Assumptions and Methods in Dose Reconstruction

Daniel J. Strom
Battelle
Richland, Washington USA
strom@pnl.gov +1 (509) 375-2626
Professional Enrichment Program 3H
52nd Annual Meeting of the Health Physics Society, Portland, Oregon
Sunday, July 8, 2007

Battelle operates Pacific Northwest National Laboratory for
the U.S. Department of Energy under
Contract Nos. DE-AC05-76RLO 1830 and 1831

Disclaimer

The opinions expressed herein are those of the author alone and have not been reviewed or approved by Battelle, the Pacific Northwest National Laboratory, the National Institute for Occupational Safety and Health, or the U.S. Department of Energy.

This presentation derives from the author's experience in dose reconstruction in a variety of contexts since the early 1970s, including contract work for DOE, NIOSH, and other clients.

TBDs for the AWE DRs & SECs for the NIOSH OCAS Under the EEOICPA Using IMBA and IREP

Rosetta Stone

- Technical Basis Documents (TBDs)
- Atomic Weapons Employers (AWEs)
- Dose Reconstruction (DR)
- Special Exposure Cohort (SEC)
- National Institute for Occupational Safety and Health (NIOSH)
- Office of Compensation and Support (OCAS)
- Energy Employees Occupational Illness Compensation Program Act (EEOICPA)
- Integrated Modules for Bioassay Analysis (IMBA)
- Interactive RadioEpidemiological Program (IREP)
- Department of Labor (DOL)
- Advisory Board on Radiation and Worker Health (ABRWH)
- Computer-Assisted Telephone Interview (CATI)
- NIOSH Occupational Claims Tracking System (NOCTS)

Technical Basis Documents for the Atomic Weapons Employers Dose Reconstructions and Special Exposure Cohorts for the National Institute for Occupational Safety and Health Office of Compensation and Support under the Energy Employees Occupational Illness Compensation Program Act Using Integrated Modules for Bioassay Analysis and Interactive RadioEpidemiological Program

Dose Reconstruction in the USA

- Atomic Veterans
- EEOICPA: Department of Health and Human Services
  - Centers for Disease Control and Prevention
    - National Institute for Occupational Safety and Health (NIOSH)
      - Office of Compensation Analysis and Support (OCAS)
        - » http://www.cdc.gov/niosh/ocas/default.html
        - » Lots of rigorously-reviewed documents
- Public: Downwinders, medical patients, radon-breathers
- Workers: Ra dial painters, miners, medical, MED/AEC/ERDA/DOE, nuclear fuel cycle, air crews

Outline 1

- Uncertainty, error, blunder; variability and covariance
- Requirements and equations for dose reconstruction in support of compensation decisions
- Methods and assumptions for generic site technical basis documents for Atomic Weapons Employers under EEOICPA
  - Paucity of data
- Use and misuse of distributions
  - lognormal
  - examples

Outline 2

- Monte-Carlo approach to analysis of uncertainty and variability in time-weighted average air concentrations
- 2 freeware computer programs
  - LOGNORM4
  - Lognormal Fitting Utility
- Independence and covariance, particularly over time
  - context: uncertainty for time series external dosimetry results and doses inferred from bioassay or workplace indicators
  - Ignoring covariance causes underestimation of uncertainty

Uncertainty, Error, Blunder; Variability and Covariance

1995 ISO Guide to the Expression of Uncertainty in Measurement (GUM)

- Extensive, well-thought-out framework for dealing with uncertainty in measurement
  - Clearly-defined concepts and terms
  - Practical approach
- Doesn't cover
  - the use of measurements in models that have uncertain
    - assumptions
    - parameters
    - form
  - representativeness (e.g., of a breathing-zone air sample)

1995 ISO GUM General Metrological Terms - 1

ISO-GUM Term	Meaning
uncertainty (of measurement)	parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand. *It is a bound for the likely size of the measurement error.*
error (of measurement)	result of a measurement minus a true value of the measurand *(i.e., the [unknowable] difference between a measured result the actual value of the measurand*.) "Error is an idealized concept and errors cannot be known exactly" (Note 3.2.1)
random error	result of a measurement minus the mean that would result from an infinite number of measurements of the measurand carried out under repeatability conditions
systematic error	mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand

1995 ISO GUM General Metrological Terms - 2

ISO-GUM Term	Meaning
correction	value added algebraically to the uncorrected result of a measurement to compensate for systematic error
correction factor	Numerical factor by which the uncorrected result of a measurement is multiplied to compensate for systematic error

- In dose reconstruction for decision facilitation, significant effort is put into determining corrections and correction factors

1995 ISO GUM Additional Terms & Concepts

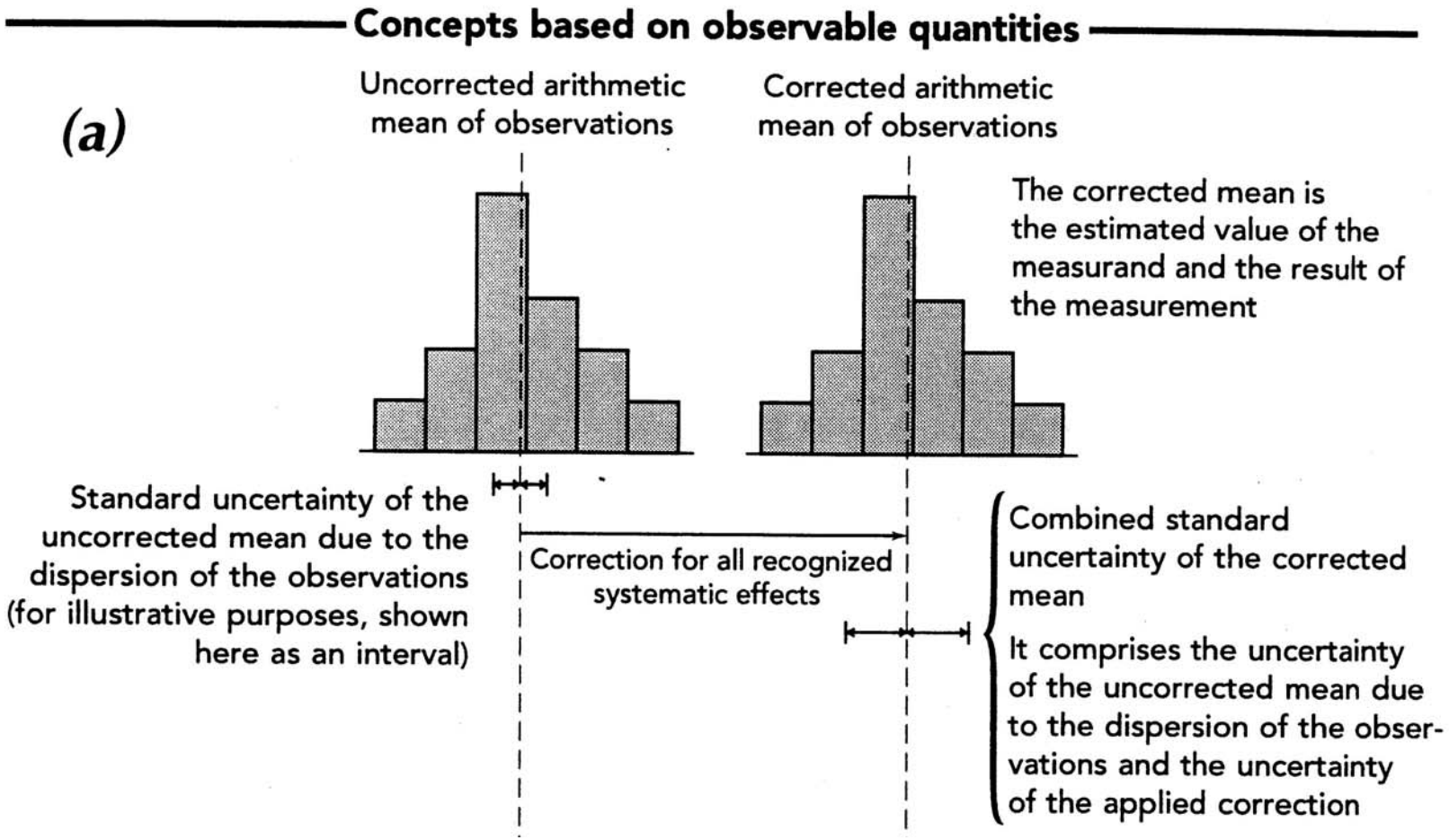
ISO-GUM Term	Meaning
blunder	"Blunders in recording or analyzing data can introduce a significant unknown error in the result of a measurement. Large blunders can usually be identified by a proper review of all the data; small ones could be masked by, or even appear as, random variations. Measures of uncertainty are not intended to account for such mistakes." (3.4.7) Other terms include mistake and spurious error.

