LA-UR- 08-6980

Approved for public release; distribution is unlimited.

Title:	Uncertainties of Mayak Urine Data
Author(s):	Guthrie Miller, Vadim Vostrotin, Vladimir Vvdensky
Intended for:	Publication in the journal Radiation Protection

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

Uncertainties of Mayak Urine Data

(LA-UR-08-xxxx)

Guthrie Miller (guthrie@lanl.gov) [1], Vadim Vostrotin[2], and Vladimir Vvdensky[2]

[1] Los Alamos National Laboratory
[2] Southern Urals Biophysics Institute

October 20, 2008

To be submitted to Radiation Protection Dosimetry as Scientific Note.

Abstract

A method of parameterizing the uncertainty of Mayak urine bioassay measurements is described. The Poisson lognormal model is assumed and data from 63 cases (1099 urine measurements in all) are used to empirically determine the lognormal normalization uncertainty, given the measurement uncertainties obtained from count quantities. The natural logarithm of the geometric standard deviation of the normalization uncertainty is found to be in the range 0.31 to 0.35 including a measurement component estimated to be 0.2.

Introduction

For internal dose calculations for the Mayak worker epidemiological study [need reference 1], quantitative estimates of uncertainty of the urine measurements are necessary. Some of the data consist of measurements of 24h urine excretion on successive days (e.g. 3 or 4 days). In a recent publication[2], dose calculations were done where the uncertainty of the urine measurements was estimated starting from the statistical standard deviation of these replicate measurements. This approach is straightforward and accurate when the number of replicate measurements is large, however, a Monte Carlo study showed it to be problematic for the actual number of replicate measurements (median from 3 to 4).[3] Also, it is sometimes important to characterize the uncertainty of a single urine measurement. Therefore this alternate method has been developed.

A very helpful discussion of the available information regarding uncertainties of Mayak urine data are contained in the 2002 technical report for project 2.4 [4]. The paper by Kranenbuhl et. al. provides an overall description of measurement techniques that have been used.[5] Further discussion of urine measurement uncertainties appears in the correspondence by Miller[6], which refers to the paper by Bess et. al. [7].

Poisson-lognormal Model

The Poisson-lognormal uncertainty model [3] is assumed, and the Poisson measurement uncertainties are characterized by a standard deviation σ_m, rather than making exact Poisson-lognormal likelihood calculations using count quantities, although this might be important when the counts are small.

The fundamental measurement quantities may be taken as the lognormally distributed normalization coefficient f, the gross sample counts N, the background counts N_B, the background scaling factor R, and a lognomally distributed environmental/analysis laboratory background b. The measurement result y and uncertainty σ_m not related to normalization uncertainty (uncertainty of f) are given by,

$$y = f(N - N_B / R) - b$$
$$\sigma_m = \sqrt{f^2(N + N_B / R^2) + \sigma_b^2} \qquad (1)$$

in terms of measurement quantities. Using information about measurement quantities contained in the Mayak database and discussed in the 2002 technical report [4] (e.g., about count times and measured count rates), it is possible to determine all the parameters in this representation of the data.

In normal-lognormal model, data uncertainty is specified by σ_m and the lognormal uncertainty of f, which will be denoted by S (Geometric Standard Deviation = exp (S)). The quantity S depends on the collection protocol, for example true 24-hr collection rather than a spot sample, as well as measurement details. An example of dependence on measurement details is when the chemical recovery is estimated, rather than being measured with a tracer. There is then a significant normalization uncertainty associated with lack of knowledge of the actual chemical recovery.

