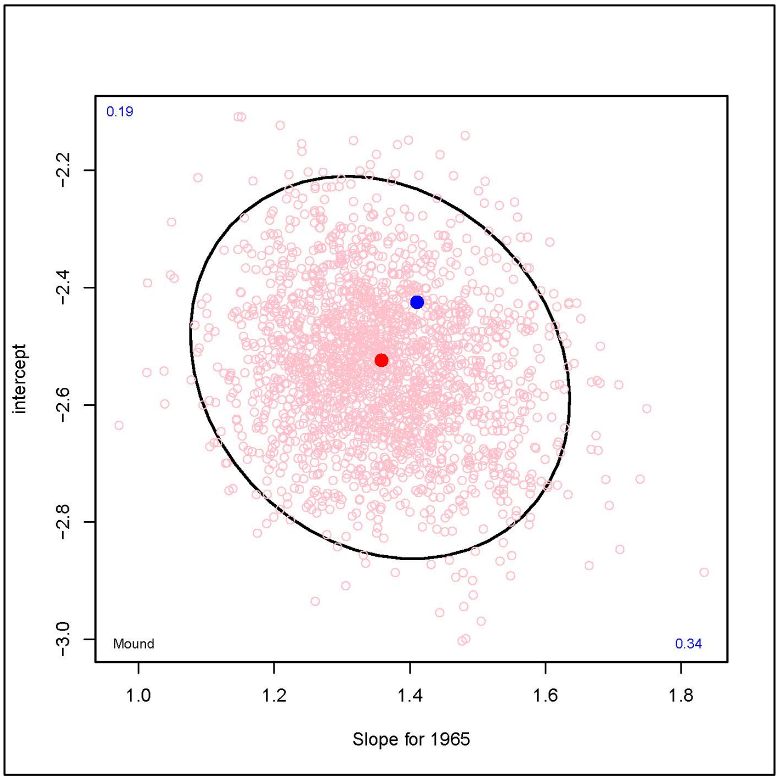


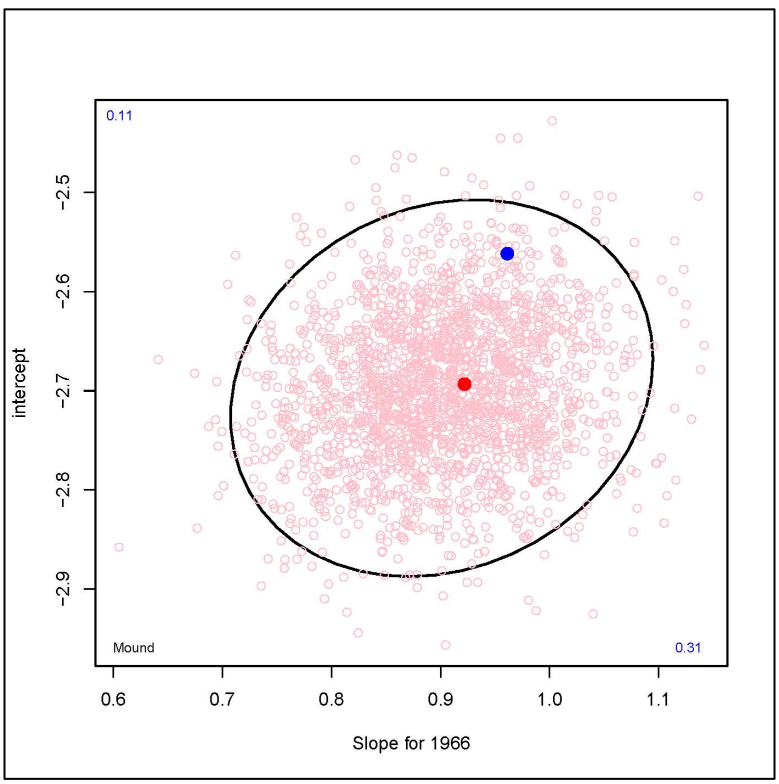
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



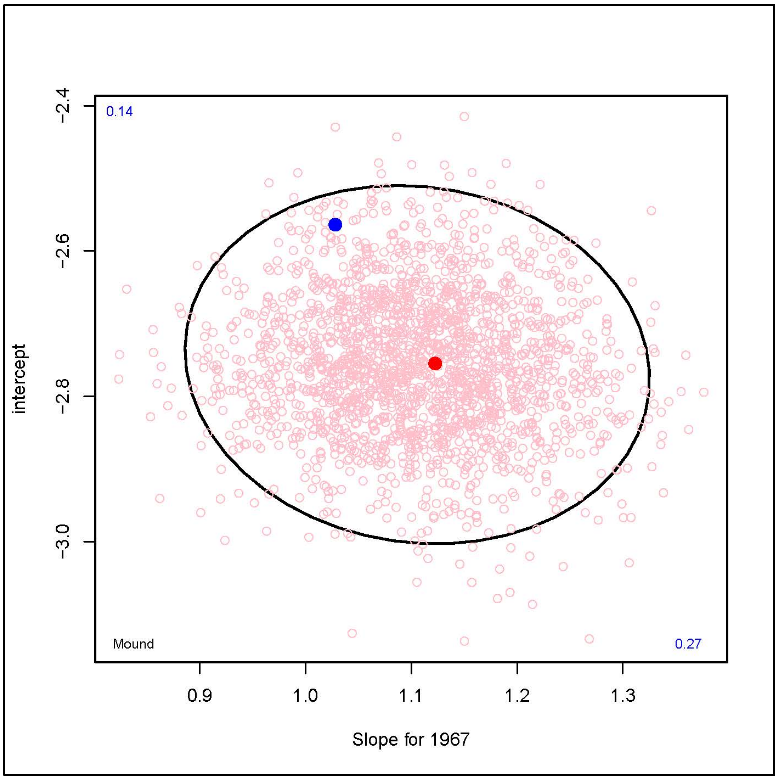


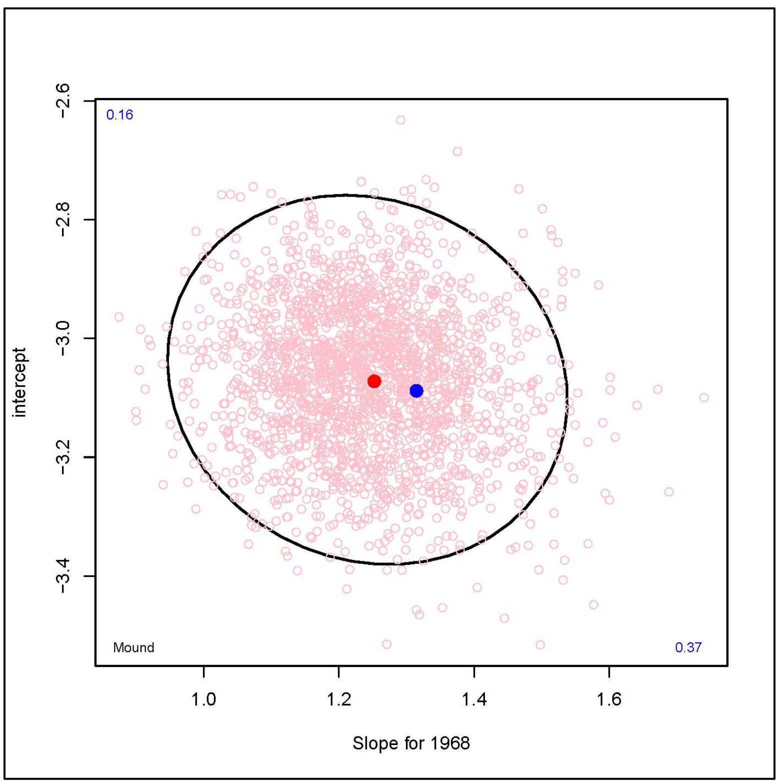
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



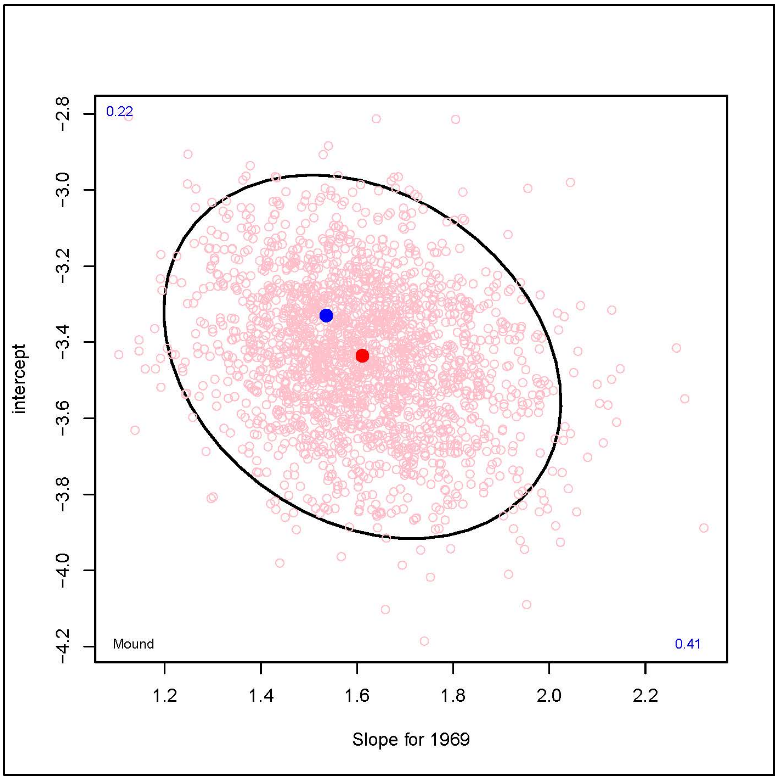


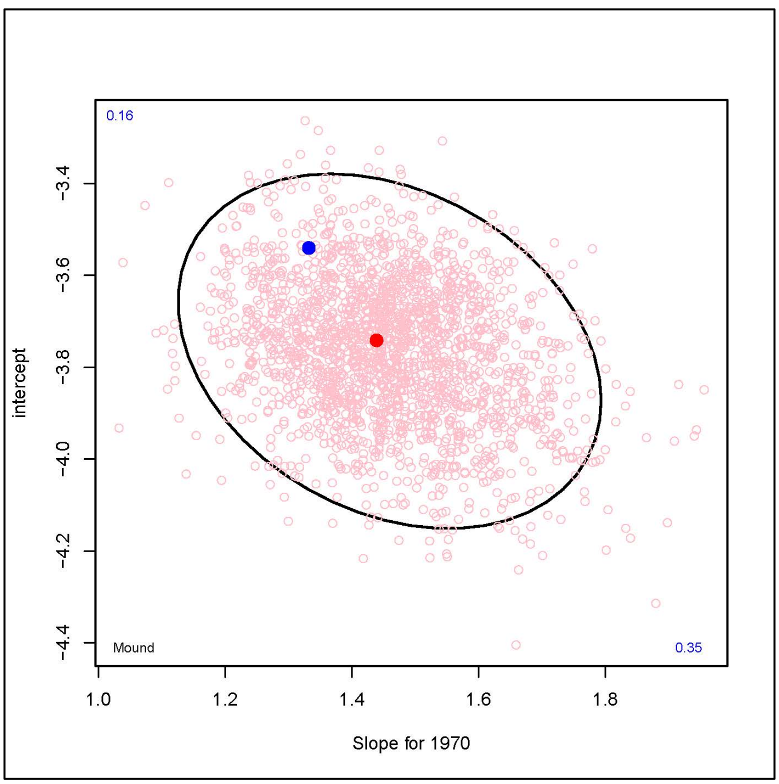
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



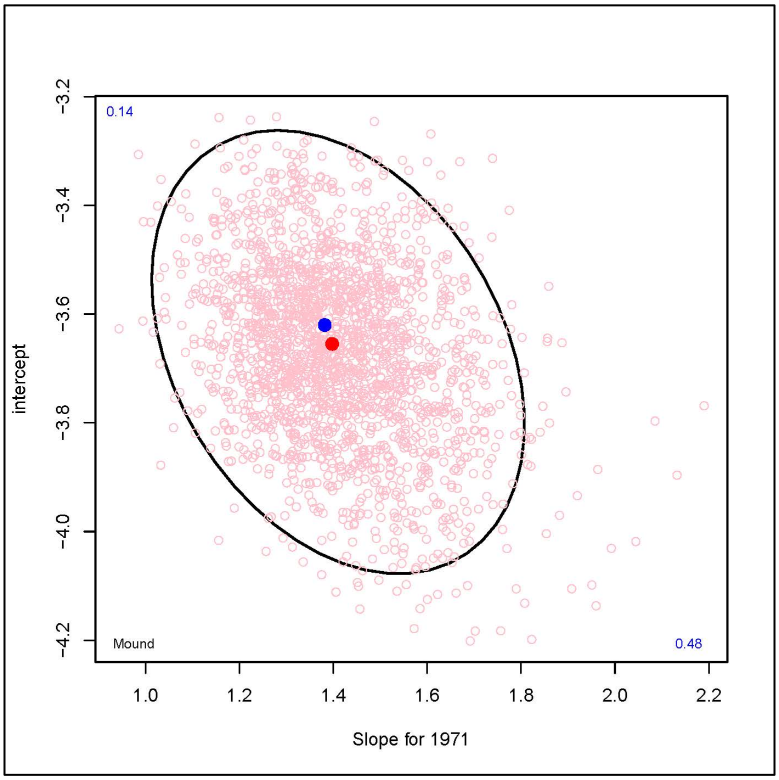


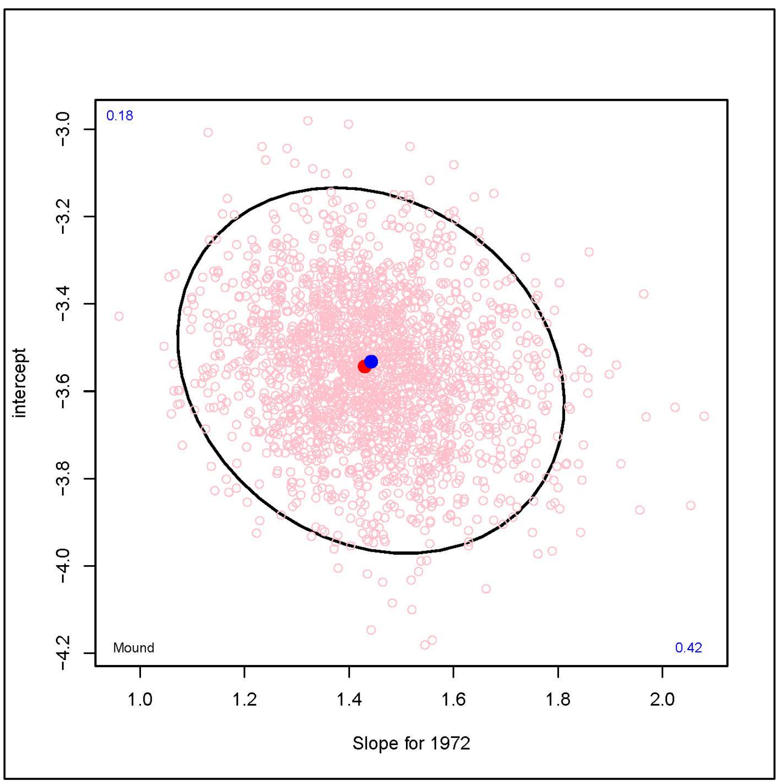
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