- "Mistake" is a synonym for "blunder"

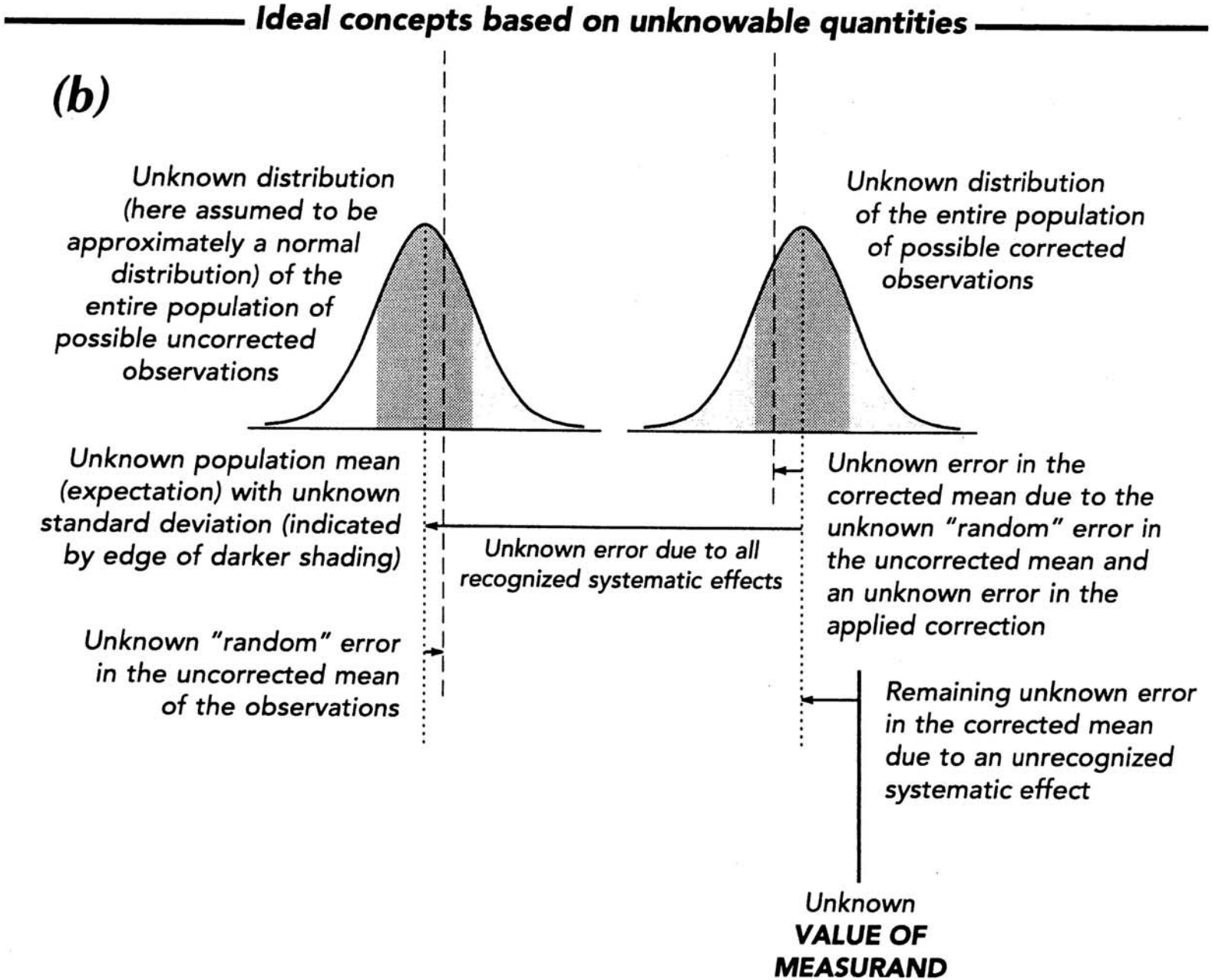
Terms: Error, Uncertainty, Variability

- "The difference between error and uncertainty should always be borne in mind."
- "For example, the result of a measurement after correction can unknowably be very close to the unknown value of the measurand, and thus have negligible error, even though it may have a large uncertainty."
- "Error bars?" No! "Uncertainty bars" is what we should say
- Variability is the range of values for different individuals in a population
  - e.g., height, weight, metabolism

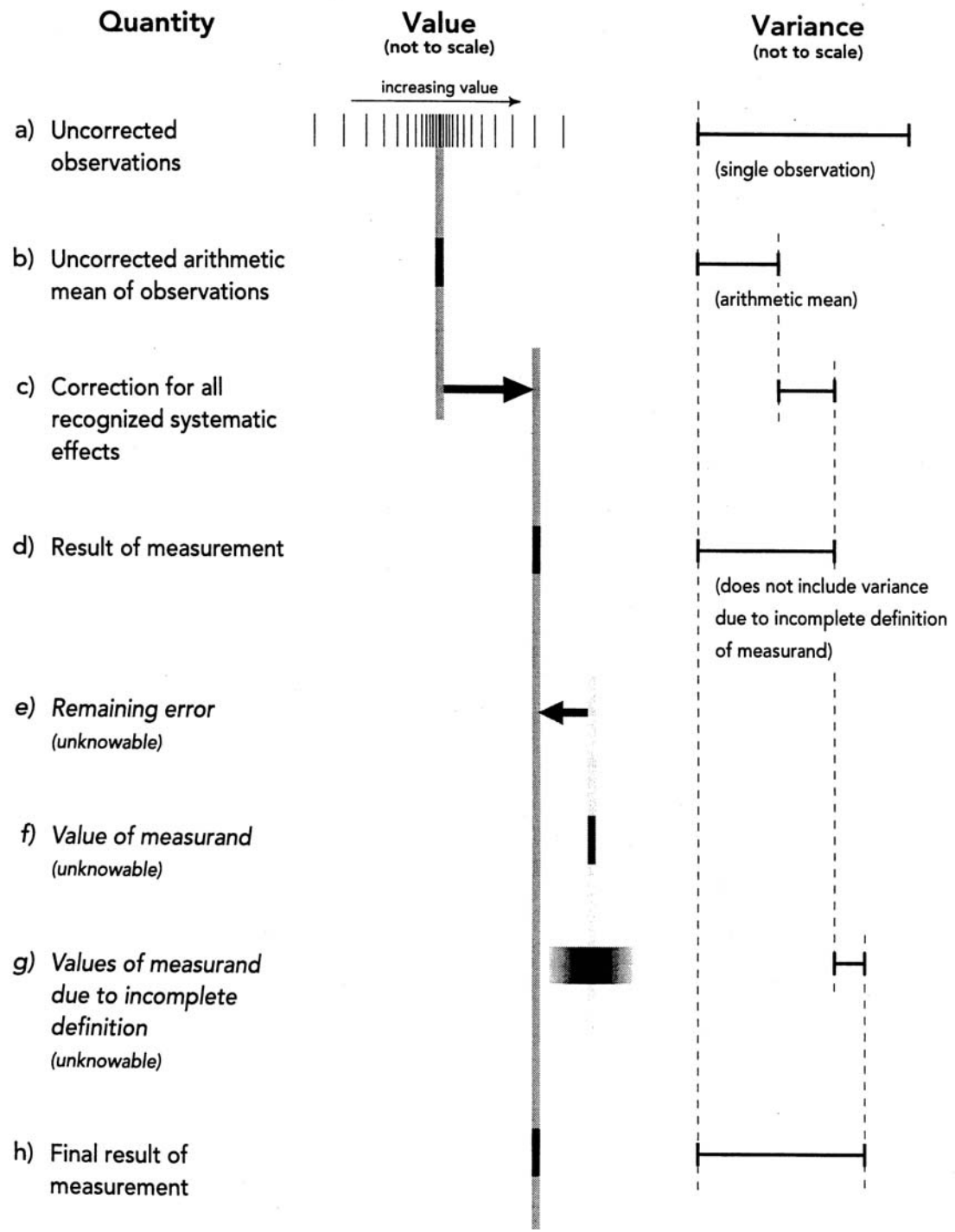
Graphical Illustration of Value, Error, and Uncertainty



Graphical Illustration of Value, Error, and Uncertainty



Graphical Illustration of Value, Error, and Uncertainty



1995 ISO GUM Additional Terms & Concepts - 1

ISO-GUM Term	Meaning
“Type A” uncertainty evaluation	uncertainty that is evaluated by the statistical analysis of series of observations
“Type B” uncertainty evaluation	uncertainty that is evaluated by means other than the statistical analysis of a series of observations

Random and Systematic Uncertainty versus Type A and Type B Uncertainty Evaluation

- GUM: There is not always a simple correspondence between the classification of uncertainty components into categories A and B and the commonly used classification of uncertainty components as “random” and “systematic.”
- The nature of an uncertainty component is conditioned by the use made of the corresponding quantity, that is, on how that quantity appears in the mathematical model that describes the measurement process.
- When the corresponding quantity is used in a different way, a “random” component may become a “systematic” component and vice versa.

