Dose Estimation For The NCI Study of Thyroid Disease in Kazakhstan: Accounting For Dosimetric Uncertainties

Steven L. Simon, PhD
U.S. National Cancer Institute

IARC SEMI-NUC Dissemination meeting
Hemholtz Zentrum, Munich, Germany
09 November 2015
Outline of Today’s Presentation

1) **Understanding the types of errors** addressed in NCI study of thyroid disease in Kazakhstan

2) **Understanding the Two Dimensional Monte Carlo method (2DMC)**

3) **NCI Study of Thyroid Disease in Kazakhstan – a short summary**

4) **Application of the 2DMC and Bayesian Model Averaging (BMA) methods** to the estimation of risk of thyroid disease in the NCI study of thyroid disease in Kazakhstan
Estimation of radiation dose in the Kazakhstan study (as in many others) involves effects of imprecision and variability as well as bias. The uncertainty in dose estimates is an aggregate of the systematic error and the random errors.

Uncertainty encompasses a wide array of concepts including:

(i) measurement imprecision,
(ii) incomplete information (missing data),
(iii) exposure model simplification (compared to reality),
(iv) natural variability, and
(v) lack-of-knowledge about the true values of the dose model parameters.
Part 1

Understanding Error Types:

- Unshared errors (Berkson & Classical)
- Shared Errors
Repeated measurements (or estimates) of dose to a single individual vary at random about the true but unknown dose for that individual.
Models of Unshared Uncertainty

Berkson error model

Unknown individual true doses are assigned a common value assumed to be equal to the mean value for the group, hence, their unknown true values vary at random about the assigned value.
Repeated measurements or estimates of dose to a single individual vary at random about the true but unknown dose for that individual.

Unknown individual true doses are assigned a common value assumed to be equal to the mean value for the group, hence, their unknown true values vary at random about the assigned value.
Where Classical & Berkson errors originate

Often a result of natural variation of measurements. Repeated measurements vary around true but unknown value. Unknown individual true doses are assigned a common value assumed to be equal to the mean value for the group.

CLASSICAL Error

BERKSON or Assignment Error
Model of Shared Uncertainty

Unknown individual true doses are assigned a common value that may differ substantially from the group mean, hence, there is systematic error of unknown magnitude associated with each of the individual dose estimates.
Model of Shared Uncertainty:
Many groups are possible – each with differing magnitude and direction of shared error
Unknown individual true doses are assigned a common value that may differ substantially from the group mean.
Why do we care about Classical Uncertainty?
True exposures (radiation doses) would give the true dose-response, but we rarely (if ever) have the true doses.
Complex exposure models estimate doses with considerable degree of uncertainty -> resulting in dispersion of the dose estimates from the truth, the degree of dispersion depends on the uncertainty.
The consequence of uncertainty in estimated doses is distortion, dilution, or even over-estimation of the dose-response.
Why Do We Care About Uncertainties?

**CLASSICAL ERROR**
- Will attenuate dose-response towards null.

**BERKSON ERROR**
- Preferred by statisticians since it should result in an unbiased estimate of dose-response, though can result in too narrow of confidence intervals.

**SHARED ERRORS**
- $\pm$ Error in steepness of slope of dose response due to bias in doses, may also effect statistical significance of slope.
## Summary of Effects of Uncertainty

<table>
<thead>
<tr>
<th>TYPES OF UNCERTAINTY</th>
<th>SHARED</th>
<th>BERKSON</th>
<th>CLASSICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description of errors</strong></td>
<td>Uncertainty in models, data, parameters, affecting doses to members of groups of individuals sharing similar exposure attributes</td>
<td>Random uncertainty of unknown true values about an unbiased assigned value</td>
<td>Random uncertainty of estimated, measured or observed values</td>
</tr>
<tr>
<td><strong>IMPACTS OF ERRORS</strong></td>
<td>± Error in steepness of slope of dose response, may effect statistical significance of slope</td>
<td>No slope attenuation, confidence interval may be too narrow</td>
<td>Attenuation of slope towards null.</td>
</tr>
</tbody>
</table>
SUMMARY OF EFFECTS OF DOSIMETRIC UNCERTAINTY

- Estimates of health risk must account for uncertainty in exposure to give honest and unbiased risk estimates.
- Over-confidence in risk estimates, by less than adequate uncertainty analysis, can lead to erroneous understanding of health risk.
The Two Dimensional Monte Carlo (2DMC) Method: A New Paradigm in Exposure Assessment
Monte Carlo Methods: A widely accepted strategy for assessing uncertainty of exposure

Traditional Monte Carlo–based uncertainty analysis samples from probability density functions (PDFs) describing uncertainty or variability of exposure-related parameters.

Example PDFs of Exposure Parameters

Traditional MC methods, do not account for shared errors and treat each study subject independently.