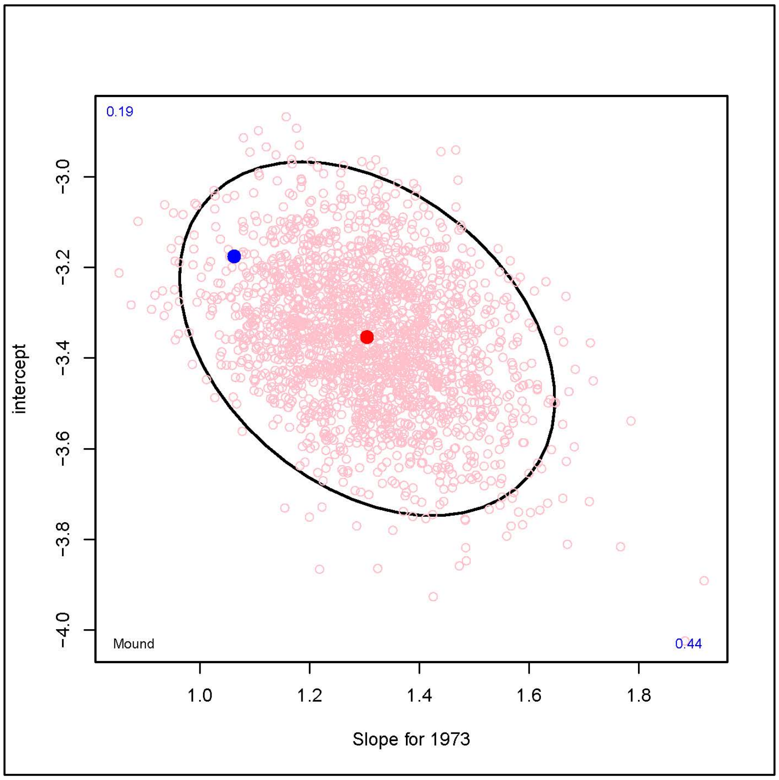


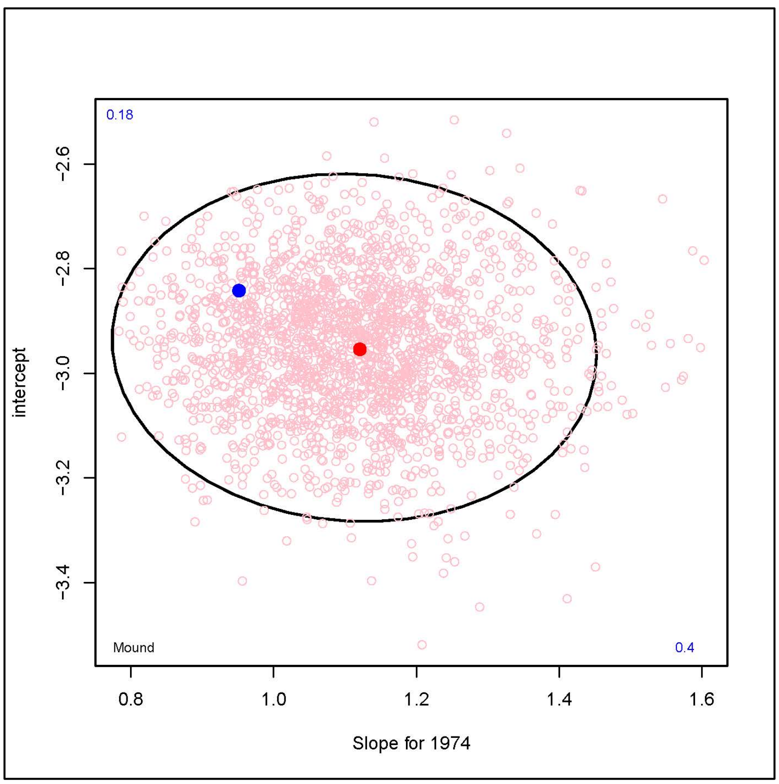
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



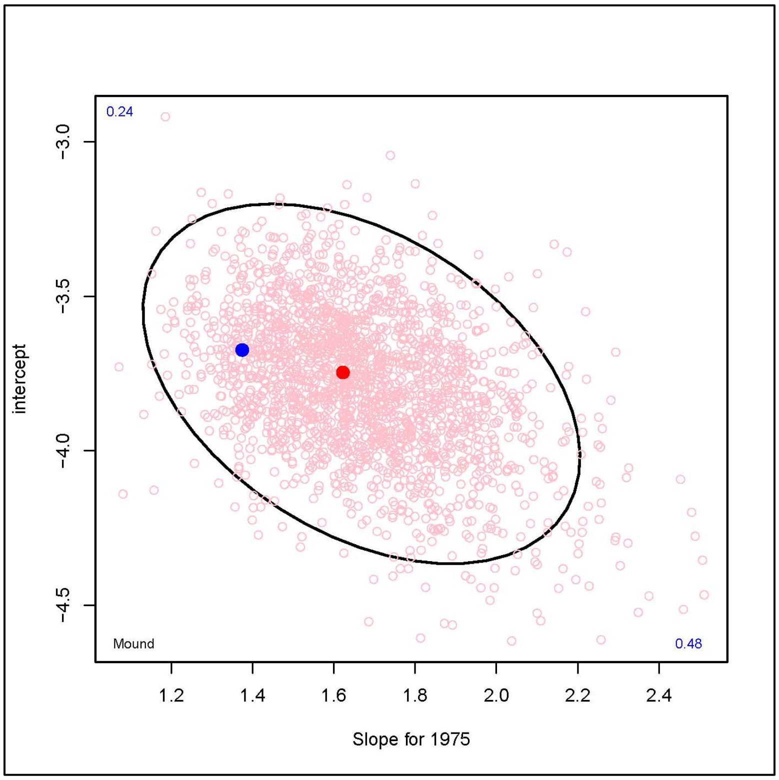


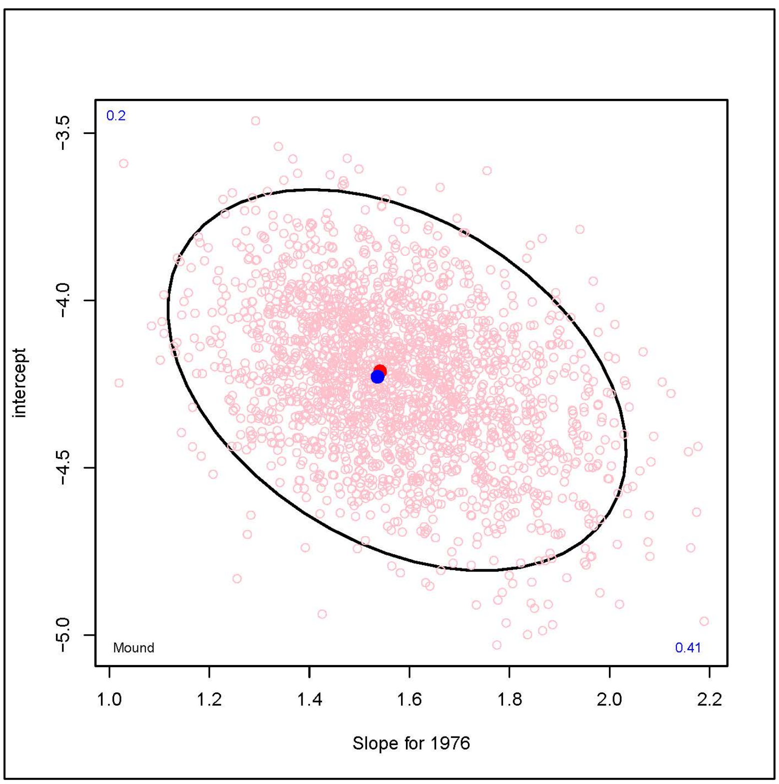
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



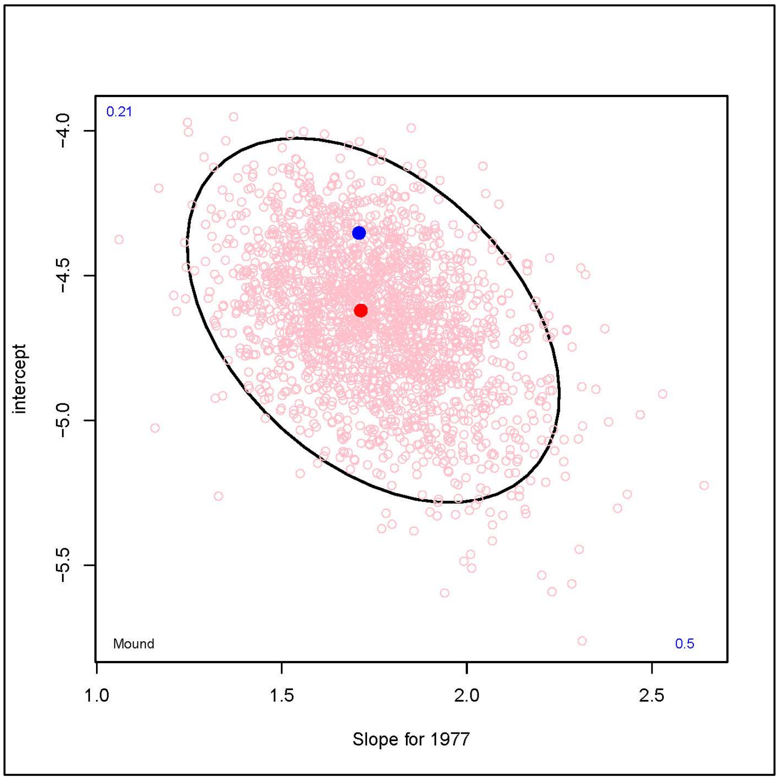


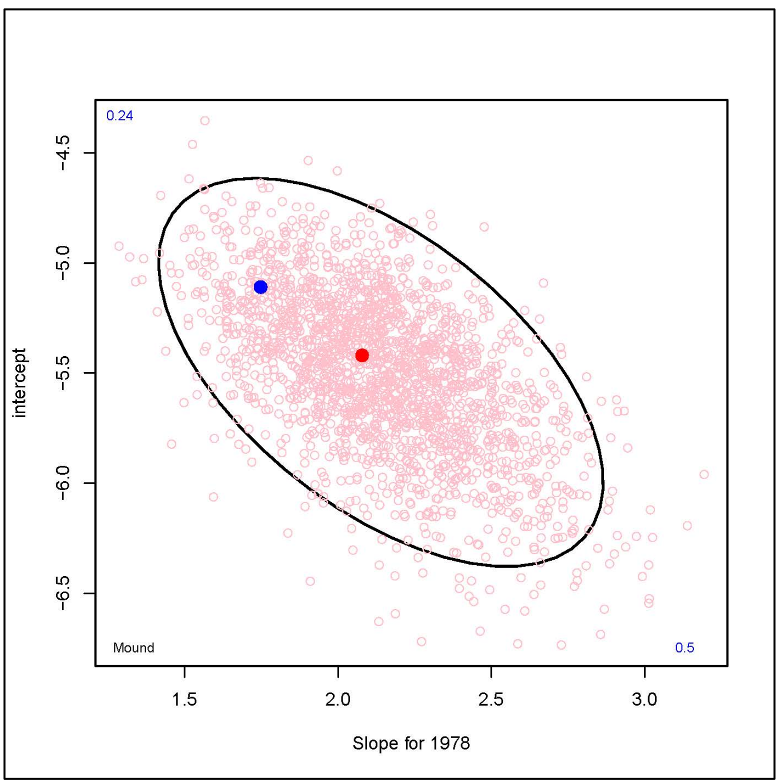
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



