



An exact method for analysing two-stage phase II clinical trials in cancer

Gordana Jovic, John Whitehead

Medical and Pharmaceutical Statistics Research Unit, Lancaster University, UK

Institute of Medical Statistics and Biometry, University of Milan, Italy

Phase II clinical trials in oncology

- aim: to test if the experimental treatment is active enough to proceed with its development
- participants: cancer patients for whom the other treatments have failed
- single arm trial (no control)
- short-term response (tumour shrinkage)
- small sample size 20-50 patients

Objectives

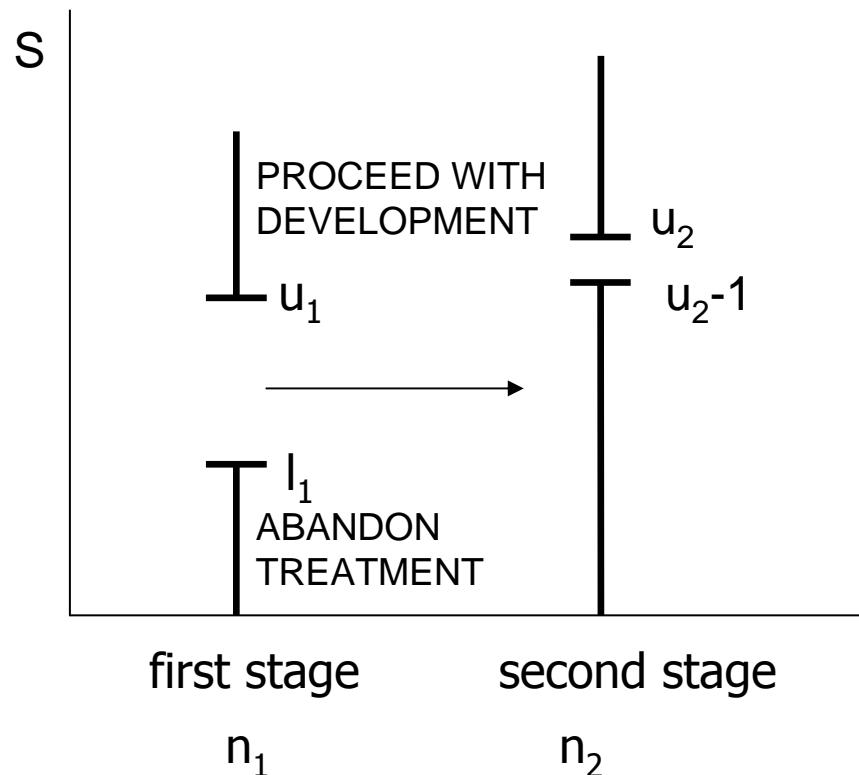
- Develop an exact method for estimation after two-stage design
- Using the exact calculation evaluate any method for estimation after (single-stage and) two-stage design
- Compare the evaluations of the methods for estimation
- Simon's optimal, Simon's minimax, Fleming's and other two-stage procedures were used to design the trials

Test

- p is the probability of success
- test $H_0: p = p_0$ vs $H_1: p = p_1$
- p_0 response rate with standard treatment
- $p_1 (>p_0)$ the response rate that would be clinically relevant

Two-stage designs

- $(l_1, u_1, n_1, u_2, n_2)$ design parameters



FIRST STAGE:

$n_1 \quad l_1 \quad u_1$

\mathbf{S}_1 : no of successes after the first stage

\mathbf{s}_1 : observed number of successes after the first stage

SECOND STAGE:

$n_2 \quad u_2$

\mathbf{S}_2 : no of successes after the second stage

\mathbf{s}_2 : observed number of successes after the second stage

Two-stage designs

- five unknowns (l_1, u_1, n_1, u_2, n_2)

- two constraints

$$p_{H_0}(S_1 \geq u_1) + p_{H_0}(S_1 \in (l_1, u_1) \text{ and } S_2 \geq u_2) \leq \alpha$$

$$p_{H_1}(S_1 \geq u_1) + p_{H_1}(S_1 \in (l_1, u_1) \text{ and } S_2 \geq u_2) \geq 1 - \beta$$

- other constraints can be imposed

Definition of p-value functions

discrete case

- $P(p)$ is the probability of observing a sample at least extreme as the one observed or more extreme in favour of the experimental treatment, when the probability of success is equal to p .
- $Q(p)$ is the probability of observing a sample strictly more extreme in favour of the experimental treatment

