One Person, One Sample (OPOS) Approach to Coworker Modeling

SC&A

SEC Work Group Meeting

April 7, 2014

OPOS Road Map

- What is OPOS?
- How does it differ from earlier coworker model procedures?
- What are the reasons for using OPOS?
- How extensive is the problem?
- How is the mean excretion rate related to the intake?
- How well does OPOS estimate the mean excretion rate?
- Problems with Implementation

SC&A Review of OPOS Methodology

- What is OPOS?
  - "One Person, One Sample" (OPOS)
  - "One Person, One Statistic" is a better description
  - OPOS is the arithmetic average of a worker's bioassay results in the time period

- Why use OPOS?
  - Introduced by NIOSH to address problems of *Data Dominance* and *Correlation*
  - SC&A examined extent of problem with data dominance and correlation at SRS and FMPC

Review – Coworker Bioassay Data Before Introduction of OPOS

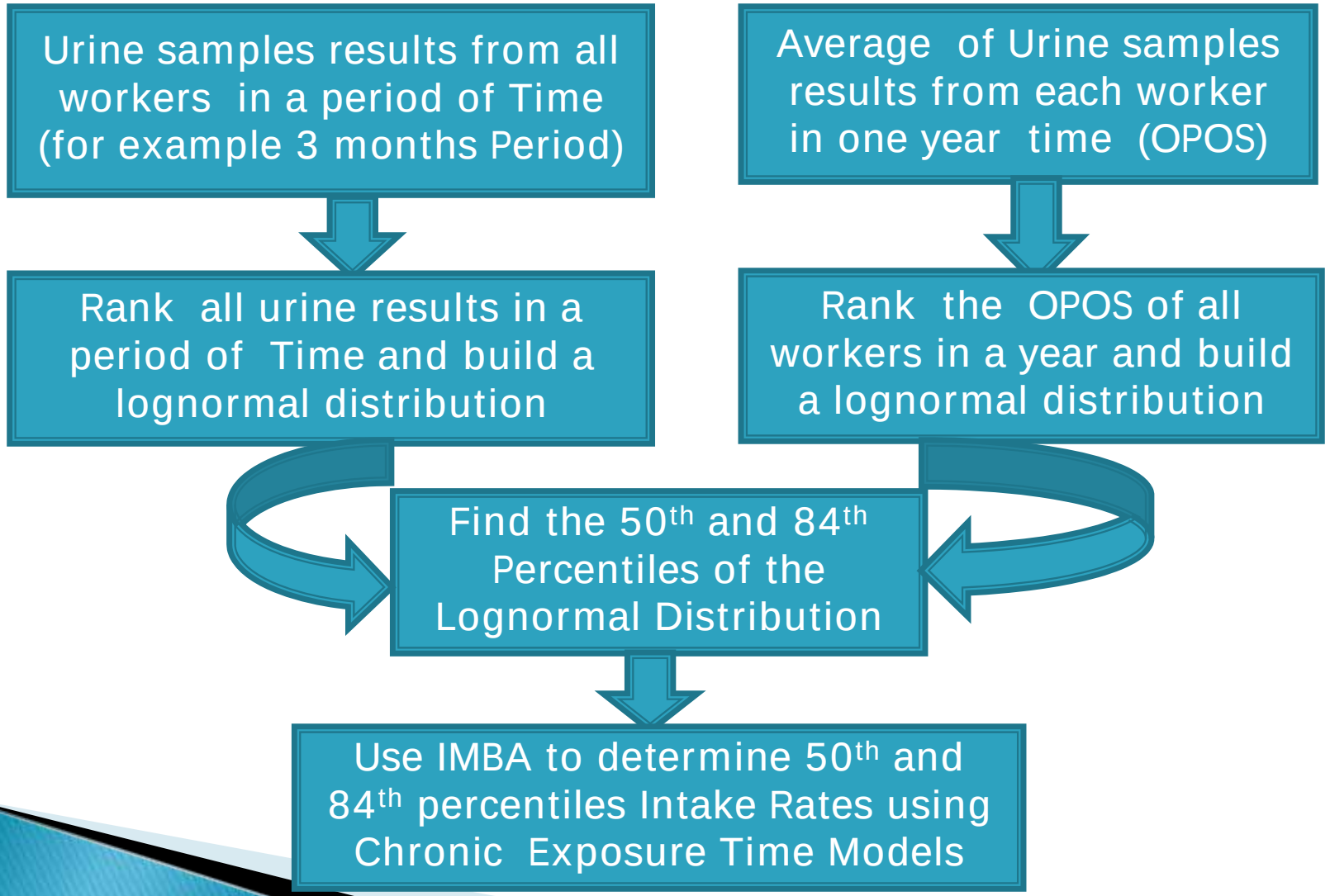
ORAUT-OTIB-0019, Rev. 1, 2005;ORAUT-PROC-0095, 2006