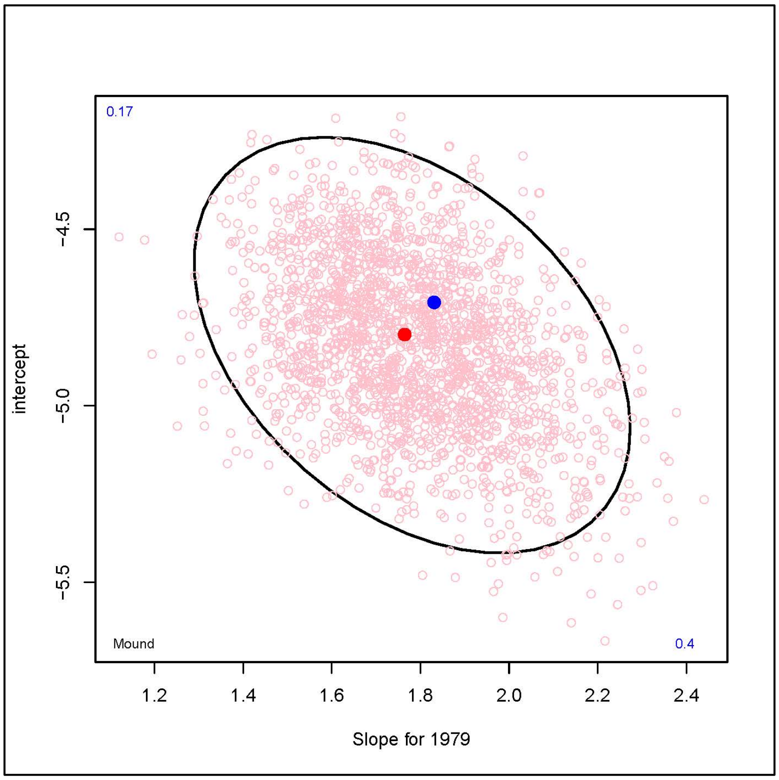


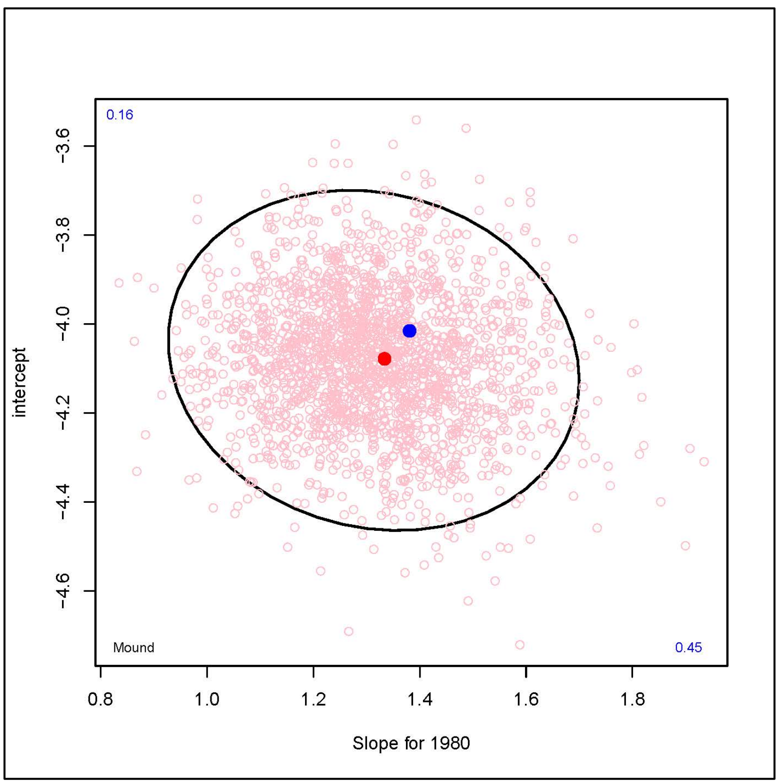
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



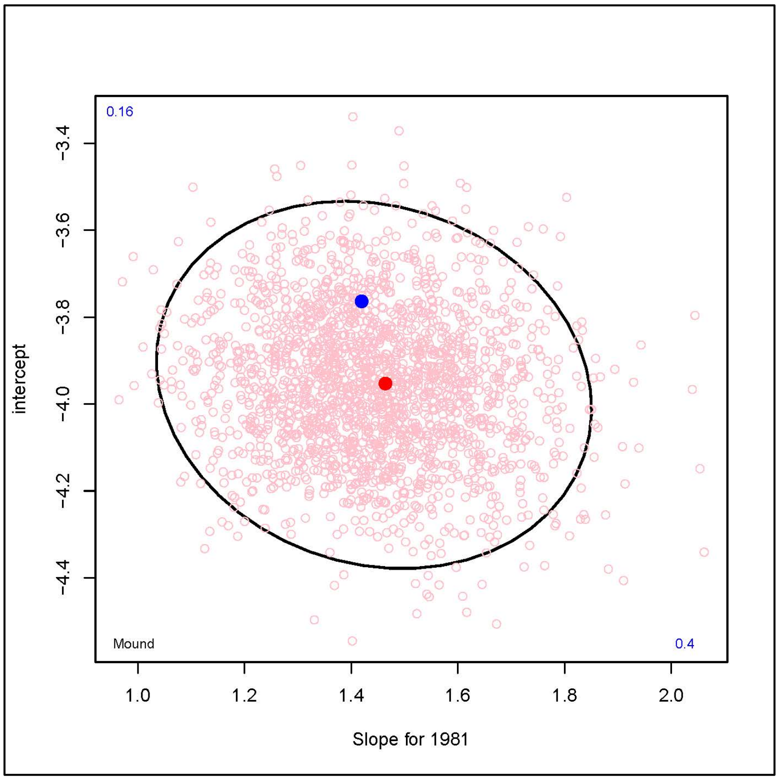


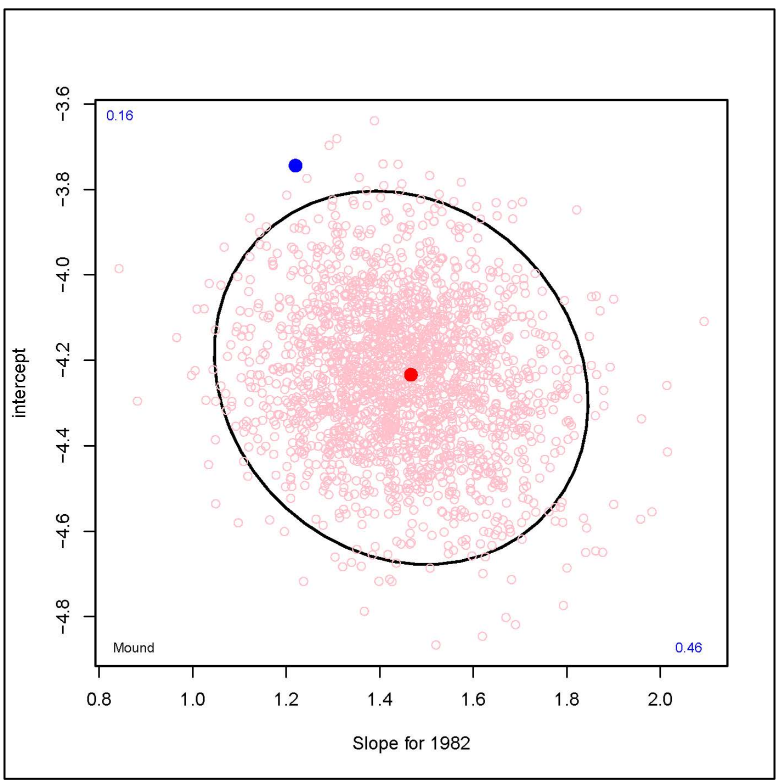
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



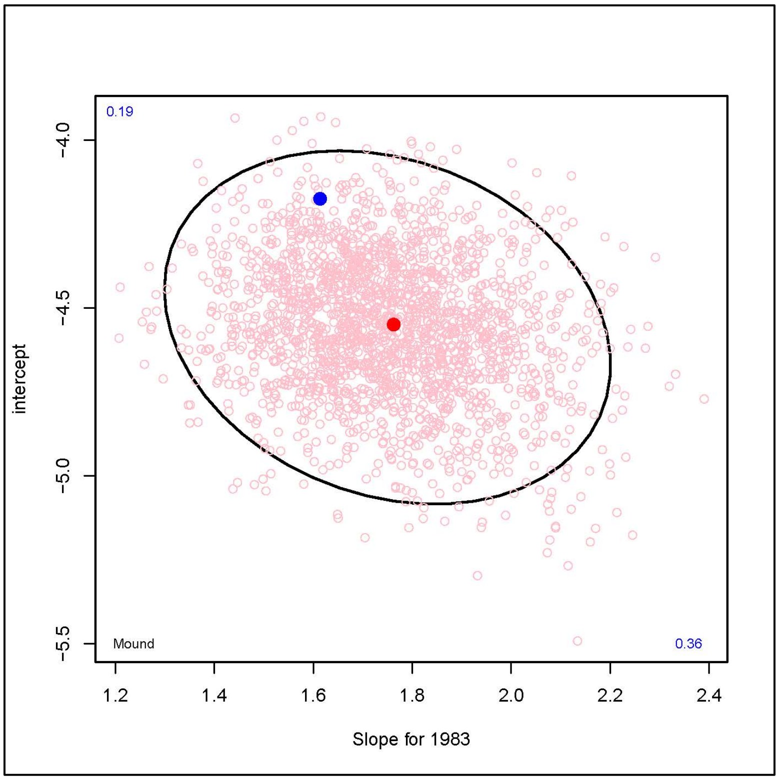


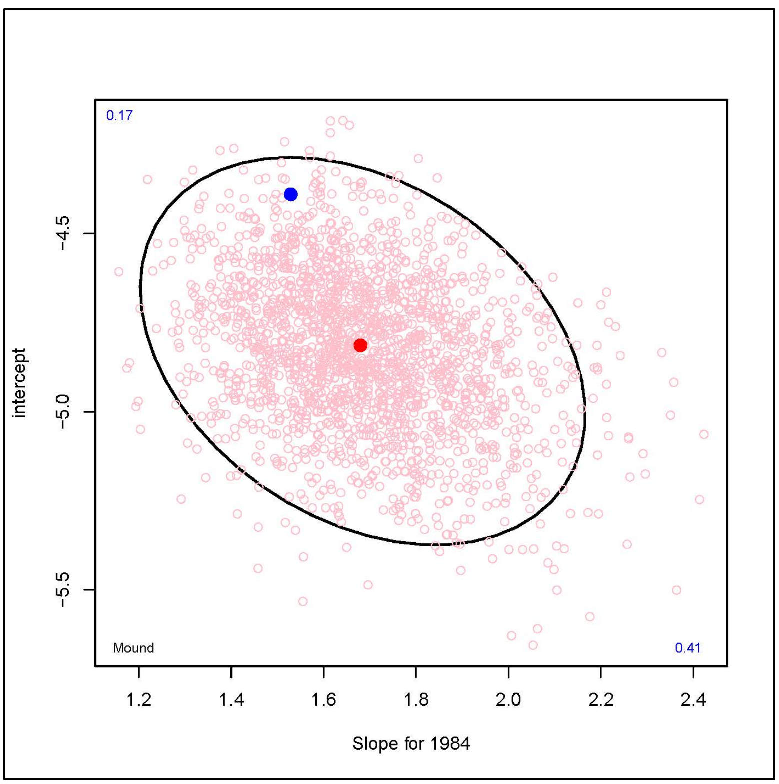
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



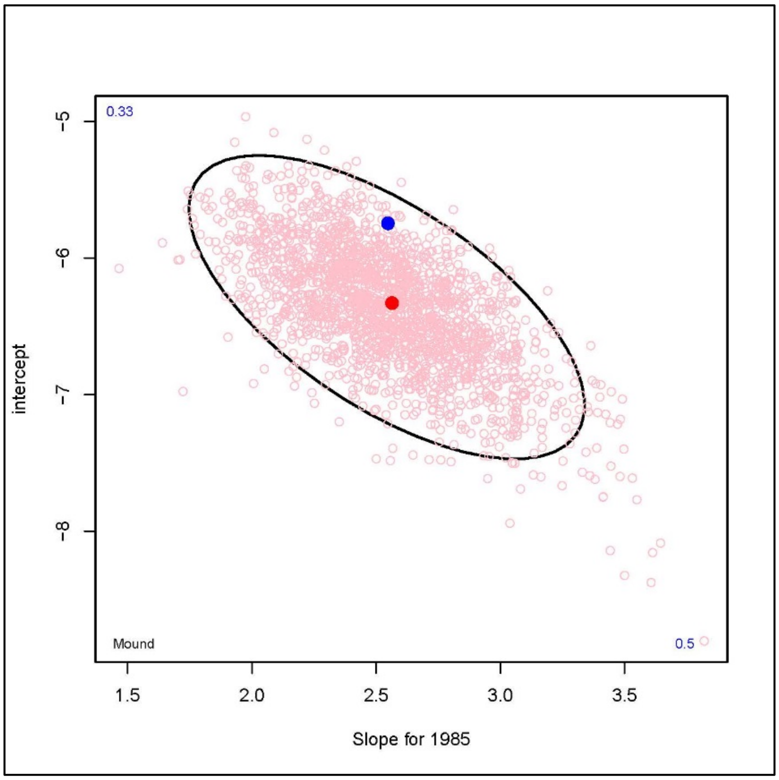


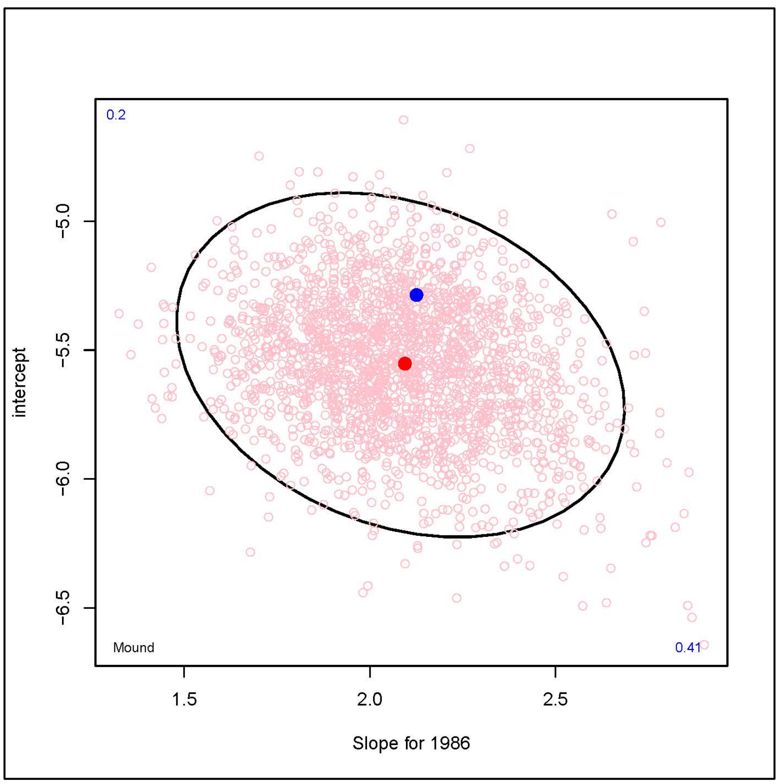
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