Ordering of outcomes

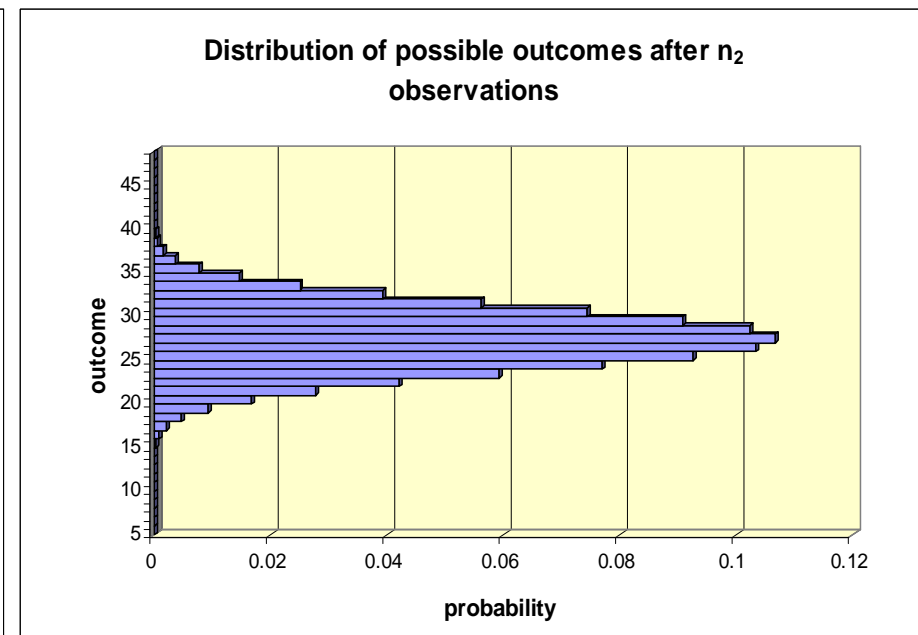
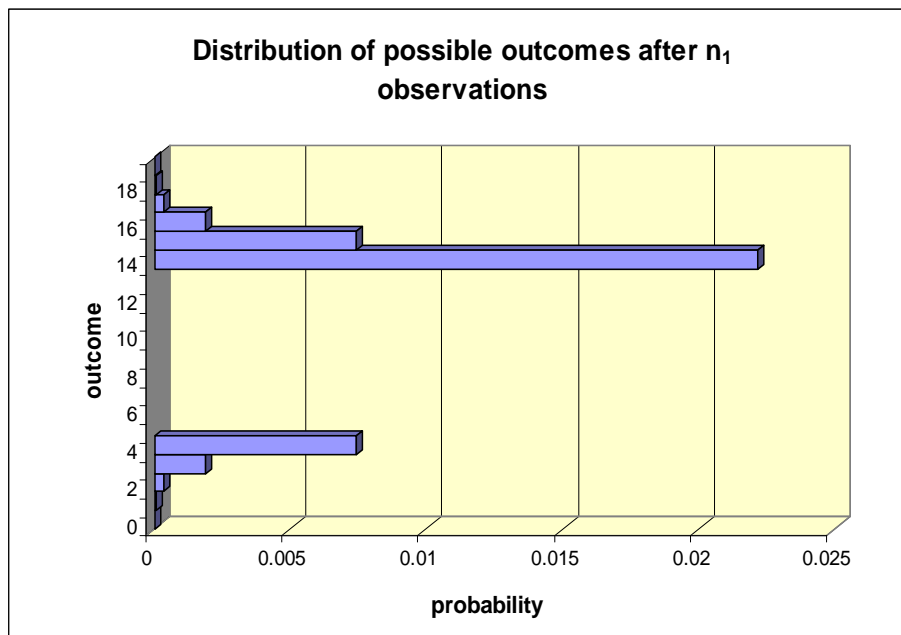
Fairbanks and Madsen (1982)

- if the actual trial leads to **rejection of H_0 after n_1 observations**, the only way in which a replicate trial could produce stronger evidence would be if $S_1 > s_1$
- if the trial continued to **n_2 observations**, then a replicate trial could produce stronger evidence by rejecting H_0 after n_1 observations or by continuing to n_2 observations with $S_2 > s_2$
- if the trial **stopped after n_1 observations with non-rejection of H_0** , then a replicate trial could produce stronger evidence by doing the same with $S_1 > s_1$, continuing to n_2 observations with any final value of S_2 , or by rejecting H_0 after n_1 observations

Determination of the p-value functions

Exact design for testing: $H_0: p_0 = 0.20$ vs $H_1: p_1 = 0.40$ $l_1=4, u_1=14, n_1=19, u_2=16, n_2=54$

