Inference For Binary Observations Following Adaptive Dose Allocation.

Nancy Flournoy & Assaf Oron
University of Missouri & Institute for Disease Mapping

flournoyn@missouri.edu @ University of Missouri

May 1, 2018
CIRM, France

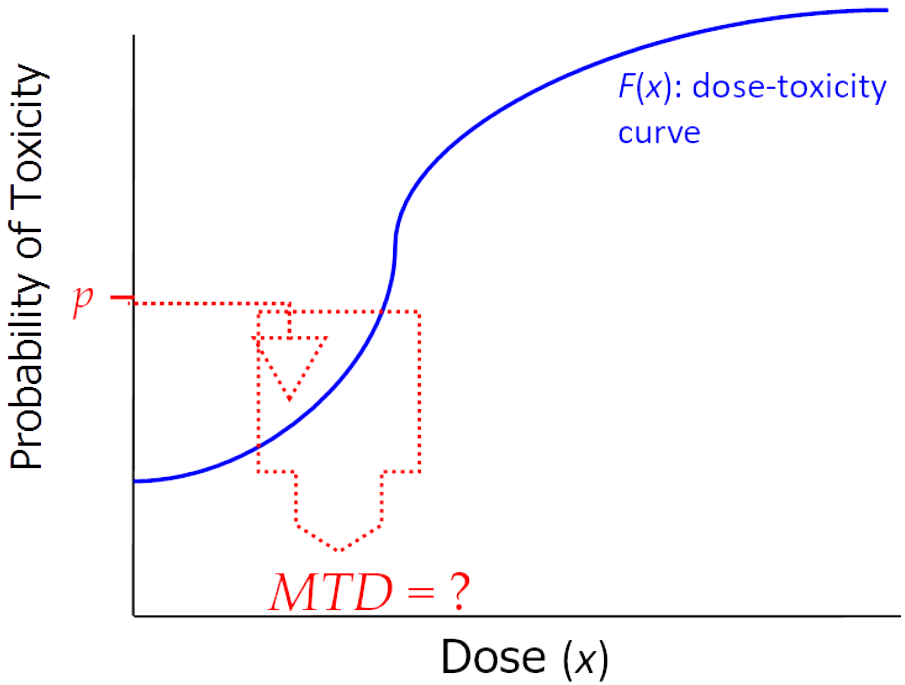
View from Home in Bellingham, Washington, USA

Nancy Flournoy & Assaf Oron
University of Missouri & Institute for Disease Mapping

Home in Vitaljina, Croatia

Context: Binary Regression Following Sequential Informative Selection of Doses

Examples assume $P\{toxicity\}$ increases with dose.



Consider inverse estimation of a target quantile, as opposed to a dose-selection (Tsutakawa, 1980).

Some Adaptive Allocation Procedures

Short Memory

- Classical Up-and-Down Design (Markov chain)
- Biased Coin Design (BCD, Markov chain)
- K-in-a-row (Geometric) (Krow, Markov chain)
- 3+3 (seat of the pants)

Long Memory

- Continual Reassessment Method (CRM, Bayesian)
- EWOC (Bayesian design)
- Interval Designs (CCD Frequentist & mTPI Bayesian)
- Adaptive Optimal Design (AO, Frequentist & Bayesian)

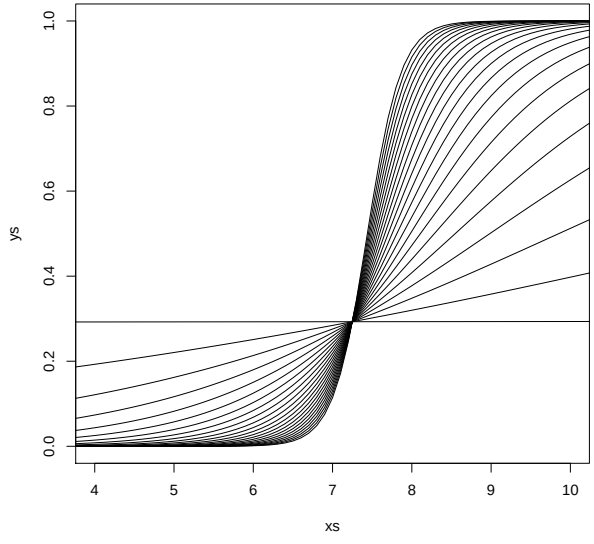
Outline: Inference Following Informative Dose Assignments

1. Model-based versus isotonic regression:
   This will warm up the talk and provide context
2. Thoughts about the Likelihood and Conditioning
3. Bias in Observed Toxicity Rates and their Variances:
   What can and cannot (?) be done about it.

Fundamental Challenges Were Recognized by Robbins (1954) & Wetherill (1963)

1. **Parametric Regression – MLE, LSE or Bayes**
   With $F[(x_i - \alpha)/\beta]$, $\widehat{\mu} = \widehat{\beta} F^{-1}(\Gamma) + \widehat{\alpha}$.
   - Concentrated allocations yield poor estimates of $\widehat{\beta}$.
   - In small studies,
     - Bayes estimates will depend heavily on priors;
     - MLEs frequently do not exist.
2. **Isotonic Regression**
   - Eliminates need to estimate slope parameter;
   - Quality of estimate depends on sample sizes at doses neighboring the target;
   - Centered Isotonic (CIR) (Oron & Flournoy, 2017).
3. **Last dose is the estimated target – SA, CRM, ...**
   - goes with trying to put all subjects on the target dose μ;
   - if "successful", can't estimate response function slope.

