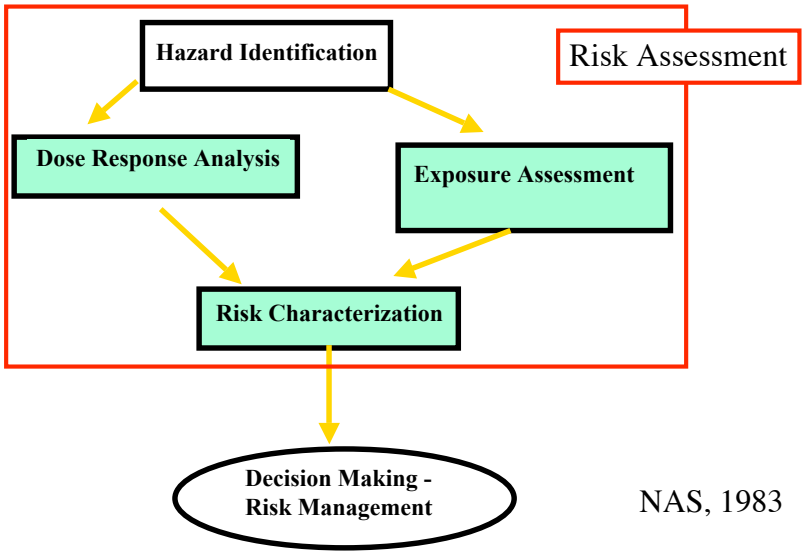
Basic Microbial Dose Response

Charles N. Haas
Drexel University
haas@drexel.edu

Outline

- What are the models
- How do we fit them
- Comparing models
- Confidence limits

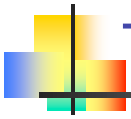
The Risk Analysis Process



NAS, 1983

Why do we need a DR model?

- We can (never) do a direct study (even with animals) to assess dose corresponding to an acceptably low risk
- We use a model to (extrap)(interp)olate to low dose



The Dose

- Average administered to a population
- Actual number an individual experiences
- Retention
- *In vivo* body burden after multiplication

Plausibility of Models

- should consider discrete (particulate) nature of organisms (high variability at low dose)
- based on concept of infection from one or more “survivors” of initial dose (birth-death models)

Derivation of Exponential DR Model

- Poisson distribution of organisms among replicate doses (mean # in dose=d).
- One organism is capable of producing an infection if it arrives at an appropriate site.
- Organisms have independent and identical probability of surviving to reach and infect at an appropriate site (k).

$$p = 1 - \exp(-kd)$$

Derivation of Beta-Poisson Model (assumptions)

- Same as the exponential model except nonconstant survival and infection probabilities
- Survival probabilities (k) are given by the beta distribution
- Slope of dose response curve more shallow than exponential

Comparison of Exponential and Beta-Poisson (I)

Beta-Poisson Model

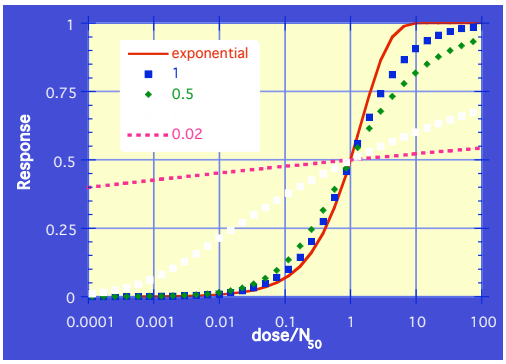
Original Form

$$P = 1 - \left(1 - \frac{d}{\beta}\right)^{-\alpha}$$

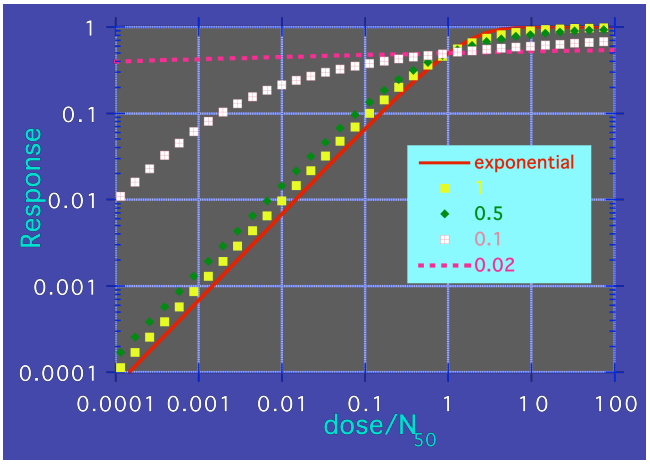
Revised Parameterization

$$P = 1 - \left[1 + \frac{d}{N_{50}}\left(2^{1/\alpha} - 1\right)\right]^{-\alpha}$$

