Evaluation of Program Success for Programs with Multiple Trials in Binary Outcomes

Meihua Wang, G. Frank Liu, Jerald Schindler

Merck Research Laboratory

12/15/2015

MERCK

Outline

- Background
- Methods
  - -- Probability of success (POS)
  - -- Probability of program success (POPS)
  - -- Confidence intervals of POS and POPS
- Simulation Results
  - -- Variation of POS and POPS
  - -- Effect of analysis time on POS and POPS evaluation
- Applications
- Discussions

Background

- Probability of success (POS)

  -- average power or average conditional power (predictive power)

  -- accounting for uncertainties of the design parameters

- POS for the entire clinical program

  -- “Probability of program success (POPS)”

  -- probability of at least 1 (or 2) phase III trial being successful among all ongoing phase III trials in the clinical program

  -- may abandon the program early if the POPS estimated is very low

MERCK

Methods
---- Basic notations

Let x_j, n_j be # of success and sample size at **interim** for group j

- Posterior (with beta priors):

$$p_j | x_j, n_j \sim Beta\left(\left(a_j + x_j\right), \left(\beta_j + n_j - x_j\right)\right)$$

Let N_j be the final sample size, y_j be the # of success after interim for group j

- Predictive distribution:

$$y_j | x_j, n_j, N_j \sim Beta - Binomial(N_j - n_j, a_j + x_j, \beta_j + n_j - x_j)$$

- Posterior distribution of p at final:

$$p_j | x_j, y_j \sim Beta\left(a_j + x_j + y_j, \beta_j + N_j - x_j - y_j\right)$$

MERCK
Be well

Methods
---- POS/POPS

- POS for a single study after interim:

$$\Pr\left(Z > Z_{(1-\alpha/2)} \middle| x_1, x_2, n_1, n_2, N_1, N_2\right) = \iint I\left[Z > Z_{\left(1-\alpha/2\right)}\right] f(p_1) f(p_2)$$

$$= \sum_{y_1=0}^{N_1-n_1} \sum_{y_2=0}^{N_2-n2} I\{Z > Z_{(1-\alpha/2)} | x_1, x_2, n_1, n_2, y_1, y_2\} f\left(y_1 \middle| x_1, n_1, N_1\right) f\left(y_2 \middle| x_2, n_2, N_2\right)$$

where $f(y_1)$ and $f(y_2)$ are Beta-Binomial distributions.

- POPS: probability of at least T trials being successful among all *K* ongoing phase III trials in the program.

$$POPS = \iint I\left[\sum_{k=1}^{K} I\{Z_k > Z_{(1-\alpha/2)}\} \geq T\right] f(p_1) f(p_2)$$

where $f(p_1)\ and\ f(p_2)$ are the posterior distributions incorporating prior and interim data from all ongoing phase III trials.

$$f(p_i) \sim Beta\left((\alpha_i + x_{i1} + x_{i2} + \cdots x_{iK}), (\beta_i + n_{i1} - x_{i1} + n_{i2} - x_{i2} + \cdots n_{iK} - x_{iK})\right)$$

  - POPS can be calculated through Monte Carlo Method.

MERCK

Methods
---- Confidence intervals of POS/POPS

- Research problems
  - Confidence measures of POS/POPS for a real clinical program
  - Appropriate time frame to perform POS/POPS evaluation

- *Consider a bootstrap approach*

  - A*ccount for uncertainty* in historical data
  - Generate prior using a bootstrap sample from the historical data
  - Calculate POS or POPS
  - Obtain empirical distribution of POS or POPS

Methods
---- Confidence intervals of POS/POPS

Computation procedures:

Step 1: Draw **prior** (α_j, β_j) from **Bootstrapping** methods (based on historical data). j=1,2

Step 2: Calculate **POS/POPS** based on **prior** (from Step 1) , **observed interim data** (x_{jk}, n_{jk}), and **final sample size** (N_{jk}). POS – direct summation of all possibilities
POPS –Monto Carlo Method

Step3: Repeat step 1 and 2 5000 times, get median and quantiles of POS/POPS.

MERCK

Results
---- Simulation Setup

- Simulation Setup:
  - Consider a program with 3 trials, each has 2 treatment groups
    - Sample size $N_{1k} = N_{2k}$ = 210 k=1,2,… K; K=3
  - Simulated 32 scenarios: combinations of four types of response rates, two priors, and four analysis times.
    - Treatment response rates:

(1): p_1= 30% vs. p_2 = 30% (5% power)

(2): p_1= 35% vs. p_2 = 30% (20% power)

(3): p_1= 43% vs. p_2 = 30% (80% power)

(4): p_1= 45% vs. p_2 = 30% (90% power)

Results
---- Simulation Setup

- Prior: (1): Non-informative prior ($beta(1,1)$)