The Two Dimensional Monte Carlo (2DMC) properly partitions shared and unshared errors.
Shared and Unshared errors: Examples

Scenario: Internal exposure – drinking contaminated milk.

Many parameter values are unique to each subject, hence, they are unshared. Some are in common with others and are shared.
Shared errors: Require an understanding of the exposure conditions

If a single parameter estimate for the radioactivity contamination applies to all subjects, there is a potential for shared \textbf{(systematic)} error.

\[
Dose = \frac{\text{radioactivity}}{\text{liter}} \times \text{liters consumed} \times \frac{\text{Dose}}{\text{Unit activity}}
\]

- **Shared**
- **Unshared**
- **Shared**
When the study cohort or part of the cohort (a subgroup) is assigned the same value of a parameter in a dose assessment model, they not only share the value, but they share the error on the value!
The 2DMC provides a simple “language” for assigning shared and unshared status to exposure parameters and the necessary Monte Carlo sampling rules.

Exposure-related variable types defined in 2DMC method:

- $S_c$ (shared among the cohort)
- $S_s$ (shared among subgroups)
- $U_c$ (unshared within the cohort)
- $U_s$ (unshared within subgroups)
- + combinations

Monte Carlo sampling rules:

- Each shared parameter is sampled ONCE for each subgroup per realization of exposure for the study population.
- Each unshared parameter is sampled ONCE for each study subject per realization of exposure for the study population.
Within each realization, each shared parameter is only sampled once for the entire subgroup, while each unshared parameter is sampled once for each subject.

In each successive realization, each shared parameter is again sampled once for the entire subgroup and each unshared parameter is again sampled for each person.

Within each realization, each Shared/Unshared parameter must be defined (or redefined) by sampling a new mean and variance from pre-assigned uncertainty distributions.

Once done, individual values of the Shared/Unshared parameter are sampled in that realization for each subject.

In each successive realization, a new mean and variance are sampled to determine the uncertainty distribution of the Shared/Unshared parameter and new values for each subject are sampled again.
Two Dimensions: *Shared* (systematic errors) and *Unshared* (true stochastic variability and unshared errors)
Realization uncertainty and Individual dose uncertainty are both available

Each column is an alternative realization of possibly true doses for the study group as a whole.
Realization uncertainty and Individual dose uncertainty are both available

Each row contains the alternative estimates of dose for a single subject.
The **OLD PARADIGM** for Exposure Assessment

Best estimate of dose per subject; usually assumed to be an unbiased estimate of the mean dose per subject *(but often could not be confirmed as such)*.

\[
\begin{bmatrix}
\bar{D}_1 \\
\bar{D}_2 \\
\bar{D}_3 \\
\bar{D}_4 \\
\bar{D}_5 \\
\bar{D}_6 \\
\bar{D}_7 \\
\bar{D}_8 \\
\bar{D}_9 \\
\bar{D}_{10}
\end{bmatrix}
\]
**NEW PARADIGM:** Multiple Realizations of the cohort dose distribution provided by 2DMC method

