Risk Characterization

Patrick L. Gurian, Ph.D.
Drexel University
The Risk Assessment Framework

1. **Hazard Identification**
2. **Exposure Assessment**
   - Plug exposure into the dose-response function
3. **Dose Response**
4. **Risk Characterization**
   - Specific exposures in the scenario of concern
   - Literature dose-response function
Point Estimate

- Single numeric value of risk
  - May correspond to best estimate of risk
  - May be maximum reasonable exposure

- Use parameter values of exposure and dose response parameters corresponding to point estimate of interest
Example: Anthrax

- What is the risk of Anthrax attack?
- Best fit dose-response is Beta-Poisson model
  \[ \text{Alpha} = 0.974 \text{ and } N50 = 62817 \text{ (Haas unpublished)} \]
  \[ \text{Risk} = 1 - \left(1 + \left(\frac{\text{dose}}{62817}\right) \left(2^{\left(\frac{1}{0.974}\right)} - 1\right)\right)^{-0.974} \]

- If 1 spore of *B. antracis* is inhaled
  \[ \text{Risk} = 1 - \left(1 + \left(\frac{\text{dose}}{62817}\right) \left(2^{\left(\frac{1}{0.974}\right)} - 1\right)\right)^{-0.97} \]
  \[ \text{Risk} = 1.6 \times 10^{-5} \]
  Note: this is the fatality risk
Example: Cryptosporidium Risk

• Cryptosporidium is present in a surface water
• What is risk of swimming in this water?

• Let’s calculate a point estimate for our best estimate of risk
  – Use most likely exposure and dose response parameter values
Exposure Analysis

• Assume 10 infective oocysts/liter
• 0.13 liters consumed per swim, 7 swims per year (Lodge et al. 2002)
• Dose = contact rate x concentration
• Dose = 0.13 liters/swim x 10 oocyst/liter
• Dose = 1.3 oocysts/swim
Dose-Response

- Exponential with \( r = 0.004191 \)
- Table 14.13 Gerba
- Risk = 1 - exp(-dose \times 0.004191)
Risk Characterization

- Risk = 1-\exp(-\text{dose} \times 0.004191)
- Dose = 1.3 oocysts/swim
- Risk = 1-\exp(-1.3 \times 0.004191)
- Risk= 1-\exp(-0.0054483)
- Risk =1-0.9946
- Risk=0.0054
- Note this is risk of infection per swim
Morbidity and Mortality

- Often view risk of illness and death as independent of dose given that infection has occurred

- Based on Haas et al. 1999
  \[
  \text{Prob[illness|infection]} = 0.39 \\
  \text{Prob[death|illness]} = \sim 0.001
  \]
Risk of Illness and Death

- Risk of illness
  \[\text{Risk of illness} = \text{Prob}[\text{illness}|\text{infection}] \times \text{Prob}[\text{infection}]\]
  \[= 0.39 \times 0.0054 = 0.0021\]

- Risk of death
  \[\text{Risk of death} = \text{Prob}[\text{death}|\text{illness}] \times \text{Prob}[\text{illness}]\]
  \[= 0.001 \times 0.0021 = 2.1 \times 10^{-6}\]
Annual Risk

• Treat swims as discrete trials with discrete outcomes: infected vs. not infected, ill vs. healthy, dead vs. alive  
  – Binomial distribution

• Risk occurs when infections occurs on 1 or more trials

• No risk occurs when all trials have non-infection outcomes  
  – Easier to calculate
Mathematics of Converting Daily to Annual Risk

AnnualRisk = 1 – prob[no infection in N trials]
Prob[no infect. in N trials] = prob [no infect]^N
Prob[no inf. in N trials] = prob[1-DailyRisk]^N

AnnualRisk = 1 – prob[1-DailyRisk]^N
Annual Risk of Infection

- AnnualRisk = 1 - prob[1 - DailyRisk]^N
- AnnualRisk = 1 - prob[1 - 0.0054]^7
- AnnualRisk = 1 - prob[0.9946]^7
- = 1 - 0.963 = 0.037
Annual Risk of Illness

- AnnualRisk = $1 - \text{prob}[1 - \text{DailyRisk}]^N$
- AnnualRisk = $1 - \text{prob}[1 - 0.0021]^7$
- AnnualRisk = $1 - \text{prob}[0.9979]^7$
- $= 1 - 0.985 = 0.015$
Probabilistic Uncertainty Analysis

- Risk assessments are often subject to large uncertainties

- We often model these uncertainties probabilistically (as if uncertain quantity were subject to random variability)

- Propagate these uncertainties through our model
“Smearing out” parameter estimates

Now it is our most likely value, but not the only possible value.
What are the Goals of Uncertainty Analysis?