Poisson-lognormal Model in more detail

In Reference [3], the effect of a lognormally distributed background b is considered in detail. The expectation value and variance of the number of gross counts are found to be given by (Eq. A.2 of Ref. [3])

$$E(N) = \mu_B / R + \mu e^{S_n^2/2} e^{S_m^2/2} + \frac{\hat{b}}{f} e^{S_b^2/2} e^{S_m^2/2}$$

$$\begin{aligned} Var(N) &= E(N^2) - E(N)^2 \\ &= \mu_B / R + \mu e^{S_n^2/2} e^{S_m^2/2} + \left(\mu e^{S_n^2/2} e^{S_m^2/2}\right)^2 \left(e^{S_n^2 + S_m^2} - 1\right) \\ &\quad + \frac{\hat{b}}{f} e^{S_b^2/2} e^{S_m^2/2} + \left(\frac{\hat{b}}{f} e^{S_b^2/2} e^{S_m^2/2}\right)^2 \left(e^{S_b^2 + S_m^2} - 1\right) \\ &\quad + 2\mu \frac{\hat{b}}{f} e^{S_b^2/2} e^{S_n^2/2} e^{S_m^2} \left(e^{S_m^2} - 1\right) \end{aligned} \qquad (2)$$

where *E(.)* and *Var(.)* denote the expectation value and the variance. The quantity $\hat{b}$ is the median value of the background. In this model the normalization coefficient f is the product of two factors, f_n, the normalization coefficient associated with the excretion time, and f_m, the normalization coefficient associated with other measurement quantities

$$f = f_n f_m \tag{3}$$

where

$$f_n = \frac{1}{\Delta t_{ex}} \tag{4}$$

and

$$f_m = \frac{1}{\varepsilon_{chem}\,\varepsilon_{count}\,\Delta t_{count}} \tag{5}$$

where Δt_{ex} is the excretion time associated with the sample, ε_{chem} is the efficiency of chemical recovery, and ε_{count} is the counting efficiency. In Eq. (2), the expectation value of sample gross counts N has three terms. The first is the contribution from background counts given by μ_B / R, where μ_B is the true average value of the number of background counts in the background counting period, which is R sample counting periods. The second is the contribution from the amount of radioactive material excreted in urine, $\mu = \psi / f$, where ψ is the quantity of interest. The third term is the contribution from the background $\hat{b}$.

The quantities S_n, S_m, and S_b are the lognormal uncertainties of f_n, f_m, and f_b.

The measured quantity of interest y (the 24h urine excretion) is given by Eq. (1). In this equation N and N_B represent randomly varying measurement quantities, while f and b are single nominal values used in the calculation of the quantity of interest. The variance of y therefore results from the variances of N and N_B and is given by

$$Var(y) = f^2\left(Var(N) + Var(N_B)/R^2\right) = f^2(N + N_B/R^2) + \sigma_b^2 + \left(\frac{\psi\sigma_f}{f}\right)^2 + 2\psi A(S_{nm})\hat{b}A(S_{bm})\left(C_V(S_m)\right)^2 \tag{6}$$

using Eq. (2) and making the substitutions $E(N) = N$ and $Var(N_B) = \mu_B = N_B$. The quantities

$$A(S) = \exp(S^2/2) \tag{7}$$

And

$$C_V(S) = \sqrt{\exp(S^2) - 1} \tag{8}$$

are the lognormal mean to median ratio and standard deviation to mean ratio, so that the standard deviations of f and b are given by

$$\begin{aligned} \sigma_f &= fA(S_{nm})C_V(S_{nm}) \\ \sigma_b &= \hat{b}A(S_{bm})C_V(S_{bm}) \end{aligned} \tag{9}$$

With

$$\begin{aligned} S_{nm} &= \sqrt{S_n^2 + S_m^2} \\ S_{bm} &= \sqrt{S_b^2 + S_m^2} \end{aligned} \tag{10}$$

The last equation shows how uncertainties of f_n (excretion-time-related normalization) and f_m (measurement-related normalization) combine.

Equation 6 is identical to Equation 1 except for the last term, which represents the covariance of f and b and depends on f_m being variable ($C_V(S_m) > 0$).

Determination of S

Urine data collected on some number of successive days with a standard collection procedure or protocol are available for 63 cases [2]. The central value of the urine measurement is defined as the average or logarithmic average of the measurements on successive days. If all other uncertainty parameters are known, the normalization uncertainty S_n can be determined empirically from a large dataset of such data by using the relationship

$$\chi^2 = \sum_{i=1}^{N_{data}} \frac{\left(y_i - \langle y \rangle_i\right)^2}{\sigma_{mi}^2 + \left(\langle y \rangle_i C_V(S_{nm})\right)^2 + 2\langle y \rangle_i b\left(C_V(S_m)\right)^2} \cong NDF = N_{data} - N_{avg} \quad , \tag{11}$$

where NDF is the number of degrees of freedom, N_{data} is the total number of urine data, N_{avg} is the number of average urine values, and $\langle y \rangle_i$ is the average value associated with urine measurement y_i.

