Making Clinical Trials Smarter: How to Do More Faster with Less

Nancy Flournoy

Affiliation: University of Missouri
Residence: Bellingham, Washington
flournoyn@umsystem.edu

December 11, 2025

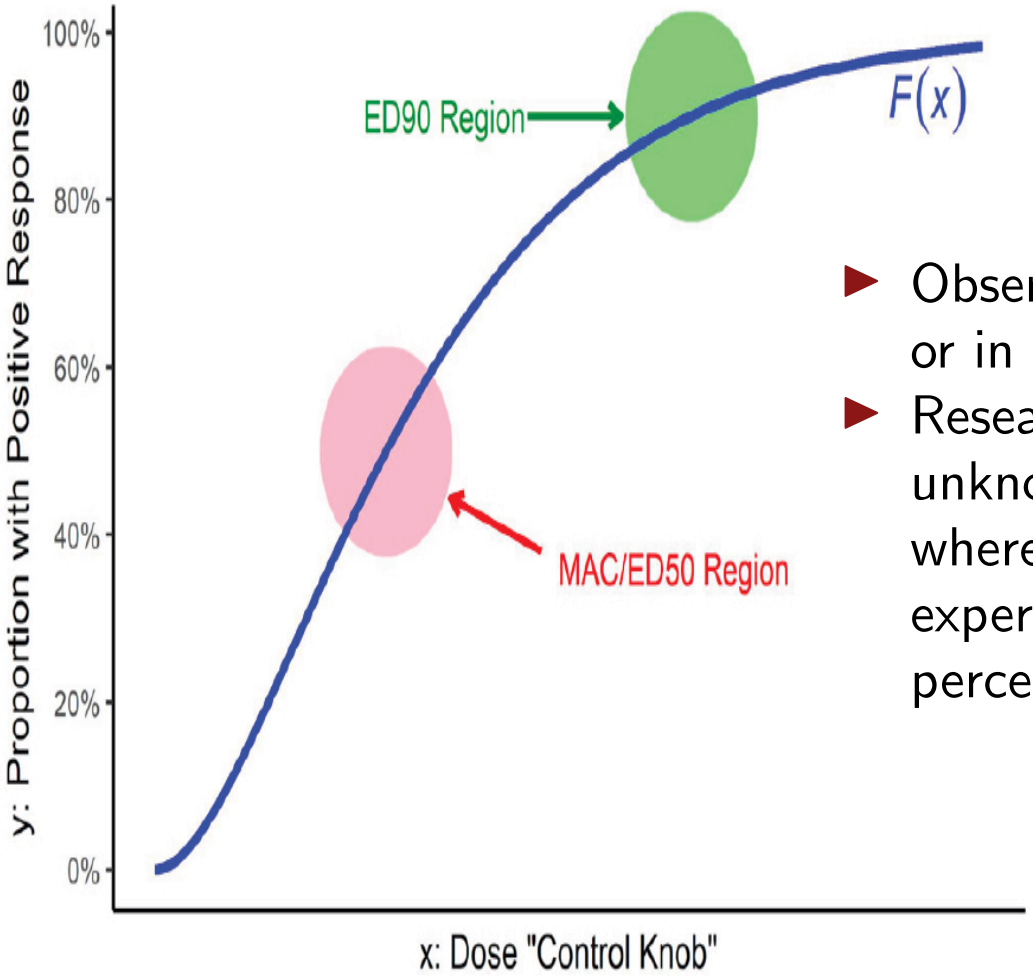
Comments on Two-Arm Clinical Trials

- **For Confirming Results:** The "Gold Standard" two-arm randomized clinical trial does the job as advertised,

  but it is
  Terrible for Learning!

- **For Learning:** A series of small experiments is much better than a single big one!
  - Do small trials square with large ones?
    (Flournoy and Olkin 1995, *Lancet*)
  - A vignette of discovery. *Past, Present and Future of Statistical Science, in Celebration of the COPSS 50th Anniversary.*
    (Flournoy 2014, *Committee of Presidents of Statistical Societies*).
- Also beware, the standard two arm trial philosophy can be, and has been, corrupted
  (Flournoy 2013, *World Academy of Art and Science*).