Random and Systematic Uncertainty

- "Random uncertainty" and "systematic uncertainty" can be misleading when generally applied
- Alternative nomenclature
  - "component of uncertainty arising from a random effect"
  - "component of uncertainty arising from a systematic effect"

  where
  - a random effect is one that gives rise to a possible random error in the current measurement process
  - a systematic effect is one that gives rise to a possible systematic error in the current measurement process
- In principle, an uncertainty component arising from a systematic effect may in some cases be evaluated by method A while in other cases by method B, as may be an uncertainty component arising from a random effect

1995 ISO GUM: Covariance

ISO-GUM Term	Meaning
covariance	The covariance of two random variables is a measure of their mutual dependence. The covariance of random variables y and z is defined by $\mathrm{cov}(y,z) = \mathrm{cov}(z,y) = E\{[y - E(y)][z - E(z)]\}$ which leads to $\mathrm{cov}(y,z) = \mathrm{cov}(z,y)$ $= \iint (y-\mu_y)(z-\mu_z)p(y,z)\,\mathrm{d}\,y\,\mathrm{d}\,z$ $= \iint yzp(y,z)\,\mathrm{d}\,y\,\mathrm{d}\,z - \mu_y\mu_z$ where $p(y, z)$ is the joint probability density function of the two variables y and z. The covariance cov(y, z) [also denoted by $v(y,z)$] may be estimated by s(y_i,z_i) obtained from n independent pairs of simultaneous observations y_i and z_i of y and z, $s(y_i, z_i) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})$

1995 ISO GUM: Correlation Coefficient, Independence

ISO-GUM Term	Meaning
correlation coefficient	The correlation coefficient is a measure of the relative mutual dependence of two variables, equal to the ratio of their covariances to the positive square root of the product of their variances. Thus, $$\rho(y,z)=\rho(z,y)=\frac{\upsilon(y,z)}{\sqrt{\upsilon(y,y)\upsilon(z,z)}}=\frac{\upsilon(y,z)}{\sigma(y)\sigma(z)}$$ with estimates $$r(y_i,z_i)=r(z_i,y_i)=\frac{s(y_i,z_i)}{\sqrt{s(y_i,y_i)s(z_i,z_i)}}=\frac{s(y_i,z_i)}{s(y_i)s(z_i)}$$ The correlation coefficient is a pure number such that $-1\le\rho\le+1$ or $-1\le r\le+1$.
independence	Two random variables are statistically independent if their joint probability distribution is the product of their individual probability distributions.

Requirements and Equations for Dose Reconstruction in Support of Compensation Decisions

Dose Reconstruction under EEOICPA

3 customers of the Technical Basis Document (TBD):

1. the dose reconstructor
2. the scientist and software engineer who creates the DR spreadsheet tool
3. the claimant, the public, NIOSH OCAS, DOL, and the ABRWH

- The mechanics of dose reconstruction…

Possible Outcomes of TBD Development Based on NOCTS, CATI, and Site Survey Research

- Adequate information for TBD for individual dose reconstruction
- Not enough information on individuals
  - part of work history
  - all of work history

  → use generic site data
- No information or not enough information on a particular site or particular topic on a site
  - part of site history
  - all of site history
  - if no info, must state "In the documents supplied by NIOSH, no information was found on [name topic]."

  → use generic TBD (ore, refining, metalworking)

Outcomes of Dose Reconstructions

- No question – it's compensable
  - even with low-ball dose assumptions
- No question – it's not compensable
  - even with high-ball dose assumptions
- Lots of hard work – borderline compensable
  - can result from great uncertainty

Sources of Information for Dose Reconstruction

- Project-wide information
  - 42 CFR Parts 81-83
  - Implementation Guides
  - Technical Information Bulletins (TIB)s and default assumptions
  - IMBA
- Site-specific Information
  - TBDs and Site-Specific Appendices
  - General monitoring results
  - Amounts of material in process, e.g., tons of U ore
  - Job-Title or co-worker monitoring results
  - Details from all CATI data for a site
- Claimant-specific information
  - NOCTS info on claimant, e.g., job title(s), ICD-9 code(s), various dates
  - CATI data from claimant, survivors, or both
  - Individual monitoring results, including TWAs

"Monitored," "Missed," and "Unmonitored" Doses

- Monitored: measured for an individual
  - film badge, TLD, pocket ion chamber
  - in vivo counting, urinalysis, fecal analysis, exhalation
- Missed: Monitored doses recorded as 0 or "less than" or "M" (minimal)
  - Film couldn't detect < 30 mrem, so weekly film badges could miss (30 mrem/week) × (50 weeks/year) = 1,500 mrem/year
  - Usually use ½ limit of detection (LOD) with a triangular distribution from 0 to LOD
- Unmonitored: no personal measurements (doesn't necessarily mean no measurements whatsoever)

A Site-Specific Spreadsheet Tool for Dose Reconstruction

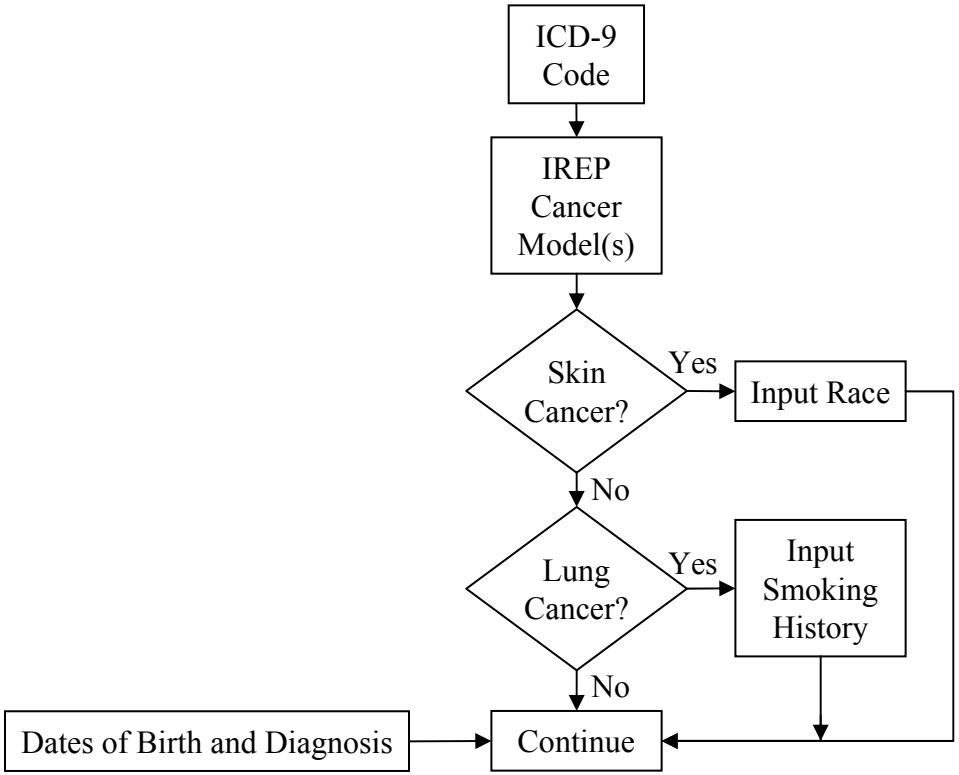
- Is an "executable" version of the TBD
- Incorporates all assumptions and default values in TBD
  - occupational medical, environmental, internal, external doses
- Is "as automated as reasonably achievable" (AAARA)
- Requires
  - understanding of TBD and DR process
  - claimant-specific information
  - professional judgment concerning job titles and default exposures
- Is peer-reviewed with assured software quality

The DR Spreadsheet Tool…

- Produces a correct, defensible, documented spreadsheet for each claimant for input to IREP (Interactive RadioEpidemiological Program)
  - IREP ignores information needed for other purposes
- creates summary data for reports to NIOSH, DOL, and Claimant or Survivors
  - reports will automatically extract personal data and values
- documents the DR process for
  - peer review
  - archives

