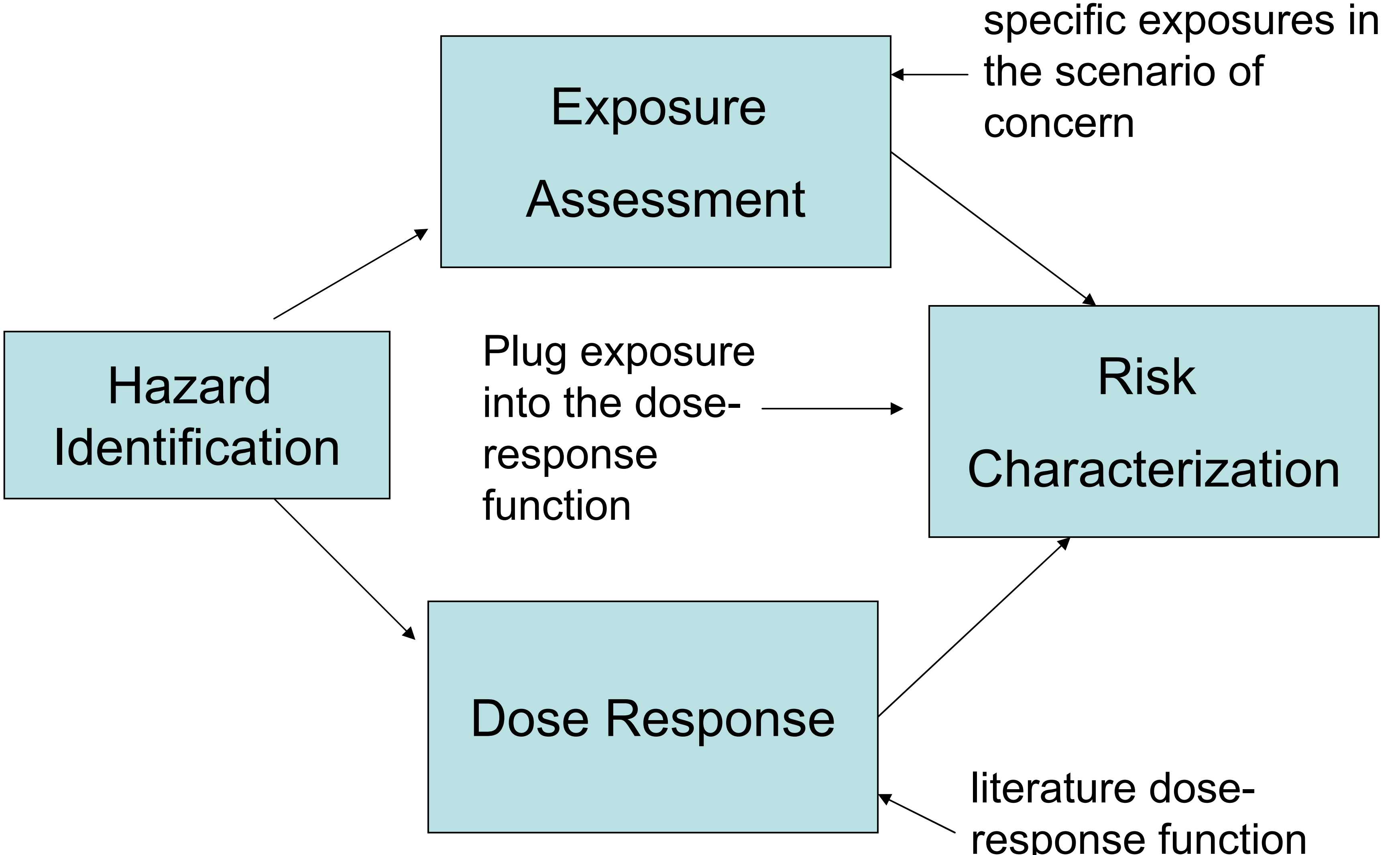
Risk Characterization

Patrick L. Gurian, Ph.D.
Drexel University

Center for Advancing Microbial Risk Assessment

The Risk Assessment Framework



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

  Alpha = 0.974 and N50 =62817 (Haas unpublished)

  Risk = 1-(1+(dose/62817)*(2^(1/0.974)-1))^-0.974

- If 1 spore of *B. antracis* is inhaled

  Risk = 1-(1+(dose/62817)*(2^(1/0.974)-1))^-0.97

  Risk = 1.6×10^{-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 x 0.004191)