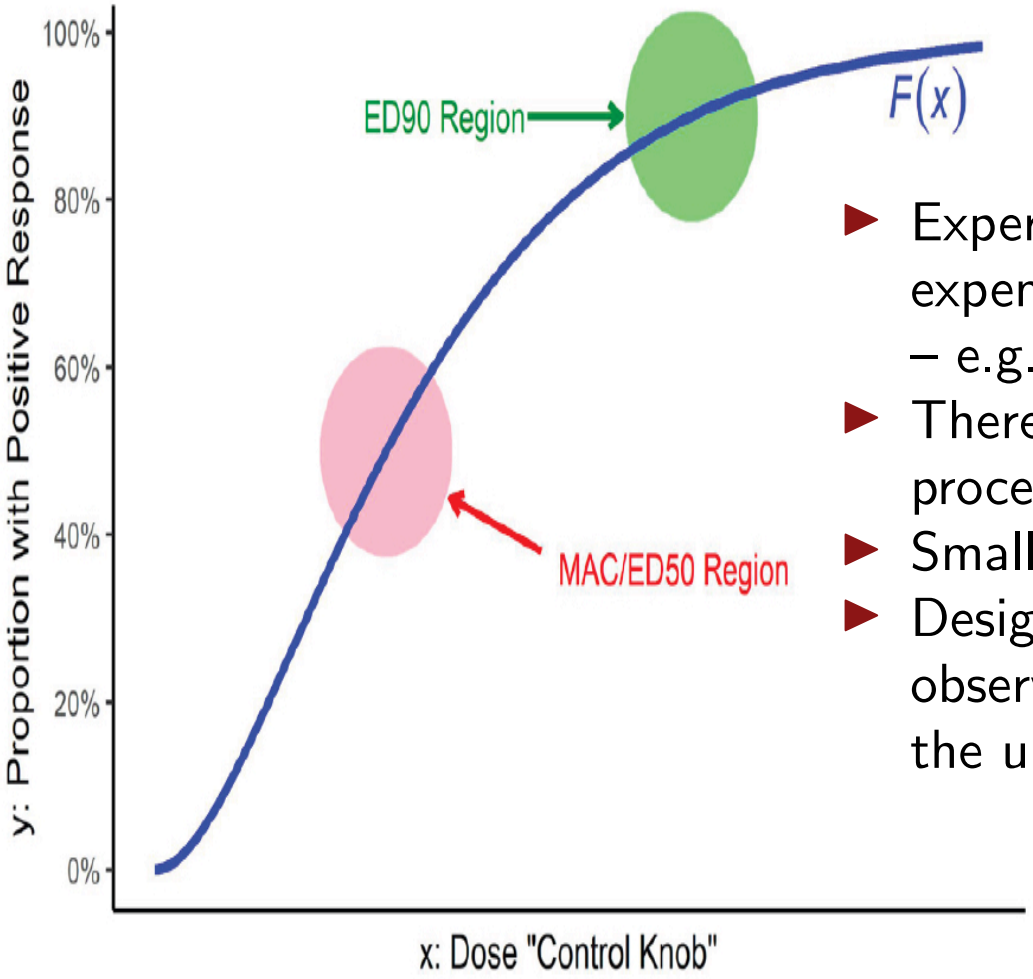
The Dose-Finding Environment



- Observations come sequentially or in batches.
- Researchers aim to estimate the unknown dose where the population treated experiences a prespecified percent of positive responses.

This "quantal" dose-response curve demonstrates the average effect of a stimuli, as a function of its "dose", in a population of individuals.

The Dose-Finding Experiment



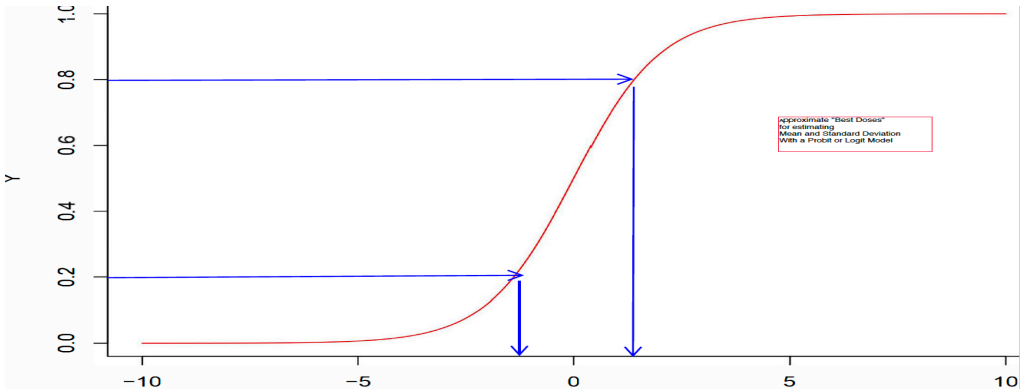
- Experimental units are expensive (i.e., very valuable) – e.g., airplanes or people;
- Therefore, risk-adverse research procedures are desirable.
- Small sample sizes are desirable.
- Design aim is to concentrate observations in the region of the unknown dose of interest.

Basic Requirements for Dose-Finding on a Grid

1. Doses are administered to a *sequentially* to a sequence of subjects on whom responses are observed.
2. Responses are *binary;*
   e.g., effective/ineffective.
3. Doses administered are restricted to a *discrete set*: $d_1 < \cdots < d_M$. Each dose level is a specific magnitude value of the continuous dose variable x.
4. The expected chance of a positive response increases with increasing dose.

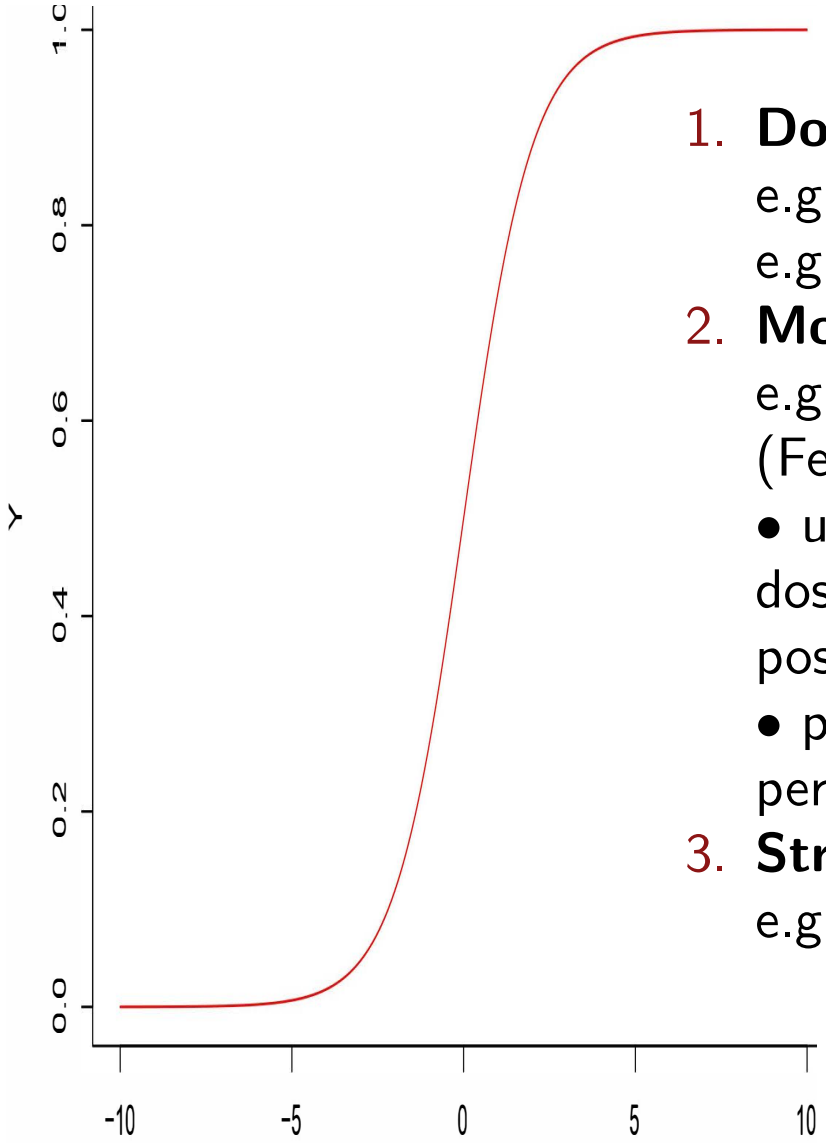
Designs for continuous responses and/or a continuous range of doses permitted are studied with completely different mathematics.