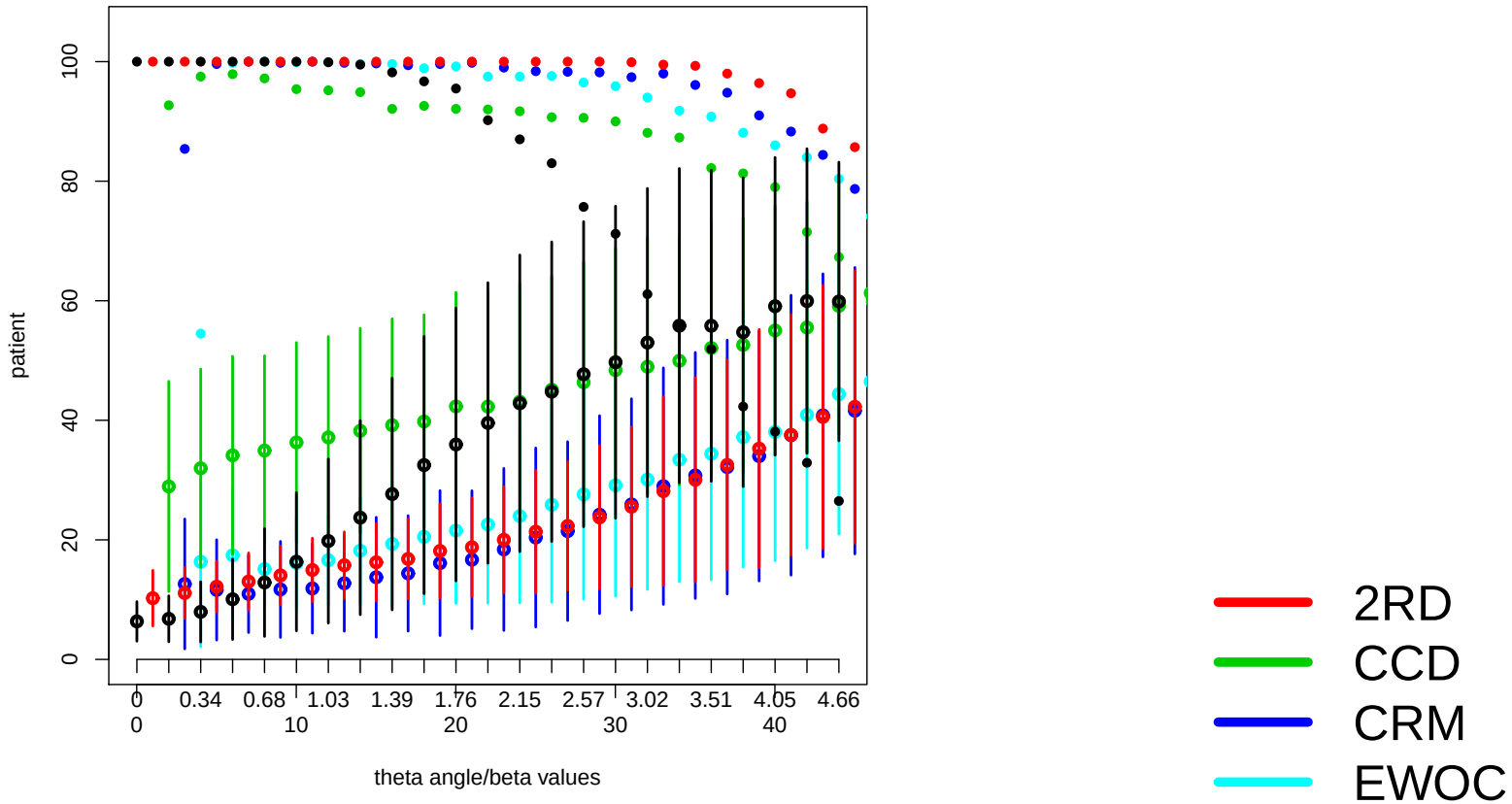
Consider Simulations from the Logistic Model: $logit(F) = \alpha + \beta x$



θ	β	Dose 6	Dose 7	Dose 8
0.01	0.0008	0.2927	0.2928	0.2930
1.01	0.09	0.2714	0.2885	0.3063
2.01	0.17	0.2510	0.2842	0.3199
4.01	0.34	0.2134	0.2757	0.3481
8.01	0.68	0.1505	0.2590	0.4081
12.01	1.02	0.1029	0.2427	0.4723
18.01	1.57	0.0550	0.2186	0.5735
24.01	2.15	0.0274	0.1948	0.6752
28.01	2.57	0.0164	0.1790	0.7398
32.01	3.02	0.0094	0.1630	0.7994

