A likelihood-based approach to selecting doses with acceptable toxicity in a standard phase I algorithm

Cody Chiuzan, Elizabeth Garrett-Mayer

Department of Public Health Sciences
Medical University of South Carolina

May 20, 2013

'3+3' Design

Starts the dose escalation with 3 patients treated at dose j. If:

- 0 DLTs (dose-limiting toxicities), escalate to dose $j+1$
- 1 DLT, treat 3 more patients at dose j
- $>$1 DLT, de-escalate to dose $j-1$

The maximum tolerated dose (MTD) is usually defined as the hightest dose at which 0 or 1 DLTs are observed in 6 patients.

'3+3' - Standard Algorithm for Phase I Trials

Pros

- Transparent and easy to implement
- Comparable properties with the Continual Reassessment Method (CRM) for ≤ 5 doses[1]
- Better than model-based designs when none of the doses is close to the true MTD[2]

Cons

- Short memory
- Conservative escalation, resulting in a large proportion of patients allocated to subtherapeutic doses
- No ability to select a dose limiting toxicity rate, e.g., 33%[3]

'3+3' - Heuristic, but still Popular

- In the last two decades, about 99% of the dose-finding cancer trials implemented variations of this algorithm[4]
- Relies on empirical reasoning
- Limited capabilities of describing and accounting for uncertainties in the observed data

Our method offers . . .

- A probabilistic support for analyzing the behavior of the '3+3' design
- A likelihood-based approach for computing the operating characteristics under different:
  - Hypotheses
  - Levels of evidence
  - True (best guessed) toxicity rates
- An alternative for identifying the highest acceptably safe dose in phase I trials or for defining stopping rules in phase II/III

Framework - Likelihoodism

- Method based on the **Evidential Paradigm** (R. Royall[5], J. Blume[6])
- Concept first introduced by Hacking[7] in 1965, by formally stating the Law of Likelihood
- Provides the likelihood ratio (LR) of two hypotheses as an objective measure of the strength of evidence, decoupled from the uncertainty

Likelihoodism (con't)

Given $H_0 : p = p_0$ and $H_1 : p = p_1$, calculate $LR = \dfrac{L(p_1; x)}{L(p_0; x)}$

Using a certain benchmark k, interpret the strength of evidence as:

- Weak evidence, if $\frac{1}{k} < LR < k$
- Evidence in favor of H_1, if $LR \geq k$
- Evidence in favor of H_0, if $LR \leq \frac{1}{k}$

Prob. of misleading evidence ($\approx$ type I error) is low and bounded:

$$P_0(LR \geq k | H_0) \leq \frac{1}{k}, \text{ with } k \text{ fixed over } n$$

In the '3+3' context . . .

For any prespecified $dose_j$, consider the hypotheses:

$$H_0^j : p_j = p_0 \text{ (DLT rate is unsafe)}$$
$$H_1^j : p_j = p_1 \text{ (DLT rate is acceptable)}$$

For n_j patients and y_j toxicities at $dose_j$:

$$LR_j = \left(\frac{p_1}{p_0}\right)^{y_j} \left(\frac{1-p_1}{1-p_0}\right)^{n_j - y_j} ; \; j = 1, 2, \ldots, D$$

Operating characteristics

Four hypotheses: p_0 - toxic DLT, p_1 - safe DLT

A:(0.40, 0.15); B:(0.50, 0.10); C:(0.15, 0.05); D:(0.50, 0.30)

Likelihood-ratio threshold $k \in \{1, 2, 8\}$:

1. Probability of weak evidence
2. Probability of favoring H_0 under $H_0 \approx 1$ - type I
3. Probability of favoring H_1 under $H_1 \approx 1$ - type II

Identify acceptable/unacceptable doses

- For cohorts of 3, the escalation rules are clear for 0/3, 1/3, 2/3, etc.
- For 1/4 or 2/5, what is considered safe enough to escalate ?
- The k value with the best operating characteristics is chosen as threshold for the LR:

$$LR \geq k\text{; dose is acceptably safe}$$

$$LR < k\text{; dose is unacceptably toxic}$$

Likelihood method vs. '3+3'