Challenges for Dose-Finding



1. Optimal design theory says to get the most precise estimates one must spread out the doses
   – but this is not acceptable.
2. Concentrating observations in the region of interest should provide good information about the response function there.
3. Challenge: how to concentrate information around an unknown dose and still get good estimates of the dose of interest?

Major Dose-Finding Design Paradigms



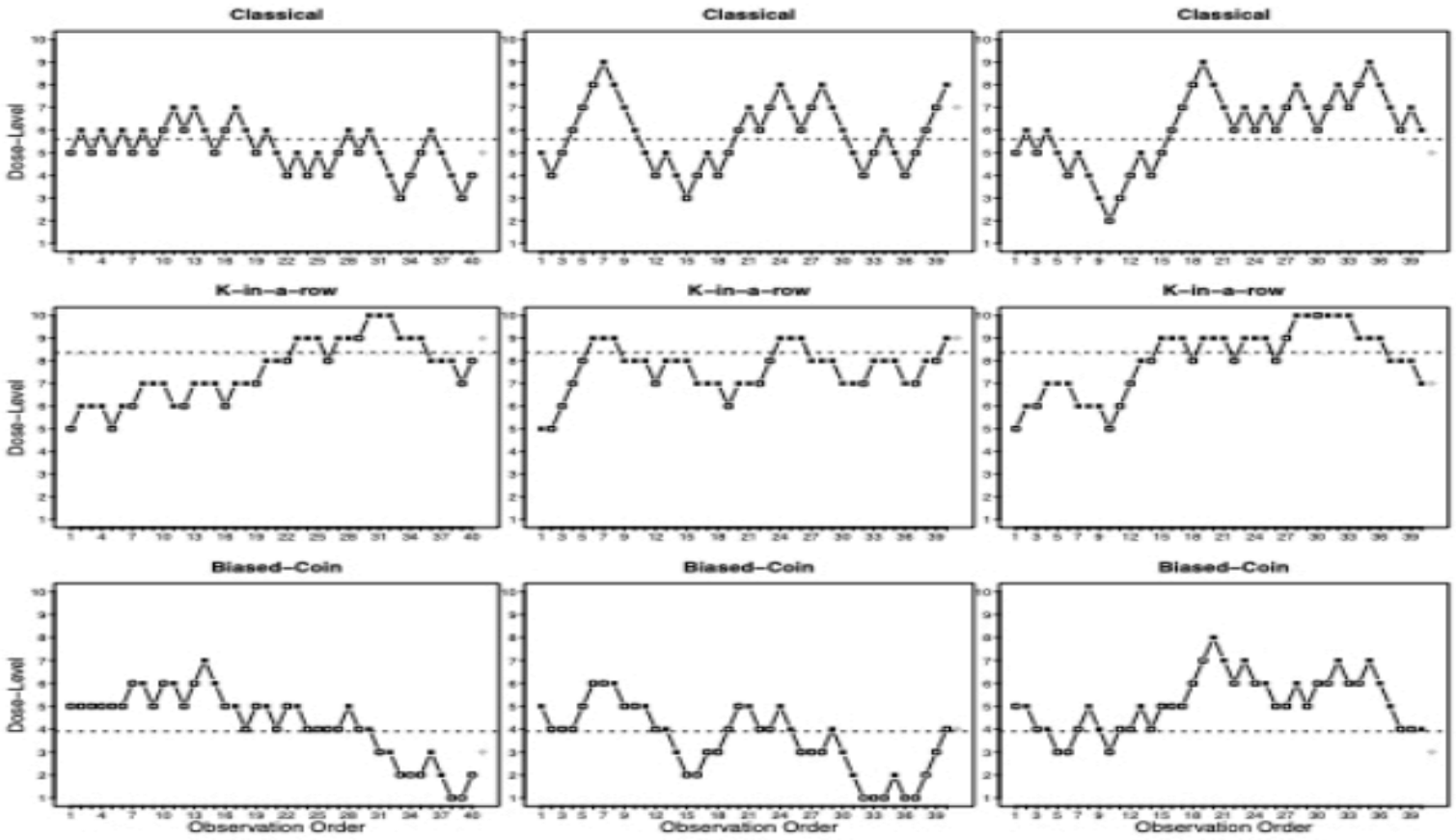
1. **Dose Escalation**
   e.g., increase to first positive response;
   e.g., 3+3;
2. **Model-Based**
   e.g., *Best Intention Designs*
   (Fedorov and Leonov 2013)
   • use statistical models to estimate the dose with prespecified probability of positive response, and
   • place the next subject at the nearest permissible dose.
3. **Strategic Random Walks**
   e.g., Up-and-Down

Introduction to Up-and-Down Designs (UDDs)

- UDDs are ubiquitous in a wide variety of scientific, engineering, and clinical fields.
- UDDs' solid mathematics (Flournoy and A. Oron 2015) support their robust tractable behavior and straightforward usage.
- UDD's good dose-finding performance has won the trust of practitioners and their consulting analysts across fields and continents.
- A. P. Oron, Souter, and Flournoy 2022b have a good practical overview in *Anesthesiology*.
- Comparisons to other dose-finding designs can be found in
  - "Up-and-Down: The Most Popular, Most Reliable, and Most Overlooked (by statisticians) Dose-Finding Design." (A. P. Oron and Flournoy 2024)
  - "Supplement" in *Anesthesia*. (A. P. Oron, Souter, and Flournoy 2022a)

Dose Assignments from Sample UD Experiments

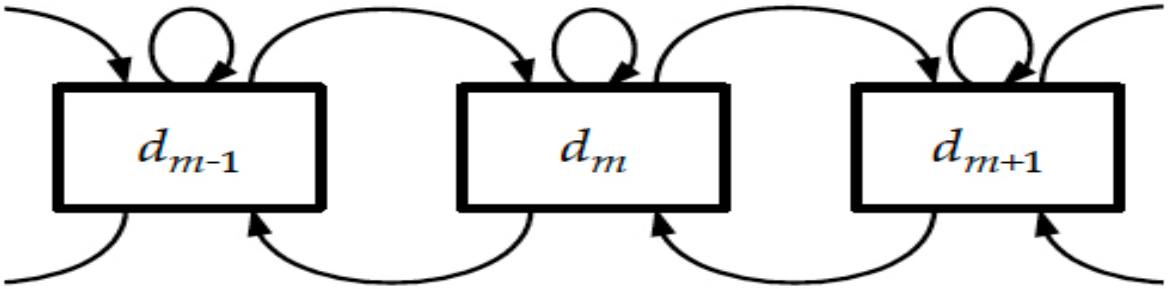
With a few simple rules,
UD experiments generate a "random walk" around the unknown dose which has a pre-specified probability of positive response.