- Assumes that bioassay results for the group of workers have a lognormal distribution.
- Sort ALL the data (FROM ALL WORKERS) for each identified period from low to high results.
- Rank all the data and determine the 50th and 84th percentile values for the ranked data.
- Log-transform the data, calculate the *z*-score for each transformed data point, and plot the *z*-scores on the *x*-axis and the natural logarithms of their respective data on the *y*-axis.
- Use a line equation to calculate the 50th percentile, the GSD, and the 84th percentile value for each period.
- Determine intake rates by performing fits of the two data bioassay data sets (50th and 84th percentile results) associated with each radionuclide. For most data sets, intakes are assumed to be chronic.

Introduction of OPOS

- Assumes that bioassay results for a group of workers have a lognormal distribution.
- For each worker, calculate the average of all bioassay results in ONE YEAR (OPOS) theoretically using the face value of the censored results.
- Sort ALL the OPOS results for EACH YEAR from low to high results (each worker should have one OPOS result in the year being analyzed).
- Rank all the data and determine the 50^{th} and 84^{th} percentile values for the ranked data.
- Log-transform the data, calculate the *z*-score for each transformed data point, and plot the *z*-scores on the *x*-axis and the natural logarithms of their respective data on the *y*-axis.
- Use a line equation to calculate the 50^{th} percentile, the GSD, and the 84^{th} percentile value for each period.
- Determine intake rates by performing fits of the two data bioassay data sets (50^{th} and 84^{th} percentile results) associated with each radionuclide. For most data sets, intakes will be assumed to be chronic.

Main Differences Between OPOS Approach and OTIB-0019 Approach



NIOSH's Position

- OPOS was designed to address problems of:
  - Data dominance: a large fraction of the samples being submitted by small fraction of the individuals.
  - Correlated data: multiple samples submitted by an individual can be correlated, which complicates the use of statistical tests.

SC&A's Principal Finding

- The use of OPOS on an annual (or other fixed-period) basis as a general matter does not appear to be scientifically justified. *The use of pooled individual bioassay data is recommended despite its known drawbacks. When there is clear evidence of data dominance, the samples related to a particular incident may be averaged to provide a single composite data point to be inserted into the distribution of pooled data, resulting in a "mixed" model.*

- SC&A notes that even in this limited context, there is a high degree of uncertainty in the estimated OPOS value due to irregularly spaced collection times, the Regression through the Origin (RTO) hypothesis, the assumption that the variance of the residual error is in direct proportion to the magnitude of intake retention function, and the use of weights inversely proportional to the variance of the measurement.

DATA Dominance

- How relevant is the problem of data dominance?
- Will a large number of incident-related samples from a few workers skew the distributions used for coworker modeling?
- How frequently do we find data dominance in DOE Facilities?

DATA DOMINANCE: Frequency Distribution of Pu SRS Bioassay Results per Claimant (1950–1991)

Number of Bioassay per Time Interval	Frequency per quarter	Cumulative % per quarter	Frequency per year	Cumulative % per year
1	15,273	86.99%	7,402	62.43%
2	1,644	96.35%	2,445	83.05%
3	346	98.32%	867	90.36%
4	148	99.16%	591	95.34%
5	47	99.43%	219	97.19%
6	41	99.66%	111	98.13%
7	17	99.76%	65	98.68%
8	14	99.84%	50	99.10%
9	7	99.88%	31	99.36%
10	9	99.93%	19	99.52%
>10	12	100%	57	100%

In over 95% of the cases where OPOS would be applied at SRS, the workers have no more than four Pu bioassays in the period.

Data Dominance at Fernald

- *ORAUT-OTIB-0078, Rev. 2, 2012*
  OPOS methodology: Inclusion of Codes 50 samples (special study-large number of samples from several workers).

- *ORAUT-OTIB-0078, Rev. 1, 2010*
  NO Codes 50 samples.

- Both versions of OTIB-0078 take into account results labeled 40 and 49, which are incident-related samples.