N_{50}= organisms for 50 % infectivity



Comparison of Exponential and Beta-Poisson (II) - low dose extrapolation

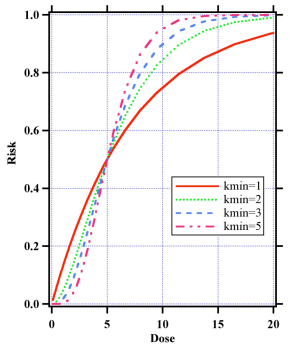


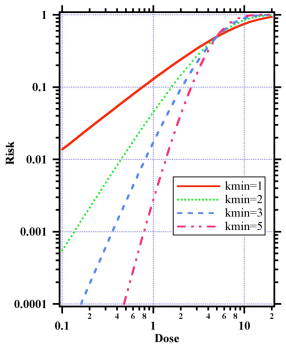
A Generalized Framework

$$P_I(d) = \sum_{k_{min}=1}^{\infty} \sum_{k=k_{min}}^{\infty} \sum_{j=k}^{\infty} P_1(j|d) P_2(k|j) P(k_{min})$$

- $P(k_{min})$: fraction of subjects that require k_{min} original organisms to survive in order to become infected (point; truncated Poisson, etc.)
- $P_1(j|d)$: fraction of subjects ingesting from an average dose d who actually ingest j organisms (Poisson...)
- $P_2(k|j)$: fraction of subjects ingesting j organisms in which k organisms survive (binomial; beta-binomial)

"Threshold (>1)" Models





Median dose fixed at 5

- threshold models (k_{min}>1) yield steeper slopes and non-linear low dose models
- no human data sets yet examined justify these models

Empirical Models

- obviously others as well
- but these do not take into account the "particle" nature of organisms
- give nonlinear low-dose behavior

- Log probit

$$P_I = \Phi\left(\frac{1}{q_2}\ln\left(\frac{d}{q_1}\right)\right)$$

- Log logistic

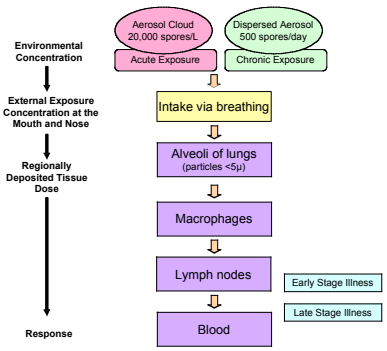
$$P_I = \frac{1}{1+\exp\left[q_1 - q_2\ln(d)\right]}$$

- Weibull

$$P_I = 1 - \exp\left(-q_1 d^{q_2}\right)$$

PBDRM's

- Requires insight into biological/physical mechanisms leading to infection/disease
- May be more complex than extant data justify



Thran, personal comm.

Fitting of DR Models

Experimental Protocol

- Animals/subjects divided (randomly) into k groups
- In group "i" (i=1..k)
  - All subjects exposed to (poisson average) dose d_i
  - Of the T_i total subjects, P_i are "positive"

- Quantal
- Poisson average dose
- Binomial variability

Mechanics of Fitting (I)

- each dose of our bioassay is a sample from a binomial distribution (with T_i) total organisms and an unknown positive probability (of adverse outcome) of π. so from binomial relationship, we would have:

$$f(P_i) = \frac{T_i!}{P_i!(T_i - P_i)!} \pi_i^{P_i} (1 - \pi_i)^{T_i - P_i}$$

Mechanics of Fitting (II)

- but we have multiple doses (i>1, including control), and so if we use the likelihood criteria

$$\ln(L) = \sum_{i=1}^{N} \ln(f_i(P_i))$$

- we would have

$$\ln(L) = \text{constant} + \sum_i \left[P_i \ln(\pi_i) + (T_i - P_i)\ln(1 - \pi_i)\right]$$

- the best possible fit (maximum value of ln L) we could have is when our dose response predictor precisely goes through the observed data, i.e.,

$$\pi_i^o = \frac{P_i}{T_i}$$

Any dose-response model must give a fit no better (i.e., ln L would be smaller --- more negative).

Mechanics of Fitting (III)

- it is convenient to look at the fit of some model versus the best possible, and also to multiply by -2 (to transform to minimization of a positive value, and recall χ^2 confidence limit behavior for likelihoods)
- obtain best fit parameters by finding $\hat{\Theta}$ (parameter vector) that minimizes Y:

$$\min Y = -2\sum_{i=1}^{N}\left[P_i \ln\left(\frac{\pi_i}{\pi_i^0}\right) + (T_i - P_i)\ln\left(\frac{1-\pi_i}{1-\pi_i^0}\right)\right]$$