<table>
<thead>
<tr>
<th>D_{1,1}</th>
<th>D_{1,2}</th>
<th>D_{1,3}</th>
<th>D_{1,4}</th>
<th>D_{1,5}</th>
<th>D_{1,6}</th>
<th>D_{1,7}</th>
<th>D_{1,8}</th>
<th>D_{1,9}</th>
<th>D_{1,10}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_{2,1}</td>
<td>D_{2,2}</td>
<td>D_{2,3}</td>
<td>D_{2,4}</td>
<td>D_{2,5}</td>
<td>D_{2,6}</td>
<td>D_{2,7}</td>
<td>D_{2,8}</td>
<td>D_{2,9}</td>
<td>D_{2,10}</td>
</tr>
<tr>
<td>D_{3,1}</td>
<td>D_{3,2}</td>
<td>D_{3,3}</td>
<td>D_{3,4}</td>
<td>D_{3,5}</td>
<td>D_{3,6}</td>
<td>D_{3,7}</td>
<td>D_{3,8}</td>
<td>D_{3,9}</td>
<td>D_{3,10}</td>
</tr>
<tr>
<td>D_{4,1}</td>
<td>D_{4,2}</td>
<td>D_{4,3}</td>
<td>D_{4,4}</td>
<td>D_{4,5}</td>
<td>D_{4,6}</td>
<td>D_{4,7}</td>
<td>D_{4,8}</td>
<td>D_{4,9}</td>
<td>D_{4,10}</td>
</tr>
<tr>
<td>D_{5,1}</td>
<td>D_{5,2}</td>
<td>D_{5,3}</td>
<td>D_{5,4}</td>
<td>D_{5,5}</td>
<td>D_{5,6}</td>
<td>D_{5,7}</td>
<td>D_{5,8}</td>
<td>D_{5,9}</td>
<td>D_{5,10}</td>
</tr>
<tr>
<td>D_{6,1}</td>
<td>D_{6,2}</td>
<td>D_{6,3}</td>
<td>D_{6,4}</td>
<td>D_{6,5}</td>
<td>D_{6,6}</td>
<td>D_{6,7}</td>
<td>D_{6,8}</td>
<td>D_{6,9}</td>
<td>D_{6,10}</td>
</tr>
<tr>
<td>D_{7,1}</td>
<td>D_{7,2}</td>
<td>D_{7,3}</td>
<td>D_{7,4}</td>
<td>D_{7,5}</td>
<td>D_{7,6}</td>
<td>D_{7,7}</td>
<td>D_{7,8}</td>
<td>D_{7,9}</td>
<td>D_{7,10}</td>
</tr>
<tr>
<td>D_{8,1}</td>
<td>D_{8,2}</td>
<td>D_{8,3}</td>
<td>D_{8,4}</td>
<td>D_{8,5}</td>
<td>D_{8,6}</td>
<td>D_{8,7}</td>
<td>D_{8,8}</td>
<td>D_{8,9}</td>
<td>D_{8,10}</td>
</tr>
<tr>
<td>D_{9,1}</td>
<td>D_{9,2}</td>
<td>D_{9,3}</td>
<td>D_{9,4}</td>
<td>D_{9,5}</td>
<td>D_{9,6}</td>
<td>D_{9,7}</td>
<td>D_{9,8}</td>
<td>D_{9,9}</td>
<td>D_{9,10}</td>
</tr>
<tr>
<td>D_{10,1}</td>
<td>D_{10,2}</td>
<td>D_{10,3}</td>
<td>D_{10,4}</td>
<td>D_{10,5}</td>
<td>D_{10,6}</td>
<td>D_{10,7}</td>
<td>D_{10,8}</td>
<td>D_{10,9}</td>
<td>D_{10,10}</td>
</tr>
</tbody>
</table>
2DMC: Provides alternative possibly-true sets of doses
Common source of shared errors radiation dose reconstruction: Internal dosimetry

Internal dosimetry requires a model to estimate movement of radioactive materials within the body and organ masses.

Rarely, are rate constants or organ masses known on an individual basis.

If the rate constants or masses are derived from literature or from a subsample of the population and applied to the rest, then shared errors are introduced.

2DMC allows us to account for uncertainty by simulating alternative (shared) values of the rate constants or organ masses.
The Problem of Excess Variation in the 2DMC

- A potential problem in Monte Carlo is excessive variation in individual doses.

- May be due to over-estimation of unshared uncertainty in 2DMC dose estimation

- Excess random unshared uncertainty produces the equivalent of *Classical error*

- Classical error will generally bias the estimate of the dose-response towards the null
How Can We Remove Excess Classical Error from 2DMC Simulations?
Recall that every dose $D_{i,j} = f($shared and unshared parameters$)$

To remove excessive variability for every dose $D_{i,j}$, resample the unshared parameters 100x, maintaining the sampled shared parameter estimates and calculate a mean (or median) dose to replace $D_{i,j}$

Because each mean (or median) is "conditioned" on the value of the shared parameters, the new value is termed a **conditional mean** (or **conditional median**) dose
Replace 2DMC Simulated Individual Doses With Conditional Mean and Median Doses

- Excess random unshared uncertainty is removed from each realization of cohort doses by use of conditional means (or medians) to replace each individual simulated dose.
- Remaining uncertainty results primarily from shared conditions.
- Potential for slope dilution due to classical error is largely removed.
- Within each vector of conditional mean or median doses, mainly Berkson error remains.
- The risk analysis method only has to find the most plausible vector of doses.
2DMC Method: Summary

• A means to unambiguously partition shared and unshared errors.
• Produces alternative, possibly-true realizations (sets) of exposures (doses) to the entire study population.
• Produces uncertainty of exposure on an individual basis.
• Powerful tool for estimating exposures when data are missing and subjects share exposure-related attributes.

Limitations:

Does not guarantee better results unless exposure assessors (dosimetrists) have deep understanding of all dose-related parameters.

Presents a more complex data fitting problem to statisticians.
Part 3

Summary of NCI Study of Thyroid Disease in Kazakhstan
Original study

• Cohort of ~3,000 persons in 8 villages exposed to radioactive fallout from Soviet nuclear testing (1949-1962) at Semipalatinsk, Kazakhstan.

• Thyroid screening by ultrasound + interview in 1998.

• Thyroid radiation doses estimated for external plus internal exposure (by ingestion).

• Uncertainty not characterized or accounted for.

• Risks were derived by conventional regression and published in 2008 (Land et al. *Radiat Res*).