  (2): Informative prior (M_1=M_2=100)

- Analysis time:

  (1): 30% (n_{1k}=n_{2k}=63)

  (2): 50% (n_{1k}=n_{2k}=105)

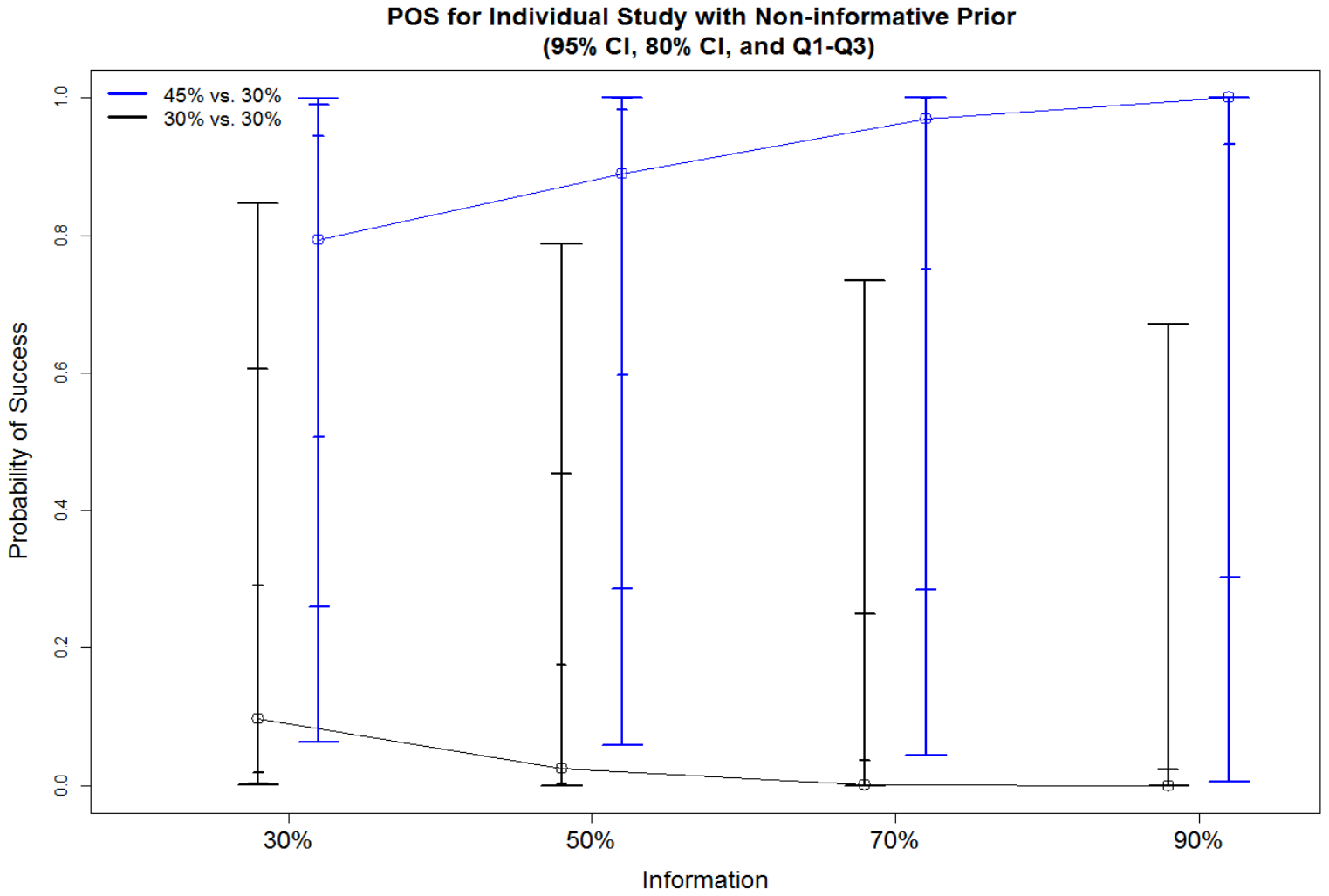
  (3): 70% (n_{1k}=n_{2k}=147)

  (4): 90%(n_{1k}=n_{2k}=189)

- Simulated data from binomial distribution with given n (or M) and the true response probability (p_1, p_2),
- Evaluate POS and POPS (defined as 2 or more studies successful)