Risk Characterization

- Risk = 1-exp(-dose x 0.004191)
- Dose = 1.3 oocysts/swim
- Risk = 1-exp(-1.3 x 0.004191)
- Risk= 1-exp(-.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

  Prob[illness|infection]=0.39

  Prob[death|illness]=~0.001

Risk of Illness and Death

- Risk of illness
  = Prob[illness|infection] x Prob[infection]
  = 0.39 x 0.0054 = 0.0021

- Risk of death
  = Prob[death|illness] x Prob[illness]
  = 0.001 x 0.0021 = 2.1×10^{-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$[1\text{-DailyRisk}]^N$
- AnnualRisk =1-prob$[1-0.0054]^7$
- AnnualRisk =1-prob$[0.9946]^7$
- =1-0.963=0.037

Annual Risk of Illness

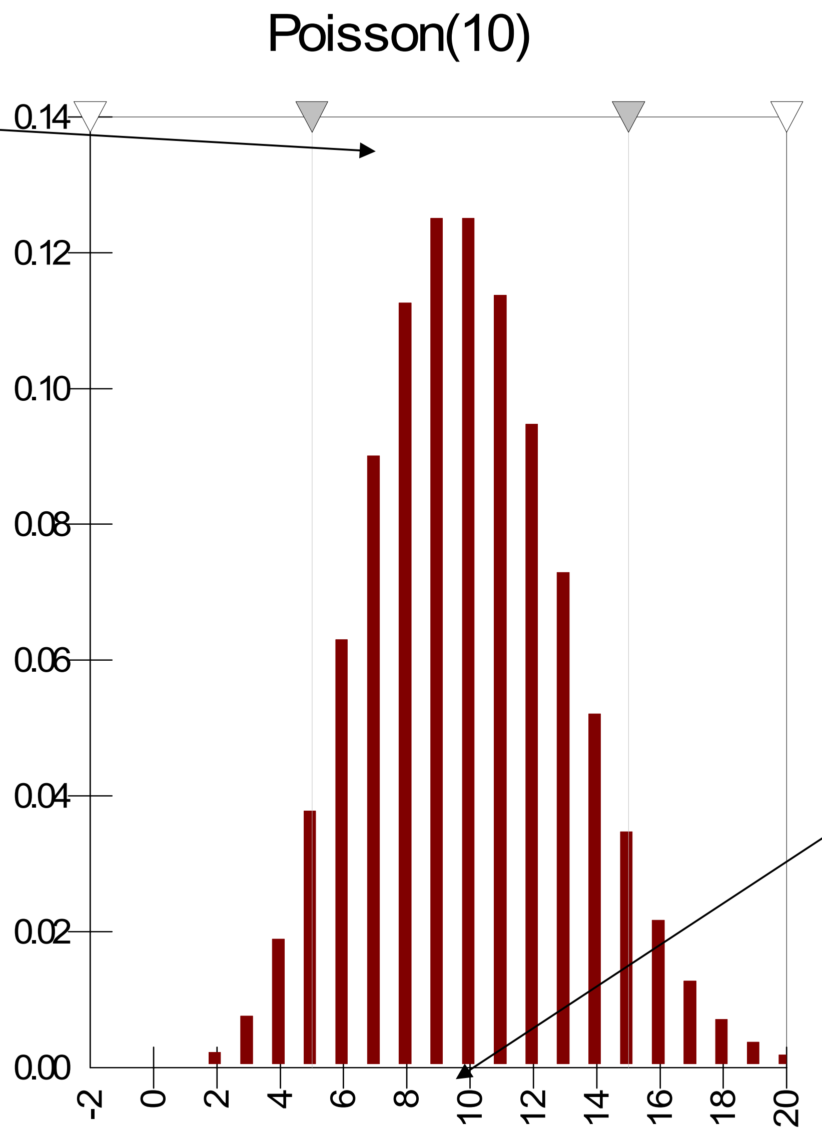
- AnnualRisk =1-prob[1-DailyRisk]^N
- AnnualRisk =1-prob[1- 0.0021]^7
- AnnualRisk =1-prob[0.9979]^7
- =1-0.985=0.015

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



This was our point estimate

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" X1 and X2
- Calculate corresponding Y = F(X1, X2)
- Repeat N times, each Y value equally plausible prob [Yi] = 1/N