NOT BINOMIAL

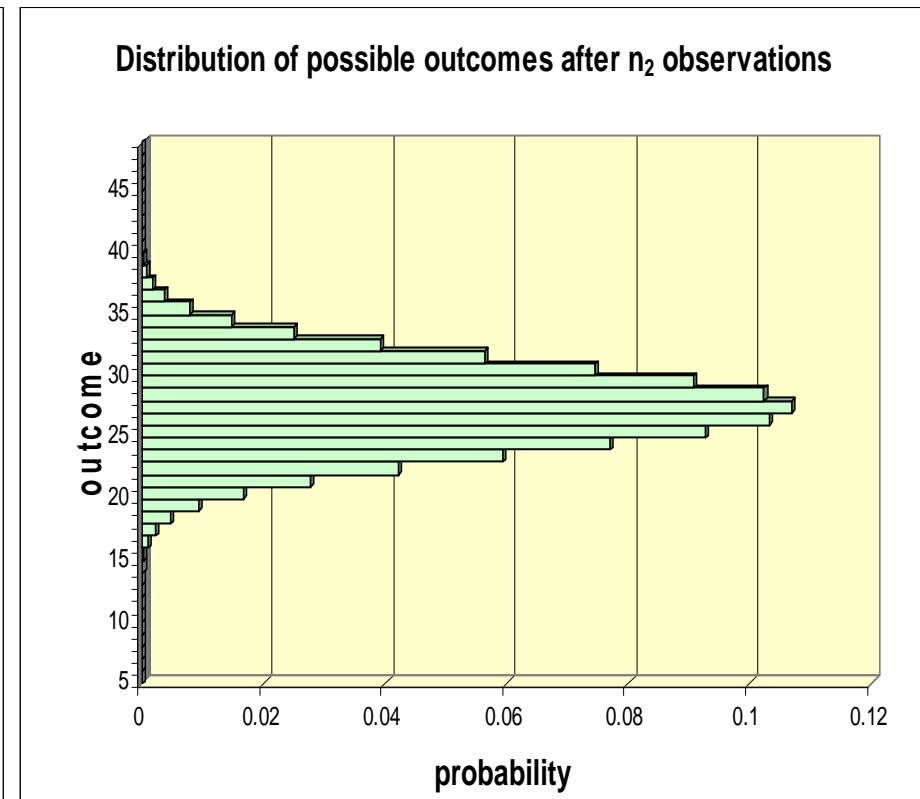
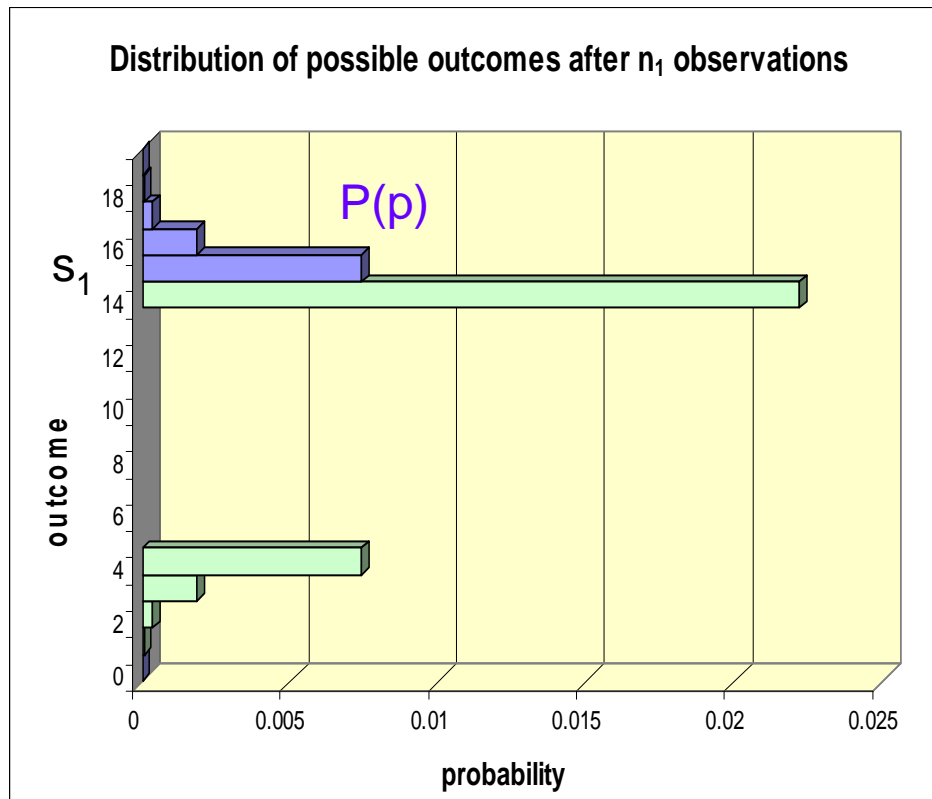


$$p(S_2 = s_2) = \sum_{s_1=l_1+1}^{u_1-1} p(S_2 = s_2 | S_1 = s_1) p(S_1 = s_1)$$

$$p(S_2 = s_2 | S_1 = s_1) = \binom{n_2 - n_1}{s_2 - s_1} p^{s_2 - s_1} (1 - p)^{(n_2 - n_1) - (s_2 - s_1)}$$