• Thyroid nodule prevalence found to be associated with increasing radiation dose, but internal dose alone was not significant.
Exposures in Kazakhstan from radioactive fallout result from all conventional exposure pathways. Because this work focused on thyroid disease and thyroid exposure, the doses were primarily considered to be a result of food-chain transport of radioiodines.
Determine exposure rates by nuclear test and by location

Estimate ground deposition of I-131

Account for interception on plants, retention, solubility, decay

Estimate concentrations of I-131 in milk, water, air, etc.

Type of house, age, occupation and time spent outdoors

Questionnaire data:
- where were you?
- milk consumption rate?
- origin and type of milk?

Metabolic data:
Age
Mass of thyroid, Dietary iodine
Uptake of I

Internal Thyroid Dose

External Thyroid Dose
1999 study identified need for better information to estimate internal dose

**Strategy:** Obtain information from village elders about common practices by the use of focus groups design to stimulate individual memory and reduce individual recall bias (2005).

**Purpose:** Develop age-, gender- and ethnic-specific distributions of intake-related parameters and descriptions of lifestyle
Questions framed in relation to recognizable concepts and quantities about food intakes.
The utility of focus group interviews to capture dietary consumption data in the distant past: dairy consumption in Kazakhstan villages 50 years ago

M. Schwerin¹, S. Schoenfeld², V. Drozdovitch³, K. Akimshianov⁴, D. Aldyugurov⁵, A. Bouville⁶, C. Land⁷, N. Luckyanov⁸, K. Makuchi⁹, Y. Semenova⁹, S. Simon⁶, A. Tokareva⁸, Z. Zhumadilov⁸ and N. Potishman²

¹RTI International, Research Triangle Park, Durham, NC, USA
²National Cancer Institute – Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA
³Government Semipalatinsk State Medical University, Semipalatinsk, Kazakhstan
2DMC dosimetry findings for Kazakhstan study: External and Total Dose

5,000 2DMC dose vectors

The spread of the CDFs of dose vectors reflects shared errors
Cohort dose distribution uncertainty provided by 5,000 realizations of exposure for entire cohort

2,376 Subject-specific uncertainty provided by individual geometric standard deviation (GSD) across realizations

Replace 2DMC Simulated Individual Doses With Conditional Median Doses

- Excess random unshared uncertainty is removed.
- Potential for slope dilution due to classical error is largely removed.
- Remaining uncertainty results only from shared conditions and is primarily Berkson-type error which facilitates fitting of dose-response.
Part 4

Application of Bayesian Averaging Method (BMA) to NCI Study of Thyroid Disease in Kazakhstan
Bayesian-based dose vector weights for Kazakhstan thyroid disease study

Bayesian-based Markov Chain Monte Carlo (MCMC) method used to fit 5,000 dose vectors.

Dose vector weights (based on relative goodness-of-fit) of each of 5,000 dose distributions.

Shows importance of only 4% of cohort dose realizations.

Comparison of Risk Analysis for Kazakhstan Study Using Conventional Regression and 2DMC & Bayesian Analysis (1/2)

MALES

Conventionally-derived dose response*

BMA**

*Conventional regression on mean dose per person

**2DMC (5,000 cohort realizations) coupled with Bayesian Averaging Method

Land et al., Radiat Res, 2014
Comparison of Risk Analysis for Kazakhstan Study Using Conventional Regression and 2DMC & Bayesian Analysis (2/2)

Note that lower bounds on both external and internal dose are significant using 2DMC and BMA methods.

The application of 2DMC and BMA method resulted in new scientific conclusions:

- **Confidence intervals** that properly reflect exposure uncertainty,
- **Higher** excess risk per unit radiation dose,
- Dose responses for both exposure pathways (external dose and internal dose) are each statistically significant,
- Risk of internal dose and external dose are statistically compatible (new finding for females).
Steps in Kazakhstan Dosimetry and Risk Analysis

• Assign all dose-model parameters to Shared, Unshared, or Shared/Unshared status.

• Simulate multiple vectors of doses using 2DMC (each vector representing a unique set of ‘shared’ exposure conditions)

• Remove Classical Error by Conditional Mean (or Conditional Median) Value

• Search for best fitting vectors to the dose-response model by BMA method
Recent Publications

1. Two-dimensional Monte Carlo Method

2. Bayesian Dose-response Analysis For Epidemiological Studies With Complex Uncertainty In Dose Estimation

3. Application of 2DMC & Bayesian methods to study of thyroid disease in Kazakhstan following exposure to radioactive fallout
Collaborating group on development of uncertainty methods

Me          Deukwoo Kwon        Owen Hoffman        Charles Land

Eduard Hofer        Bob Weinstock        Brian Moroz
(deceased)
The End