Algorithms Used by a Dose Reconstruction Spreadsheet Tool

Mapping ICD-9 Code to IREP Cancer Model



Organ Dose from External Irradiation

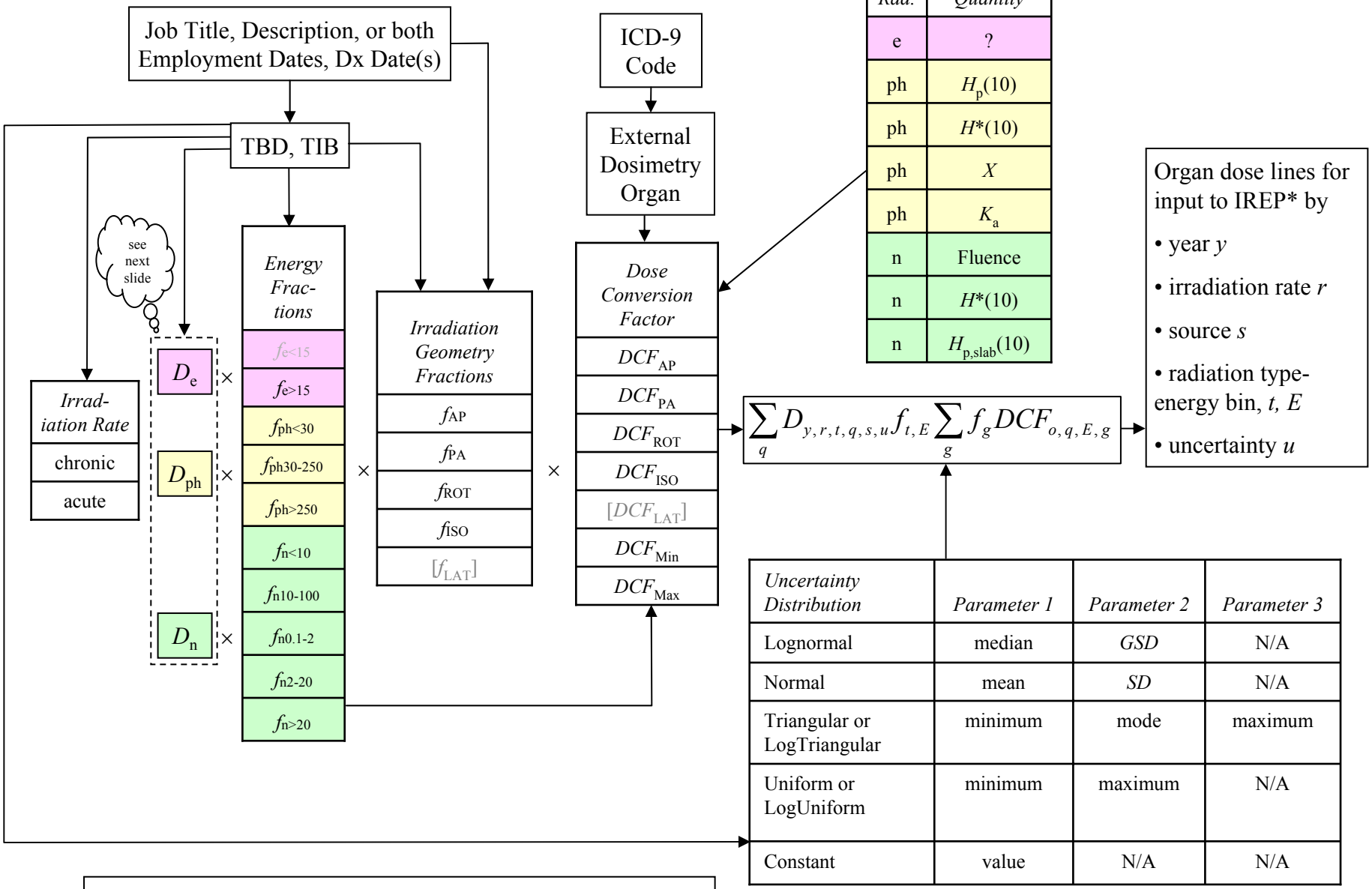
$$Organ\ Dose(y, r, t, E, s, u) = \sum_{q} D_{y,r,t,q,s,u} f_{t,E} \sum_{g} f_g DCF_{o,q,E,g}$$

- *y* (**y**ear): year dose received
- *r* (irradiation **r**ate): chronic, acute
- *t* (radiation **t**ype): electron (e), photon (ph), neutron (n), alpha (a)
- *E* (**e**nergy range or bin):
  - e: <15 or >15 keV
  - ph: < 30, 30-250, >250 keV
  - n: <10, 10-100 keV; 0.1-2, 2-20, >20 MeV
  - a: all
- *s* (**s**ource): occupational medical, environmental, internal (monitored, missed, unmonitored), external (monitored, missed, unmonitored)
- *u*: **u**ncertainty type: normal, lognormal, triangular, logtriangular, uniform, loguniform, constant; as well as uncertainty parameter(s)
- *D*: a "**d**ose"-like variable of one kind or another
- $f_{t,E}$: **f**raction of "dose" *D* of radiation type *t* in energy bin *E*
- *g* (**g**eometry): AP, PA, ROT, ISO, [LAT], Max, Min; Max, Min over all *g*; external irrad. only
- f_g: **f**raction of *D* in received in geometry *g*; external irradiation only
- *DCF*: **d**ose **c**onversion **f**actor (lookup); external irradiation only
- *o*: external dosimetry **o**rgan (may have to repeat calculation if multiple organs are needed)
- *q* (dosimetric **q**uantity):
  - e: unspecified
  - ph: $H_p(10)$, $H^*(10)$, X, K_a
  - n: Fluence, $H^*(10)$, $H_{p,slab}(10)$

Correction of External Dose Results

- All external dose measurements (p, e, n) must be corrected for instrument or dosimeter response to the appropriate quantity such as exposure or fluence (listed above)
- Neutron measurements must be corrected by the ratio of w_R/Q, where
  - w_R is the current ICRP radiation weighting factor for neutrons
  - Q is the quality factor that was used at the time the measurement was recorded

External Dose Spreadsheet Lines for Input to IREP



Rad.	*Quantity*
e	?
ph	$H_p(10)$
ph	$H^*(10)$
ph	X
ph	K_a
n	Fluence
n	$H^*(10)$
n	$H_{p,slab}(10)$

Uncertainty Distribution	*Parameter 1*	*Parameter 2*	*Parameter 3*
Lognormal	median	*GSD*	N/A
Normal	mean	*SD*	N/A
Triangular or LogTriangular	minimum	mode	maximum
Uniform or LogUniform	minimum	maximum	N/A
Constant	value	N/A	N/A

*IREP can accept more than one line for each year for a given exposure rate, rad type, energy bin, and uncertainty.

Origins of "Dose"-Like Data for External Irradiation

D_e, D_{ph}, and D_n may have various origins, and all will vary over time

	Data Source	Nature	Examples
1	Personal Dosimeter	M & m	Film Badge, TLD, criticality monitors, activation of personal items
2	Personal Monitors	M & m	Pocket Ionization Chambers
3	Co-Worker Data	M,m,U	Film Badge, TLD, Pocket Ionization Chambers, etc.
4	Area Monitoring	U	Workplace radiation surveys, duration of exposure
5	Source Term	U	Source strength, distance from source, duration of exposure, and shielding information
6	Radiation Control Limits	U	Generally, workplace posting has been required when the dose rate exceeded 0.025 mSv/hr
-	Soil Concentration, Material in process	U	(Environmental, e.g., FUSRAP data) + FGR-12 Dose Coefficients + Exposure Times
-	X Ray TIB	U	PA Chest radiograph

Nature: M = monitored; m= missed; U = unmonitored.