• Find range of possible outcomes
• Determine if the uncertainty matters
• Determine which inputs contribute the most to output uncertainty
• Compare range of outcomes under different decisions, policies
  – Inform risk management
Propagating Uncertainty

• Usually use the same formulae as your point estimate

• Parameters are not single values but probability distributions
Uncertainty Propagation (a little more formally)

• Model $F(x)$ where $x$ is a vector of model inputs (parameters)

• Given probability distributions for $x$, what is distribution of $F(x)$?
  – Propagation of uncertainty through model
  – From inputs to outputs
Monte Carlo Uncertainty Analysis

- The work horse of probabilistic risk assessment (PRA)
- Algorithms exist to generate random numbers
- Generate or “sample” $X_1$ and $X_2$
- Calculate corresponding $Y = F(X_1, X_2)$
- Repeat $N$ times, each $Y$ value equally plausible
  $\text{prob } [Y_i] = 1/N$
Monte Carlo Results

- Have a discrete distribution of Y that approximates true distribution of Y
- $E[Y] = \sum Y_i / N$
- $Var[Y] = \sum (Y_i - E[Y])^2 / (N-1)$
- True percentiles of Y $\sim$ percentile of Yi values
- Typically summarize by mean, median, upper bound, and lower bound
Monte Carlo Sensitivity Analysis

• Calculate Correlation of \((Y, X1)\) and \((Y, X2)\) in samples
• Larger (absolute value of) correlation indicates more important influence on \(Y\)
• May wish to do this based on rank order correlations (order all \(Y\) values from 1 to \(N\), all \(X1\) and \(X2\), correlate ranks) to avoid influence of outliers, non-linearities
• Always good to look at scatter plots of \(Y\) vs. \(X\)
Implementing Monte Carlo Analysis

• Need large N
• How large? How many samples/iterations?
• Run until you get convergence
• Answer does not change much as you continue to do additional simulations
• As a rule of thumb 1000 is bare minimum
• 10,000 is recommended (see Kammen and Hassenzahl, Burmaster)
Monte Carlo implementation

• Add on software packages for Excel exist such as @risk and Crystal Ball

• Can be done in Excel without these packages
  – Make each column a variable
  – Each row a realization of your model with different inputs sampled by random number generator
Excel Random Number Generator

- “Tools” select “Add ins”
- Make sure “Analysis Toolpak” is checked
- Then select “Data Analysis” from the “Tools” menu and pick “Random Number Generation”.
- This will bring up a dialogue box and you can enter the appropriate distribution type and parameter values.
From Point Estimate to PRA

- Risk = 1 - exp(-r x ingestion x concentration)
- Choose input distributions that reflect plausible spread in these values
- Ingestion = 0.13 l/swim
- Concentration ~ Poisson(10)
- \( \ln(r) \sim N(-5.5, 0.35^2) \)
  - generate \( \ln(r) \) from normal generator
  - \( r = \exp(\text{generated number}) \)
Here's What It Looks Like in Excel

<table>
<thead>
<tr>
<th>Dose</th>
<th>Ln r generated</th>
<th>r generated</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>-5.584469832</td>
<td>0.00375574</td>
<td>0.003898352</td>
</tr>
<tr>
<td>4</td>
<td>-5.183225496</td>
<td>0.005609883</td>
<td>0.002912888</td>
</tr>
<tr>
<td>14</td>
<td>-5.096604994</td>
<td>0.00611748</td>
<td>0.011072063</td>
</tr>
<tr>
<td>9</td>
<td>-5.562297033</td>
<td>0.003839946</td>
<td>0.004482659</td>
</tr>
<tr>
<td>6</td>
<td>-5.755909526</td>
<td>0.003164028</td>
<td>0.002464899</td>
</tr>
<tr>
<td>9</td>
<td>-4.786115781</td>
<td>0.008344808</td>
<td>0.009715917</td>
</tr>
<tr>
<td>13</td>
<td>-5.911354677</td>
<td>0.002708515</td>
<td>0.00456693</td>
</tr>
<tr>
<td>15</td>
<td>-5.494658538</td>
<td>0.004108659</td>
<td>0.007979876</td>
</tr>
</tbody>
</table>
Presenting Results

• Present both point estimates and distributions, as appropriate
• Give an estimate of central tendency (mean/median) or risk
• Generally want plausible upper bound for risk
  – Not assume people drink nothing but wastewater for 70 years
  – Reasonably maximally exposed individual
• Consider susceptible subpopulations
• Identify major contributors to output variance
  – Are these uncertain? Variable? Both?
Specific Statistics to Present

• Mean, median, standard deviation
• 5th percentile, 95th percentile
• Histogram of output
• Correlations of inputs with output
  In Excel =correl(input column, output column)
  Where input column is A1:A1000 or similar
• Scatter plots of inputs with output
Now we know that risk could plausibly be twice as high.

Point estimate was at the median.
Scatterplot: Risk vs. Dose (correl=.64)
Scatterplot: Risk vs. r (correl=.74)
Risk Characterization

• After all the effort of a Monte Carlo analysis, in practice people want a number
• Tendency to collapse distribution to most likely number (or conservative, protective number)
• What do we really want to get out of our analysis?
• Not just a number but to inform multiple decisions
  – Is risk acceptable? How bad could it be?
  – Can the risk be reduced?
  – What do we need to know to improve management of this risk?
  – Are there subpopulations we should be concerned about?
Informing Risk Management

• What protective action is needed to reduce best estimate of risk to a target value? To reduce upper bound of risk to the target value?

• How much will different risk management actions cost and what risk reductions will they achieve? How certain are we?
Arsenic in Drinking Water Example: Distribution of Costs under 3 Policy Scenarios

Median costs are fine, here we look at upper bound

This is the “acceptable” cost