Monte Carlo Results

- Have a discrete distribution of Y that approximates true distribution of Y
- $E[Y] = \Sigma Yi/N$
- $Var[Y] = \Sigma\{Yi – E[Y]\}^2 /(N-1)$
- True percentiles of Y ~= 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) ~N(-5.5, 0.35^2)

  generate LN (r) from normal generator

  r = exp(generated number)

Here's What It Looks Like in Excel

Dose generated from Poisson with lambda=10	Ln r generated from N(-5.5, .35^2)	r =exp(C5)		Risk = 1-exp(-D5*0.13*B5)
8	-5.584469832	0.00375574		0.003898352
4	-5.183225496	0.005609883		0.002912888
14	-5.096604994	0.00611748		0.011072063
9	-5.562297033	0.003839946		0.004482659
6	-5.755909526	0.003164028		0.002464899
9	-4.786115781	0.008344808		0.009715917
13	-5.911354677	0.002708515		0.00456693
15	-5.494658538	0.004108659		0.007979876

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?

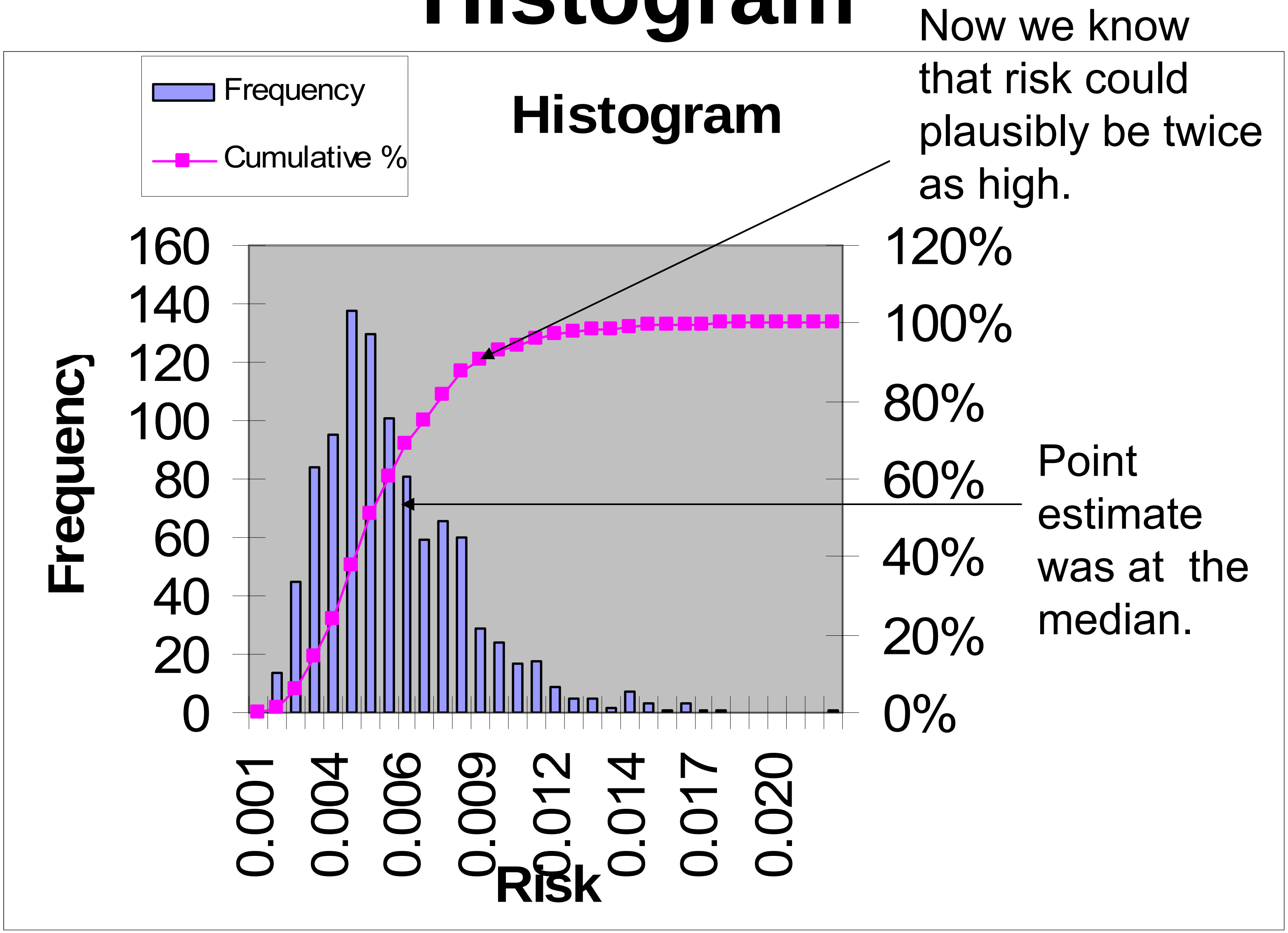
Specific Statistics to Present

- Mean, median, standard deviation
- 5^{th} percentile, 95^{th} percentile
- Histogram of output
- Correlations of inputs with output

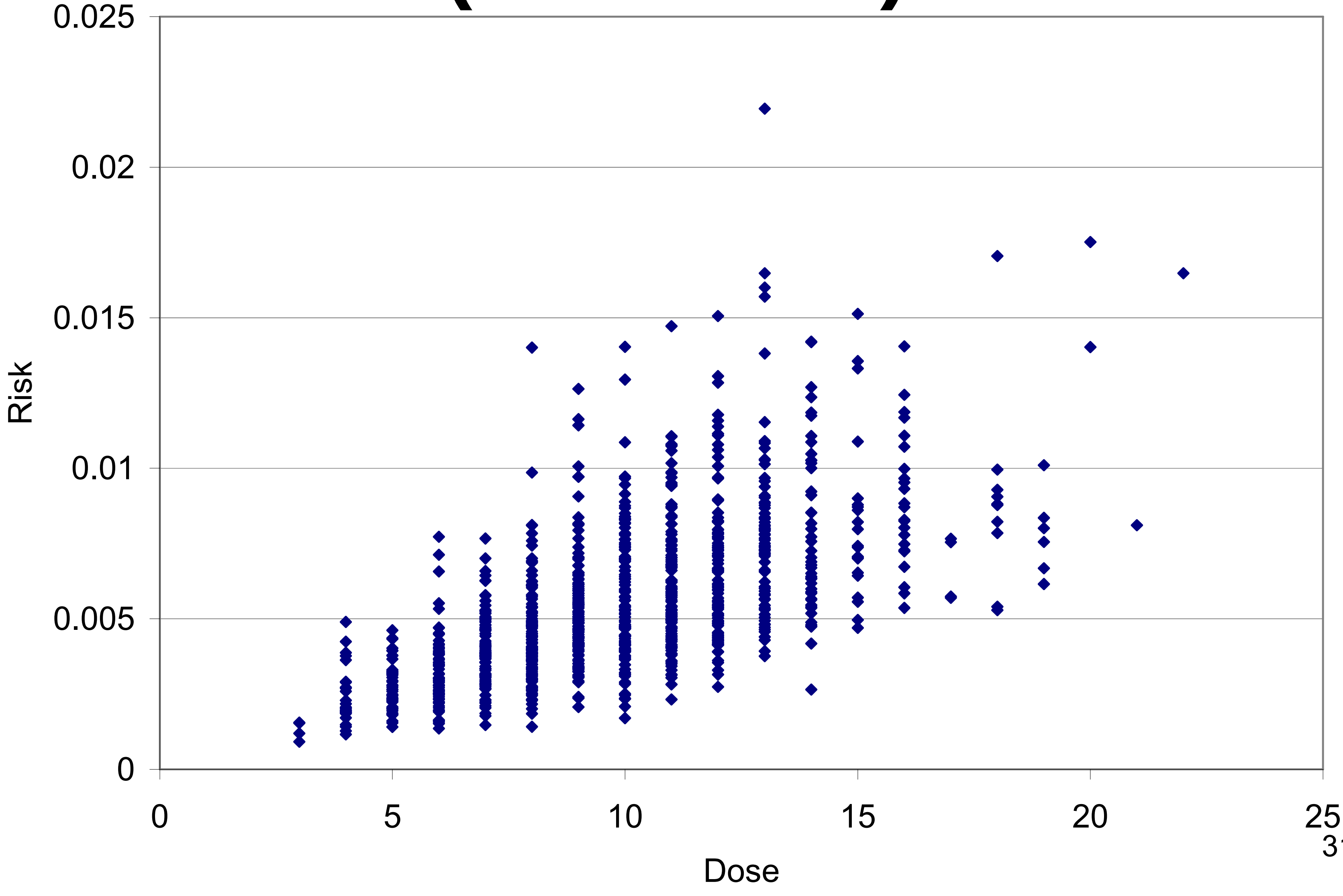
  In Excel =correl(input column, output column)

