Contributions of CDP to Methods for Monitoring RCTs

SCT May 20, 2019

David DeMets, PhD
University of Wisconsin-Madison
demets@biostat.wisc.edu

Declaration of Helsinki

- Original Declaration of Helsinki (1964)
- Ethical imperative included
  - *'Measures to minimise the risks must be implemented. The risks must be continuously **monitored**, assessed and documented by the researcher.'*
  - *'When the risks are found to outweigh the potential benefits or when there is conclusive proof of definitive outcomes, physicians must assess whether to continue, modify or immediately stop the study'*

1974 Belmont Report

- Primary purpose to protect subjects in clinical research
- Commissioned by the US Government in response to ethical failures in medical research, such as the Tuskegee Syphilis Study
- Three basic principles
  - Respect
  - Beneficence
  - Justice
- CDP already underway

"GREENBERG REPORT"

- National Heart Institute constituted a committee charged with developing recommendations regarding conduct of clinical trials
- Committee chaired by Bernard Greenberg, Chair of Biostatistics at UNC
- Report issued in 1967; published in 1988 in *Controlled Clinical Trials; applied to CDP*
- Report included recommendations on trial monitoring
  - "A mechanism must be developed for early termination"
  - only with the advice and on the recommendation of independent **consultants**.

Greenberg Report Recommendations

- Develop a mechanism to terminate early if
  - Question for benefit or harm already answered
    - **Serious Toxicity**
    - **Established Benefit Convincingly**
  - Trial can't achieve its goals - **futility**
  - Unusual circumstances
    - **Logistical / design failures**
  - Hypothesis no longer relevant
- Otherwise DSMB Recommendations
  - **Continue Protocol Unmodified**

DMC Considerations for Early Termination
(Canner, CDP, 1981,CCT)

1. Comparability
2. Bias
3. Compliance
4. Main effect vs. Potential side effects
5. Internal Consistency
    a. Outcome measures
    b. Subgroups
    c. Centers
6. External Consistency
7. Impact: current vs future patients
8. Statistical Issues

Statistical Challenges in Data Monitoring

- How does a DMC decide if emerging data is convincing for
  - Benefit or harm?
  - Futility; that is, little to no chance of demonstrating benefit
- Statistical methods exist to help
  - Guides, not rules
  - Still require clinical & statistical experts
  - Statistical methods tools to help us think but not an excuse to not think

Coronary Drug Project (CDP)

- Randomized double blind multicenter trial (53 centers)
- Over 8,000 patients with a recent myocardial infection
- Five treatment groups vs. a placebo group (2 levels of estrogen, clofibrate, dextrothyroxine, nicotinic acid)
- CDP became a prototype for many subsequent NIH trials
- Data Monitoring Committee Decisions to Recommend

  1. Stop high dose (5 mg) estrogen for increased CV events
  2. Stop dextrothyroxine (DTH) for increased CV events
  3. Stop low dose (2.5 mg) estrogen for increased CV events
  4. Continue clofibrate despite a nominally significant early trend

- Details described by Canner & CDP Group (1981) *Controlled Clinical Trials*

DMC Interim Analysis Challenge

- **Ethical, scientific and financial reasons for monitoring**
- **However, repeated analysis of accumulating data causes a statistical problem**
- **Problem similar to multiple testing of several variables**
- **That is, comparisons may be significant by chance alone at a greater frequency than the standard false positive error rate**

Three Procedures for Conservative Interim Monitoring

A. Group Sequential

   A modification of classical sequential

B. Curtailed Sampling/Conditional Power

C. Bayesian Relative Betting Odds (RBO)

All THREE METHODS HAVE PRIMITIVE VERSIONS IN CORONARY DRUG PROJECT

Repeated Significance Testing

- **Repeated testing increases Type I error or false postive conclusions AMR (*JRSS*, 1969)**
- **Example of using pvalue of 0.05 (two sided) at each analysis:**