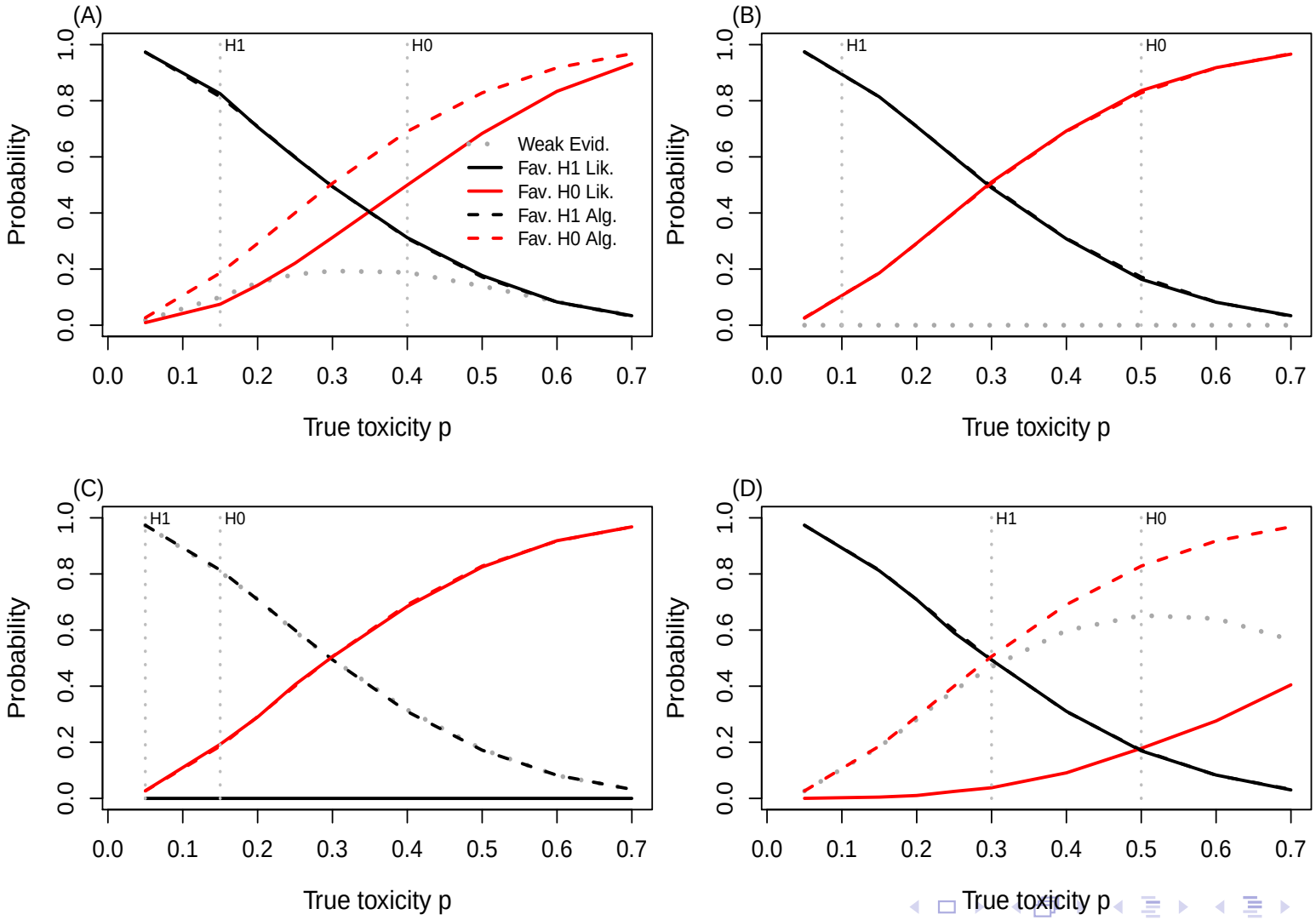
$$H_0 : p_0 = 0.40 \quad H_1 : p_1 = 0.15$$

$DLT(n)^*$	'3+3' rule	$L(p_1)/L(p_0)$	$k = 1$	$k = 2$	$k = 8$
$0(3)$	acceptable	$LR = 2.84$	acceptable	acceptable	weak
$1(3) + 0(3)$	acceptable	$LR = 2.14$	acceptable	acceptable	weak
$2(3)$	toxic	$LR = 0.20$	toxic	toxic	toxic
$3(3)$	toxic	$LR = 0.05$	toxic	toxic	toxic
$1(3) + 1(3)$	toxic	$LR = 0.57$	toxic	weak	weak
$1(3) + 2(3)$	toxic	$LR = 0.15$	toxic	toxic	weak
$1(3) + 3(3)$	toxic	$LR = 0.04$	toxic	toxic	weak