Determination of the p-value functions

- $s_1 \geq u_1$: finish after n_1 observations and reject H_0

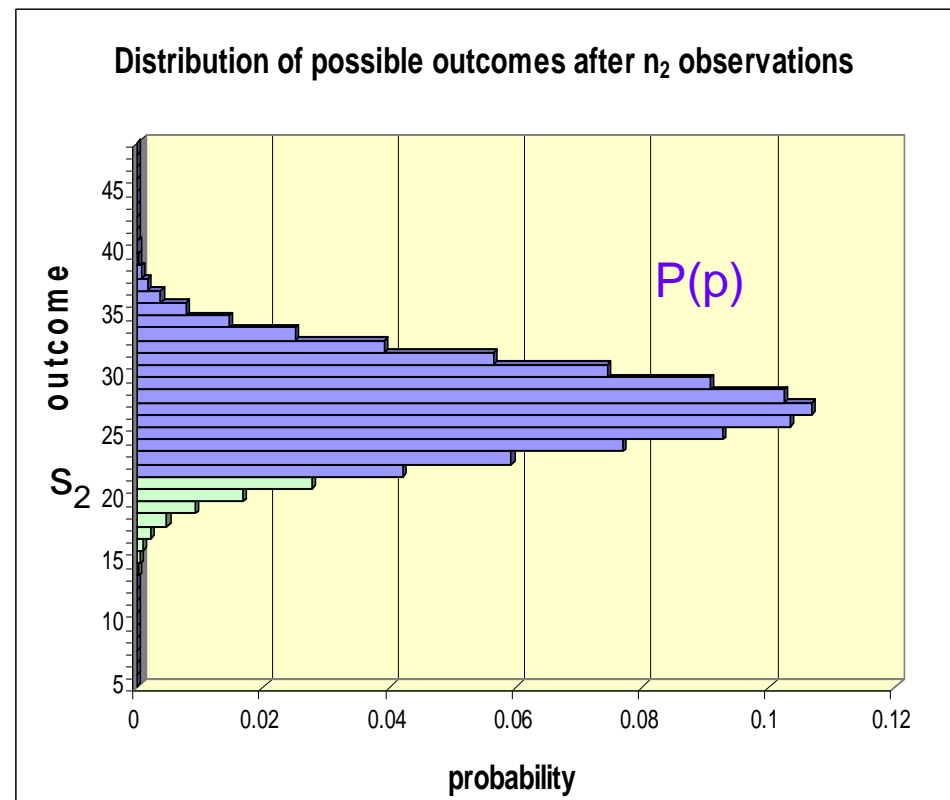
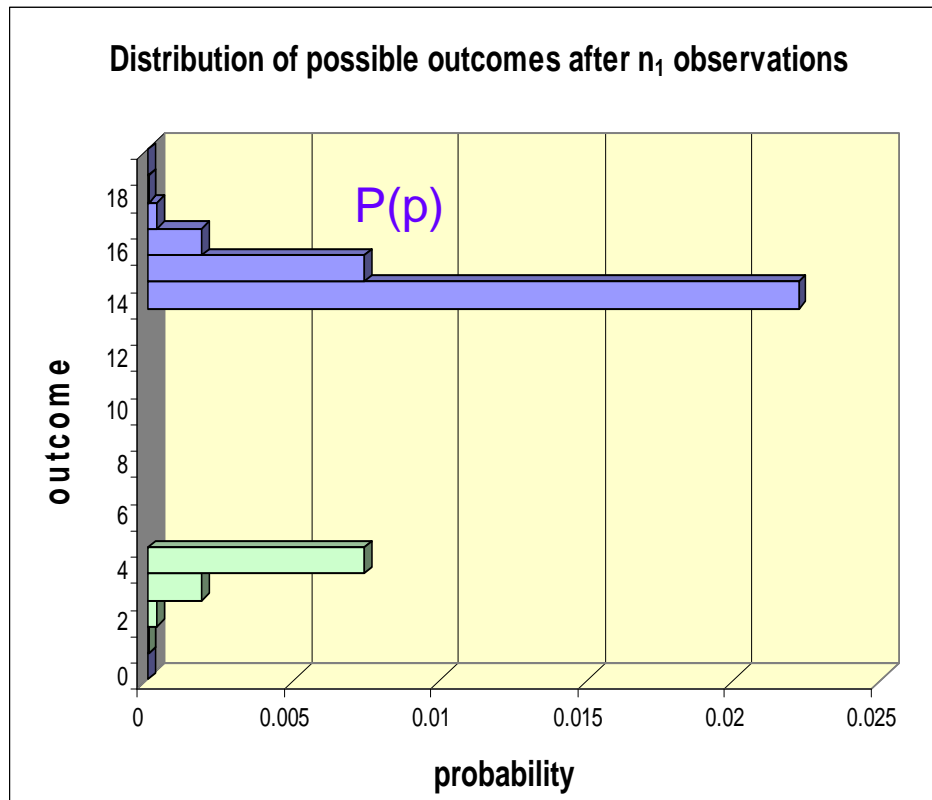


