SIGNIFICANCE AND POWER OF SEQUENTIAL BERNOULLI TRIALS

David J. Doolette and Wayne A. Gerth
Navy Experimental Diving Unit, Panama City, Florida, U.S.A.

INTRODUCTION

In sequential trials, data are analyzed after each group of results become available and the trial stopped when the treatment effect first becomes significant. Early stopping saves resources and reduces exposure of subjects to unnecessary risk. For instance, a trial of a single decompression procedure, in which the outcome for each subject is decompression sickness (DCS) or no-DCS (Bernoulli trial), could be stopped when the observed DCS incidence first exceeds a high limiting value or first drops below a low limiting value with 1-α confidence, see Figure 1.

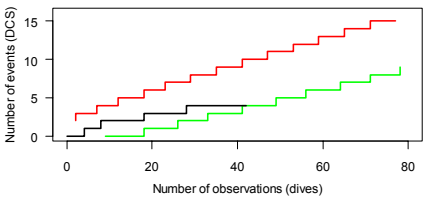


Figure 1. Sequential trial and stopping rules. The black line shows the cumulative DCS (y-axis) after each group of two dives (x-axis) is completed. In this example the trial stops early when the cumulative DCS intersects the stop-low (green) line, indicating the upper 75% binomial confidence limit (α=25%) of the observed DCS incidence is less than 15%. The stop-high (red line) indicates DCS incidences with lower 75% binomial confidence limit greater than 15%.

In this type of design, α is not the significance (probability of Type I error) of the sequential trial. Significance and power are defined in terms of conditional probabilities in Figure 2.

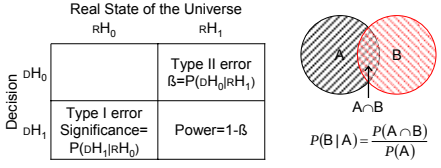
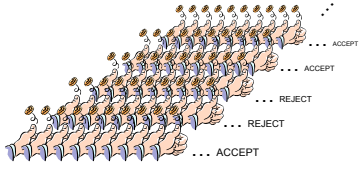


Figure 2. Significance, power, and conditional probability. H_0 and H_1 represent experimental decisions (D) or reality (R) consistent with the null and alternative hypotheses, respectively.

We present a Monte Carlo method to estimate significance and power of sequential Bernoulli trials that extends and corrects a previous description (Ball R, Parker EC. Aviat Space Environ Med 2000;71:102-8).

METHODS

The method involves generating a large number of Monte Carlo simulations of Bernoulli trials, under specified maximum trial size and trial stopping rules, and counting the relative frequency of each trial outcome.



This simulation is repeated at selected values over the range of possible real probabilities DCS (0≤p≤1), from which estimated probability distributions of accept, reject, and indeterminate trial outcomes are generated, see Figure 3.

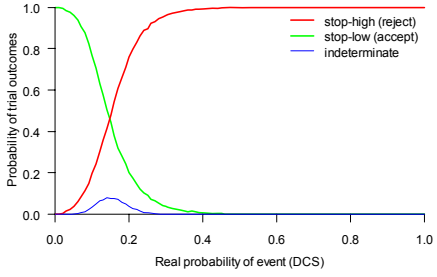


Figure 3. Probability distributions of possible outcomes for the trial design illustrated in Figure 1 with maximum trial size n=80. Estimated probabilities of trial outcomes (y-axis) are the relative frequencies from 10000 trial simulations at each selected value of the real probability of DCS (x-axis).

Probabilities of stopping the trial in error are estimated from appropriate areas under the distributions, see Figures 4 and 5.

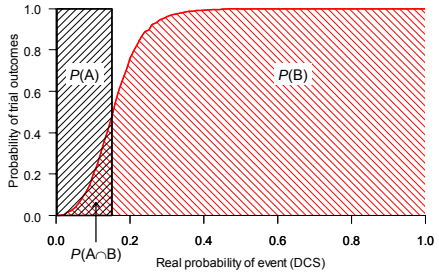


Figure 4. Conditional probability of a stop-high (reject) trial outcome given real probability of DCS<15%. The area inside the black rectangle is the probability of all possible trial outcomes for real probability of DCS<15% [*P*(A), 15%]. The area under the red distribution is the probability of a stop-high trial outcome for all real probabilities of DCS [*P*(B)]. The intersection of these two areas (red and black cross-hatching) is the probability of a stop-high trial outcome for real probabilities of DCS<15% [*P*(A∩B)]. The conditional probability of a stop-high trial outcome given real probability of DCS<15% [*P*(B|A)] is the red and black cross-hatched area divided by the black hatched area [*P*(A∩B)/*P*(A)].

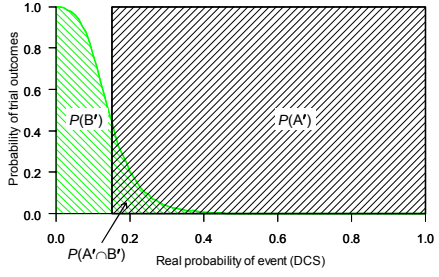


Figure 5. The conditional probability of a stop-low (accept) trial outcome given the real probability of DCS>15% is the green and black cross-hatched area divided by the black hatched area.

Conditional probabilities can be similarly calculated for indeterminate trial outcomes. How the conditional probabilities relate to Type I and Type II errors depends on how the null hypothesis (H_0) is stated and how indeterminate trial outcomes are treated.

For the design:

- H_0: the decompression procedure is acceptable
- Accept with stop-low
- Reject with stop-high or indeterminate

the probability of a Type II error (ß=1-power) is that illustrated in Figure 5. The probability of a Type I error (trial significance level) is that illustrated in Figure 4 plus the conditional probability of an indeterminate trial outcome given real probability of DCS<15% [*P*(indet|A), not illustrated]. These probabilities are tabulated to illustrate the effects of sample size and stopping rule α on significance and power.

RESULTS

Increasing sample size has only minor effect on power but improves significance level, primarily by deceasing the probability of an indeterminate trial outcome (Table 1).

Table 1. Effect of maximum trial size (n) on significance and power

n	*P*(B|A)	*P*(indet|A)	Significance Sum	1-Power *P*(B′|A′)
40	14.1%	9.7%	23.8%	2.9%
80	14.8%	2.5%	17.3%	3.2%
120	15.1%	1.3%	16.4%	3.3%
160	15.1%	0.7%	15.8%	3.4%

Increasing stopping rule confidence (1-α) modestly increases trial power but worsens significance level, primarily by increasing the probability of an indeterminate trial outcome (Table 2).

Table 2. Effect of stopping rule α on significance and power

1-α	*P*(B|A)	*P*(indet|A)	Significance Sum	1-Power *P*(B′|A′)
75%	15.0%	2.6%	17.6%	3.2%
80%	11.5%	6.1%	17.6%	2.9%
85%	9.8%	10.8%	18.6%	1.9%
90%	5.6%	19.7%	25.3%	1.0%
95%	3.5%	30.3%	33.8%	0.5%

CONCLUSIONS

Increasing stopping rule α and sample size decreases the probability of indeterminate trial outcome but has minor effect on probabilities of trial outcome errors.