- The comparison of the 50th and 95th percentile intake rates derived in Rev. 1 (2010) and Rev. 2 (2012) of OTIB-0078 has shown that in some years Rev. 1 produced higher results, while in other years Rev. 2 produced higher results, **showing that neither methodology has a systemic bias that will always yield the higher intake rates.**

Correlation

- The fact that some workers have more samples than other workers in a given time period is not in itself a basis to establish correlation.
- When workers are exposed in accidents or special work assignments, it is not clear why correlation problems end from one year to the other.
- At FEMP, the use of OPOS methodology has not resolved the dependence of monitoring results from different and sequential intake periods. OTIB-0078 Rev. 2 (2012) explicitly exemplifies that for the 1994-2006 period, earlier intake rates significantly biased later intake rates for all solubility Types of uranium compounds.

Influence of Frequency of Monitoring on OPOS

- The computation of OPOS for the year averages urine activities collected from periods of no intakes lumped together with activities from periods with intakes. **The consequence is a strong dependence on the frequency of monitoring, in addition to the number of significant exposures**.
- The annual individual OPOS result at Fernald calculated for workers involved in the same incident was shown to be influenced by the worker's frequency of monitoring.
- For the same reason, when OPOS is applied to compare two groups of workers, it is necessary that the monitoring protocols be the same for the two groups in order to make a meaningful comparison.

Data ⇔ OPOS ⇔ Intake ?

- The cornerstone of NIOSH's defense of OPOS is the proportional relationship of the mean excretion rate and intake estimate when using weighted least square regression to calculate the intake. This raises two technical questions:
  1. When is the mean excretion rate proportional to the intake?
  2. How well does OPOS estimate a worker's mean excretion rate?

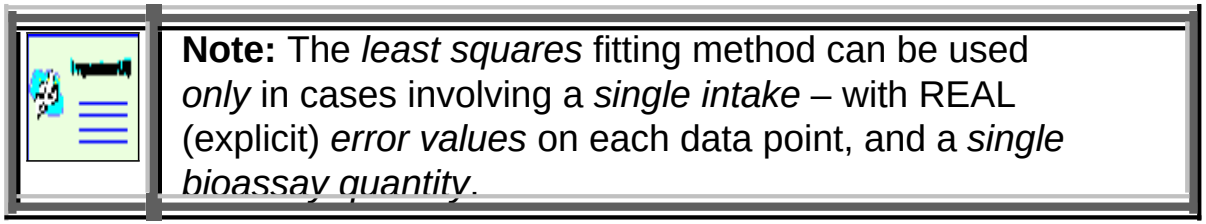
Using Weighted Least Squares to Justify OPOS

- **The least square regression result is only valid to calculate the intake when using the excretion results in urine that are a consequence of that intake.**
- NCRP 164, 2013:

  "This appendix provides a summary of the least-squares method formulas that can be used to derive the intake starting from measurements of activity in bioassay samples. **The formulas assume only one intake,** no prior knowledge about the magnitude of the intake (i.e., uniform prior in the Bayesian formulation of the intake derivation problem), the biokinetic model and its parameters are known perfectly, and that all measurements are independent, and properly normalized (e.g., all urine data represent excretion of activity in 24 h)."

Using Weighted Least Squares to Justify OPOS, Continued

- From the IMBA User Manual:



Note: The *least squares* fitting method can be used *only* in cases involving a *single intake* – with REAL (explicit) *error values* on each data point, and a *single bioassay quantity*.