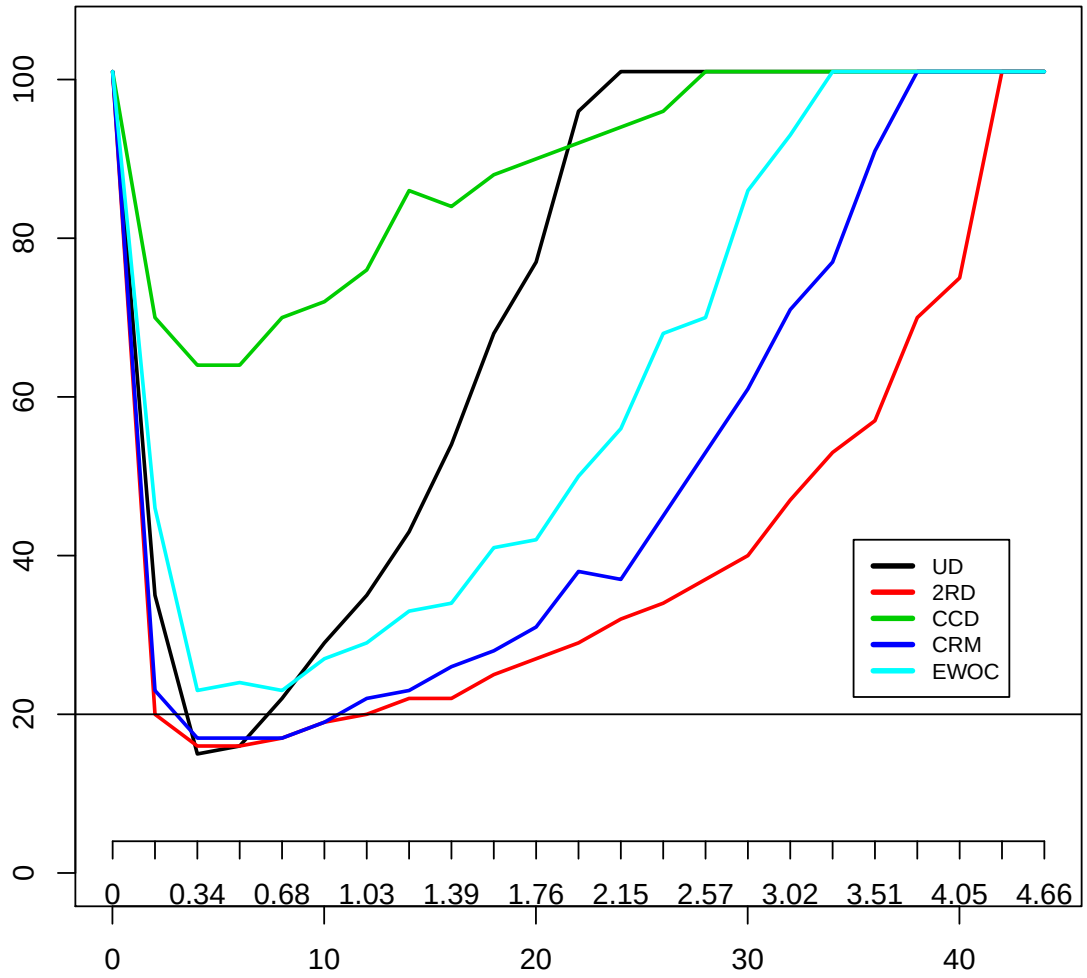
Logistic toxicity probabilities for selected angles θ such that $F(7.25) = 0.2929$ and $F'(7.25) = tan(\theta)$.

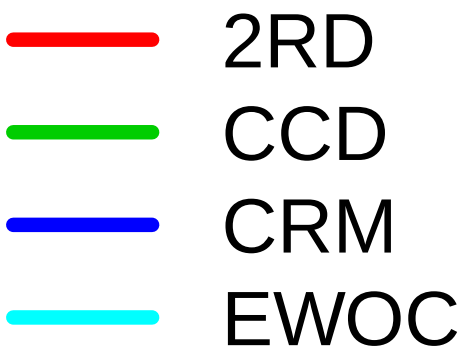
First Patient of $n = 100$ for Whom MLEs Exist, i.e., Silvapule Conditions are Met



Dots are percent of time that MLEs exist (Moler, 2018). There are many Bayesian and frequentist ”fixes”, but if conditions are not met, experiment is a failure and should be treated as such.

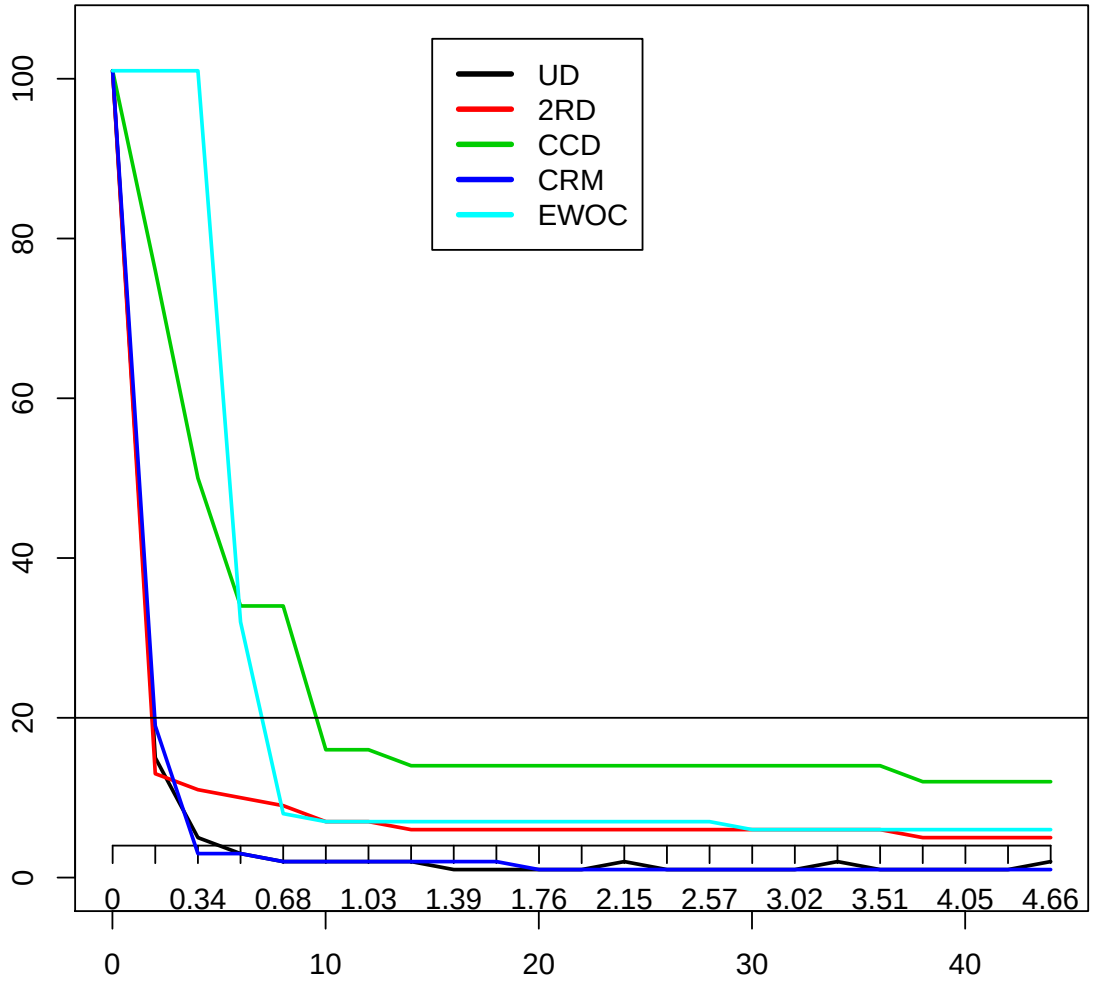
First patient for whom $\geq 90\%$ of executions provide a MLE versus β, θ