MERCK

Results
---- Measurement for Variation of POS



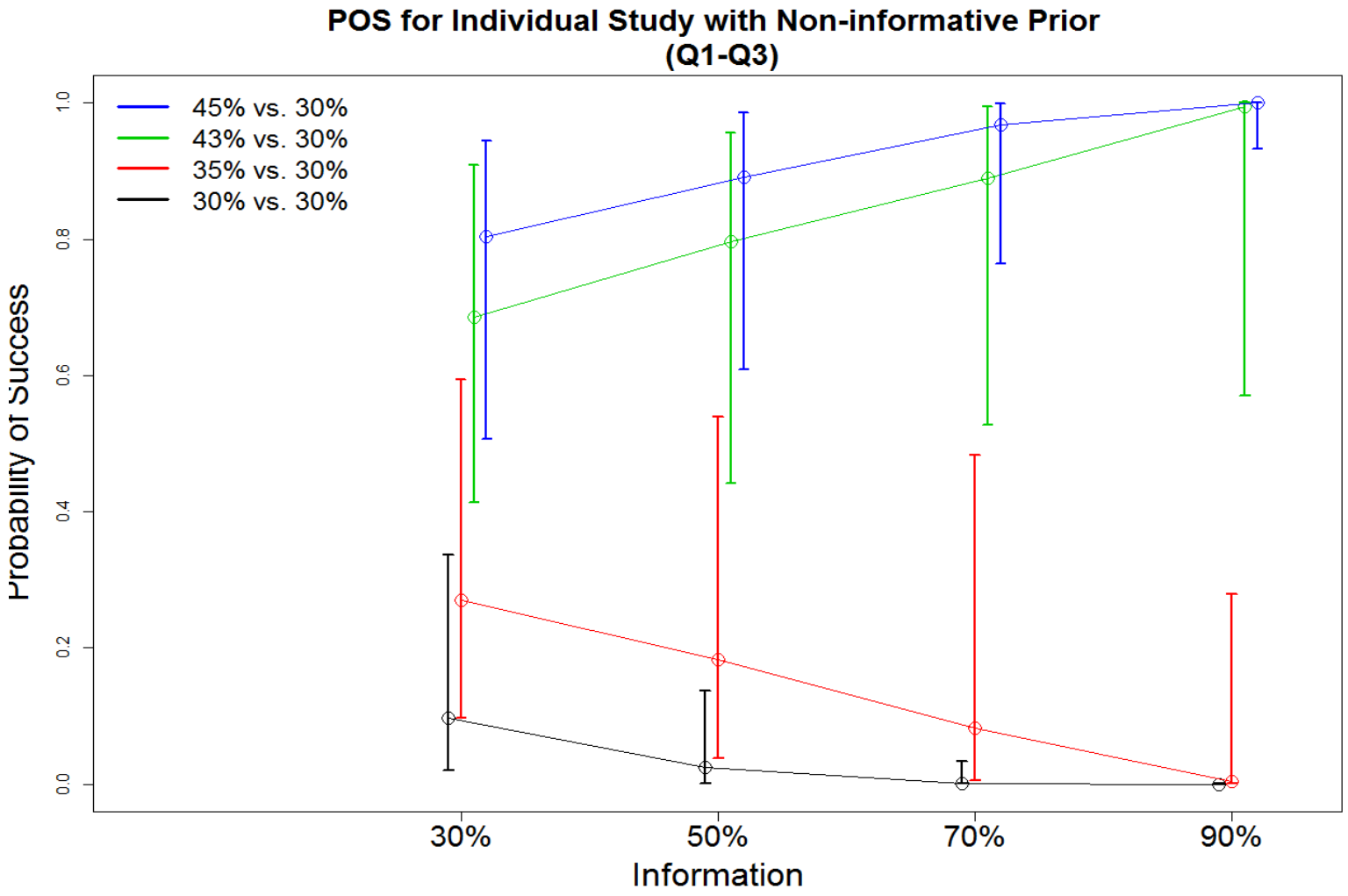
MERCK
Be well

Results
---- Measurement for Variation of POS

- The plots under both null and alternative scenarios illustrate
  - that the distribution of POS can be very skewed
  - 95% or 80% CI can be very wide
- Q1-Q3 may be more appropriate than 95% and 80% CI to describe the variations of POS estimate.