It is also possible to use the lognormal approximation for sufficiently positive data, where sufficiently positive is defined to mean

$$\frac{y_i}{\sigma_{mi}} > \frac{x_{LN}}{S_{nm}} \quad , \tag{12}$$

with x_{LN} a large number (e.g. 3). Then for these data

$$\chi^2 = \sum_{i=1}^{N_{data}} \frac{\left(\ln(y_i) - \langle \ln(y) \rangle_i\right)^2}{S_{nm}^2 + \left(\sigma_{mi} / y_i\right)^2} \cong NDF = N_{data} - N_{avg} \quad , \tag{13}$$

where $\langle \ln(y) \rangle_i$ is the logarithmic average value associated with urine measurement y_i.

A difficulty is the large effect on χ^2 caused by a few outliers. If the summand in Eq. (11) is denoted by x^2, where x is the number of standard deviations the measurement deviates from the average value, one can consider how the sum in Eq. (11) is influenced by the exclusion of outliers with large x values, shown in Table I.

Table I—Influence of outliers on χ^2 / NDF, assuming $S_n = 0.27$ and $S_m = 0.2$. The value of S_n was adjusted to obtain χ^2 / NDF $\cong 1$ when measurements within 4 standard deviations of the average are used (last line of table).

exclude if x >	# excluded (out of 1099)	χ2/NDF
32	0	4.104
16	3	1.453
8	10	1.109
4	14	1.004

Of the 14 urine results excluded in the last line of Table I, 11 were found to have been affected by chelation, which greatly increases urine excretion.

Table II compares the results for $S_n = 0.27$ obtained using the "exact" formula given by Eq. (11) and the lognormal approximation given by Eq. (13) with $x_{LN} = 3$.

Table II—Comparison of χ^2 / NDF for $S_n = 0.27$ using the two methods used to evaluate χ^2.

Method	χ^2	N_{data}	N_{avg}	$NDF = N_{data} - N_{avg}$	χ^2 / NDF
exact	891.5	1085	197	888	1.004
lognormal	548.6	560	99	461	1.190

To estimate the statistical error of χ^2 / NDF because of small sample size in this study the dataset is divided into N_{avg} pieces looking at the statistical variations χ^2 / NDF, where *NDF* in this case is the number of successive samples minus 1. Results are shown in Table III.

Table III—Estimates of statistical uncertainty of χ^2 / NDF because of small sample size.

Method	N_{data}	N_{avg}	χ^2 / NDF
exact	1085	197	0.924 ± 0.072
lognormal	560	99	1.105 ± 0.111

In order for the Table III values of χ^2 / NDF to be within one standard deviation of 1, S_n need be within the range 0.24 to 0.27 using the exact formula, while using the lognormal formula within 0.27 to 0.29.

Composite Representation of Replicate Data

Replicate data are collected on some number *n* (e.g. 3) of successive days with a standard collection procedure or protocol. The final urine measurement is defined as the average of the *n* measurements. The question is how to parameterize the uncertainty associated with this average result.

Consider the data shown in Table IV below.

Table IV—individual replicate urine measurements

measurement uncertainty	normalization uncertainty	data variance
σ_{m1}	S	$(\sigma_{m1})^2 + (A(S)\psi C_V(S))^2$
σ_{m2}	S	$(\sigma_{m2})^2 + (A(S)\psi C_V(S))^2$
σ_{m3}	S	$(\sigma_{m3})^2 + (A(S)\psi C_V(S))^2$