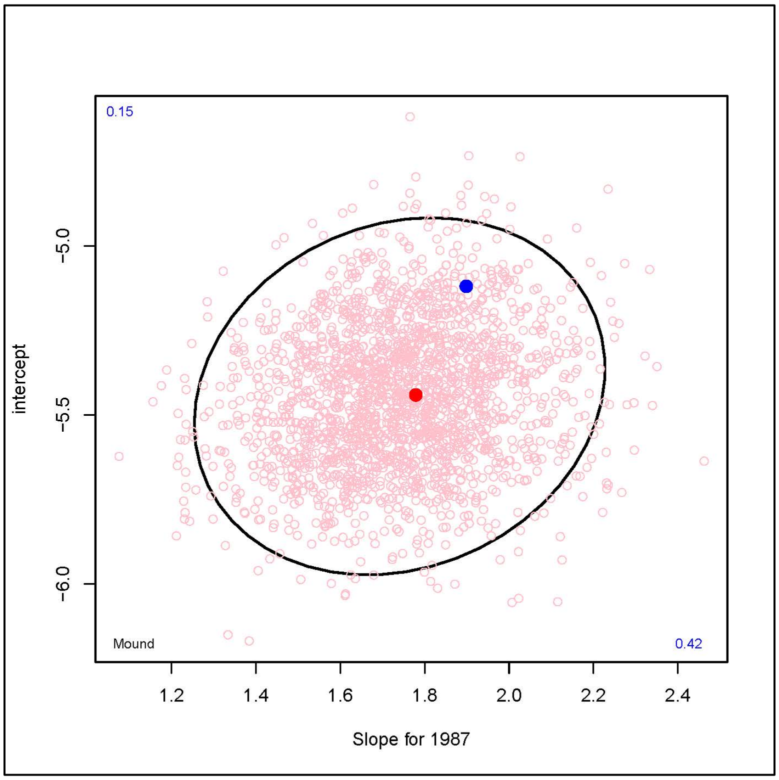


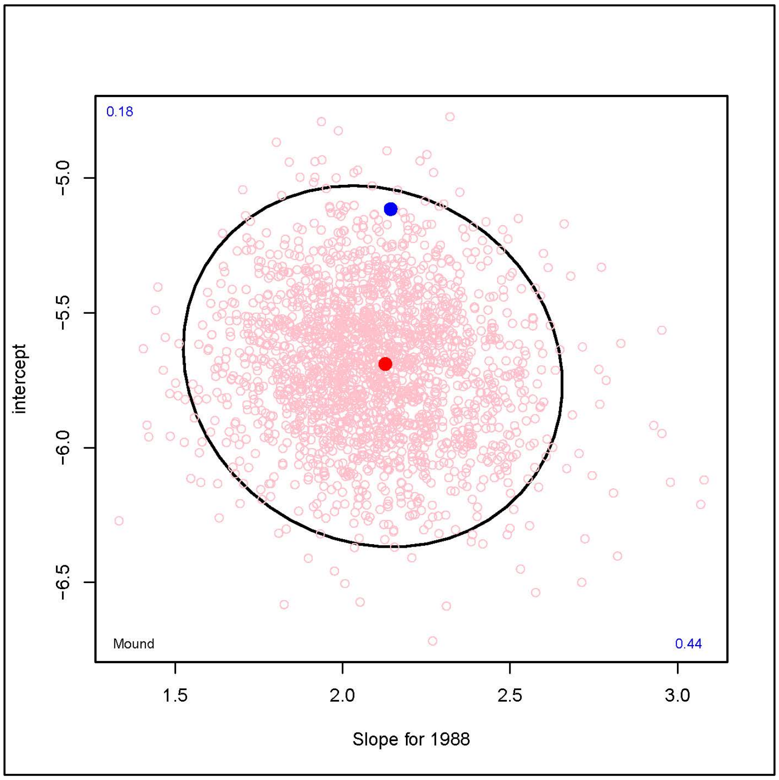
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



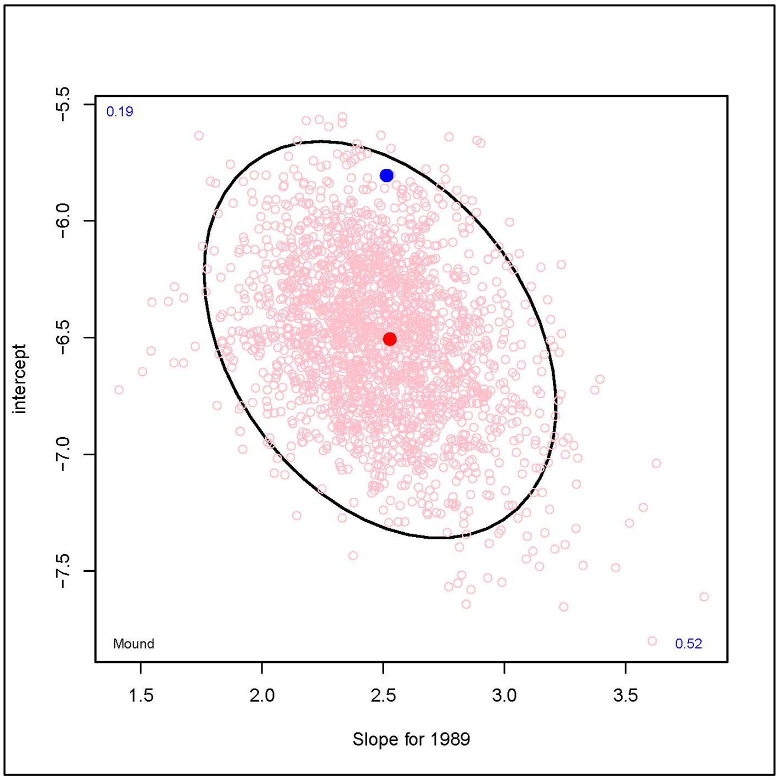


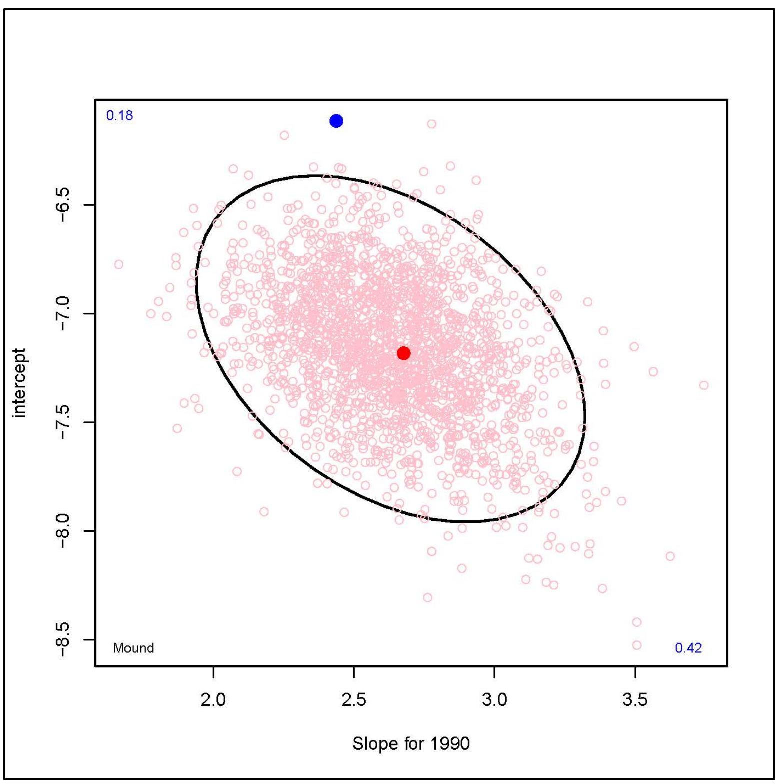
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)





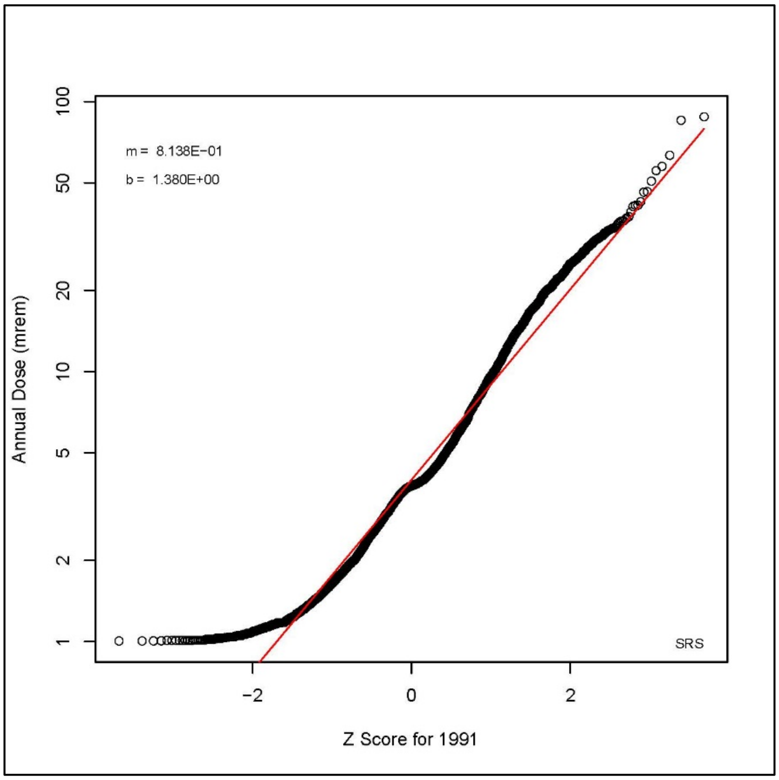
ATTACHMENT F
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE MOUND DATASET, 301 WORKERS (continued)



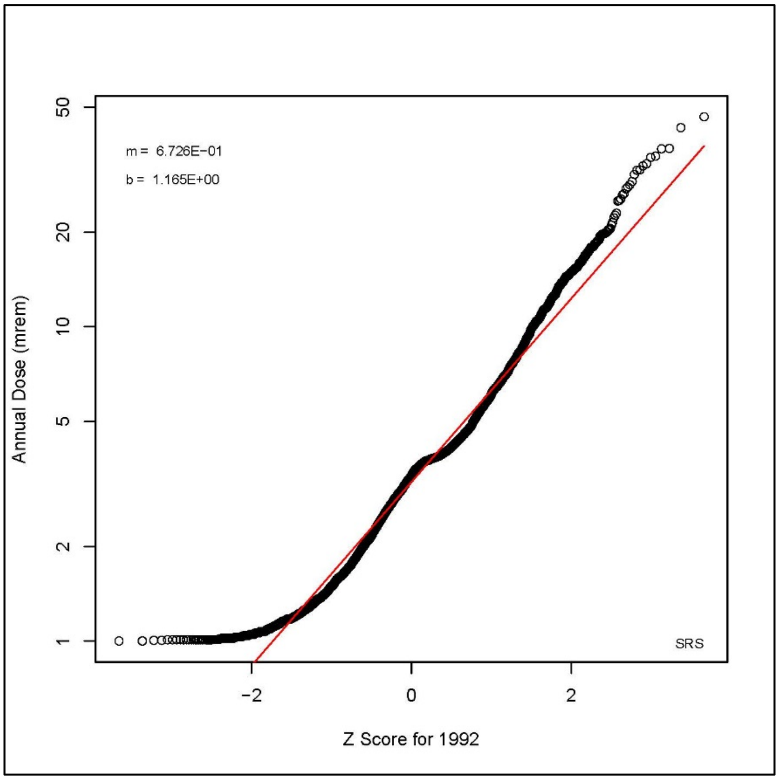


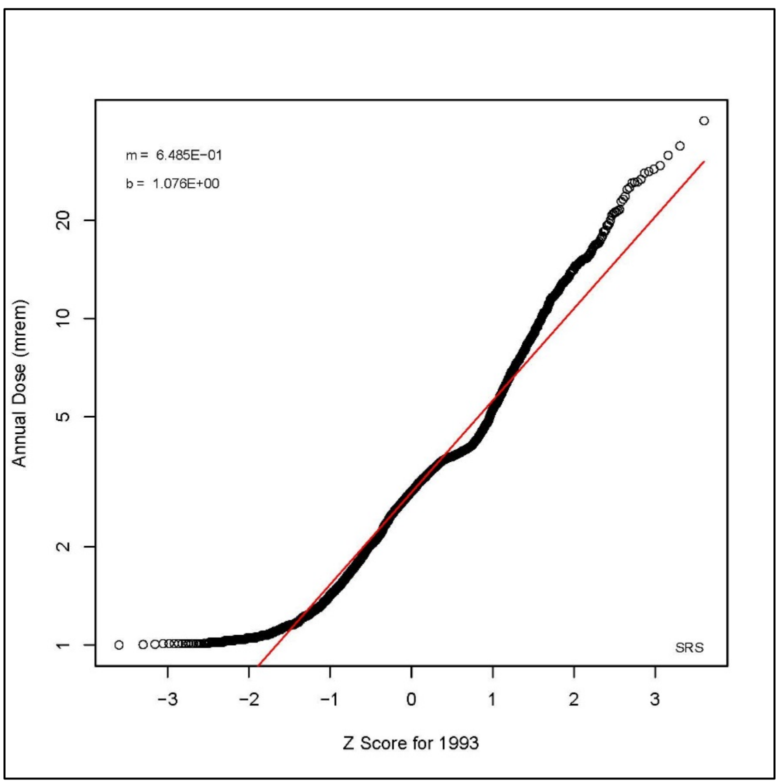
ATTACHMENT G
LOGNORMAL PROBABILITY PLOTS, COMPLETE SRS DATASET

This attachment presents lognormal probability plots for the complete SRS tritium dose dataset for 1991 to 2000. The slope and intercept for the line of best fit to the data are given on the plot.