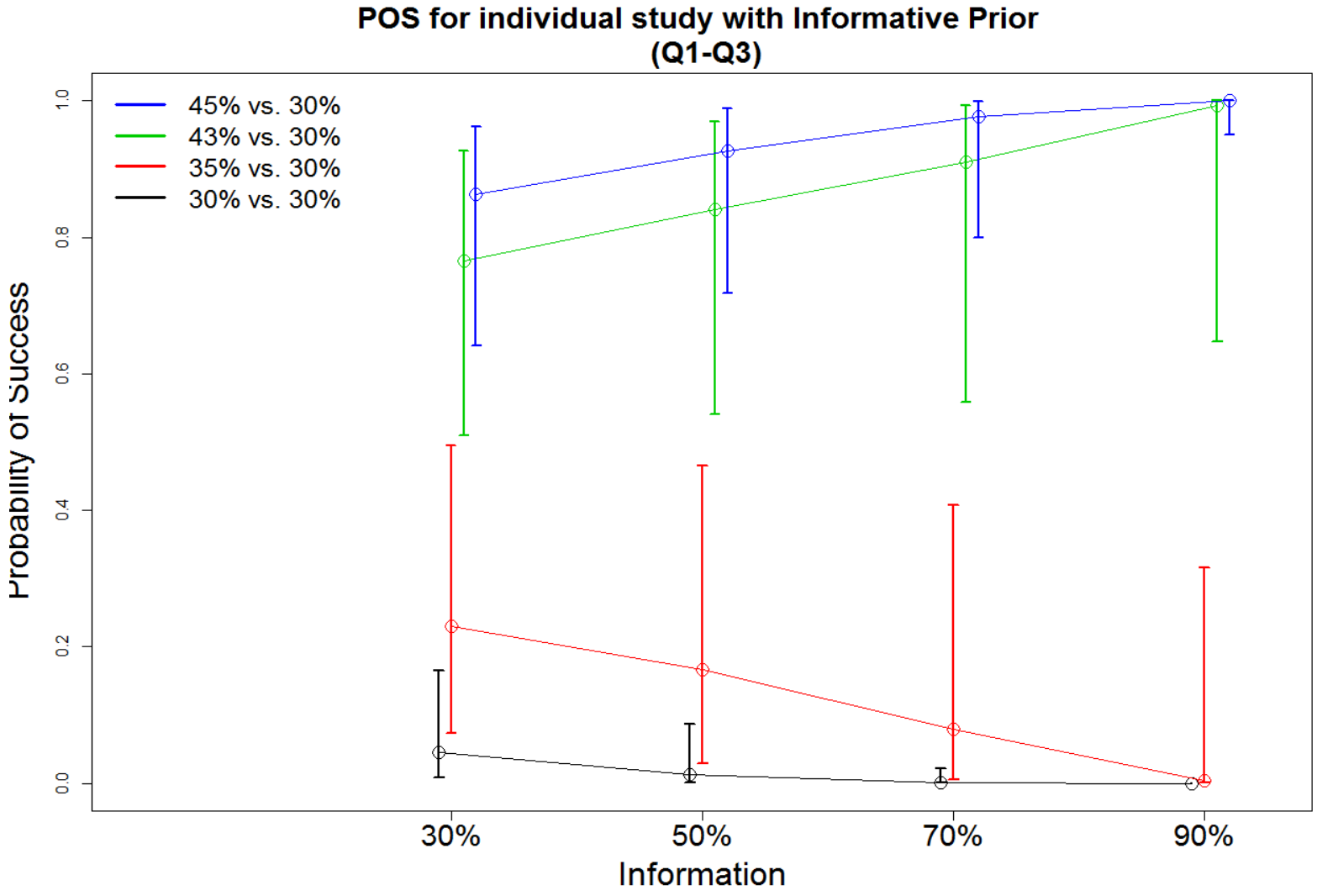
MERCK
Be well

Results
---- Interim Analysis Timing and Priors for POS

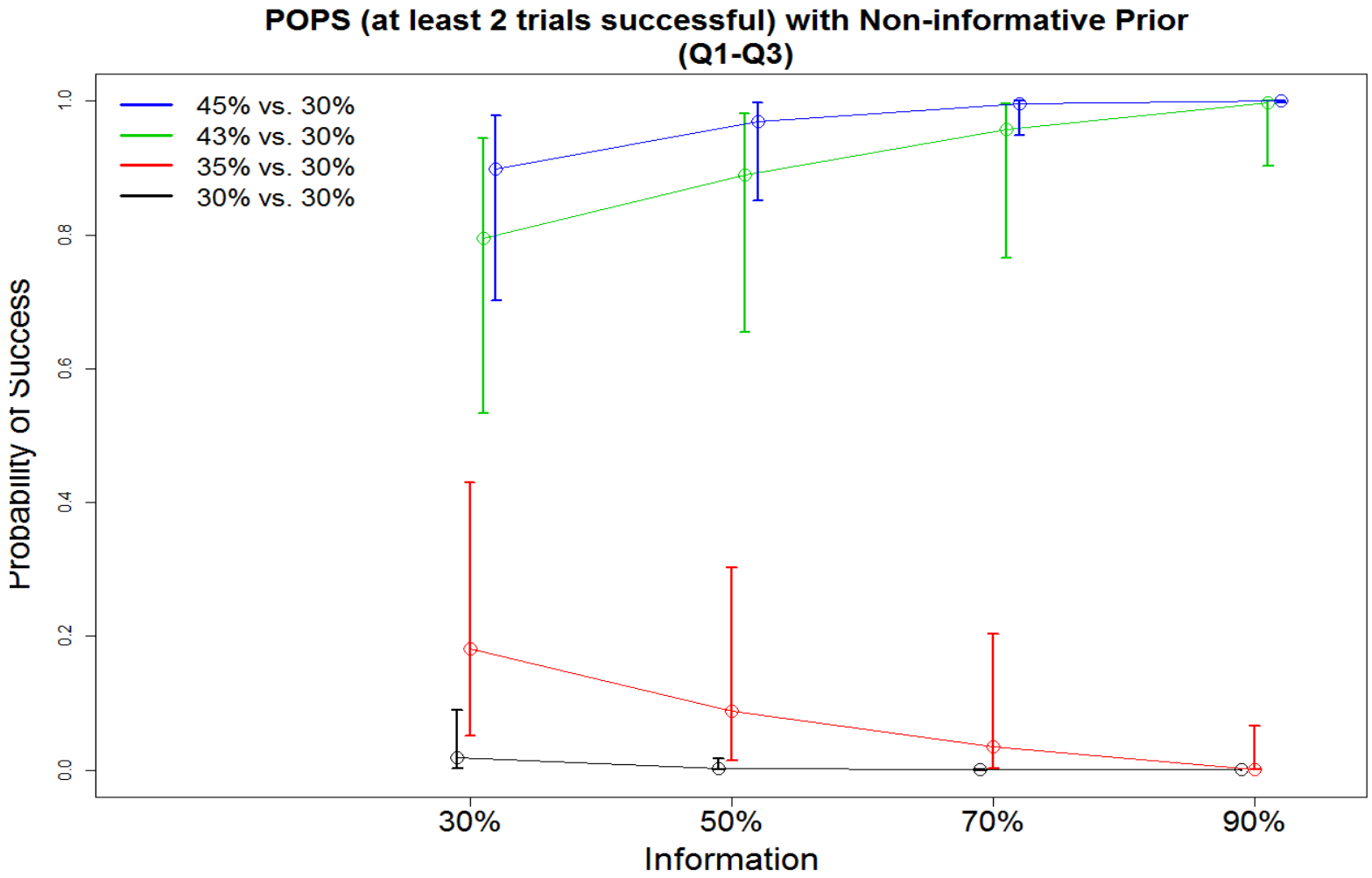


MERCK
Be well

Results
---- Interim Analysis Timing and Priors for POS

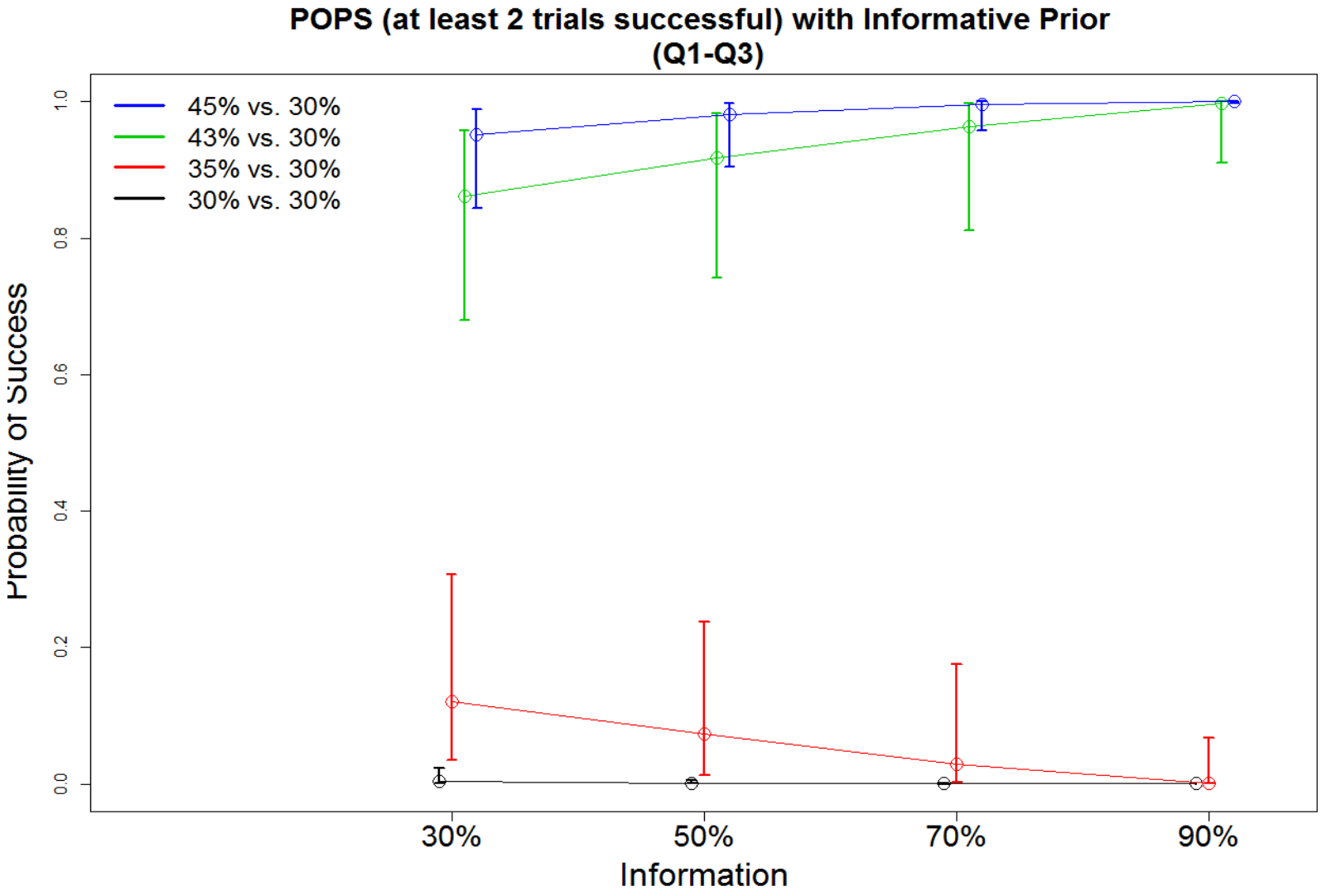


Results
---- Interim Analysis Timing and Priors for POPS



MERCK
Be well

Results
---- Interim Analysis Timing and Priors for POPS



Results

---- Interim Analysis Timing and Priors for POS/POPS evaluation

- As more trial/program information available, confidence intervals of POS/POPS got narrower.
- POPS had a narrower confidence interval than POS.
- Informative priors led to narrower confidence intervals. However, as more data from trial/program are available, the impact from prior will gradually decrease.
- Different scenarios of response rates led to different POS/POPS estimates.
  - the (Q1– Q3) of POPS from the first two hypothesis scenarios were separated from those from the two later alternative hypothesis scenarios, even at 30% information, the separation became especially prominent at 50% information.
- POPS provided reasonable estimates when 30~50% of program information is available.

MERCK

Applications

- **Prior (Phase II):**

$$p_{MK-0869} \sim beta(126, 152)$$
$$p_{active} \sim beta(120, 165$$
$$p_{placebo} \sim beta(98, 191)$$

- **Interim data and final sample size**
  - -- K=4
  - -- Suppose interim analyses were done in Aug02 (retrospective analysis)
  - -- Three studies had 30~50% patients and 1 study had 15% patients.

Group	P059 x/n/N	P060 x/n/N	P061 x/n/N	P062 x/n/N
MK-0869	27/59/150	28/60/145	28/75/139	14/26/165
Active control	37/67/148	30/57/151	42/77/137	12/15/161
Placebo	29/69/150	29/62/150	37/76/141	16/27/154