Dashed Lines are the Doses Being Sought.

Up-and-Down Designs, with Other Names

Doses are allocated to patients sequentially, only allowing the dose to be increasd by one level, decreased by one level, or repeated.

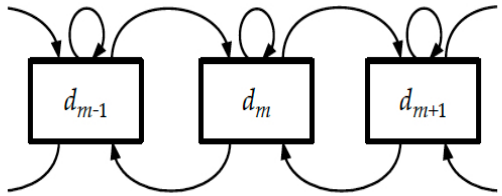


- In statistics, "Up-and-Down" (Dixon and Mood 1948)
- In sensitivity testing, sensory studies, psychophysics ... the "Staircase Method" (Anderson, McCarthy, and Tukey 1946) and sensitivity testing, the "Bruceton Method" (Stresau Jr and Starr Jr 1950)
- In statistics, the "Random Walk Design" (Durham and Flournoy 1994)

Special Features of an Up-and-Down Design

With strategic transition rules, Up-and-Down Designs have sound, well-established mathematical theory that can be well utilized to produce good, robust estimates.

Rules for Assigning the Next Dose:



1. Doses are strategically chosen to center assigned doses around the dose of interest.
2. The next dose only depends on the doses and responses of the last patient or several patients;
   This makes them nimble, and not bogged down by all patient data going back to the beginning of the experiment.
3. Dose-assignment rules do not use any estimated quantity that changes during the study.

Four Up and Down Families: Focus Today

Four Up and Down Design families have been used and studied extensively for decades:

1. The original UDD which we call "Classical":
    - Reduce the stimuli if the response is positive
    - Increase a stimulus if the response is negative

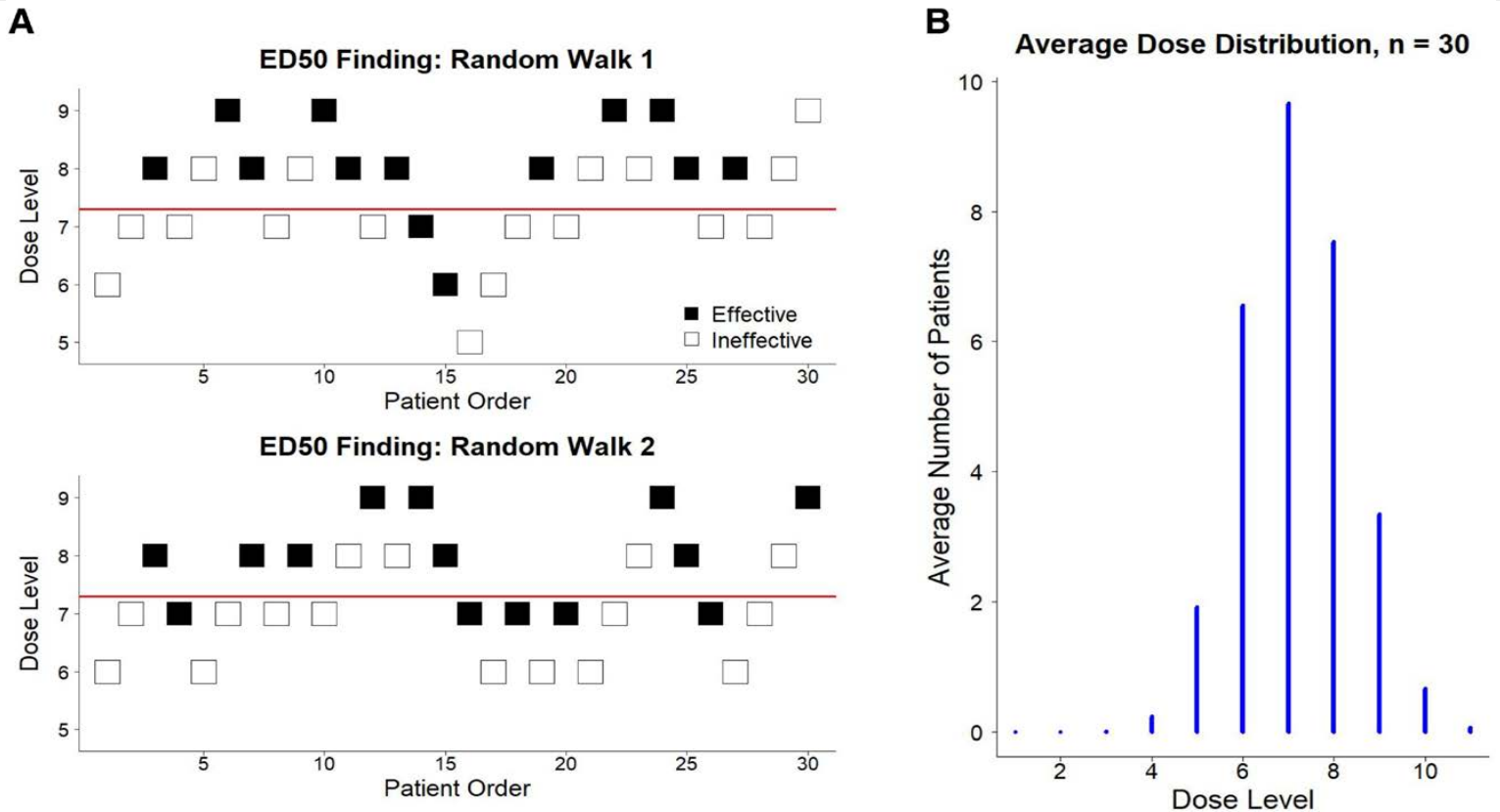
and three straightforward extensions:

2. the K-in-a-row (KRDs)
3. the Biased Coin (BCDs)
4. the Group UDDs (GUDs)

The Classical Up-and-Down Design

For the next dose assignment:

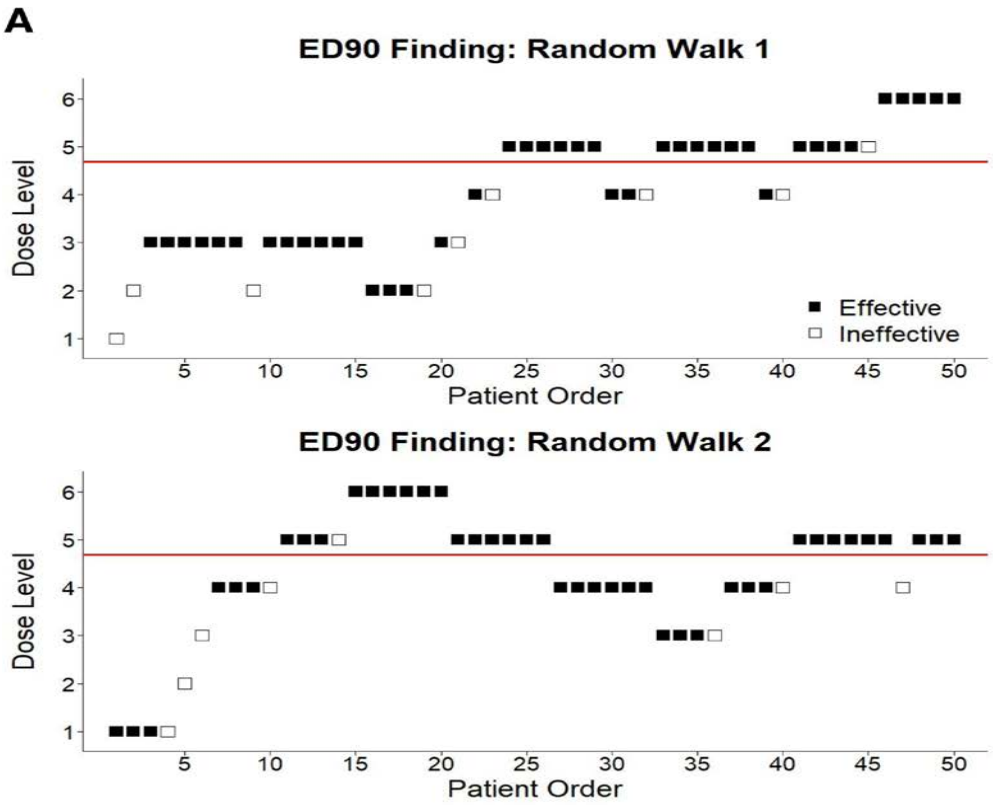
- Increase after a negative response.
- Decrease after a positive response.

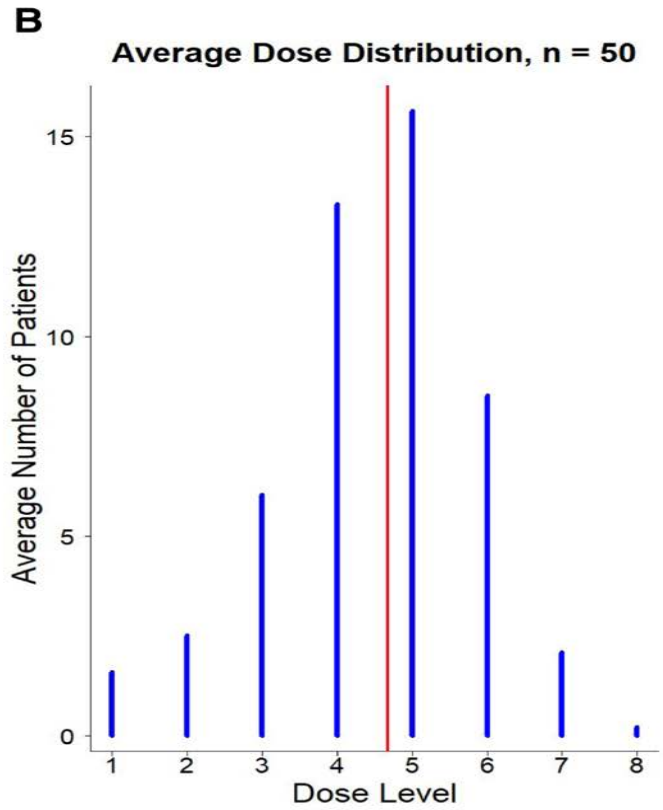


The K-in-Row Design

If the quantal of interest is > 0.5, for the next dose assignment:

- Increase after a negative response.
- Decrease only after observing K consecutive positive responses at the same dose.
- Otherwise, repeat the current dose.





The Biased Coin Design

For the next Biased Coin dose assignment:

- Increase after a negative response.
- Upon a positive response,
  "toss a biased coin" (draw a random number)
  and then for quantals of interest $\Gamma > 0.5$*:*
  - *Decrease the dose with probability* $(1 - \Gamma)/\Gamma$,
  - *Otherwise, repeat the current dose*

The Group Up-and-Down Designs

For the next dose assignment for Groups of size K, fix upper and lower bounds u and l:

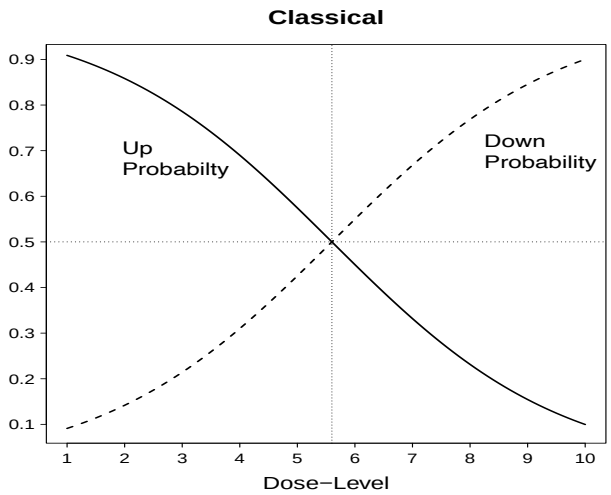
- Increase the dose if $\geq u/K$ successes.
- Decrease if $\leq l/K$ failures.
- Otherwise, repeat the current dose.