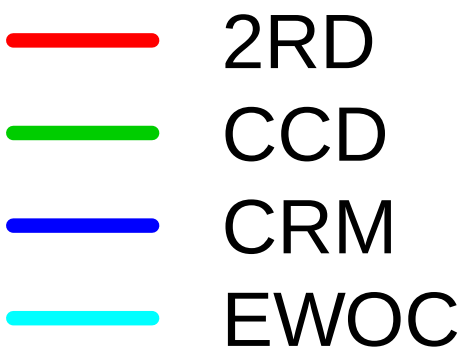




There are lots of papers on algorithmic failures in addition to non existence. (figure by Moler, 2018)

First Patient of $n = 100$ for Whom CIREs Exist versus β, θ





(Moler, 2018)

Some Basic Notation

n	Number of subjects in the study
$d_1 < \cdots < d_M$	Doses in the treatment space
$X(j)$	Dose for the jth subject, $j = 1, \ldots, n$
$\delta_m(j)$	$= \begin{cases} 1 & \text{if } X(j) = d_m \\ 0 & \text{if else.} \end{cases}$
$N_m = \sum_{j=1}^{n} \delta_m(j)$	Frequency of allocations to dose d_m;
$Y(j)$	$= \begin{cases} 1 & \text{if } j\text{th subject has toxicity;} \\ 0 & \text{if else.} \end{cases}$
F_m	$= P\{Y(j) = 1 \mid \delta_m(j) = 1\} \quad \forall n.$
$T_m = \sum_{j=1}^{n} Y(j)\delta_m(j)$	Frequency of toxicities at dose d_m;

Observed Toxicity Rates Are Fundamental Summary Statistics

Isotonic regression methods use observed toxicity rates directly.

Standard *likelihood-based methods* use observed toxicity rates indirectly: $\mathcal{L} = \prod_{m=1}^{M} F_m^{T_m}(1 - F_m)^{N_m - T_m}$.

- Likelihood is a function of dose-specific allocation & toxicity frequencies $\{T_m, N_m\}$;
- MLE of F_m is T_m/N_m when F_m is not a function of additional parameters;
- Fisher Information $(-\mathrm{E}[\frac{\partial^2}{\partial\theta^2} \log \mathcal{L}])$ is a first-order linear approximation of Var[MLEs of F] when F is nonlinear.

The Likelihood (Rosenberger, Flournoy, Durham, 1994)

$$\mathcal{L}[F_1, \ldots, F_M | (N_1, T_1), \ldots, (N_M, T_m)] = \prod_{m=1}^{M} F_m^{T_m}(1 - F_m)^{N_m - T_m}$$

Assumptions

1. Responses are conditionally Bernoulli random variables:

$$P[Y(j) | \delta_m(j) = 1 \ \& \ \text{all history}] = F_m^{Y(j)}[1 - F_m^{[1-Y(j)]}],$$

with

2. allocation rules that are independent of the past given the current dose assignment:

$P[\delta_m(j) = 1 | \text{ history}], j = 2, \ldots, n,$ does not depend on θ, and

$P[\delta_m(1) = 1]$ does not depend θ.

Should one condition on allocation frequencies $N_1, \ldots, N_M$?

Toxicity Frequencies Conditional on Dose-Allocation Frequencies

One can plug observed allocation frequencies $N_1, \ldots, N_M$ into the likelihood, and treat them as fixed, but truly

Conditional Density

$$f(T_1, \ldots, T_m | N_1, \ldots, N_M) = \frac{f(N_1, T_1, \ldots, N_M, T_m)}{f(N_1, \ldots, N_M)}$$

We have nice expression for numerator.
But for most designs denominator will be unknown.

For Markovian up-and-down designs,
$f(N_1, \ldots, N_M) \approx \pi_1^{N_1}, \ldots, \pi_M^{N_N}$, where $\pi_m = lim_{n \to \infty} N_m / n \, \forall m$.