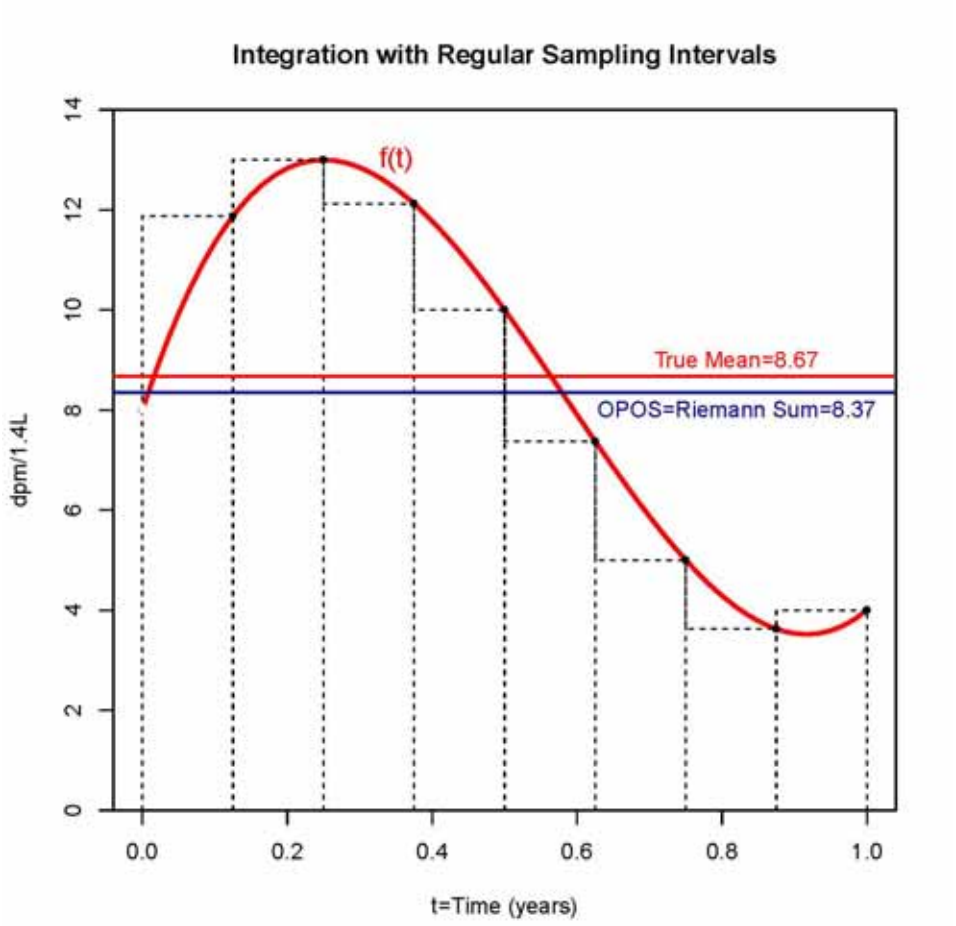
- When there are multiple intakes, the equations are different.

- NIOSH's justification for the use of a single OPOS value for each worker is based on a method that applies only to excretion results after an intake takes place and not for excretions resulting from mixed intakes, or for urine activities collected from periods of no intakes lumped together with activities from periods with intakes.

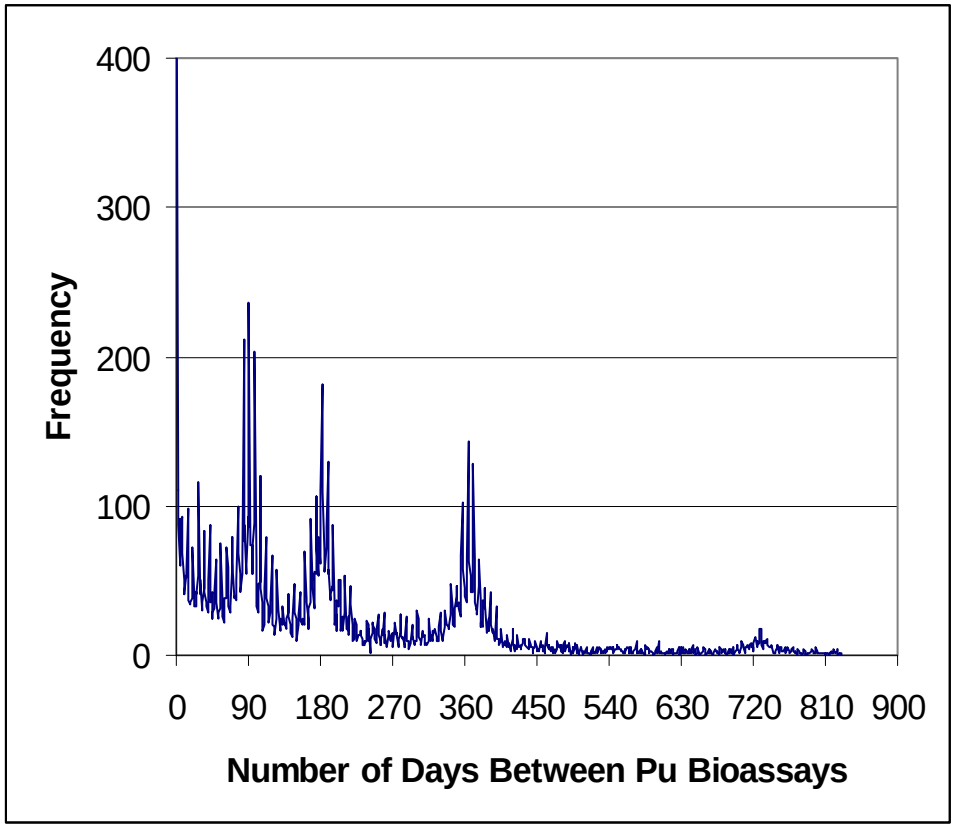
How Well Does OPOS Estimate a Worker's Mean Excretion Rate?