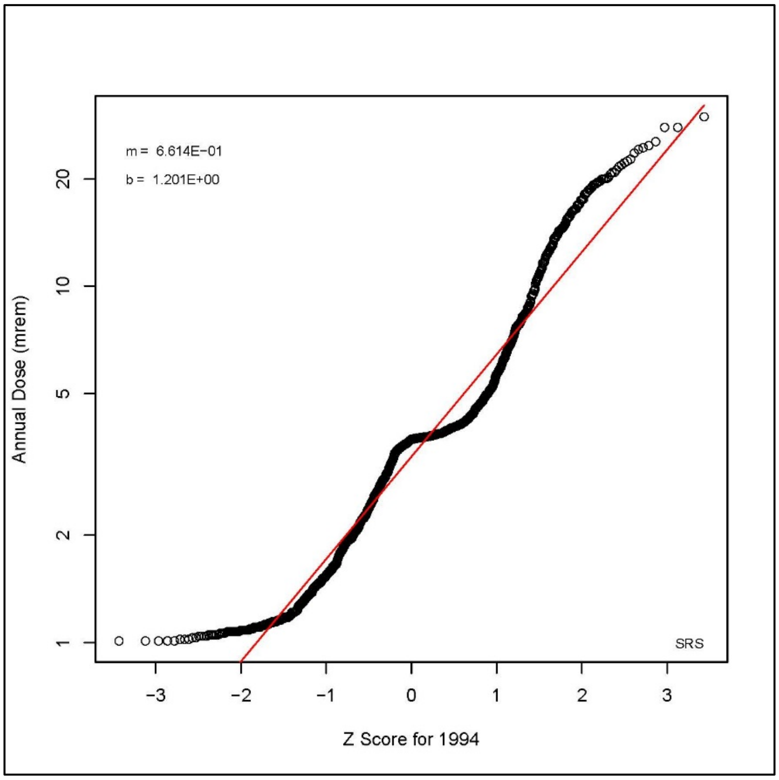


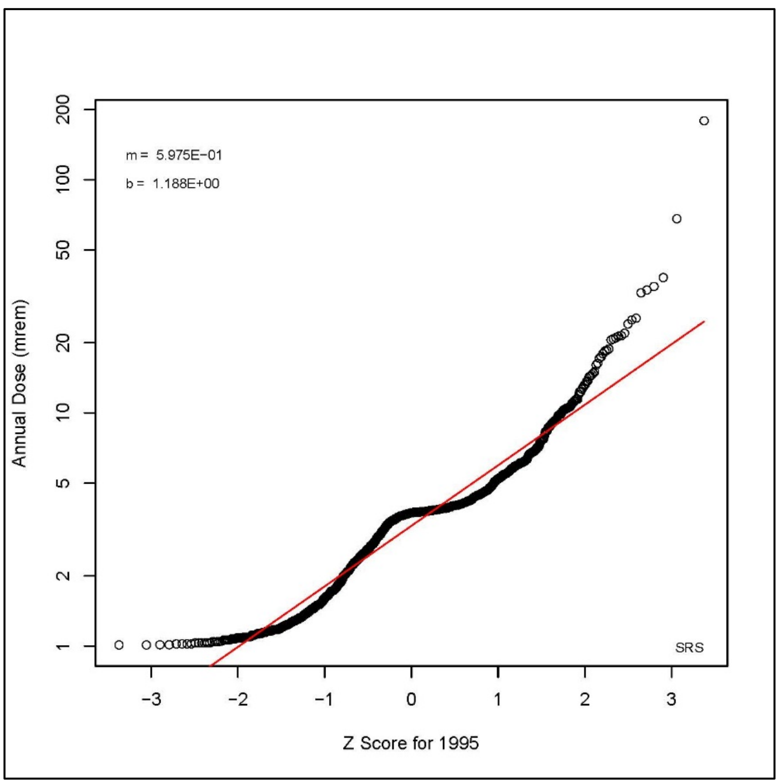
ATTACHMENT G
LOGNORMAL PROBABILITY PLOTS, COMPLETE SRS DATASET (continued)



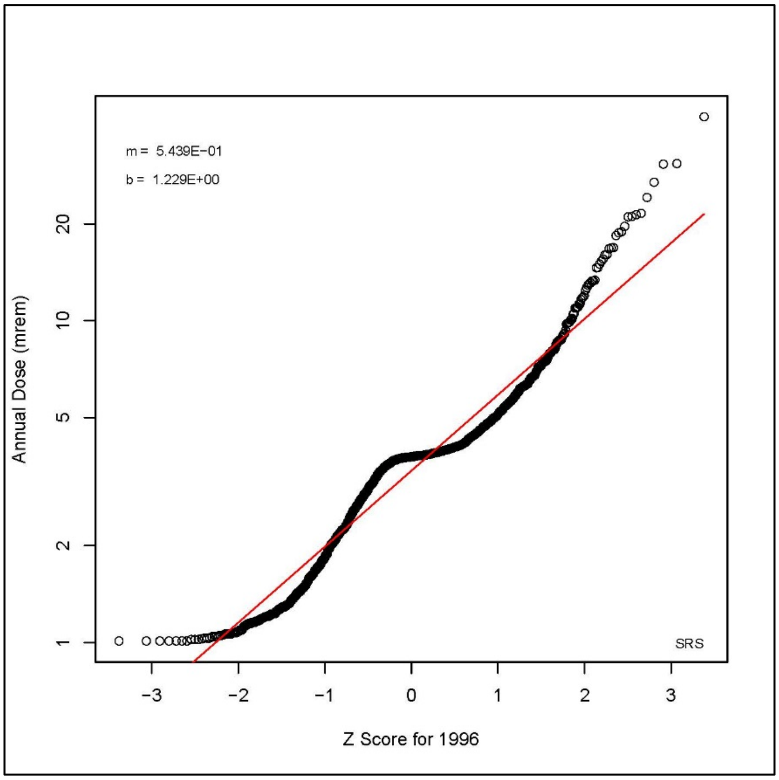


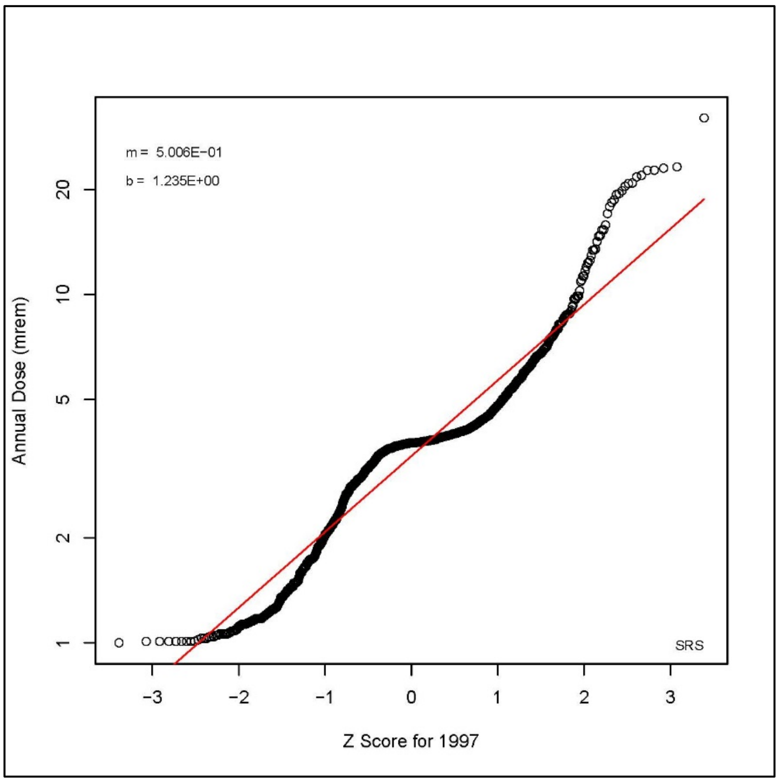
ATTACHMENT G
LOGNORMAL PROBABILITY PLOTS, COMPLETE SRS DATASET (continued)



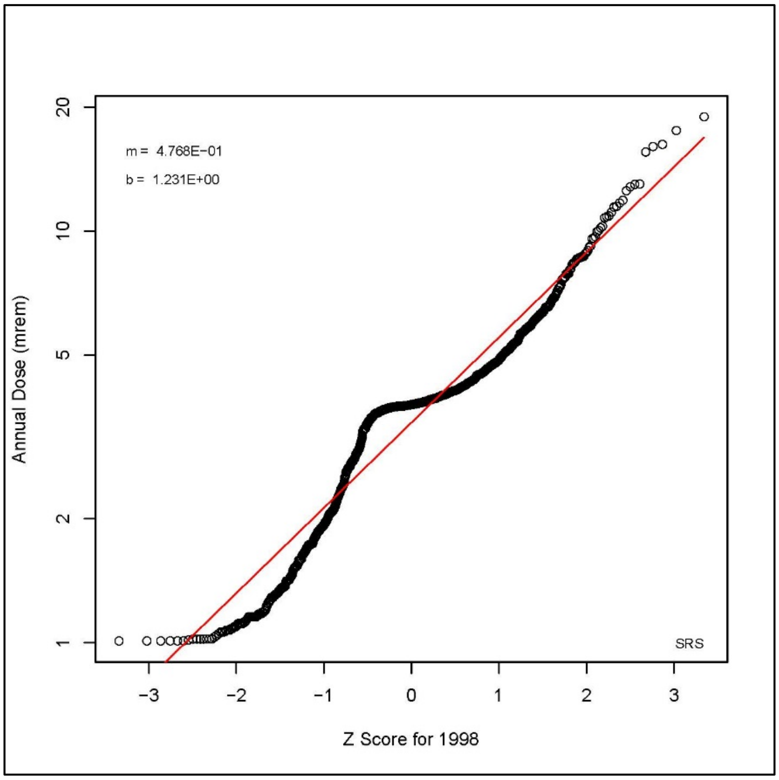


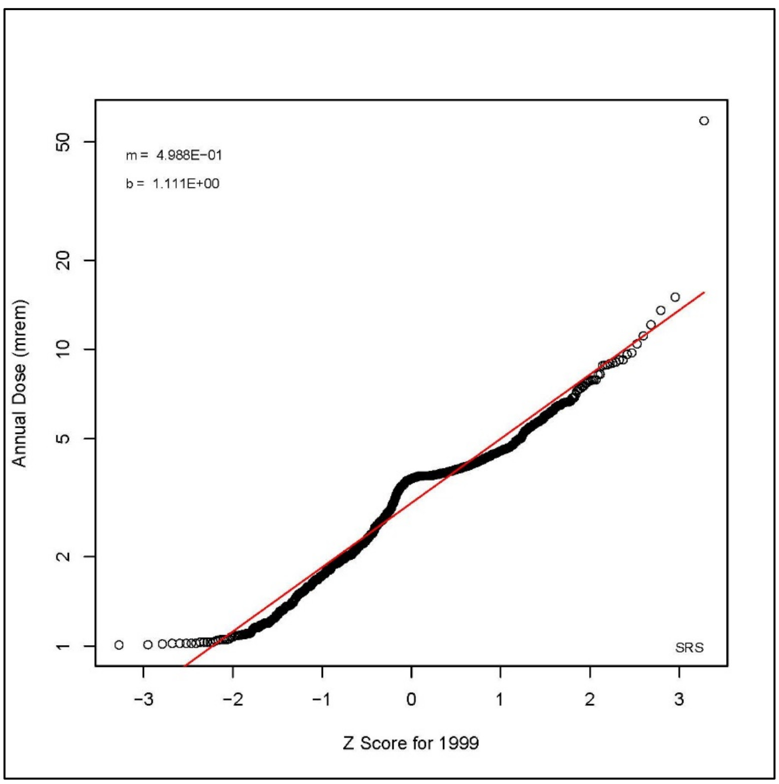
ATTACHMENT G
LOGNORMAL PROBABILITY PLOTS, COMPLETE SRS DATASET (continued)