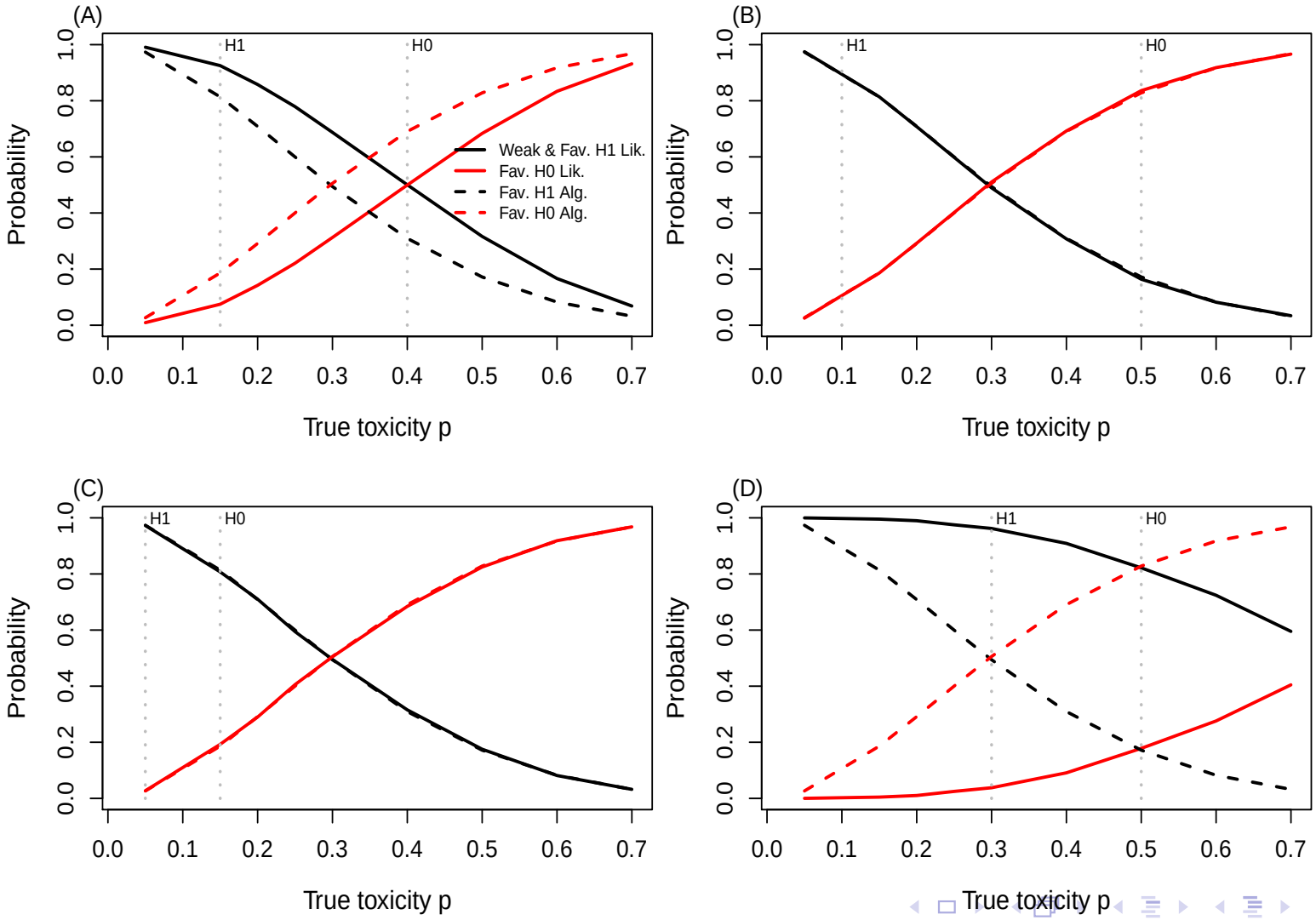
Dose limiting toxicities (cohort size)*

Acceptable: $LR \geq k$; toxic: $LR \leq 1/k$; weak: $1/k < LR < k$

Operating characteristics, $k = 2$



Dose considered safe until proven otherwise, $k = 2$



Take home points

- For $p_0 > 0.30$ and $p_1 < 0.20$, the likelihood method supports the '3+3' rules
- For low or high toxicity scenarios, the '3+3' has no probabilistic support - Should not be used!
- With maximum 6 patients/dose, the recommended threshold for the likelihood-ratio is $k = 2$
- Our likelihood method is able to identify and discard unsafe doses for settings where the optimality criteria also includes efficacy responses, e.g., cancer immunotherapy

References

[1] Iasonos A, Wilton AS, et al. A comprehensive comparison of the continual reassessment method to the standard '3+3'. *Clin Trials*, 5(5):465-477, 2008.

[2] Ji Y, Wang SJ. Modified toxicity probability interval design: a safer and more reliable method. *J Clin Oncol*, 2013.

[3] Lin Y, Shih WJ. Statistical properties of the traditional algorithmbased designs for phase I cancer clinical trials. *Biostatistics*, 2(2):203-215, 2001.

[4] Rogatko A, Schoeneck D, et al. Translation of innovative designs into phase I trials. *J Clin Oncol*, 25(31):4982-4986, 2007.

[5] Royall R. Statistical Evidence: A likelihood paradigm: Chapman Hall/CRC, 1997.

[6] Blume JD. Likelihood methods for measuring statistical evidence. Stat Med, 21(17):2563-2599, 2002.

[7] Hacking I. Logic of statistical inference: Cambridge University Press, NY, 1965.

Thank you!

Email: chiuzan@musc.edu

Work supported by the NIH grant, P01 CA 154778-01 and part by the Biostatistics Shared Resource, Hollings Cancer Center, MUSC (P30 CA138313)