In this Table, the quantity S denotes $\sqrt{S_n^2 + S_m^2}$, also ψ denotes the true (but unknown) median value of the urine excretion, which is assumed to be the same for the three days, $A(S)$ is the lognormal mean/median ratio,

$$A(S) = \exp(S^2/2) \quad ,$$

and $C_V(S)$ is the lognormal standard deviation to mean ratio,

$$C_V(S) = \sqrt{\exp(S^2) - 1} \quad .$$

The measurements are independent so the variances add, and the average result has variance

$$Var = \frac{1}{n^2}\sum_{i=1}^{n}(\sigma_{mi})^2 + \frac{1}{n}(A(S)\psi C_V(S))^2 \quad .$$

This expression is the same as the variance of a single measurement with measurement uncertainty

$$\sigma_m^{(eff)} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}(\sigma_{mi})^2} \quad ,$$

and normalization uncertainty $S^{(eff)}$ given by the solution of the equation

$$A(S^{(eff)})C_V(S^{(eff)}) = \frac{1}{\sqrt{n}}A(S)C_V(S) \quad .$$

When S is fairly small,

$$A(S)C_V(S) \cong S(1+\frac{3}{4}S^2) \quad ,$$

which leads to the approximation

$$S^{(eff)} \cong \frac{S}{\sqrt{n}}\left(1+\frac{3(n-1)}{4n}S^2\right) \quad .$$

Thus, the individual urine measurements may be replaced by a single composite urine measurement having measurement uncertainty $\sigma_m^{(eff)}$ and normalization uncertainty $S^{(eff)}$.

An example of urine data uncertainties determined in this way, urine measurements for Mayak case number 1 of the set of 63 cases discussed in Ref. 2 is given in Table V, assuming $S = 0.34$ ($S_n = 0.27$, $S_m = 0.20$), and $R = 19.2$. The reason that the background b varies is that it is based on a certain assumed background concentration of plutonium in the measured aliquot, which is then multiplied by a varying total daily urine excretion volume to obtain the 24h urine excretion. The effective normalization uncertainty is given by

$$S^{(eff)} = \frac{0.34}{\sqrt{3}}\left(1+\frac{(0.34)^2}{2}\right) = 0.208$$

Table V—Numerical example of a single composite urine result replacing 3 replicate measurements. The ratio of background count time to sample count time, R, is 19.2.

Date	y(dpm/24h)	σ_y(dpm/24h)	f(dpm/24h/count)	N(count)	N_B(count)	b(dpm/24h)	σ_b(dpm/24h)
18-Apr-81	1.14	0.25	0.0106	150	192	0.339	0.212
19-Apr-81	1.02	0.15	0.00588	216	192	0.188	0.118
21-Apr-81	2.79	0.46	0.0188	190	192	0.603	0.377

Composite: 1.65 ± 0.18 dpm/24h, $S^{(eff)}$ = 0.208

References

1.

2. Miller, G., Guilmette, R., Bertelli, L., Waters, T., Romanov, S. and Zaytseva, Y., *Uncertainties in Internal Doses Calculated for Mayak Workers—A Study of 63 Cases*, Radiat. Prot. Dosim. Available at http://rpd.oxfordjournals.org/cgi/reprint/ncn181v1 (2008).

3. Miller, G., *Statistical Modeling of Poisson-lognormal Data*, Radiat. Prot. Dosim. **124** (2) 155-163. Available at http://rpd.oxfordjournals.org/cgi/reprint/ncl544v1 (2007).

4. Technical Report for Project 2.4 4-09.2002, Appendix D

5. Kranhenbuhl, M.P., Bess, J.D., Wilde, J.L., Vostrotin, V.V., Suslova, K.G., Khokhryakov, V.F., Slaughter, D.M., and Miller, S.C., *Uncertainties Analysis of Doses Resulting from Chronic Inhalation of Plutonium at the Mayak Production Association.* Health Phys. **89**, 33-45 (2005).

6. Miller, G., *Uncertainties of Internal Dose Calculations for the Mayak Study*, Health Phys. **94,** 581-584 (2008)

7. Bess, J.D., Krahenbuhl, M.P., Miller, S.C., Slaughter, D.M., Khokhryakov, V.F., Suslova, K.G., Vostrotin, V.V., *Uncertainties Analysis for the Plutonium Dosimetry Model, Doses-2005, Using Mayak Bioassay Data*. Health Phys. **93**, 207-219 (2007).