  Critical Value = 1.96

  Reject H_0 if | Z| > 1.96

No. Of Looks (Planned)	Type I Error
1	0.05
2	0.08
...	
5	0.14
...	
10	0.20

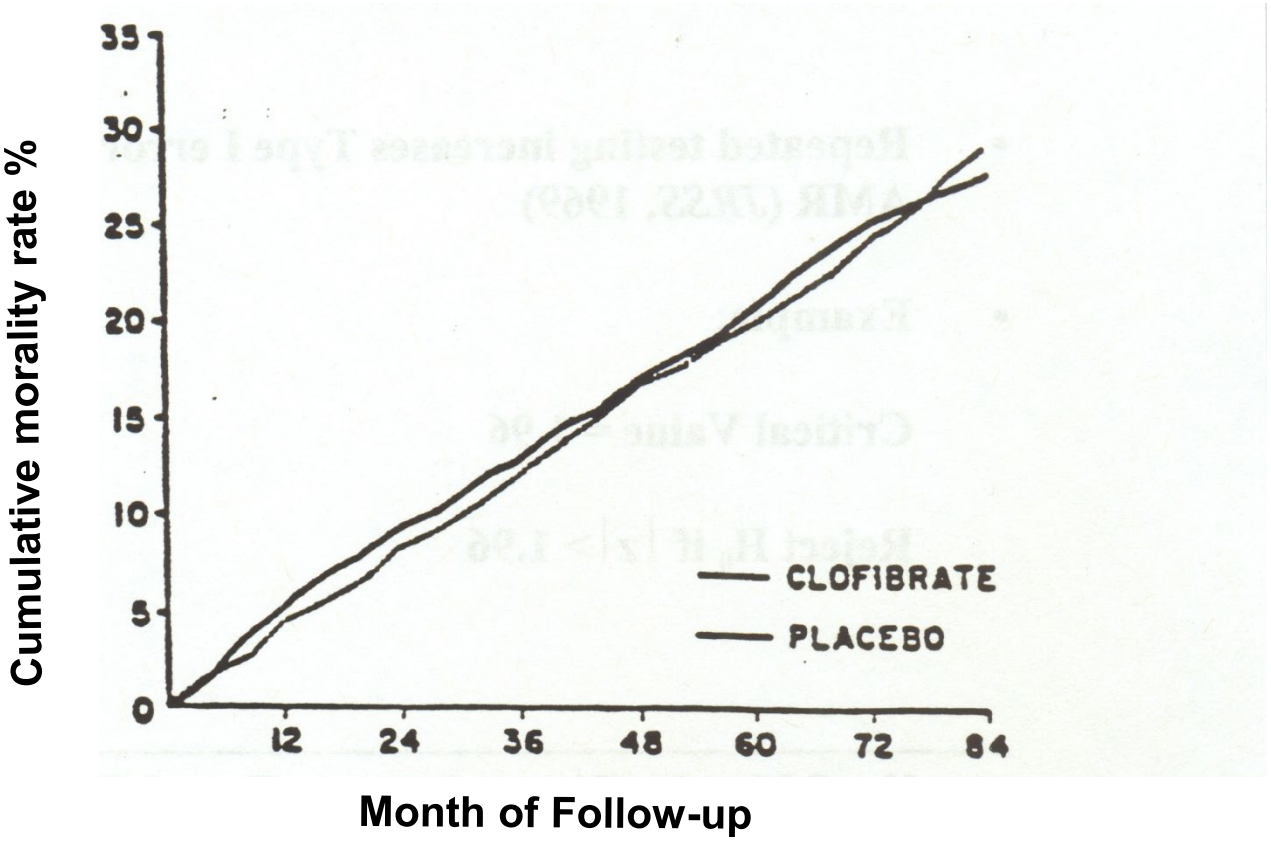
- **Must adjust interpretation of z to be conservative.**

A: CDP Sequential Monitoring

- CDP proposed a very conservative “rule” for interim results:
  - Canner, (1977) Biometrics
  - Canner, (1981) CCT
  - Canner, (1983) CCT
- Simulation Program (2000 reps)
  - 100 months of followup, 50 looks
  - 5 simultaneous comparisons of mortality
  - Overall 0.05 two sided alpha
  - Interim Upper Z value for harm 3.06
  - Interim Lower Z value for benefit, -2.79
- Note the asymmetry
- “Pocock Like” constant parallel bounds

Coronary Drug Project (CDP)

(Canner, 1981, CCT)

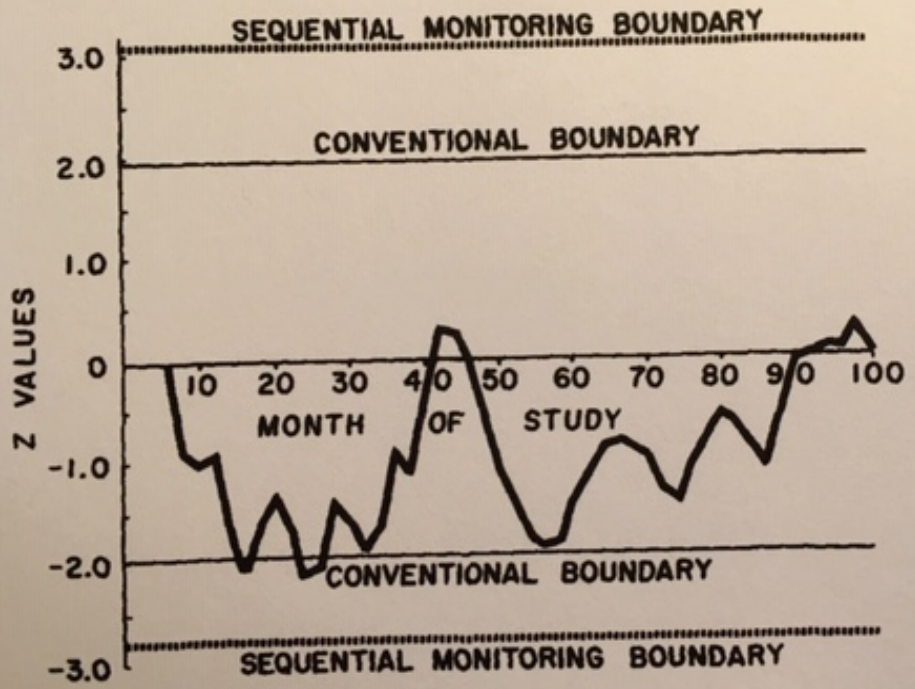


Life-table cumulative mortality rates, Coronary Drug Research Project Group

CDP Sequential Boundaries

Canner, 1983 CCT

Figure 1 Z values for clofibrate–placebo differences in proportion of deaths by calendar month since beginning of study (Month 0 = March 1966, Month 100 = July 1974), CDP.