1-6 Adapted from Table 1.1, OCAS IG-001

FUSRAP (or UMTRCA) Data

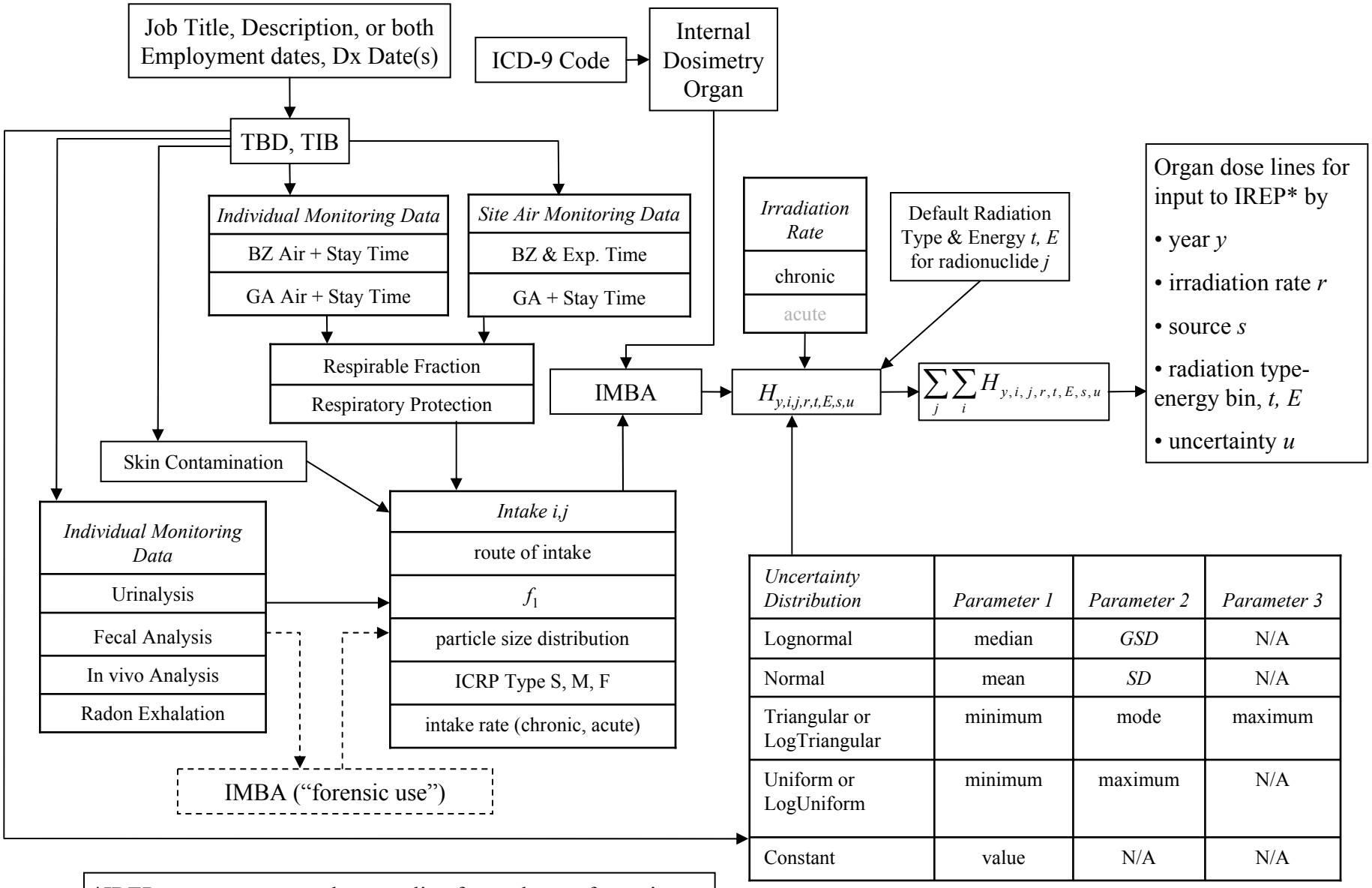
- Needed for environmental dose reconstruction
  - external
  - internal
- 2 kinds of FUSRAP soil contamination data
  - before remediation
  - after remediation
- For the period between cessation of operations (or cessation of cleanup) to the end of the FUSRAP cleanup, the initial FUSRAP site characterization data are needed
  - external dose rates in mR/hour, μR/hour, etc.
  - soil concentrations in pCi/g or other units (specify radioactive materials, e.g., natural U)
  - surface contamination measurements (dpm/100 cm^2), etc.

Organ Dose from Internal Radionuclides Except ^{220}Rn, ^{222}Rn, and Their Short-lived Decay Products

$$Organ\ Dose(y, r, t, E, s, u) = \sum_j \sum_i H_{y,i,j,r,t,E,s,u}$$

- *y* (**y**ear): year dose received; most intakes give doses over many years (up to the date of diagnosis)
- *i* (**i**ntake year): the year the intake occurred
- *j* (radionuclide): the radionuclide taken in, including route of intake; f_1; particle size distribution; ICRP transportability type S, M, F; intake rate (chronic, acute). Highest organ dose may require different choices of transportability type for different organs.
- *r* (irradiation **r**ate): chronic, acute: with rare exceptions, most intakes lead to chronic exposure rates
- *t* (radiation **t**ype): electron (e), photon (ph), neutron (n), alpha (a)
- *E* (**e**nergy range or bin):
  - e: <15 or >15 keV
  - ph: < 30, 30-250, >250 keV
  - n: <10, 10-100 keV; 0.1-2, 2-20, >20 MeV
  - a: all
- NOTE: It is a computational practice to use only one *t, E* combination for each radionuclide *j*
- *s* (**s**ource): occupational medical, environmental, internal (monitored, missed, unmonitored), external (monitored, missed, unmonitored)
- *u*: **u**ncertainty type: normal, lognormal, triangular, logtriangular, uniform, loguniform, constant; as well as uncertainty parameter(s)
- *H*: an "equivalent dose"-like variable of one kind or another; typically an annual organ equivalent dose from IMBA
- *o*: internal dosimetry **o**rgan (may have to repeat calculation if multiple organs are needed)
- *q* (dosimetric **q**uantity) is the same for e, ph, n, a: $H_{T,r,\text{annual}}$ (annual equivalent dose to internal organ *T*). NOTE: *q* is included only for completeness by analogy with external organ doses.

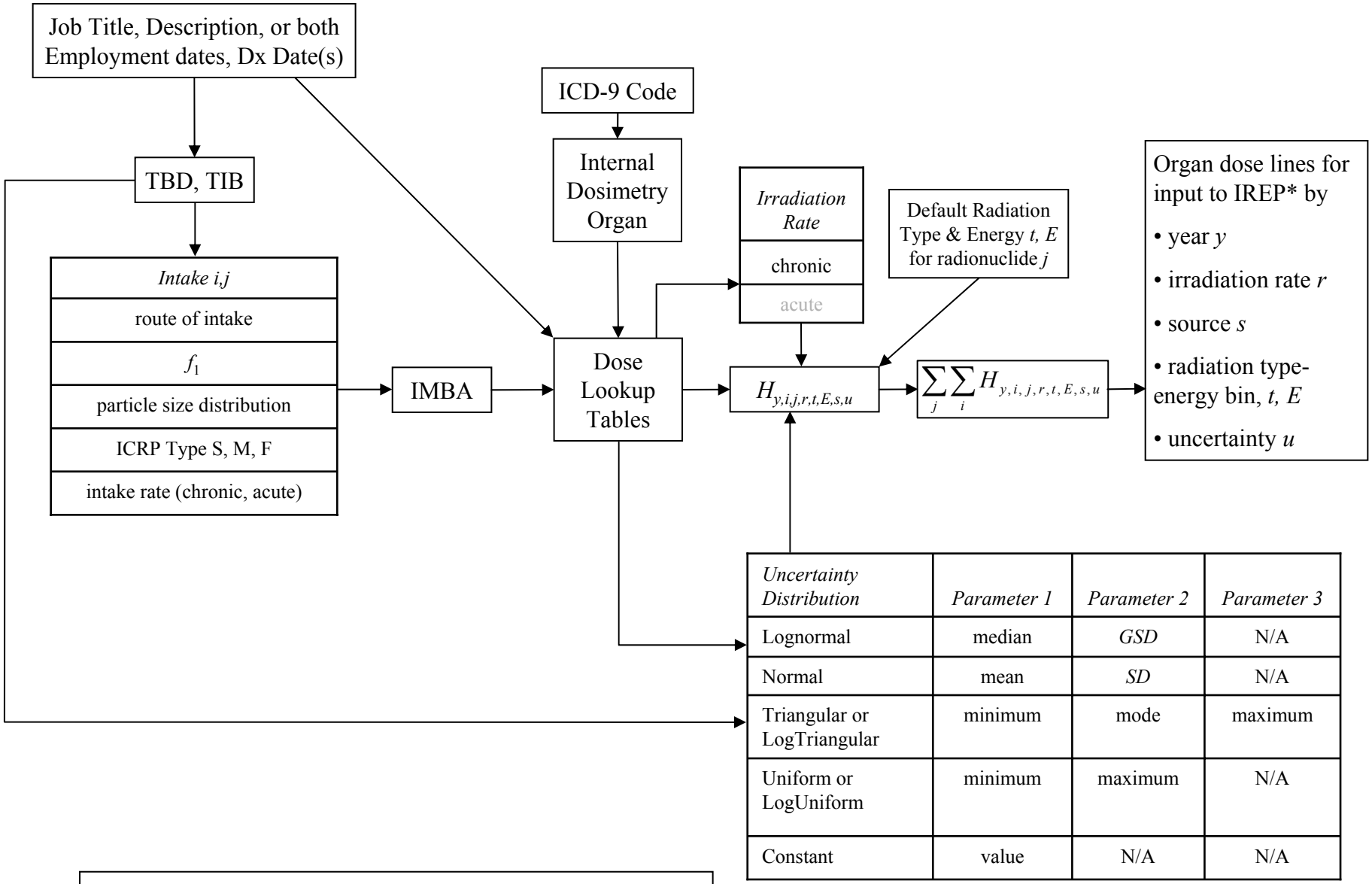
Internal Dose Spreadsheet Lines for Input to IREP: General Case



Uncertainty Distribution	*Parameter 1*	*Parameter 2*	*Parameter 3*
Lognormal	median	*GSD*	N/A
Normal	mean	*SD*	N/A
Triangular or LogTriangular	minimum	mode	maximum
Uniform or LogUniform	minimum	maximum	N/A
Constant	value	N/A	N/A

*IREP can accept more than one line for each year for a given exposure rate, rad type, energy bin, and uncertainty.