Even in this simple case, conditioning is pretty ugly $\Rightarrow$

Marginal Density of Markovian Up-and-Down Dose-Allocation Frequencies

$$f(N_1,\ldots,N_M) \approx \prod_{m=1}^{M} \pi_m^{N_m}$$

$$= \frac{\left(\frac{p_{1,2}}{p_{2,1}}\right)^{\left(\sum_{m=2}^{M} N_m\right)} \cdots \left(\frac{p_{M-1,M}}{p_{M,M-1}}\right)^{N_M}}{\left[1+\sum_{m=2}^{M}\prod_{j=1}^{m-1}\left(\frac{p_{j,j+1}}{p_{j+1,j}}\right)\right]^{n}},$$

where $p_{i,j}$ is
$P(\text{treating next subject at } d_j | \text{ current subject is treated at } d_i)$.

$\{p_{i,j}\}$ can be written in terms of $F_1,\ldots,F_M$ for a particular design $\Rightarrow$

e.g., for the Durham-Flournoy Biased Coin Design

Density of Toxicity | Allocation Frequencies

$$f(T_1, \ldots, T_m | N_1, \ldots, N_M)$$
$$\approx \frac{\prod_{m=1}^{M} F_m^{T_m}(1-F_m)^{N_m - T_m}}{\prod_{j=1}^{M-1}\left(b\frac{1-F_j}{F_{j+1}}\right)^{(\sum_{m=j+1}^{M} N_m)}\left[1+\sum_{m=2}^{M}\prod_{j=1}^{m-1} b\left(\frac{1-F_j}{F_{j+1}}\right)\right]^n},$$

where $b = \Gamma/(1-\Gamma)$ and Γ is the target toxicity rate.

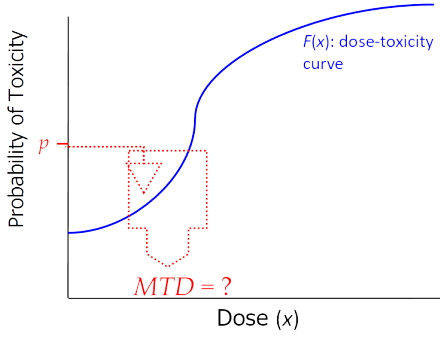
The complexity of this simplest of examples makes it clear that conditional inference is unlikely to gain support in practice regardless of conceptional arguments for and against.

Observed Toxicity Rates

$R_m = T_m/N_m, \qquad N_m \geq 1; m = 1, \ldots, M$

An erroneous presumption is widespread:

$$T_m|N_m \propto \text{Binomial}(N_m, F_m)$$



- Binomial distribution requires conditioning on the single observed allocation frequency.
- Probabilities under this conditioning are not the same as under the conditional distribution of $R_m|N_m$.

$\mathbf{E}[T_m|N_m] \neq F_m$

except in special circumstances, as we will show.

Dose-Specific Toxicity Rate: $F_m = \mathrm{E}[T_m]/\ \mathrm{E}[N_m] \neq \mathrm{E}[T_m/N_m]$

$N_m = \sum_{j=1}^{N} \delta_m(j)$ Number of allocations to dose d_m;

$T_m = \sum_{j=1}^{N} Y(j)\delta_m(j)$ Number of toxicities at dose d_m.

$$\begin{aligned}\mathrm{E}[T_m] &= \sum_{j=1}^{N} \mathrm{E}[Y(j)\delta_m(j)] \\ &= \sum_{j=1}^{N} P\{Y(j)|\delta_m(j)=1\}P\{\delta_m(j)=1\} \\ &= F_m \sum_{j=1}^{N} P\{\delta_m(j)=1\} = F_m\,\mathrm{E}[N_m].\end{aligned}$$

$$\Rightarrow F_m = \mathrm{E}[T_m]/\ \mathrm{E}[N_m].$$

Amazing Insight Comes Straight from the Definition of Covariance

Recall $R_m = T_m/N_m$.

$$\begin{aligned}\mathrm{Cov}[R_m, N_m] &= \mathrm{E}[R_m N_m] - \mathrm{E}[R_m]\,\mathrm{E}[N_m] \\ &= \mathrm{E}\left[\frac{T_m}{N_m} N_m\right] - \mathrm{E}[R_m]\,\mathrm{E}[N_m] \\ &= \mathrm{E}[T_m] - \mathrm{E}[R_m]\,\mathrm{E}[N_m].\end{aligned}$$

$$\Rightarrow$$

$$\begin{aligned}\mathrm{E}[R_m] &= \frac{\mathrm{E}[T_m]}{\mathrm{E}[N_m]} - \frac{\mathrm{Cov}[R_m, N_m]}{\mathrm{E}[N_m]} \\ &= F_m - \frac{\mathrm{Cov}[R_m, N_m]}{\mathrm{E}[N_m]}.\end{aligned}$$

Having allocations depend on outcomes induces bias in using observed toxicity rate to estimate probability of toxicity.