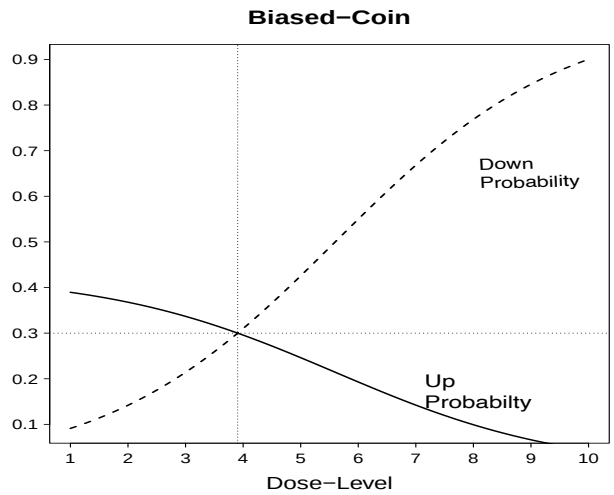
For example, with group sizes of 4,

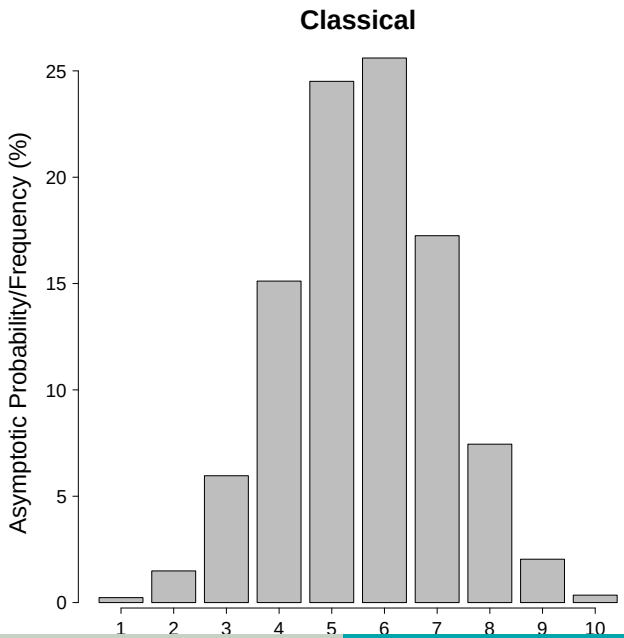
- Increase the dose if 3 or 4 successes.
- Decrease the dose if 0 or 1 failures.
- Repeat the current dose if 2 successes

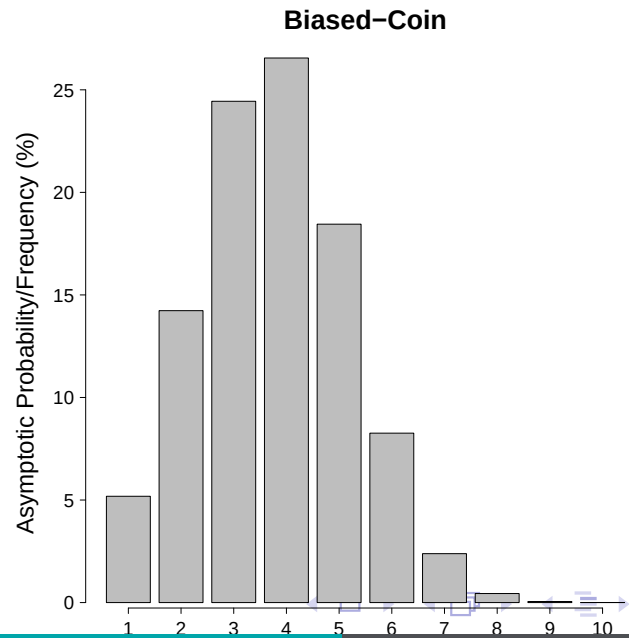
What do these designs have in common?

1. Dose assignments tend to center around the quantal of interest.



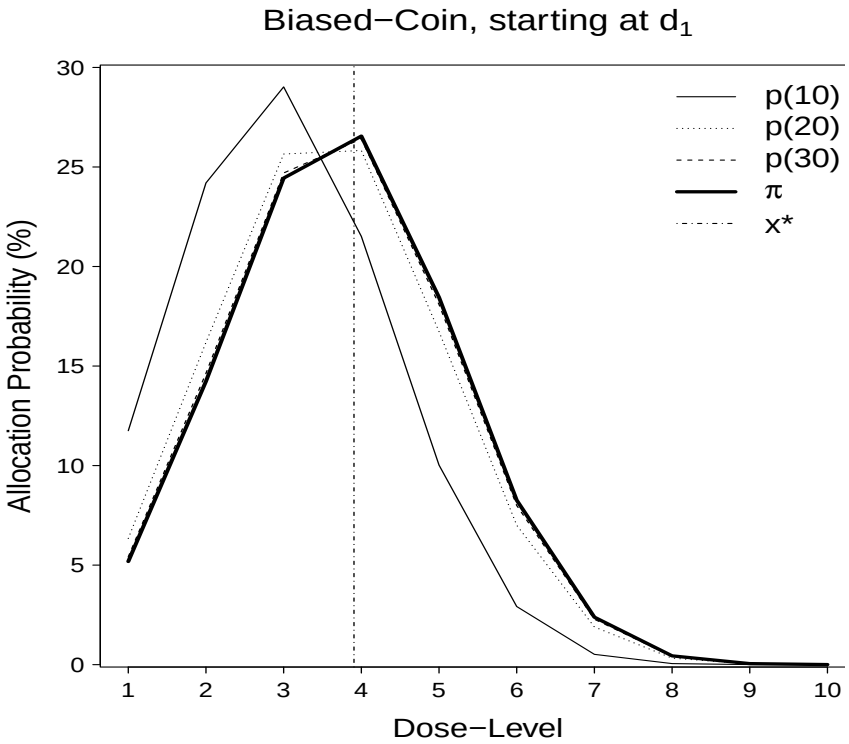






What do these designs have in common?