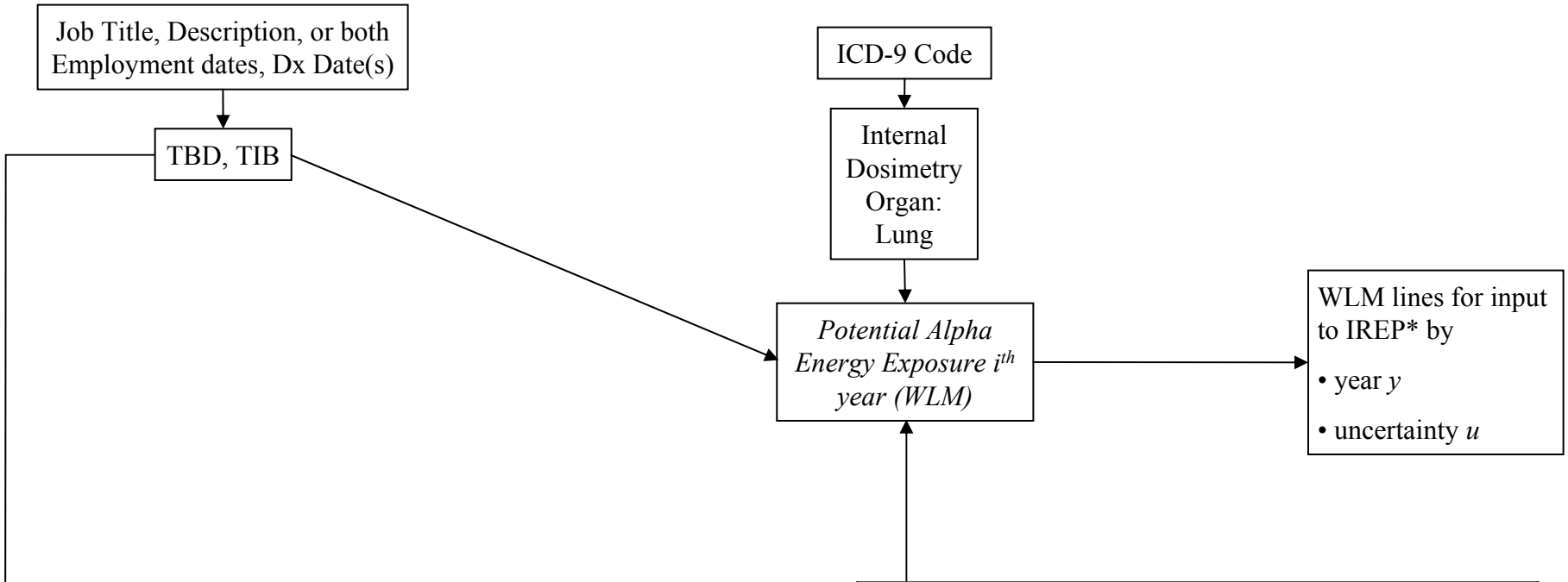
Internal Dose Spreadsheet Lines for Input to IREP: Simplified Generic Case



Uncertainty Distribution	*Parameter 1*	*Parameter 2*	*Parameter 3*
Lognormal	median	*GSD*	N/A
Normal	mean	*SD*	N/A
Triangular or LogTriangular	minimum	mode	maximum
Uniform or LogUniform	minimum	maximum	N/A
Constant	value	N/A	N/A

*IREP can accept more than one line for each year for a given exposure rate, rad type, energy bin, and uncertainty.

Internal Dose Spreadsheet Lines for Input to IREP: Radon and Thoron Decay Products Lung Cancer Only; Requires Smoking History



Uncertainty Distribution	*Parameter 1*	*Parameter 2*	*Parameter 3*
Lognormal	median	*GSD*	N/A
Normal	mean	*SD*	N/A
Triangular or LogTriangular	minimum	mode	maximum
Uniform or LogUniform	minimum	maximum	N/A
Constant	value	N/A	N/A

*IREP can accept more than one line for each year for a given exposure rate, rad type, energy bin, and uncertainty.

Origins of Data for Irradiation Due to Intakes

- All will vary over time
- For site-wide data, values may vary over multi-year intervals.

	Data Source	Nature	Examples
1	In vivo counts	M & m	Whole body counts, chest counts, wound counts
2a	Excreta measurements	M & m	Urinalysis, fecal analysis
2b	Exhalation	M & m	radon, ^{3}H
3	Air samples and exposure times	M & m	GA, BZ air in d/m/m^{3}; WLM measurements for Rn & Tn decay products
4	Contamination measurements	M & m	nasal swabs or skin or wound measurements
5	Resuspended contamination	U	Soil concentration or surface contamination, resuspension factors (FUSRAP data)

Nature: M = monitored; m= missed; U = unmonitored.

Order is of increasing uncertainty in dose

"Forensic Use" of IMBA

- For individual bioassay data, use of IMBA with various (and perhaps multiple) sets of values of
  - route of intake
  - f_1 (fraction taken up from GI tract)
  - particle size distribution
  - ICRP transportability type S, M, F
  - intake rate (chronic, acute)

  may be needed to infer intake parameters

Conclusions:
IH Site Survey Researchers should…

- identify, record the source of, and capture data for TBDs or TBD appendices that support the TBD customers
- need to record the absence of these, too!
- in collaboration with an HP, provide information in a form that can be directly used by the dose reconstructor and DR spreadsheet tool developer without further processing
- Most important: identify documents with *individual monitoring data, job titles,* and *time-weighted average exposure data*

Mission Ridge Ski Area, Wenatchee, Washington

Methods and Assumptions for Generic Site Technical Basis Documents for Atomic Weapons Employers under EEOICPA

Battelle TIB-5000

- Strom DJ. 2007. *Default Assumptions and Methods for Atomic Weapons Employer Dose Reconstruction*. Battelle TIB-5000, http://www.cdc.gov/niosh/ocas/awedocs.html

Battelle TIB-5000: Fitting Statistical Distributions to Data

- Lognormal
- Triangular
- Normal
- Rectangular
- Constant

Battelle TIB-5000: Default Assumptions 1

- External Irradiation Geometry
- 95%ile and "Constant" Uncertainty Distribution for Limited Data Sets
- Uncertainty in Biokinetic Models
- Aerosol Particle Size and Respirable Fraction
- Use of Time-period-specific, Process-based GSDs for Published Mean Aerosol Concentration Data
- Use of Distributions to Describe Multiple Populations
- Use of Time-Weighted Averages, Breathing Zone (BZ) Air Samples, and General Area (GA) Air Samples, Process (P) Air Samples, and Considerations of Sample Duration

Battelle TIB-5000: Default Assumptions 2

- Particle Solubility (ICRP 66 Transportability Classes F, M, S) and f_1 (Gastrointestinal Absorption Fractions)
- Exposure Time and Intake Calculations
- Ingestion
- Occupational Medical Doses
- External Dose Conversion Factors
- External Missed Dose When There Was Monitoring
- Internal Missed Dose When There Was Monitoring
- Environmental Dose

Battelle TIB-5000: Default Assumptions 3

- Radon and Thoron and Their Short-lived Decay Products
  - Radon and Thoron
  - Potential Alpha Energy Exposure and Concentration
  - Equilibrium Factors
  - Summary of radon and thoron quantities and conversion factors
- Radium Monitoring by Breath Radon Analysis
- Determination of the Uncertainty Distribution for Annual Organ Doses Summed Over Multiple Intakes

Battelle TIB-5000: Default Assumptions 4

- Representativeness of Air Samples
  - Inferring Representativeness by Comparing BZ with GA Samples
  - Inferring Representativeness by Comparing Excretion Rates Predicted from Air Samples with Measured Excretion Rates
- Propagation of Medians and Uncertainties for Lognormal Distributions
  - Propagation of Medians (not Means) for Products of Lognormal Distributions
  - Propagation of Uncertainties for Lognormal Distributions
- Adding Doses with Differing Distributions
- Adjusting Process-Specific Dose Rates or Air Concentrations for Time Trends over Periods of Years