MERCK
Be well

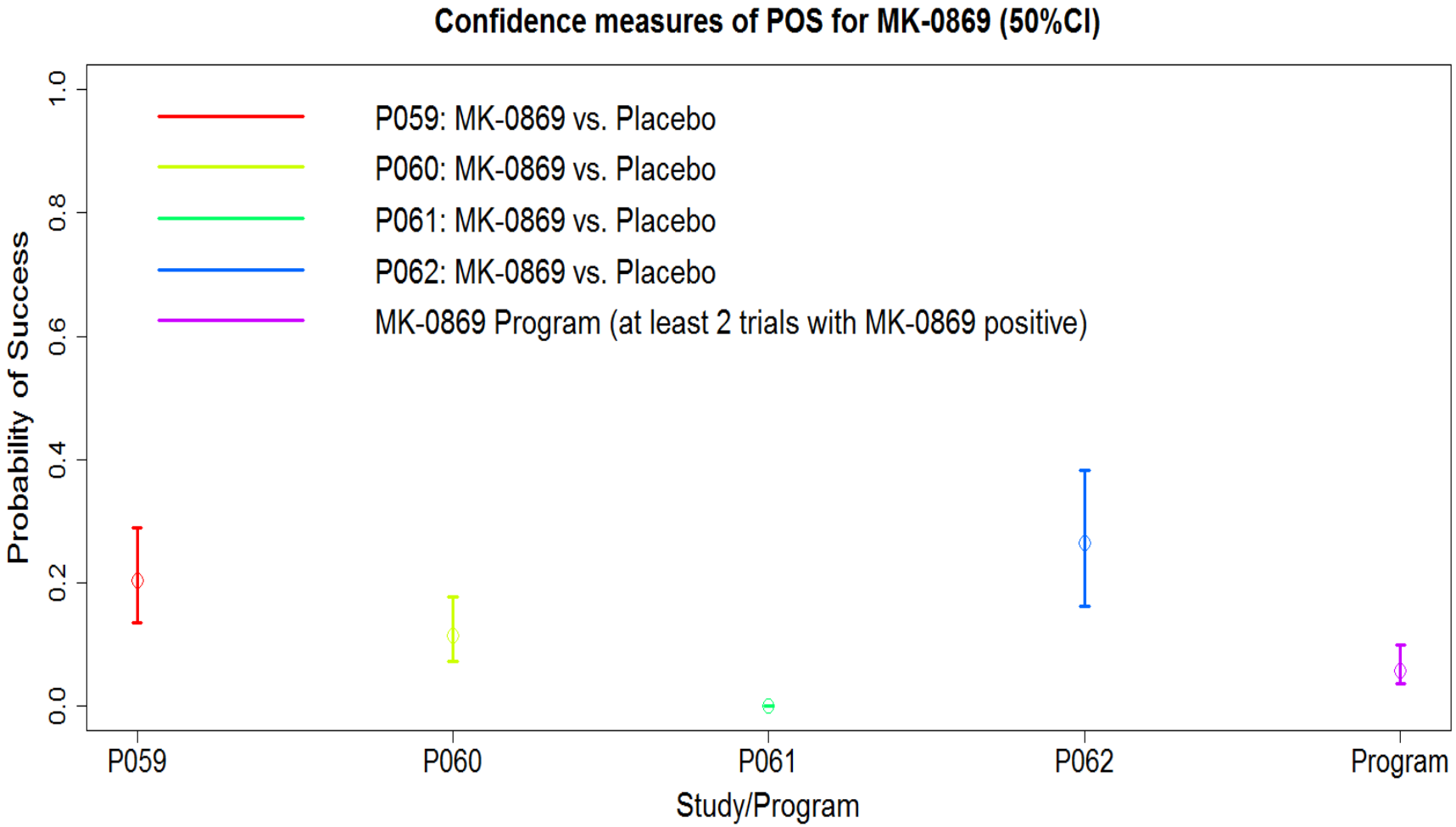
Applications

Table: Probability of Success for Mk-0869 Program

POPS	MK0869	Active Control
POPS requiring at least 1 trial positive	0.485	0.868
POPS requiring at least 2 trials positive	0.061	0.489
POPS requiring at least 3 trials positive	0.004	0.139

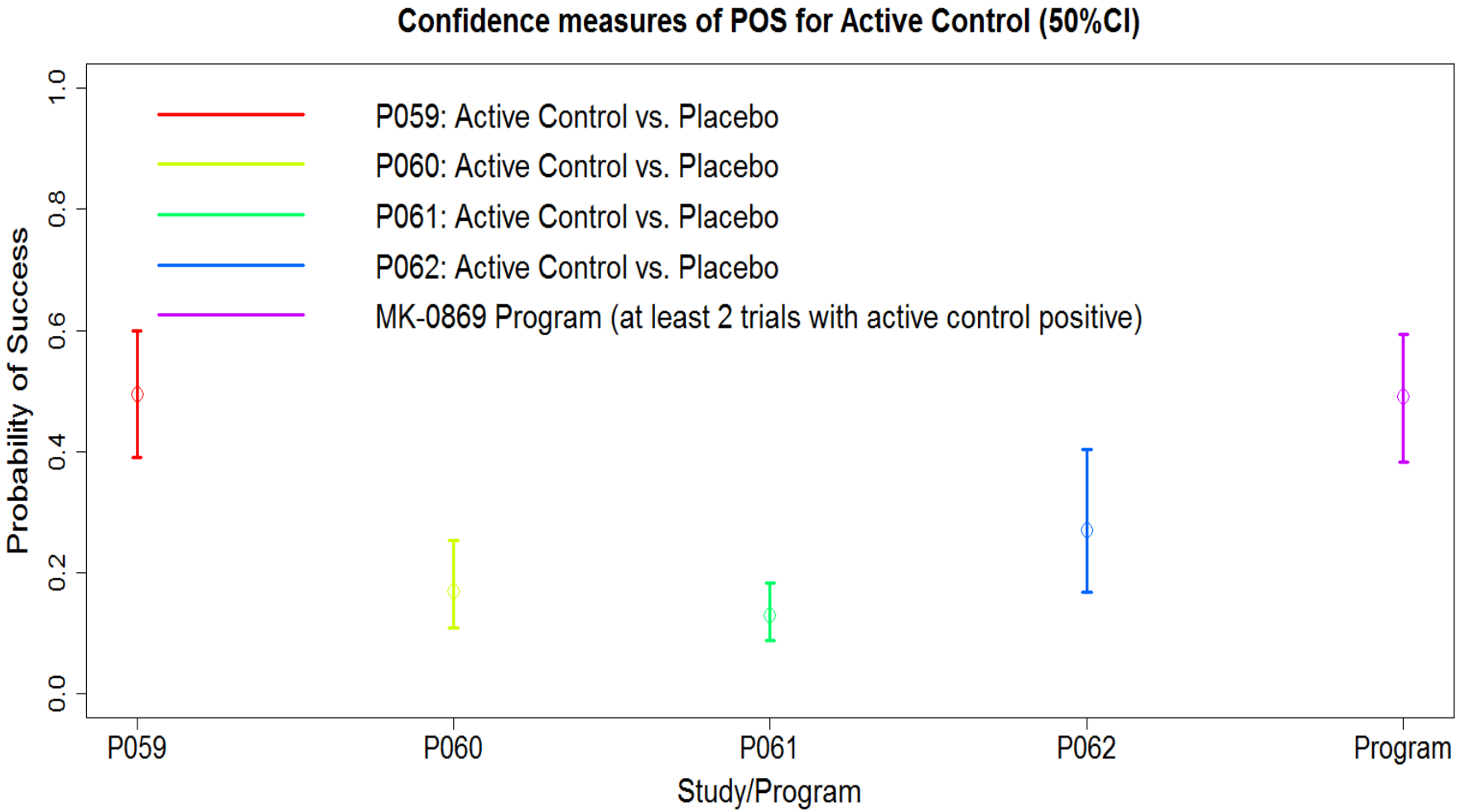
- In the completion of all 4 studies,
  - None of the studies were positive for MK0869
  - 2 studies (p059, p062) were positive for active control