$$P(p) = p(S_1 \geq s_1; p)$$

$$Q(p) = p(S_1 > s_1; p)$$

Determination of the p-value functions

- $\mathbf{s}_1 \in (l_1; u_1)$: finish after n_2 observations

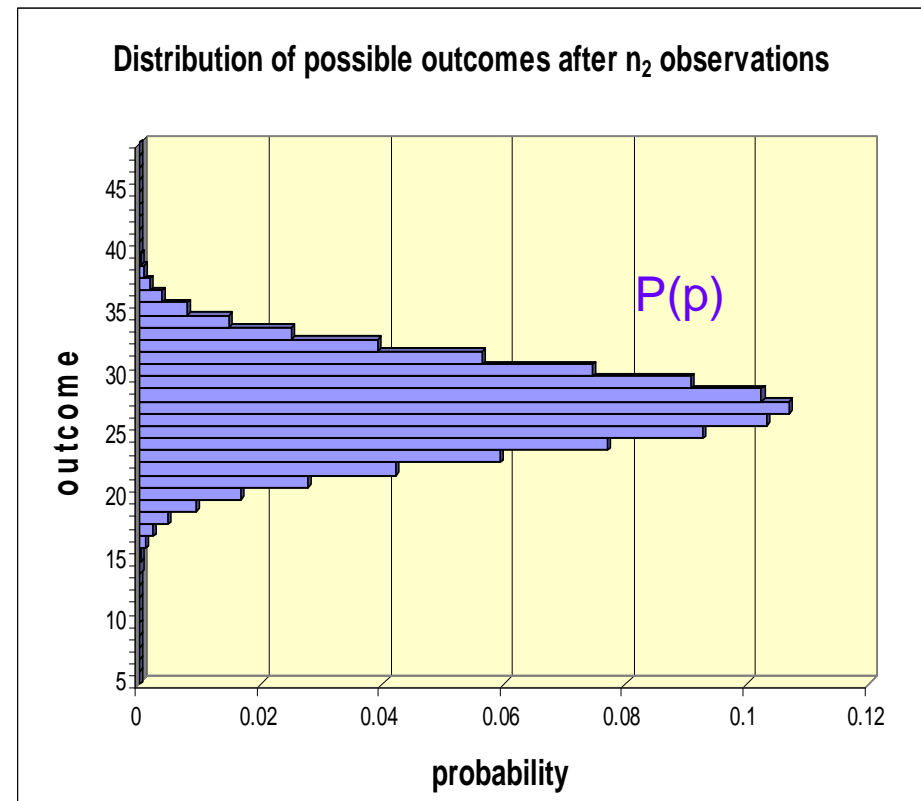
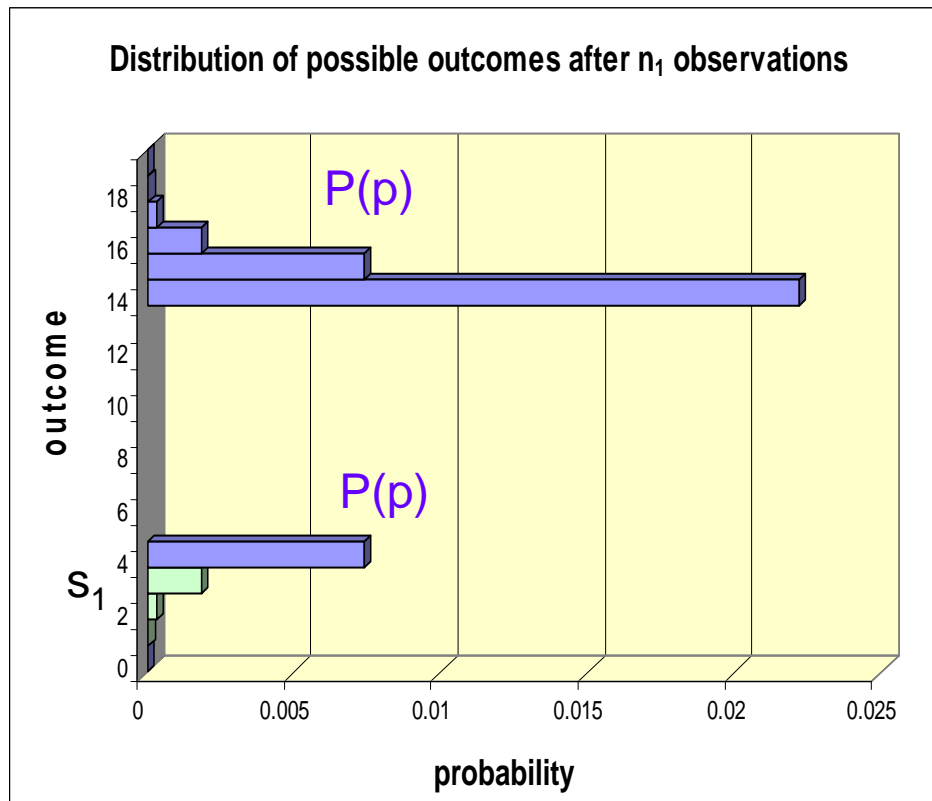