Using Lognormal Distributions in Dose Reconstruction

Parameters and Statistics of the Lognormal

Symbol	Parameter	Relationship to other parameters
μ	logarithm of the geometric mean	$\mu = \ln x_{50}$
σ	logarithm of the geometric standard deviation	$\sigma = \ln GSD$
x_{50}	geometric mean (median)	$x_{50} = \exp(\mu)$
GSD	geometric standard deviation	$GSD = \exp(\sigma)$
$\bar{x}$	arithmetic mean	$\bar{x} = \exp(\mu + \sigma^2 / 2)$
x_{mode}	mode	$x_{\text{mode}} = \exp(\mu - \sigma^2)$
CV	coefficient of variation	$CV^2 = \exp(\sigma^2) - 1$
$\text{var}(x)$	variance	$\text{var}(x) = (\bar{x} CV)^2$
SD	arithmetic standard deviation	$SD(x) = \bar{x} CV$
x_{95}	95th %ile	$x_{95} = \exp(\mu + 1.645\sigma)$
x_1, x_2	values 1 and 2	
z_1, z_2	standard normal deviates 1 and 2	
p_1, p_2	percentiles for x_1, x_2	
f_1, f_2	fractiles for x_1, x_2	

Uncensored Individual Observations

- the maximum likelihood estimator of the geometric mean (median)

$$x_{50} = \exp(\mu) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right),$$

- maximum likelihood estimator of the variance of the logarithms of the data

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\ln x_i - \mu\right)^2,$$

- geometric standard deviation, *GSD*

$$GSD = \exp(\sigma) = \exp\left(\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\ln x_i - \mu\right)^2}\right).$$

Sixteen distinct ways of determining a lognormal distribution from minimal information

1	the mean and median (or their natural logs)
2	the mean and mode (or their natural logs)
3	the median and mode (or their natural logs)
4	the median (or its natural log) and the *GSD* or sigma = ln(*GSD*)
5	the mean (or its natural log) and the *GSD* or sigma = ln(*GSD*)
6	the mode (or its natural log) and the *GSD* or sigma = ln(*GSD*)
7	a value and its percentile OR fractile OR std norm deviate and *GSD* or sigma=ln(*GSD*)
8	the median and a value with its percentile OR fractile OR std normal deviate
9	the mean and a value with its percentile OR fractile OR std normal deviate
10	the mode and a value with its percentile OR fractile OR std normal deviate
11	the median and [arithmetic] standard deviation OR coefficient of variation
12	the mean and [arithmetic] standard deviation OR coefficient of variation
13	the mode and [arithmetic] standard deviation OR coefficient of variation
14	a value and its percentile OR fractile OR std norm deviate and [arithmetic] *SD* or *CV*
15	a pair of values and their percentiles OR fractiles OR std normal deviates
16	minimum, maximum, and mean values

LOGNORM4

- Strom DJ and PS Stansbury. 2000. "Determining Parameters of Lognormal Distributions from Minimal Information." American Industrial Hygiene Association Journal 61(Nov-Dec):877-880.
- Both LOGNORM4 and the Lognormal Fitting Utility can be downloaded at http://qecc.pnl.gov/LOGNORM4.htm

Example Using Two Data Points

$$\sigma = \frac{\ln x_2 - \ln x_1}{z_2 - z_1} \text{ and}$$

$$\mu = \frac{\ln x_2 + \ln x_1 - \sigma(z_2 + z_1)}{2}.$$

- "Exposure to Soluble Uranium Compounds," reproduced from Eisenbud and Quigley (1956)

Exposure, mg/m^3	1948	1949	1950	1951	1952
0 - 0.1	9	13	38	33	55
0.1 - 0.5	13	14	62	55	48
0.5 - 2.5	44	31	--	30	22
>2.5	34	61	32	8	--

Grouped Data, 2 Usable Points

Concentration, C (mg/m^3)	Frequency	Cumulative Frequency	f_i
0.1	38	38	0.284
0.5	62	100	0.753
2.5	0	100	0.753
> 2.5	32	132	0.996

- LOGNORM4.EXE (Alt-Enter for full screen)
- Strom DJ and PS Stansbury. 2000. "Determining Parameters of Lognormal Distributions from Minimal Information." American Industrial Hygiene Association Journal 61(Nov-Dec):877-880

Using Minimum, Mean, and Maximum Values with Number of Observations to Determine the Parameters of a Lognormal Distribution

- Problem is overdetermined
- LOGNORM4 can be applied separately to each of 3 pairs
- Can also ignore number of observations and use next method

Warning! May give lousy or even impossible results if one of the extremes is an outlier.

Using Minimum, Mean, and Maximum Values without Number of Observations

- if the $x_{\min}$ and $x_{\max}$ values are symmetric about the geometric mean x_{50}, then the 3 values uniquely determine a lognormal distribution
- $f_{\min} = 1 - f_{\max}$, so that $-z_{\min} = z_{\max}$

$$\mu = \frac{\ln x_{\min} + \ln x_{\max}}{2} \text{ or}$$

$$x_{50} = \sqrt{x_{\min} x_{\max}}$$

$$\sigma^2 = 2\ln\bar{x} - \ln x_{\min} - \ln x_{\max}$$

$$\sigma = \sqrt{2\ln\bar{x} - \ln x_{\min} - \ln x_{\max}}$$

Warning! May give lousy or even impossible results if one of the extremes is an outlier.

$$z_{\min} = \frac{\ln x_{\min} - \mu}{\sigma}$$

$$\text{Check}: \; n \approx \frac{1}{2 f_{\min}}?$$

Use of Range and Average Value Data that Are Inconsistent with a Lognormal Distribution

- The range (minimum and maximum values) and average value data may be inconsistent with a lognormal distribution
  - e.g., $x_{50} \geq \bar{x}$
- Possibility: triangular distribution with the $x_{\min}$ and $x_{\max}$ as parameters 1 and 3
  - mean unlikely to be usable as parameter 2
  - This eliminates the possibility that the underlying population exhibited values outside of the quoted range
  - The latter possibility is quite significant if the number of measurements was few

Using Data that Are Inconsistent with a Lognormal Distribution: Plan B

- The average value is generally a more robust statistic than minimum or maximum values, which can be outliers
- The preferred default assumption is to ignore the minimum and maximum values, use the average (arithmetic mean) value, and *assume* a lognormal distribution with
  - a *GSD* of 5 for data describing a single process (e.g., a series of air samples), or
  - a *GSD* of 10 for data describing an entire site, plant, or factory
- The basis for assuming these *GSD*s derives from analyzing data from many facilities

Censored Individual Observations

- Sometimes values are reported as “less-than” some number or as zero
  - This is referred to as **left-censoring**
- One cannot take the logarithm of zero or a less-than value
- Simple averaging of natural logs won’t work

Fitting a Lognormal to Left-Censored Data

- Sort the data in ascending order
- Assign fractiles $f_i = (i - 0.5)/n$ to each data point
- Transform the fractiles into standard normal deviates, z_i
- Take the natural logarithm of the non-zero, non-censored values, $\ln x_i$
- Using only the logs of non-zero, non-censored observations, perform a uniformly-weighted (i.e., unweighted) linear regression of $\ln x_i$ as a function of z_i
- The slope of the linear regression is σ, and the intercept of the linear regression is μ
- This method is described by Strom (1986) and probably many others

The Lognormal Fitting Utility

- Consider this data set: 18, <2, 5, <2, 2, 3, <2, 8
- Lognormal Fitting Utility

"Reasonableness" of a Lognormal Distribution

- Generally, lognormal distributions with *GSD*s above 10 are not plausibly drawn from a single population
- In fact, any *GSD* > 5 should probably be investigated if possible
- Additional considerations
  - Use of distributions to describe multiple populations
  - Use of time-weighted averages, breathing zone (BZ) air samples, and general area (GA) air samples, process (P) air samples, and considerations of sample and operation duration
- A lognormal may still be a better choice than other distributions accommodated by IREP

Use of Distributions to Describe Multiple Populations

- Statistical distributions used to describe observations of single populations
  - e.g., the distribution of body mass of adult people
- One does not generally use distributions to model combined populations
  - e.g., body masses of infants and body masses of adults
  - body masses of microbes, insects and mammals.
- However, it may be necessary to combine all air samples for a single plant to describe the exposures of workers who are "roamers" or for whom no description of job duties or locations is available
- It is possible to combine breathing zone (BZ) air sample results for different processes and describe the combined data with a statistical distribution, but it is difficult to imagine what such a distribution signifies unless one has a specific end in mind