  Where input column is A1:A1000 or similar
- Scatter plots of inputs with output

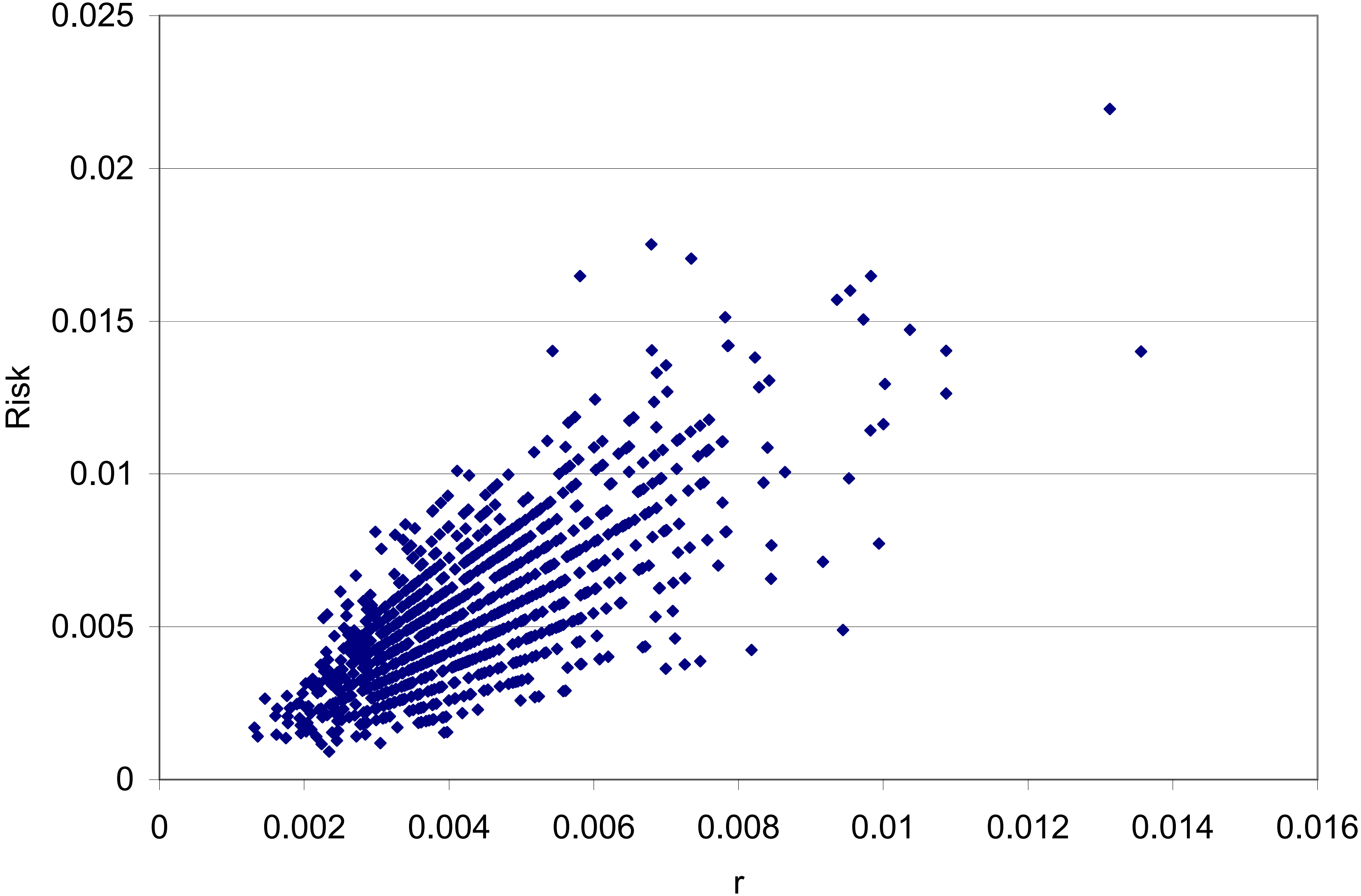
Histogram and Cumulative Histogram



Scatterplot: Risk vs. Dose (correl=.64)