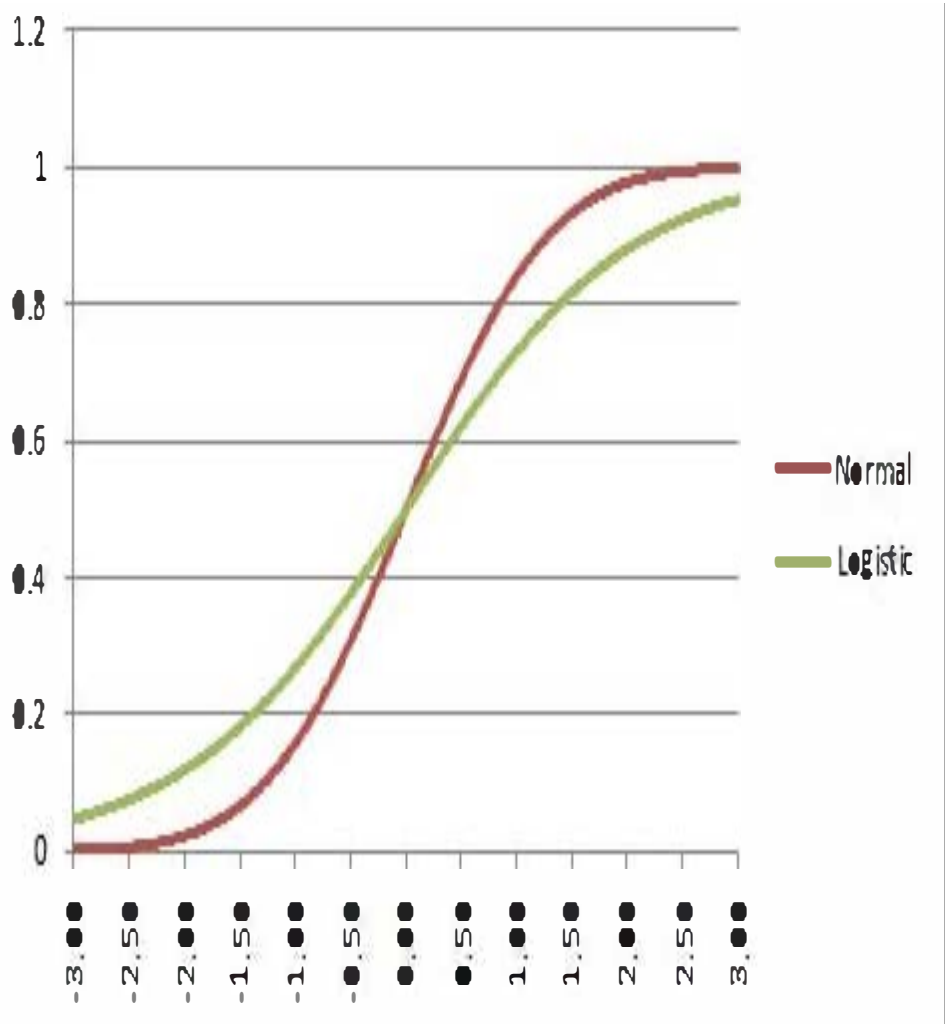
2. Given a response function, mathematical theory allows for exact calculation of the probability the jth subject will get a particular dose.



The probability the jth subject will get a particular dose converges exponentially fast.

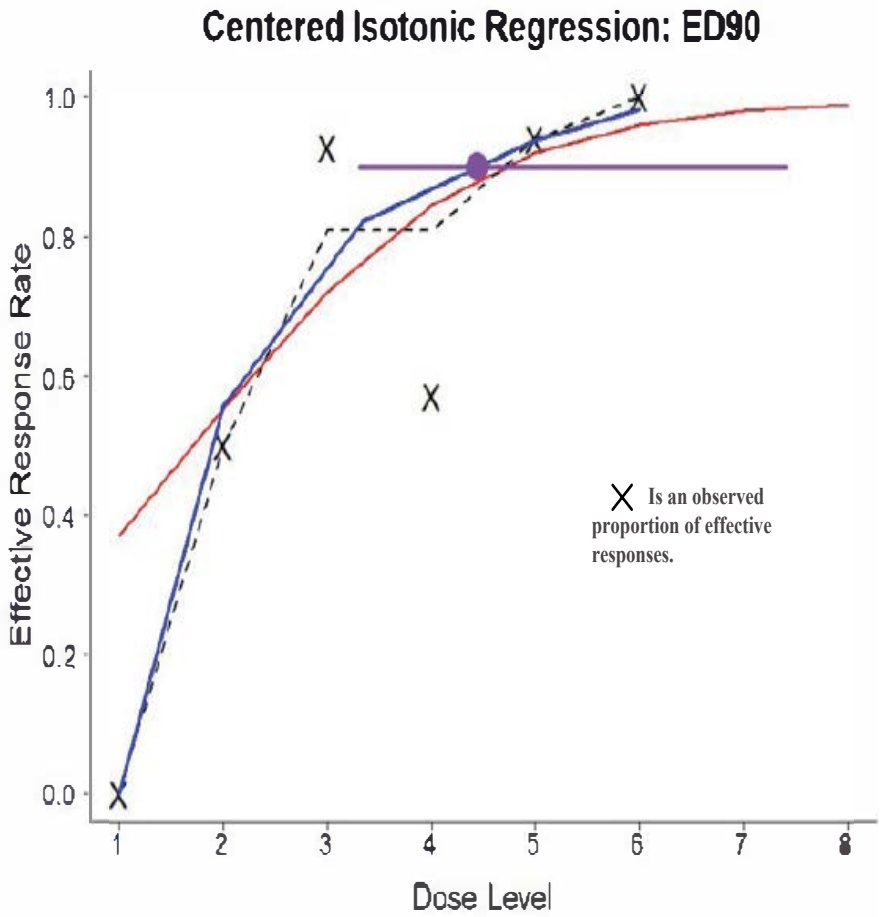
Dose-specific sample proportions converge to same numbers.

Major Estimation Strategies: 1) Parametric



- A model must be pre-specified mathematically.
- Common sigmoidal shaped models are probit and logistic.
- Parameters need to be estimated by frequentist or Bayesian methods.