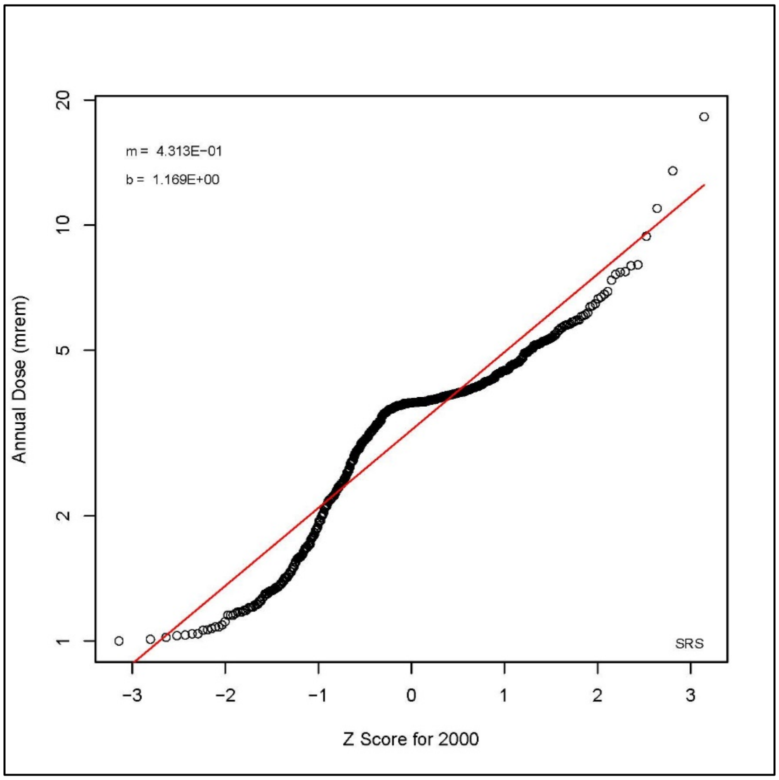


ATTACHMENT G
LOGNORMAL PROBABILITY PLOTS, COMPLETE SRS DATASET (continued)



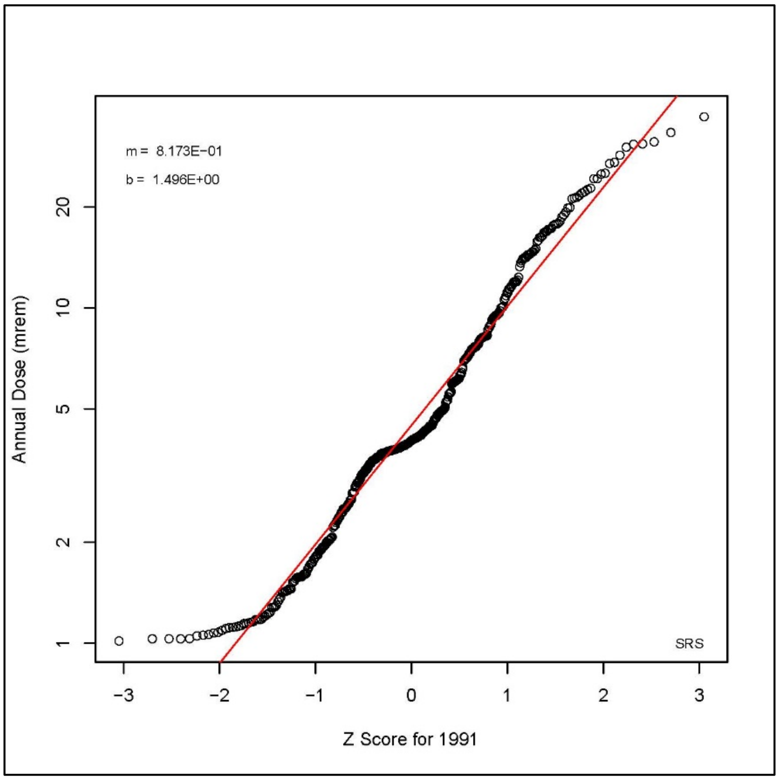


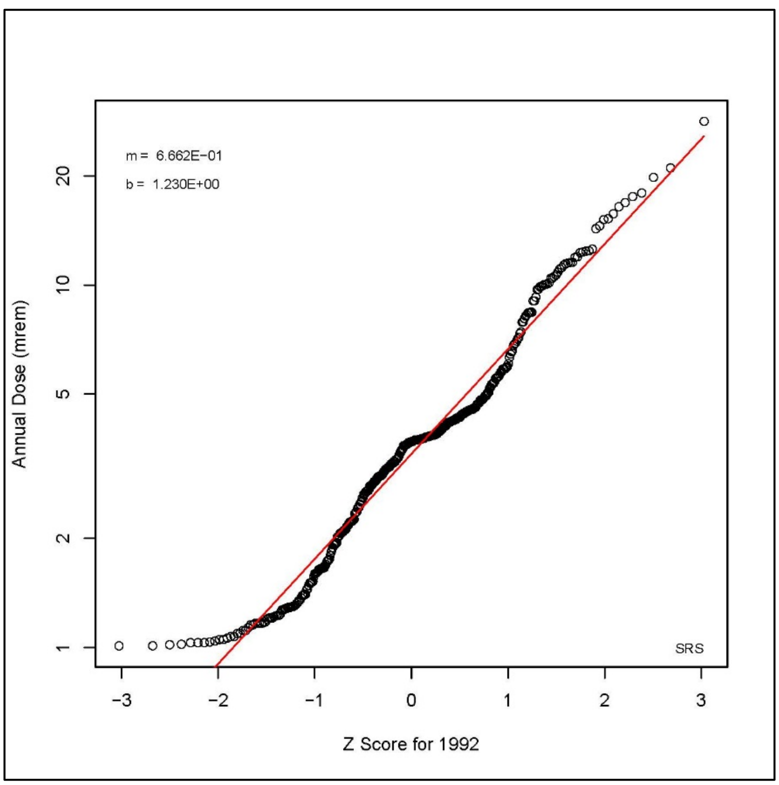
ATTACHMENT G
LOGNORMAL PROBABILITY PLOTS, COMPLETE SRS DATASET (continued)



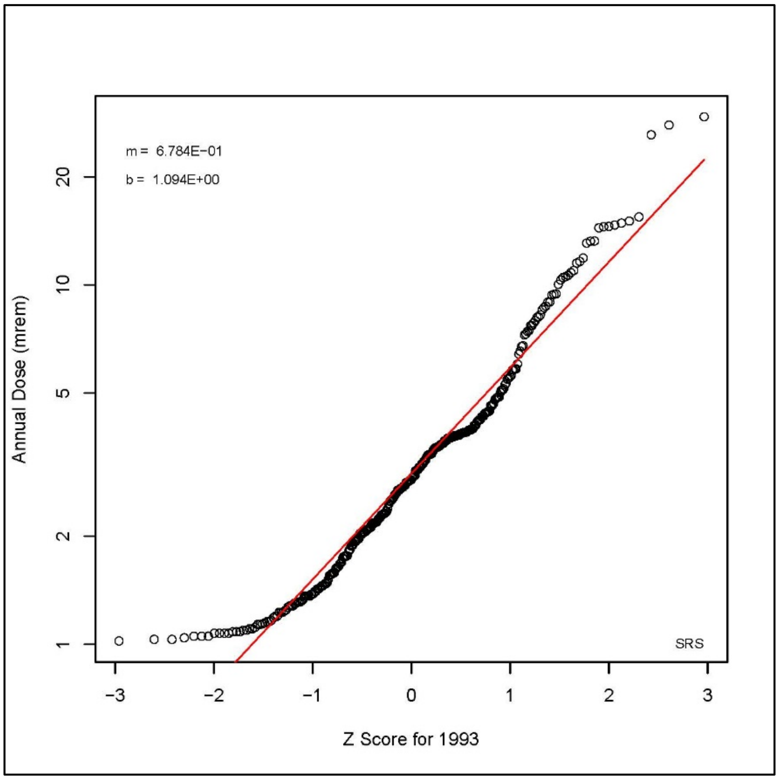
ATTACHMENT H
LOGNORMAL PROBABILITY PLOTS, CLAIMANT SRS DATASET

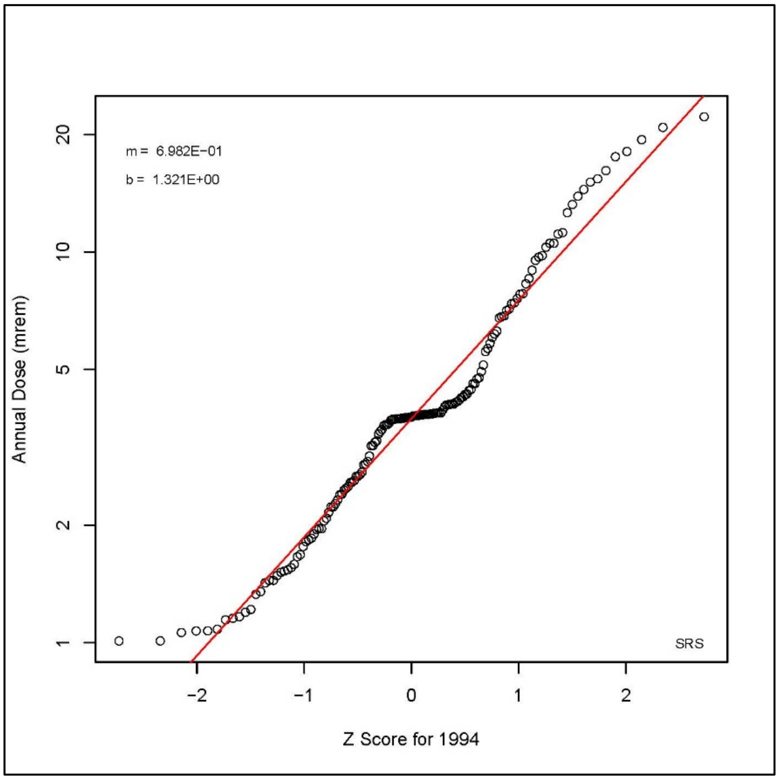
This attachment presents lognormal probability plots for the claimant SRS tritium dose dataset for 1991 to 2000. The slope and intercept for the line of best fit to the data are given on the plot.



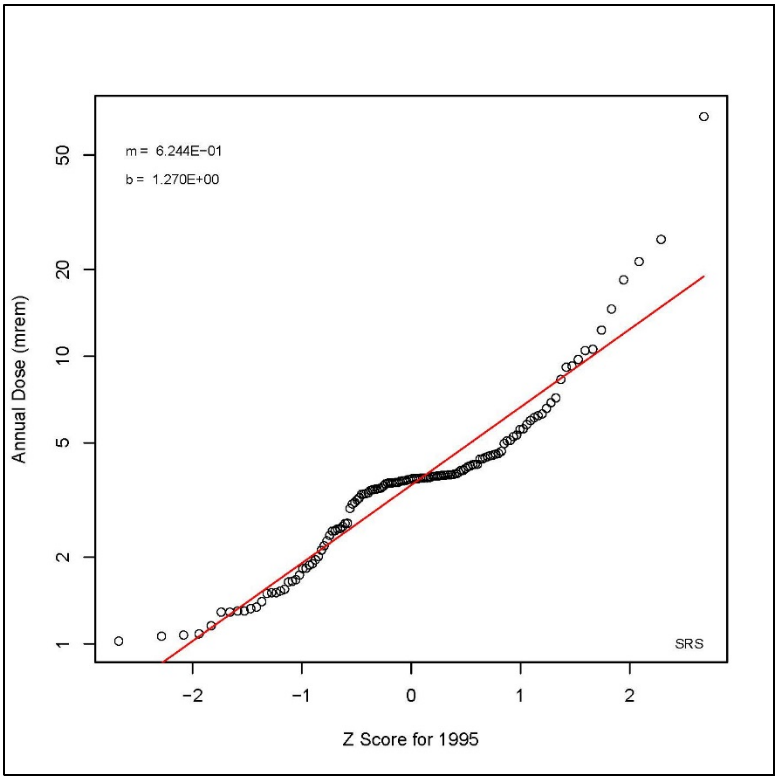


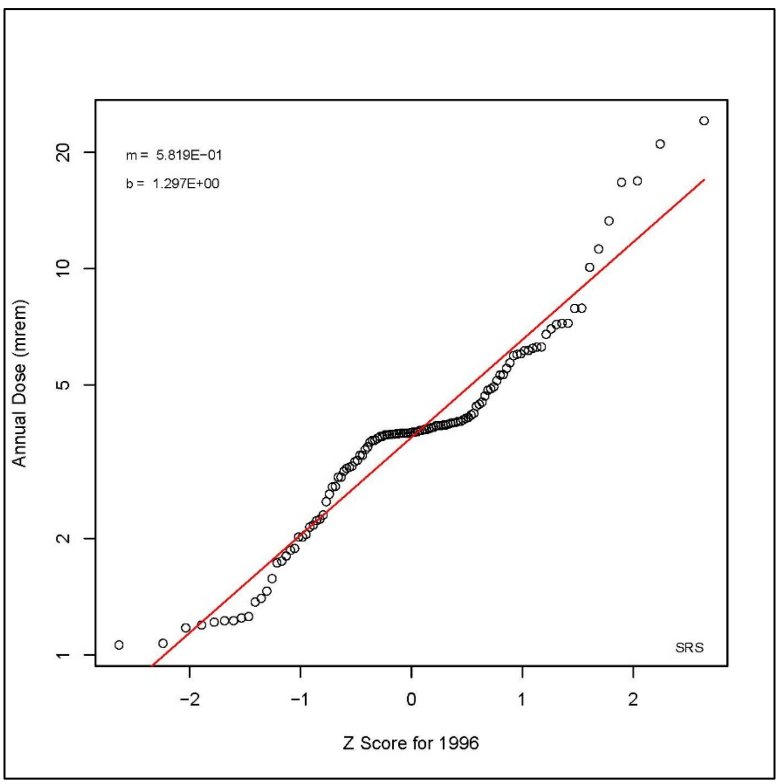
ATTACHMENT H
LOGNORMAL PROBABILITY PLOTS, CLAIMANT SRS DATASET (continued)