Scatterplot: Risk vs. r (correl=.74)



Risk Characterization

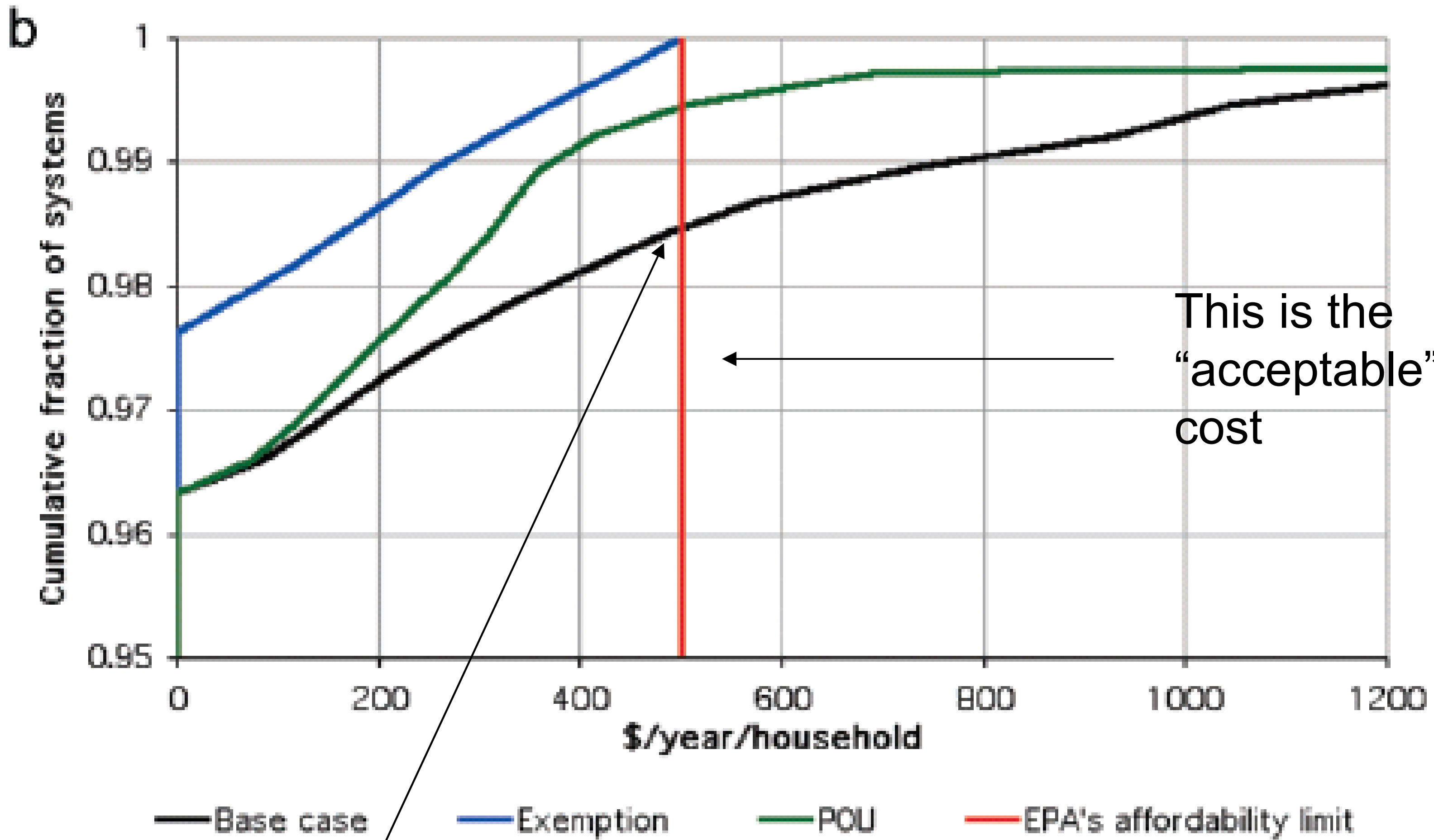
- After all the effort of a Monte Carlo analysis, in practice people want a number
- Tendency to collapse distribution to most likely number (or conservative, protective number)
- What do we really want to get out of our analysis?
- Not just a number but to inform multiple decisions
  - Is risk acceptable? 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



This is the "acceptable" cost

Median costs are fine, here we look at upper bound