Major Estimation Strategies: 2a) Non-Parametric: Centered Isotonic Regression

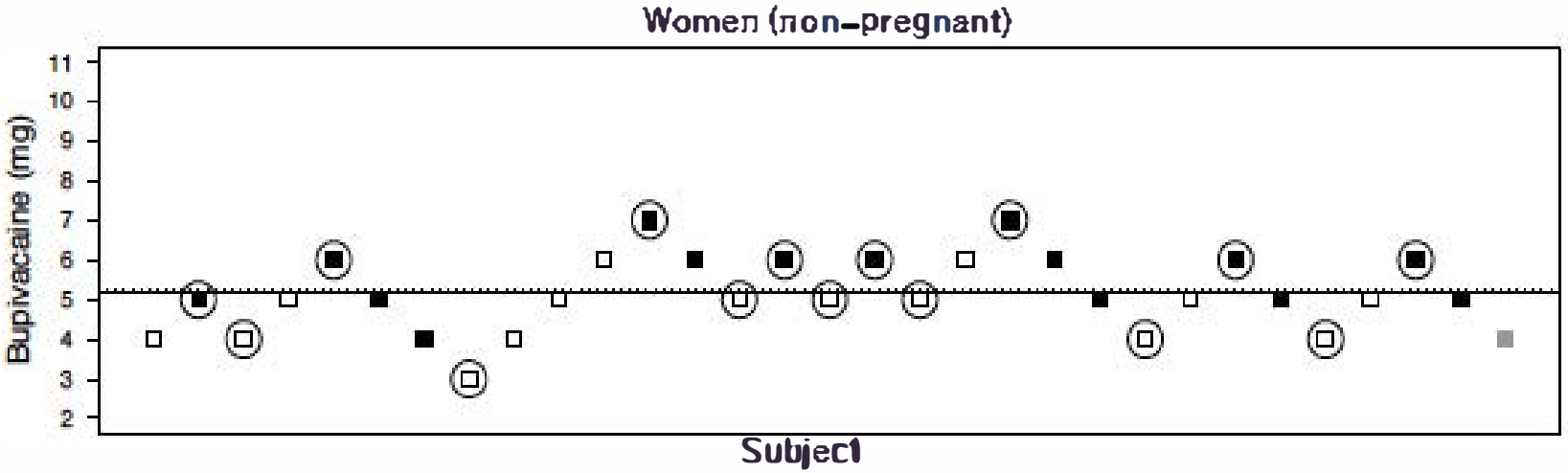


- No model needs be specified mathematically.
- Estimates only use observed response rates - adjusted to be increasing.
- We recommend this method with bias correction: (A. P. Oron and Flournoy 2017, Flournoy and A. P. Oron 2020).

Xs are observed proportions; red line is response function; blue line is centered isotonic regression.

Major Estimation Strategies: 2b) Non-Parametric: Dose Averaging

Average Doses Assigned



- Assumes only that dose assignments are rather symmetric around the quantal of interest.
- Estimates only use observed doses.
- See A. P. Oron, Souter, and Flournoy 2022b for comparisons with Centered Isotonic Regression.

Conclusions

- Don't rush into a 2-arm clinical trial.
- Use observational studies to generate hypotheses.
- Design small controlled trials with clear goals.
- Series of small dose-finding experiments can be very enlightening.
- Dose-allocations from Up-and-Down Designs have stable, robust predictable performance.
- Give UDDs a try if you haven't already.

References

Anderson, TW, PJ McCarthy, and JW Tukey (1946). "Staircase" methods of sensitivity testing. NAVORD Report 65-46". In: *Statistical Research Group, Princeton University*, pp. 1–134.

Dixon, Wilfrid J and Alexander M Mood (1948). "A method for obtaining and analyzing sensitivity data". In: *Journal of the American Statistical Association* 43.241, pp. 109–126.

Durham, Stephen D. and Nancy Flournoy (1994). "Random Walks for Quantile Estimation". In: *Statistical Decision Theory and Related Topics, V*. Ed. by Shanti S. Gupta and James O. Berger. Springer-Verlag Inc., pp. 467–476.

Fedorov, Valerii V and Sergei L Leonov (2013). *Optimal design for nonlinear response models*. CRC Press.

Flournoy, Nancy (2013). "Corruption of the Scientific Method". In: *ERUDITIO*, p. 103.

— (2014). "A vignette of discovery". In: *Past, Present and Future of Statistical Science, in Celebration of the COPSS 50th Anniversary*, pp. 349–358.

References (cont.)

Flournoy, Nancy and Ingram Olkin (1995). "Do small trials square with large ones?" In: *The Lancet* 345.8952, pp. 741–742.

Flournoy, Nancy and A.P. Oron (2015). "Up-and-down designs for dose-finding". In: *Handbook of Design and Analysis of Experiments*. Ed. by D. Bingham et al. CRC Press, Chapman Hall. Chap. 24, pp. 862–898.

Flournoy, Nancy and Assaf P Oron (2020). "Bias induced by adaptive dose-finding designs". In: *Journal of Applied Statistics* 47.13-15, pp. 2431–2442.

Oron, Assaf P and Nancy Flournoy (2017). "Centered isotonic regression: point and interval estimation for dose–response studies". In: *Statistics in Biopharmaceutical Research* 9.3, pp. 258–267.

— (2024). "Up-and-Down: The Most Popular, Most Reliable, and Most Overlooked Dose-Finding Design". In: *The New England Journal of Statistics in Data Science*, pp. 1–12.

Oron, Assaf P, Michael J Souter, and Nancy Flournoy (2022a). "Supplement to Up-and-Down "Understanding Research Methods" Article". In: *The New England Journal of Statistics in Data Science*.

References (cont.)

Oron, Assaf P, Michael J Souter, and Nancy Flournoy (2022b). "Understanding research methods: up-and-down designs for dose-finding". In: *Anesthesiology* 137.2, pp. 137–150.

Stresau Jr, RH and LE Starr Jr (1950). *Some studies of the propagation of detonation between small confined explosive charges*. Tech. rep.

Questions?

Thank you!

Stay tuned!

My book with Assaf Oron on **Up-and-Down Designs** is expected next year!