Applications



MERCK
Be well

Applications



MERCK
Be well

Applications

- For MK-0869 compound, the median POPS is 0.057 with 50% CI (0.036, 0.098);
- For Active Control, the median POPS is 0.490 with 50% CI (0.382, 0.594).
- This suggests that a real clinical program POPS evaluation is appropriate at 30~50% information available.
- Had the POPS evaluation been done, the program could have been stopped earlier.

Discussions

- It is informative to consider uncertainty in POS / POPS evaluation
- 50% Confidence interval (Q1-Q3) provides a reasonable measure for POS / POPS evaluation than the traditional 95% CI
- Informative priors lead to narrower confidence intervals for POS or POPS. However, impact is less when more data become available.
- Timing of interims: reasonable when 30~50% of program information is available.
- No universal rule for POS / POPS, generally:
  - A mean < 0.2 and Q3 < 0.5 may indicate a low POS/POPS
  - A mean > 0.5 and Q1 > 0.4 may indicate some good chance

  The choice may also depend on the disease areas and other clinical and/or public health considerations.

- Several points for considerations in the implementation:

  -- It should be with caution when incorporating prior from historical data.

  -- Tightly controlled unblinding procedures should be in place

  -- The interim POPS evaluations serve as a futility check

  -- The proposed POPS metric mainly helps the decision of phase III program continuation or termination.

  -- The application of POPS requires program-wide DMC, Charter, and a common unblinded statistician or external Statistical Center, in addition to the study-specific DMC, Charter and unblinded statistician.

- In practice, shutting down the entire program requires more discussions than relying on a single POPS metric that is obtained under certain assumptions.

References

- Spiegelhalter DJ, Freedman LS, Blackburn PR (1986). PR Monitoring clinical trials: conditional or predictive power? Cont Clin Trials 1986; 7: 8-17.
- Chuang-Stein C (2006). Sample Size and the Probability of a Successful Trial. Pharmaceutical Statistics, 5, 305–309.
- Lan K, Hu P, Proschan M (2009). A conditional power approach to the evaluation of predictive power. Statistics in Biopharmaceutical Research, 1, 131-136.
- Berry DA (1989). Monitoring accumulating data in a clinical trial. Biometrics, 45: 1197–211.
- Jennison C and Turnbull BW (2000). Group Sequential Methods with Applications to Clinical Trials, New York: Chapman & Hall.
- Dmitrienko A, Wang MD (2006). Bayesian predictive approach to interim monitoring in clinical trials. Statistics in Medicine, 25(13):2178-95.
- Stallard N, Whitehead J, Cleall S (2005). Decision-making in a phase II clinical trial: A new approach combining Bayesian and frequentist concepts. Pharmaceutical Statistics, 4: 119–128.
- Chen C, Beckman RA (2009). Optimal cost-effective designs of proof of concept trials and associated go-no go decisions. J Biopharm Stat, 19:424–36.
- Chuang-Stein C, Kirby S, French J, *et al.* (2011). A quantitative approach to go/no-go decisions in drug development. Drug Information Journal. 45: 187–202.
- Liu F (2010). An extension of Bayesian expected power and its application in decision making. Journal of Biopharmaceutical Statistics, 20:941-953.
- Jiang K (2011). Optimal sample sizes and Go/No-Go decisions for phase II/III development programs based on probability of success. Statistics in Biopharmaceutical Research, 3: 463-475.

MERCK