- OPOS ignores the times when bioassays are collected during the year
- Weighted least squares method ignores the ordering of the observations in time
- Collection times are important for constructing time-weighted average urine excretion rates

In this analysis the excretion rate f(t) varies over time. With regularly spaced collection times, OPOS provides a good estimate for the integral under the curve.



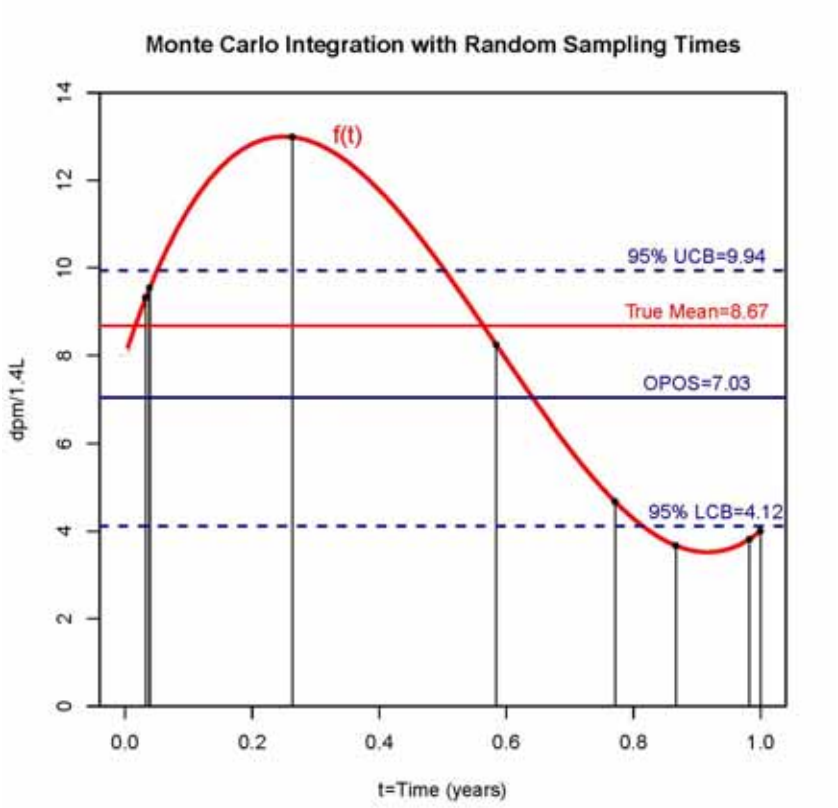
The Actual Picture is More Complicated



Frequency Distribution of Time Interval in Days between Successive Plutonium Bioassays for Claimants Working at SRS

Random Collection Times

- If bioassays are collected at random times during the year, OPOS is interpreted as a Monte Carlo integral of the function f(t) with a small number of samples



Small Sample Size → High Uncertainty in OPOS Estimate

- Student-*t* distribution with N-1 degrees of freedom (DOF) is used to find confidence bounds on the mean excretion rate
- Width of a confidence interval for the mean is derived from the variance of the *t* distribution:

$$Var(t) = \frac{DOF}{DOF - 2}, \qquad DOF > 2$$

OPOS Uncertainty is Very High Very Often

- If N is less than 4 (DOF<3), the variance of the OPOS mean estimate is infinite
- Over 90% of cases in the table on Slide 10 have N less than 4 per year!

How was OPOS Implemented in Coworker Models for SRS and FMPC?

- When nondetects are present, OPOS is calculated using the Maximum Possible Mean (MPM) algorithm as described by NIOSH
  - Step 1) Use the MDA or censoring level (CL) for data reported as <MDA.
  - Step 2) If the analysis period includes all censored data, use the mean result as a censored value for each person.
  - Step 3) If there are uncensored data during the analysis period, use the mean result as an uncensored result for each person.

How was OPOS Implemented in Coworker Models for SRS and FMPC? (continued)

- Inconsistencies were found when the MPM is applied with censored data.
- The face value of the CL should be used for all data entries explicitly reported as censored values and for all entries with recorded values that are below the CL. Although numerical values below the CL (usually zero and negative values) may be reported in the data, these entries should not be used to compute the MPM.

Questions