$$P(p) = p(S_2 \geq s_2 \text{ and } S_1 \in (l_1; u_1); p) + p(S_1 \geq u_1; p)$$

$$Q(p) = p(S_2 > s_2 \text{ and } S_1 \in (l_1; u_1); p) + p(S_1 \geq u_1; p)$$

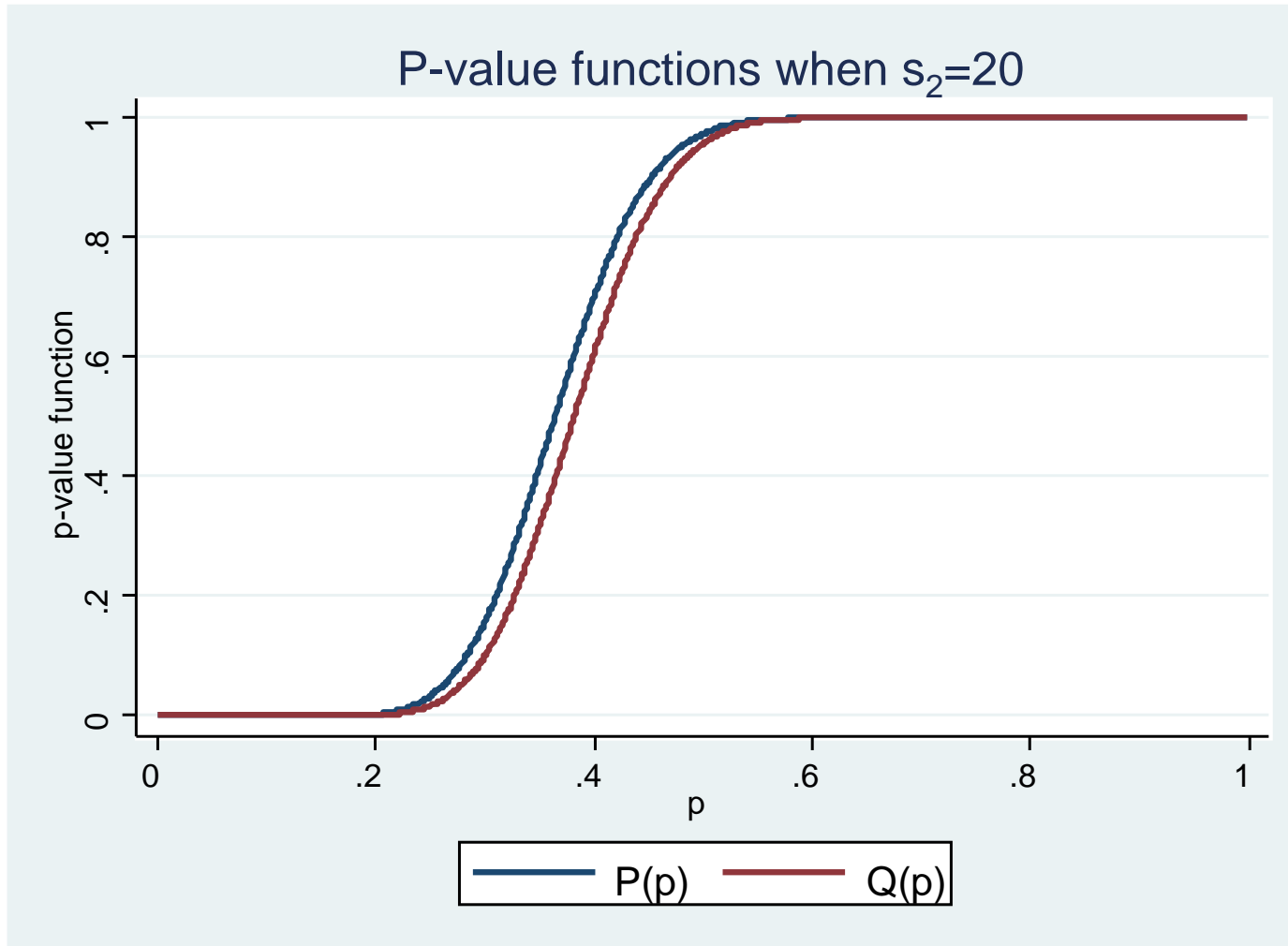
Determination of the p-value functions

- $s_1 \leq I_1$: finish after n_1 observations with non-rejection of H_0



- $P(p) = 1 - p(S_1 < s_1)$ $Q(p) = 1 - p(S_1 \leq s_1)$