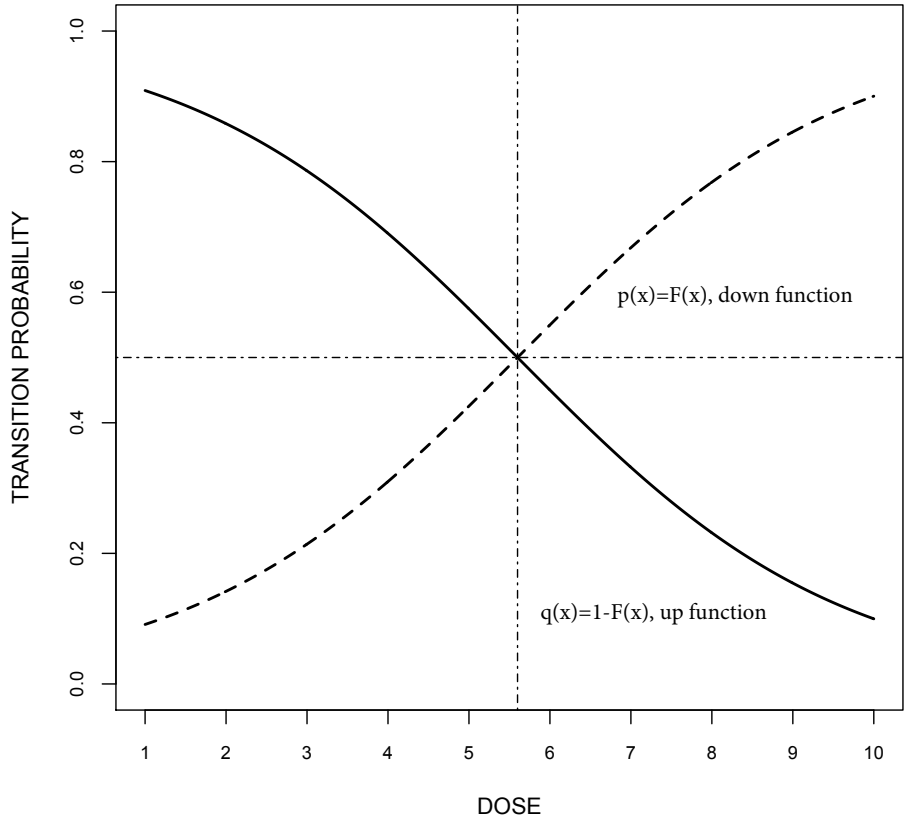
Bias of Observed Toxicity Rate for Probability of Toxicity

$$\begin{aligned} \mathrm{E}[R_m] - F_m &= -\frac{\mathrm{Cov}[R_m, N_m]}{\mathrm{E}[N_m]} \\ &= -\mathrm{Cor}[R_m, N_m]\ \frac{\mathrm{SD}[R]\ \mathrm{SD}(N)}{\mathrm{E}[N_m]} \\ &= -\mathrm{Cor}[R_m, N_m]\ \mathrm{CV}(N)\ \mathrm{SD}[R] \end{aligned}$$

Correlation (Cor) is dimensionless, 0(1).
Coefficient of Variation (CV) is dimensionless, 0(1).
Bias is same order of magnitude as $\mathrm{SD}[R]$:

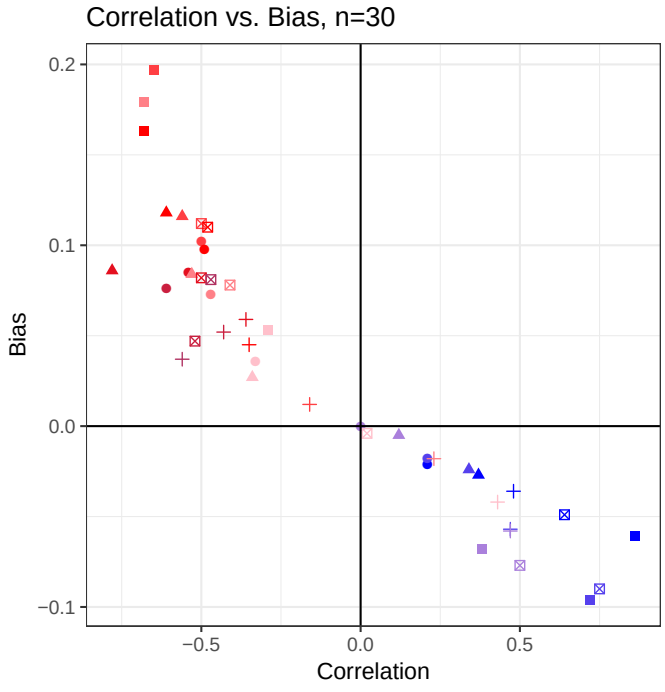
$$\mathrm{Var}[R] = \mathrm{E}_i\left[R_m^2\right] - \frac{\mathrm{Cov}_i[R_m^2, N_m^2]}{\mathrm{E}_i[N_m^2]} - \left(F_m - \frac{\mathrm{Cov}_i[R_m, N_m]}{\mathrm{E}_i[N_m]}\right)^2.$$

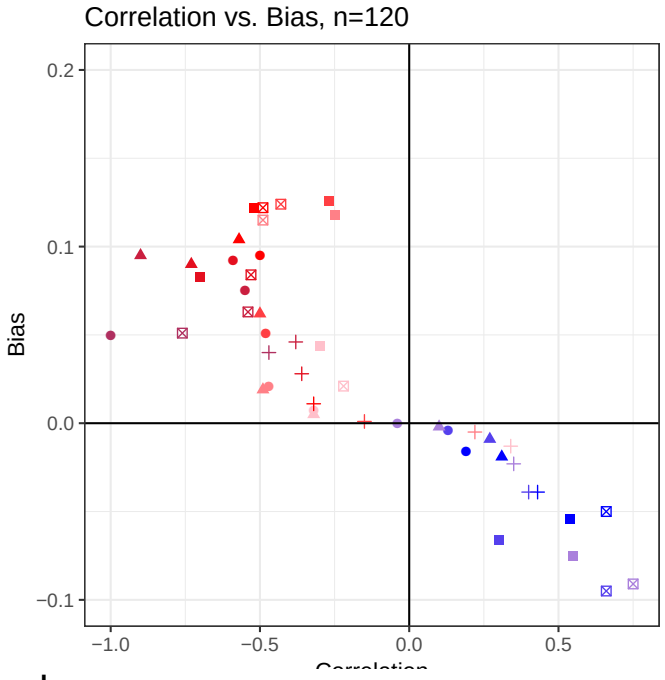
An Exemplary Logistic Dose-Response Function.