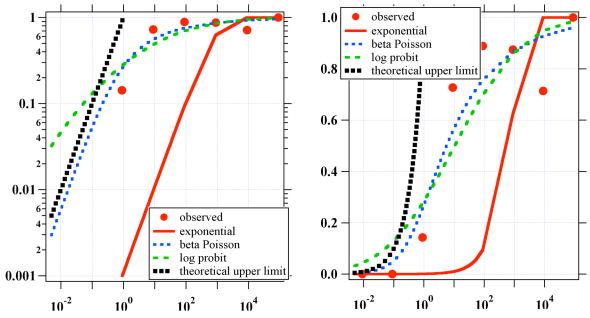
With π_i from dose-response function (function of Θ)

fit is acceptable if Y is less than the upper 5% (or 1%...) of the $\chi 2$ distribution with degrees of freedom = number of doses minus number of dose response parameters

Data Fitting Methodology

- Y provides an index of goodness of fit
  - test vs chi square doses-(# params)
- Unconstrained nonlinear optimization
  - Excel
  - R
  - (Matlab, Mathematica …)

Example of Point Estimation

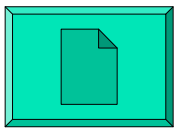


Ward, human rotavirus

- rotavirus (human)
- BP fits better than others, and is accepted as adequate

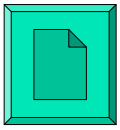
Example with Excel

- Ward rotavirus
  - Residual deviances
    - Exponential 134.5
    - Beta-Poisson 6.815
    - Log-Probit 11.874



Example with R

- Ward rotavirus fit by beta-Poisson model



Model Selection

- Model 1: Y_1, n_1 parameters
- Model 2: Y_2, n_2 parameters
- (n_2>n_1)
- Accept Model 2 iff
  - Y_1-Y_2>χ^2 (df=n_2-n_1)

Characterizing Uncertainty- Confidence Limits

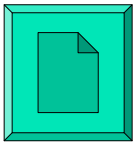
- Confidence regions determined from Likelihood Ratio approach
- all Θ in confidence region if

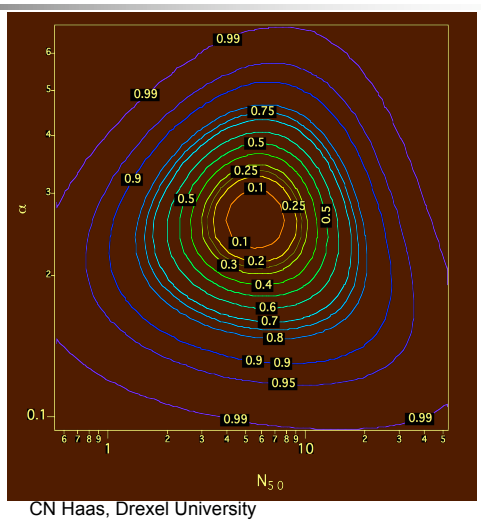
$$2\left[L(\Theta) - L\left(\hat{\Theta}\right)\right] < \chi^2$$

- need to determine n-dimensional region, which may or may not be closed
- can be done in Excel (but tedious and slow)

Example Uncertainty: Rotavirus

Here is example code in R



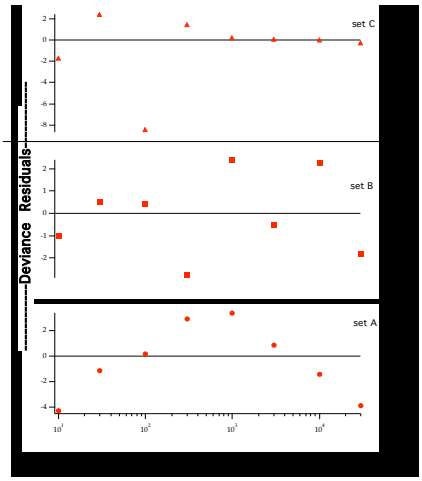


Reasons for Lack of Fit

- outlier

$$r_i = sign\left(p_i^0 - \tilde{p}_i\right) \bullet 2\left[P_i \ln\left(\frac{\tilde{p}_i}{p_i^0}\right) + \left(T_i - P_i\right)\ln\left(\frac{1-\tilde{p}_i}{1-p_i^0}\right)\right]$$

- overdispersion
- systematic deviations



Dealing with Outliers

- identification by likelihood (fit by removal of outliers and compute likelihood ratio)
  - signficance levels confirmed by Monte Carlo
- problems with multiple outliers (masking, swamping)
- not yet a well treated problem in statistics (non-normal, non-linear models)

- outlier identification is typically with respect to a model -- hence we must place “trust” in a model to identify outliers

Dealing with Overdispersion

- Replace a binomial likelihood with a beta-binomial
- This introduces an extra parameter
- Most dose-response studies do not have sufficient dose levels or replicates to truly validate this approach

Dealing with Systematic LOF

- systematic trends in deviance residuals are suggestive of need to use a different dose-response model
- perhaps one with additional parameters

CAMRA Project III Workflow