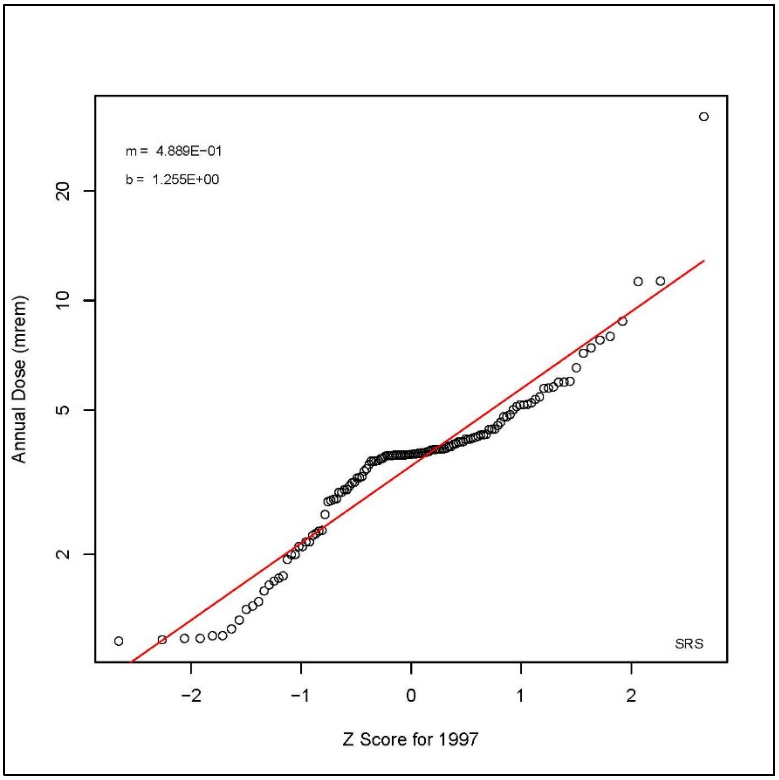


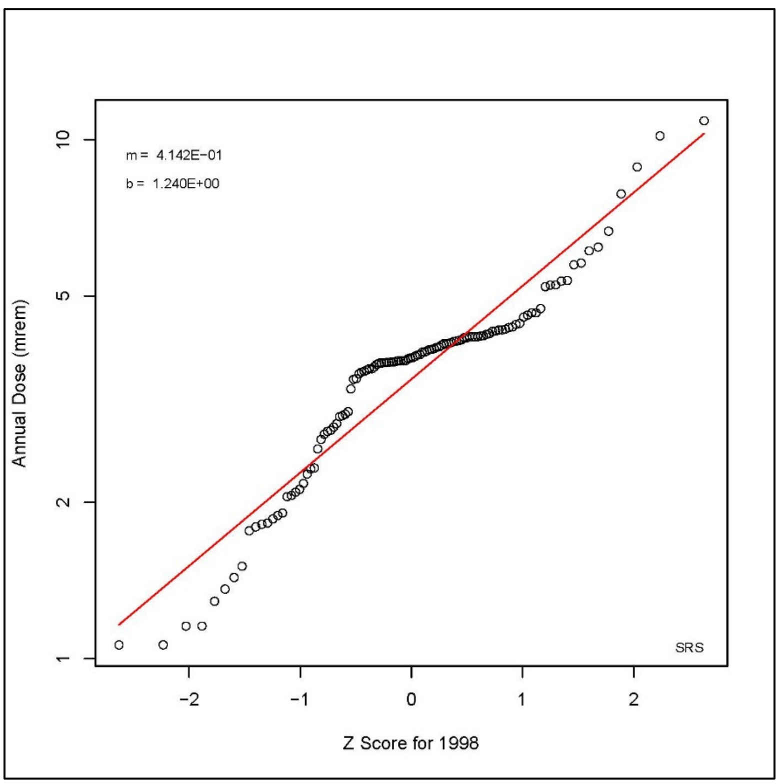
ATTACHMENT H
LOGNORMAL PROBABILITY PLOTS, CLAIMANT SRS DATASET (continued)



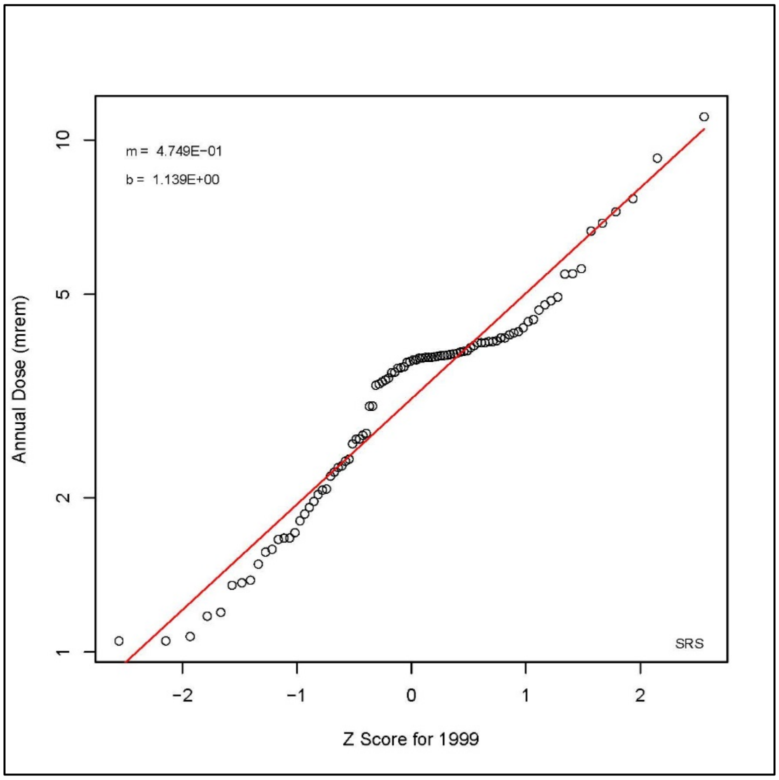


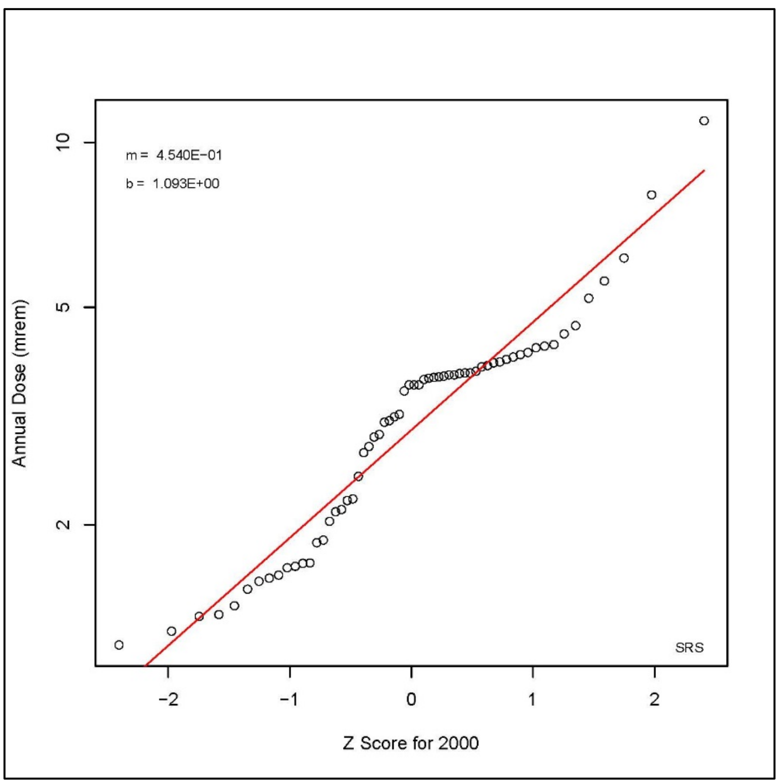
ATTACHMENT H
LOGNORMAL PROBABILITY PLOTS, CLAIMANT SRS DATASET (continued)





ATTACHMENT H
LOGNORMAL PROBABILITY PLOTS, CLAIMANT SRS DATASET (continued)



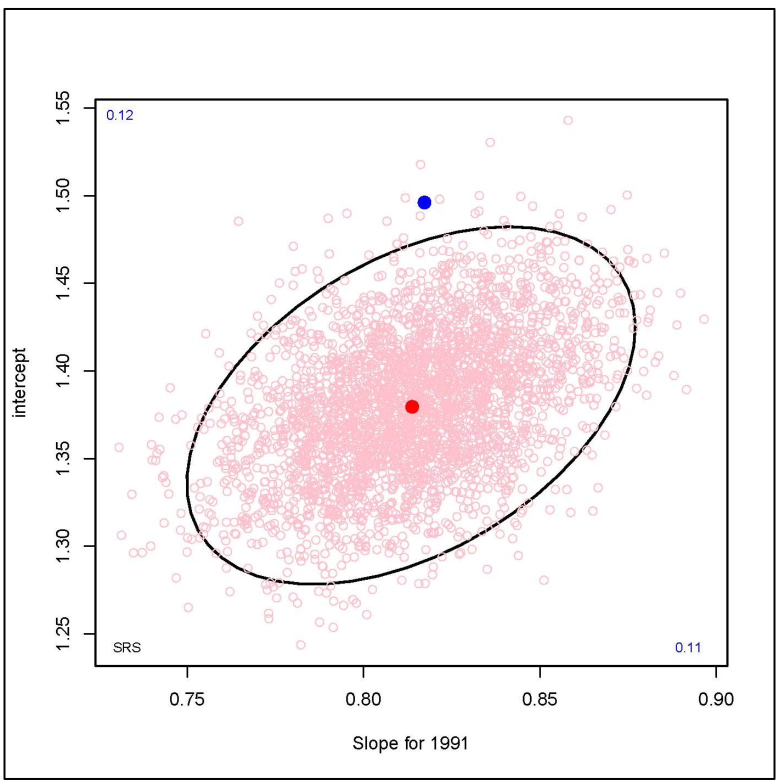


ATTACHMENT I
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE SRS DATASET, 920 WORKERS

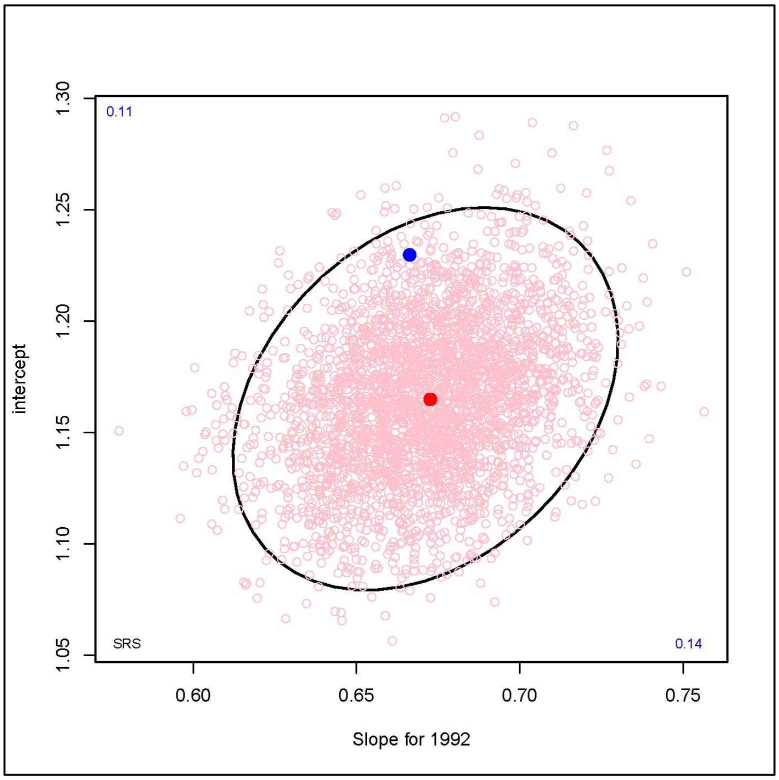
This attachment presents lognormal probability plots for a sample of 920 unique workers that was drawn at random without replacement from the 10,712 workers in the complete dataset. The slope and intercept of the plot for each year from 1991 to 2000 was calculated based on this data. This process was repeated *m* = 3,000 times to generate the 95% confidence ellipses. In each plot:

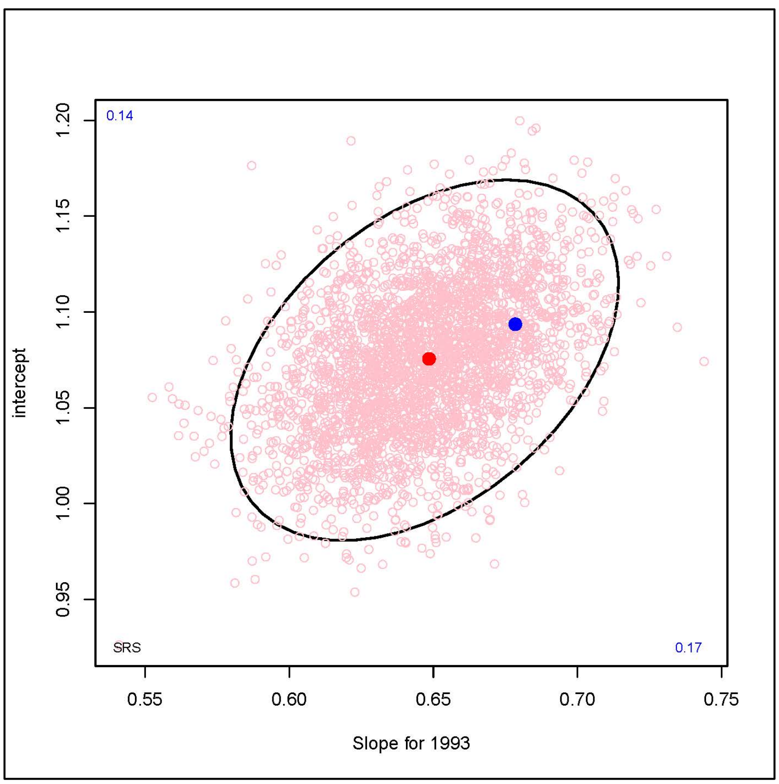
- The solid red point is the slope and intercept of the complete dataset.
- The solid blue point is the slope and intercept of the claimant dataset.
- The open pink points are the slopes and intercepts of the fits to the *m* random sets of 920 workers taken from the complete dataset.
- The ellipse is the 95% joint normal confidence ellipse.
- The numbers in the corners are the full scale of the respective axes in terms of +/- the slope or intercept of the complete dataset.