LD50=5.6,
LD30=3.9.

Toxicity Rate Bias ($\bar{R}_m - F_m$) by its Correlation with Allocation Frequency (N_m)

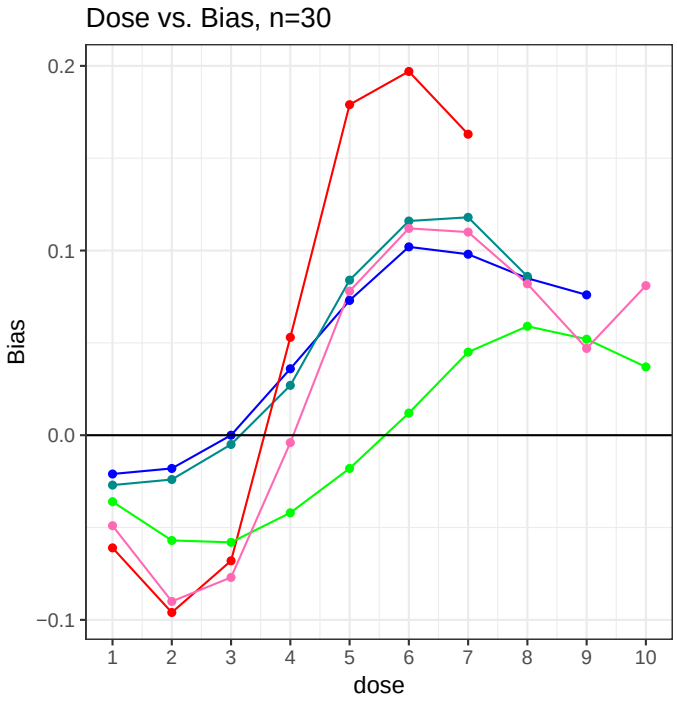


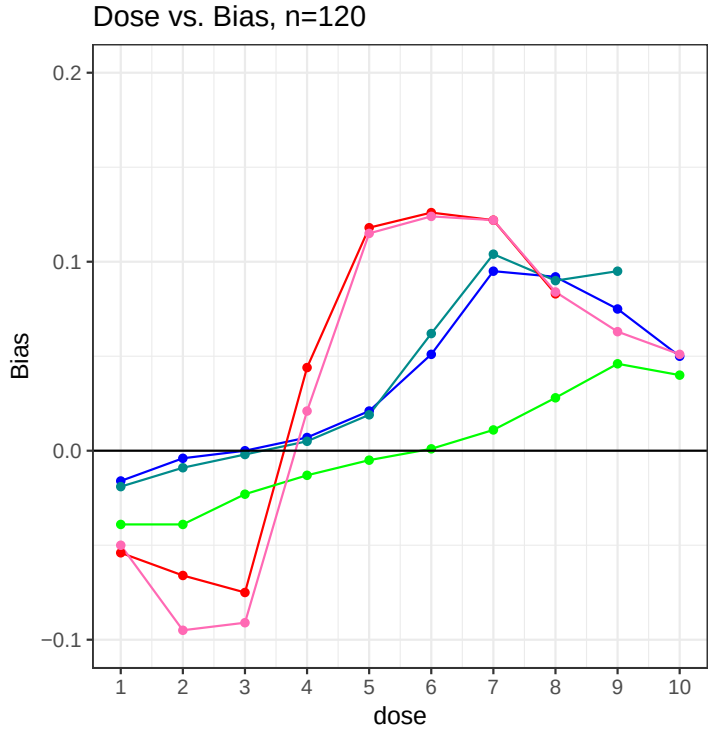


design
- 1:BCD
- 2:Krow
- 3:CRM
- 4:Classic
- 5:CCD

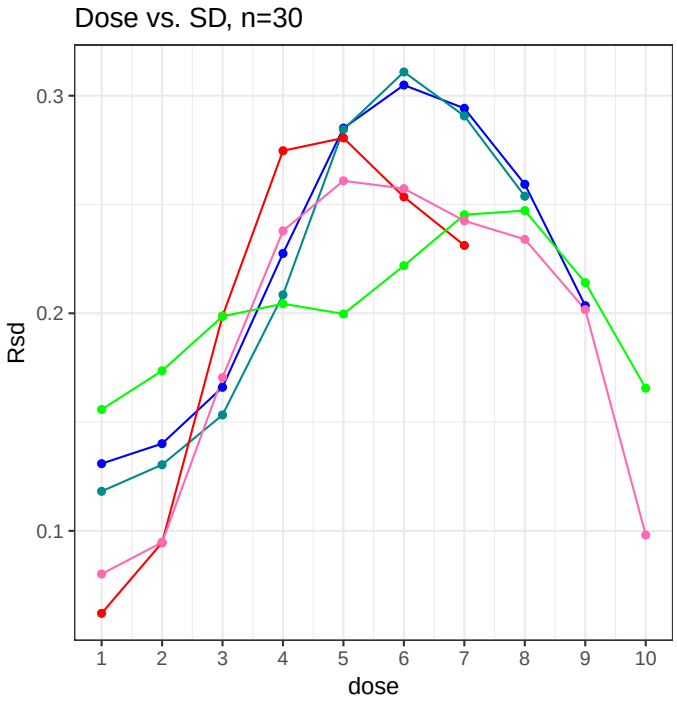
dose
10.0
7.5
5.0
2.5

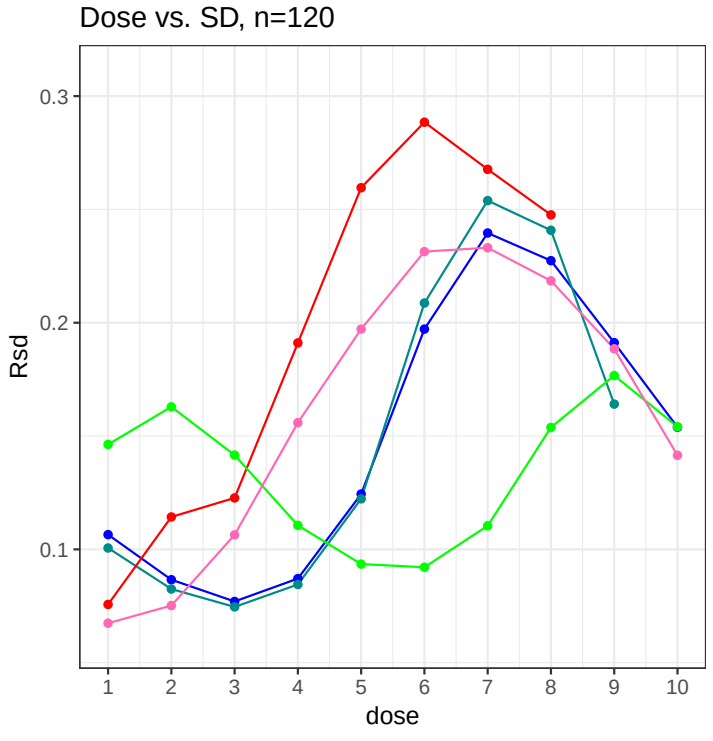
Observed Toxicity Rate Bias $(\bar{R}_m - F_m)$ by Dose. Target = LD30=3.9.





Standard Deviation of the Observed Toxicity Rates by Dose. Target = LD30=3.9.





A Couple Important Open Problems in Estimating F

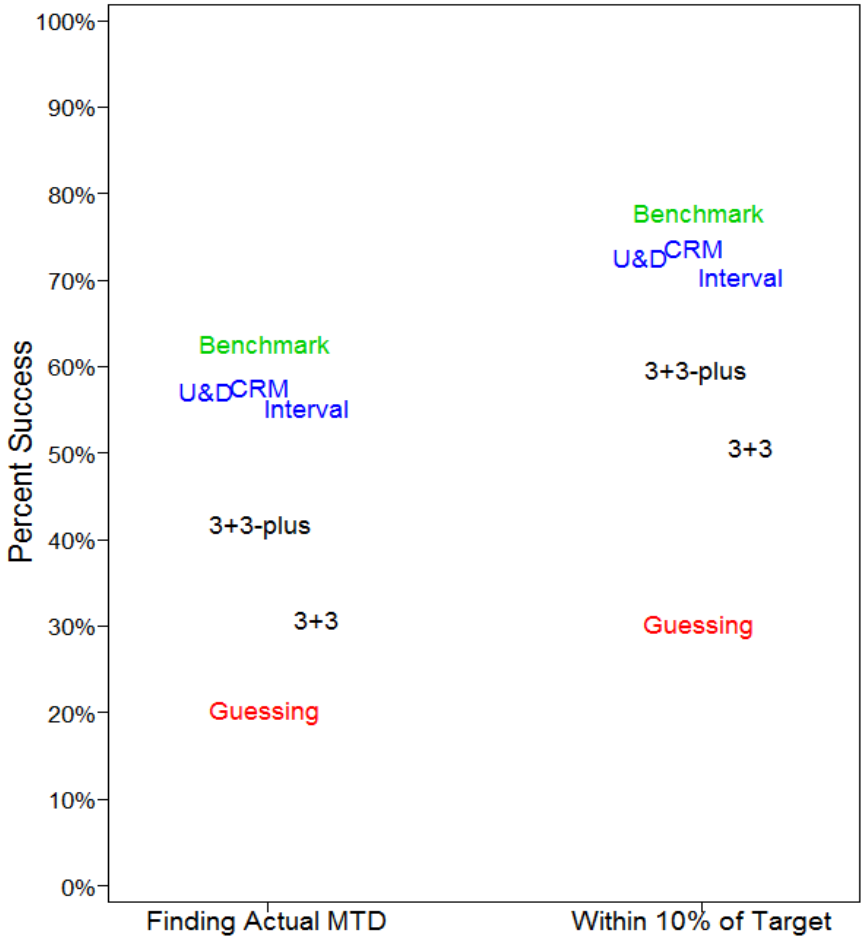
- Need a way to adjust estimates of F for bias of observed toxicity rates.
- Need better way to estimate variance of observed toxicity rates.

Warning

Don't put faith in estimates of F except in a very close neighborhood of the target.

Thank You!

Average % Correct MTD Selection from Random Sample of Curves (Oron, 2017)



Last Dose for CRM; CIR for UD & CCD Interval Design.

At least four other review papers arrive at same conclusion.

Namely, UD, CRM and Interval Methods are a toss up with respect to average % correct selection.

Choice of design will depend on other factors.

% Allocations to the MTD (Oron and Hoff, 2012)