Estimation after two-stage design

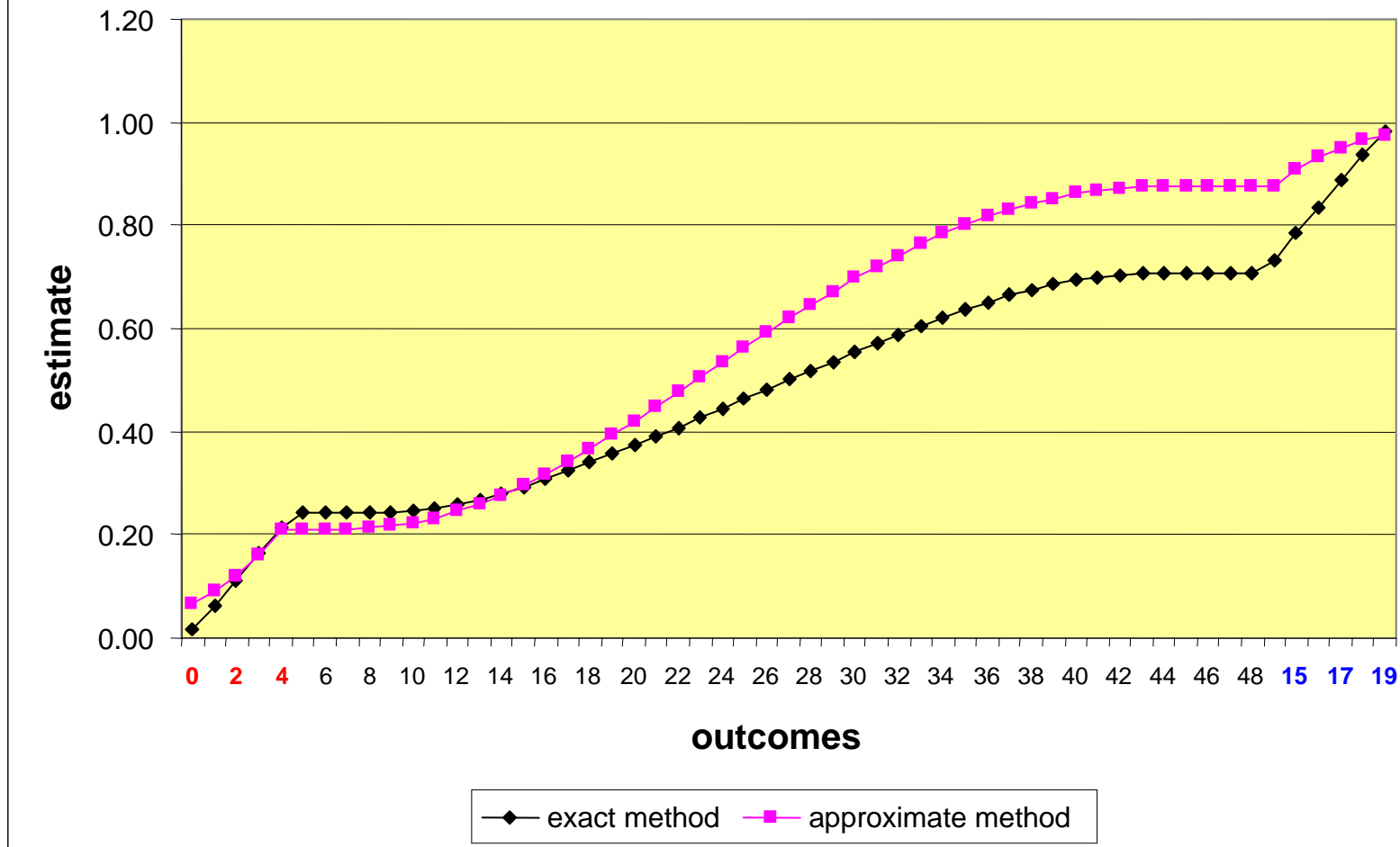


Estimation after two-stage design

statistics	P and Q functions
p-value	$P(p_0)$
median unbiased estimate	$P(p_{MP})=0.5$ $Q(p_{MQ})=0.5$ $p_{MEDIAN}=(p_{MP}+p_{MQ})/2$
lower limit of the 95%CI*	$P(p_L)=0.025$
upper limit of the 95%CI*	$Q(p_U)=0.975$