These plots are presented to support the idea that the claimant dataset is basically no different from a random draw from the complete dataset.

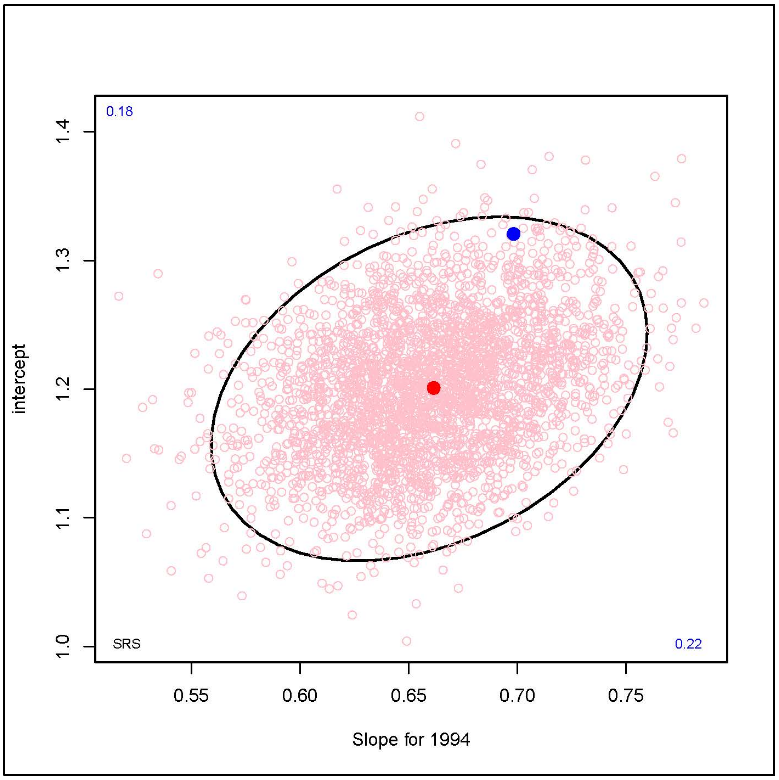


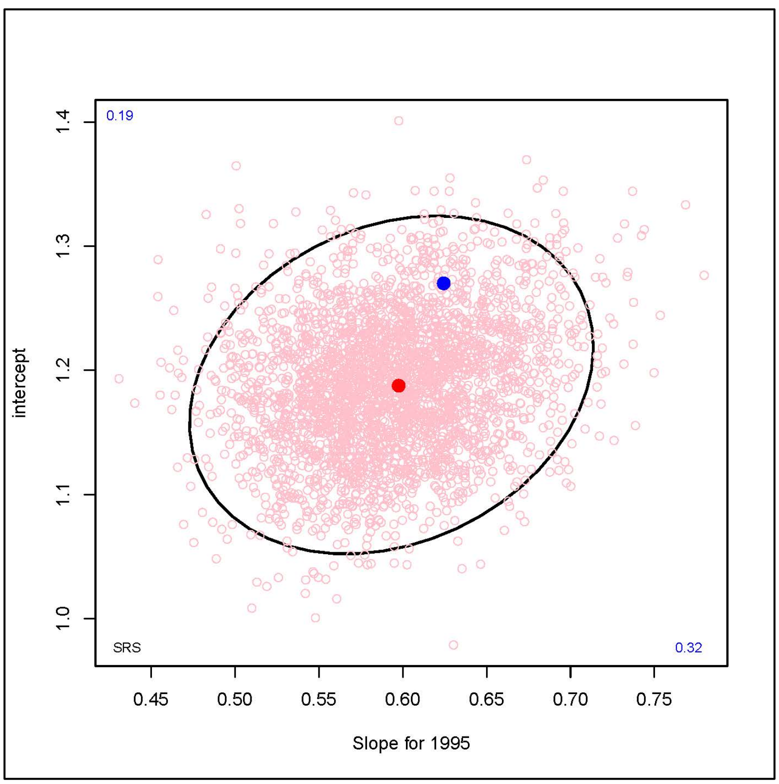
ATTACHMENT I
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE SRS DATASET, 920 WORKERS (continued)



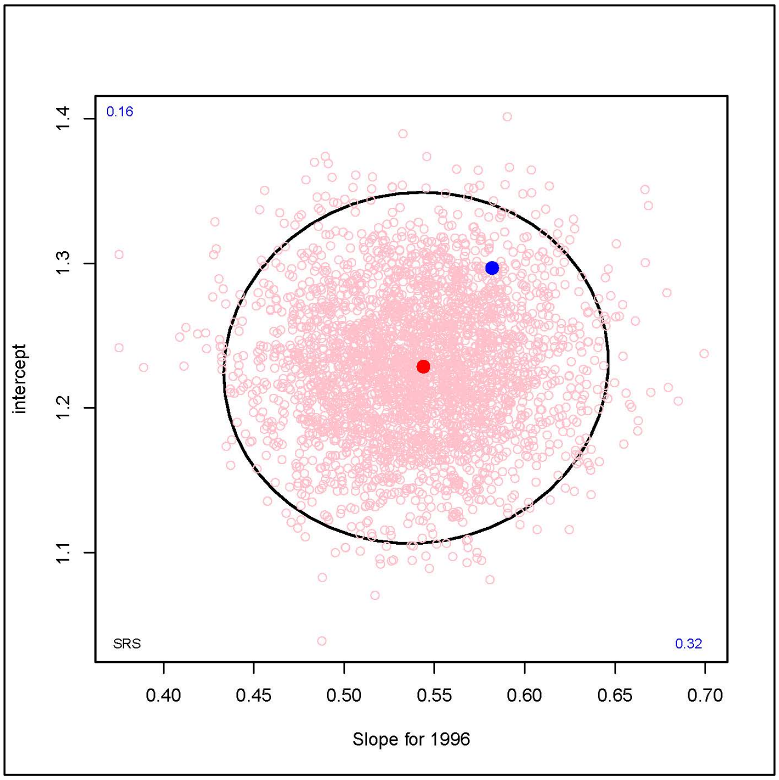


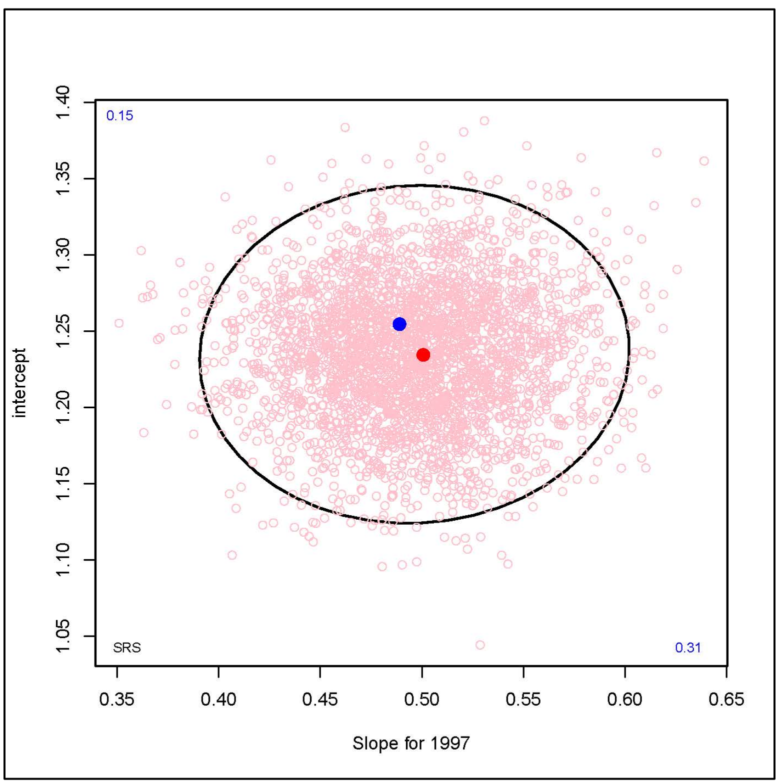
ATTACHMENT I
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE SRS DATASET, 920 WORKERS (continued)



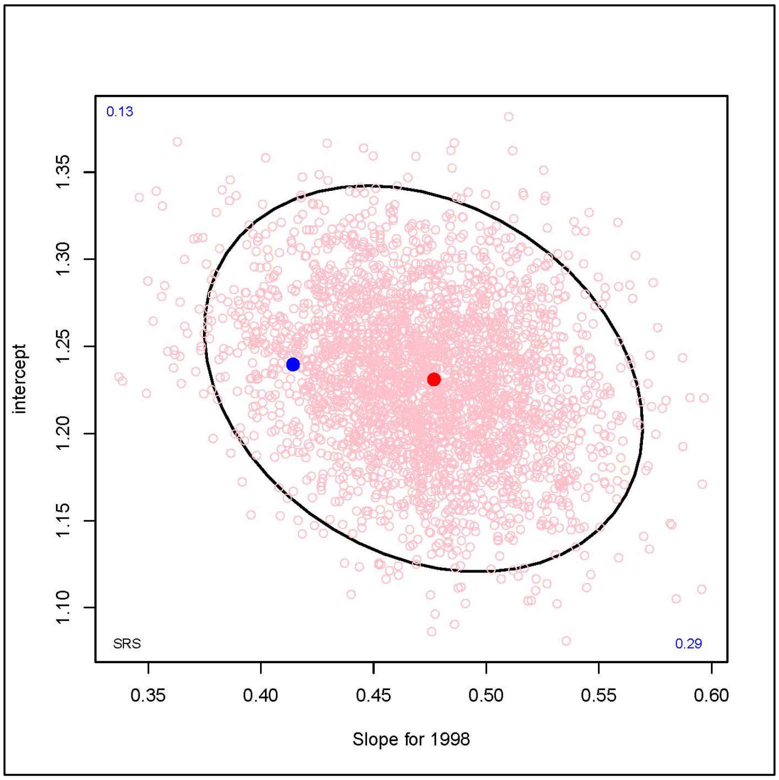


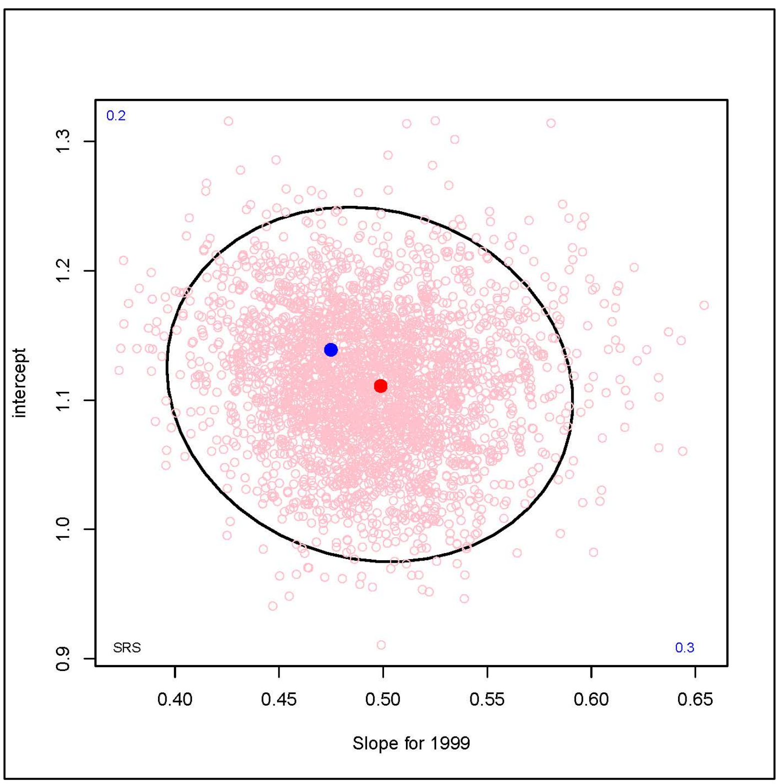
ATTACHMENT I
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE SRS DATASET, 920 WORKERS (continued)





ATTACHMENT I
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE SRS DATASET, 920 WORKERS (continued)





ATTACHMENT I
95% CONFIDENCE ELLIPSES FOR SLOPE AND INTERCEPT, COMPLETE SRS DATASET, 920 WORKERS (continued)