Use of Distributions to Describe Multiple Populations 2

- One could, for example combine BZ results for “furnace operators” and “fork lift operators” and use the result to describe exposures to “operators” of an unspecified type
- It is possible to combine
  - air samples that were taken for a fraction of a minute during the “dirtiest” part of batch processes, e.g., dumping ore from a drum into a process bin or chipping out a crucible,
  - with air samples that were taken over a period of hours during continuous processes
- Again, it is difficult to know what such a distribution of combined results would represent
- The short-duration air samples were probably taken as worst-case values, while the longer-term samples may have represented average values to which workers may have had prolonged exposures

Use of Time-Weighted Averages, Breathing Zone (BZ) Air Samples, and General Area (GA) Air Samples, Process (P) Air Samples, and Considerations of Sample Duration

- Monte-Carlo approach to analysis of uncertainty and variability in time-weighted average air concentrations

Daily Weighted Average (DWA) Airborne U Concentration

- AEC's Health and Safety Laboratory (HASL) performed air sample measurements in a variety of AWE workplaces in the 1940s and 1950s
- 1 to >20 air samples at each "operation:" measures of variability
- duration and # of times per day operation was performed were recorded for each worker
- HASL averaged concentrations for each operation
- HASL recorded minimum, average, maximum values in reports
- HASL calculated point estimate "daily weighted averages"

Uncertainty in DWA Exposures

- No information on uncertainty was reported
- Most raw data were available
- Primary source of uncertainty was variability in air sample measurements
- Using raw data for each operation for each worker, we performed Monte Carlo simulations to explore the effect of variability on uncertainty
  - Discrete method resampled raw data within each operation for each worker
  - Lognormal method fit lognormal distributions to raw data, and sampled the from the lognormal distributions for each operation for each worker

Punchline

- Adam J. Davis and Daniel J. Strom. Uncertainty and Variability in Historical Time-Weighted Average Exposure Data. Paper MPM-A.6 to be presented at 4:45 (A-105) at the 52nd Annual Meeting of the Health Physics Society, Portland, Oregon, July 9, 2007
- Manuscript has been submitted to *Health Physics*

Freeware for Dealing with Lognormal Distributions

- Both LOGNORM4 and the Lognormal Fitting Utility can be downloaded at http://qecc.pnl.gov/LOGNORM4.htm
- LOGNORM4.EXE (Alt-Enter for full Screen)
- Lognormal Fitting Utility

Independence and covariance, particularly over time

Correlation Example

- Worker has Type M ^{239}Pu inhalation in 1950
- Lifetime of bioassay data
- Dx: Liver cancer, 2007
- Annual doses to liver are not independent; they are highly correlated!
  - Uncertainty in annual dose is highly correlated
- If a Monte Carlo process picks, say, the 80th %ile of the annual dose in 1950 for the intake in 1950, it should pick roughly the 80th %ile of the annual doses for all subsequent years due to this intake
- For the annual dose in 1951 due to the intake in 1950, the 80th %ile should be chosen
- Using correlated uncertainties would result in broader lifetime dose distributions and larger PC values

Limited Independence of External Doses

- Even external doses should have some degree of correlation because
  - systematic dosimetry calibration or response errors would correlate from year to year; and
  - an individual's behavioral differences from some average would differ from year to year
    - e.g., his or her job may have required facing toward the source of radiation more than was assumed and this would be the same from one year to the next as long as the individual had the same job

Ignoring Correlation Underestimates Range of Doses

- Adding annual doses with random, independent uncertainties results in the uncertainties "canceling each other"
- Explicitly correlating doses results in a larger range of annual doses in a Monte Carlo simulation
- A larger range of doses results in a greater range of *PC*
- *SD* may be underestimated by a factor up to $\sqrt{years\ of\ exposure}$

Conclusions from Monte Carlo Simulations of the Sum of Lognormal Random Variables with Same *GSD*

- $y = \sum_i x_i$, where x_i are sampled from lognormals
- $\bar{y} = \sum_i \bar{x}_i$
- $y_{50} > \sum_i x_{50,i}$
- The distribution of the sum y is
  - very skewed
  - not even remotely Normal
  - not too different from a lognormal

This Is Not New!

- "…because of the skewness and large tail of the log-normal distribution, the convergence [to a normal] is slow and the central-limit theorem is, at best, a weak theorem for the case of the lognormal. …the distribution of the sum of log-normal variates for many cases of interest is very accurately represented by a lognormal" (Mitchell 1968)

Conclusions from Monte Carlo Simulations of the Sum of Lognormal Random Variables with Same *GSD*

- The unbiased fit of a lognormal distribution to the distribution of sums underestimates the upper 95th and upper 99th %iles of the distribution of sums and thus is not favorable to the claimant
- The unbiased estimate of the *GSD* of the resultant sum is less than the *GSD* of the random variables
- Using the arithmetic mean of the distribution of sums and the largest *GSD* of the random variables results in substantial overestimates of the upper 95th and upper 99th %iles of the distribution of sums and thus is very favorable to the claimant
- It matters whether doses to a tissue or organ during a given year are summed prior to being put into IREP, or whether they are put into IREP individually

Conclusions

- Dose reconstruction is based on many assumptions
- The larger the uncertainty, the larger the *PC*
- When there is doubt, assumptions are made that overestimate dose
- Compensation decisions under EEOICPA are made on the upper 99th percentile of *PC*
- It is urgently important that all aspects of dose reconstruction be open and transparent, and well-documented
- The lognormal distribution is very useful in dose reconstruction

References

Aitchison J and JAC Brown. 1981. *The Lognormal Distribution.* Cambridge University Press, Cambridge.

Davis AJ and DJ Strom. 2007. "Uncertainty and Variability in Historical Time-Weighted Average Exposure Data.." Health Physics (submitted).

Fix JJ, RI Scherpelz, DJ Strom, and RJ Traub. 2005. *Dose Validation for NIOSH/HERB Multi-Site Leukemia Case Control Study*. PNWD-3538, Battelle-Pacific Northwest Division, Richland, Washington.

International Organization for Standardization (ISO). 1995. *Guide to the Expression of Uncertainty in Measurement. Corrected and Reprinted, 1995*. International Organization for Standardization, Geneva.

Mitchell RL. 1968. "Permanence of the Log-Normal Distribution." Journal of the Optical Society of America 58(9):1267-1272.

Strom DJ. 1986. "Estimating Individual and Collective Doses to Groups with 'Less than Detectable' Doses: A Method for Use in Epidemiologic Studies." Health Physics 51(4):437-445.

Strom DJ. 1994. *LOGNORM4: Determining Parameters of Lognormal Distributions from Minimal Information.* http://qecc.pnl.gov/LOGNORM4.htm, Pacific Northwest National Laboratory, Richland, Washington.

Strom DJ. 2002. "Dosimetry, Dosinference, and Doswaggery." Radiation Protection Dosimetry 98(4):363-366.

Strom DJ. 2003. *A Novel Method of Evaluating Uncertainty in Intake and Effective Dose from Bioassay Measurements Using Existing Computer Codes. Presented at the 49th Annual Radiobioassay and Radiochemical Measurements Conference, Jackson, Wyoming, October 12-16, 2003*. Pacific Northwest National Laboratory, Richland, Washington.

References 2

Strom DJ. 2005. *Extreme Uncertainty: When Dose Reconstruction Becomes Doswaggery*. 1-hour Continuing Education Lecture at the 50th Annual Meeting of the Health Physics Society, Spokane, Washington, July 10-14, 2005, Pacific Northwest National Laboratory, Richland, Washington.

Strom DJ. 2006. *Lognormal Fitting Utility Program*. http://qecc.pnl.gov/Lognormal_Fitting_Utility.htm, Battelle, Richland, Washington.

Strom DJ. 2007b. *Default Assumptions and Methods for Atomic Weapons Employer Dose Reconstruction*. Battelle TIB-5000 rev. 0; http://www.cdc.gov/niosh/ocas/awedocs.html, NIOSH, Cincinnati, Ohio.

Strom DJ. 2007a. "Statistical Issues in Applied Health Physics." Chapter 6 in *Radiological Assessment: A Handbook for Detection, Identification, and Evaluation*, ed. KA Higley. Medical Physics Publishers, Madison, Wisconsin.

Strom DJ and JA MacLellan. 2001. "Evaluation of Eight Decision Rules for Low-Level Radioactivity Counting." Health Physics 81(1):27-34.

Strom DJ and JB Martin. 2004. "Uncertainty in Radiation Dose Reconstruction from Medical X-Rays. Abstract. Presentation at http://www.pnl.gov/bayesian/strom/pdfs/Strom2004G_Uncertainty_in_X-Ray_Dose_Reconstruction.pdf." Health Physics 86(6):S202.

Strom DJ and PS Stansbury. 2000. "Determining Parameters of Lognormal Distributions from Minimal Information." American Industrial Hygiene Association Journal 61(Nov-Dec):877-880.