*CI generalises Clopper and Pearson's CI from the single-stage case

Comparison of median unbiased estimates for success probability

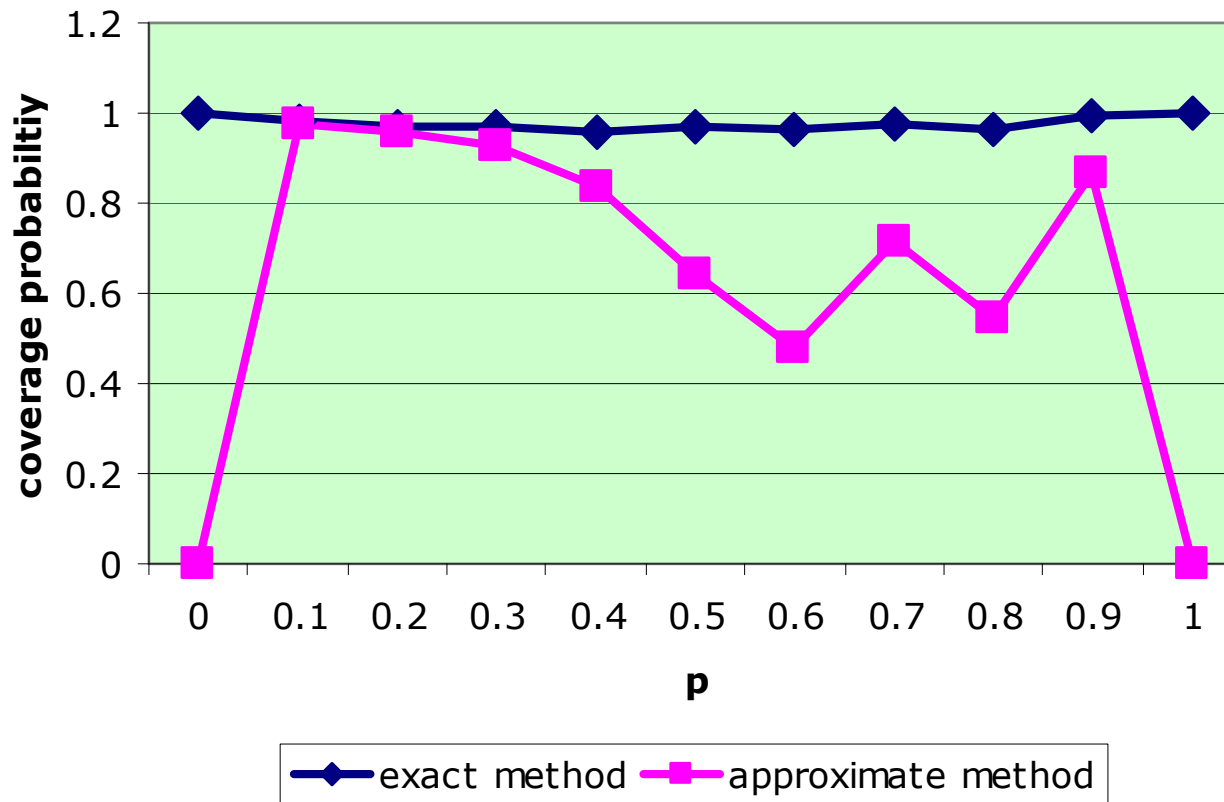


Exact design for testing: $H_0: p_0 = 0.20$ vs $H_1: p_1 = 0.40$

$l_1=4, u_1=14, n_1=19, u_2=16, n_2=54$

*the approximate method uses the parameterization with logarithm of odds-ratio $\theta = \log\left(\frac{p(1-p_0)}{p_0(1-p)}\right)$ and $Z \sim N(\theta V, V)$

Comparison of two-sided nominal 95% CI coverage probabilities



Exact design for testing: $H_0: p_0 = 0.20$ vs $H_1: p_1 = 0.40$

Conclusions

- Estimates: The approximate method overestimates the probability of success when p is different from p_0 . Using the exact method for the analysis we obtain the exact point estimates, confidence limits and p-value
- The exact evaluation: using the exact method we can evaluate any method for the analysis. When applying the exact method the coverage probability is always at least equal to the nominal value

Bibliographic references

- Whitehead J: *The design and analysis of sequential clinical trials*. Revised 2nd Ed., Chichester: Wiley, 1997.
- Simon R: Optimal two-stage design for phase II clinical trials. *Controlled Clinical Trials* 1989; 10: 1-10.
- Fleming RT: *One-sample multiple testing procedures for phase II clinical trials*. *Biometrics* 38; 143-151; 1982
- Clopper CJ, Pearson ES: *The use of confidence or fiducial limits illustrated in the case of the binomial*. *Biometrika* 1934; 26: 404-413.
- Fairbanks K, Madsen R: *P-values for tests using a repeated significance test design*. *Biometrika* 69; 69-74; 1982

- Acknowledgment
- Professor PierFranco Conte, University of Modena and Reggio Emilia Italy