CDP Sequential Boundaries
Canner 1983

- Canner discussed possible problem with CDP boundaries
  - If final z value were -2.5, for example, not significant by the boundaries but still nominal p value <<0.05
- Suggested considering
  - sloping boundaries
  - Very wide interim boundaries with conventional critical value at the end

B.1: Group Sequential Boundaries
Basic Idea

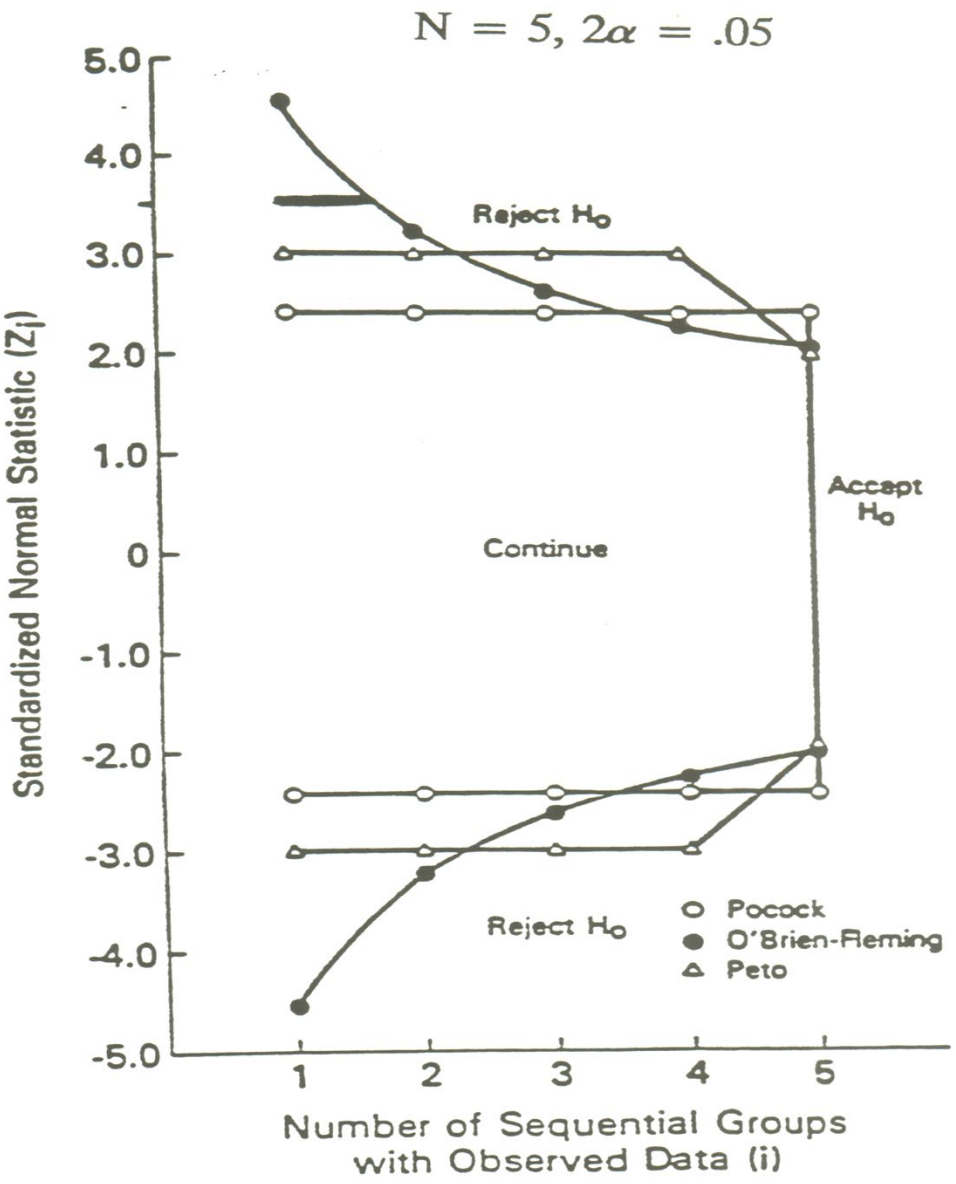
- **Compute summary statistic at each interim analyses, based on additional group of new subjects (events)**

- **Compare statistic to a conservative critical value $2\alpha = 0.05$ overall**

Various Methods

- Haybittle-Peto (1971, 1976)
- Pocock (1977)
- O'Brien-Fleming (1979)
- Slud and Wei (1982)
- Lan and DeMets (1983)

Three Common Group Sequential Boundaries

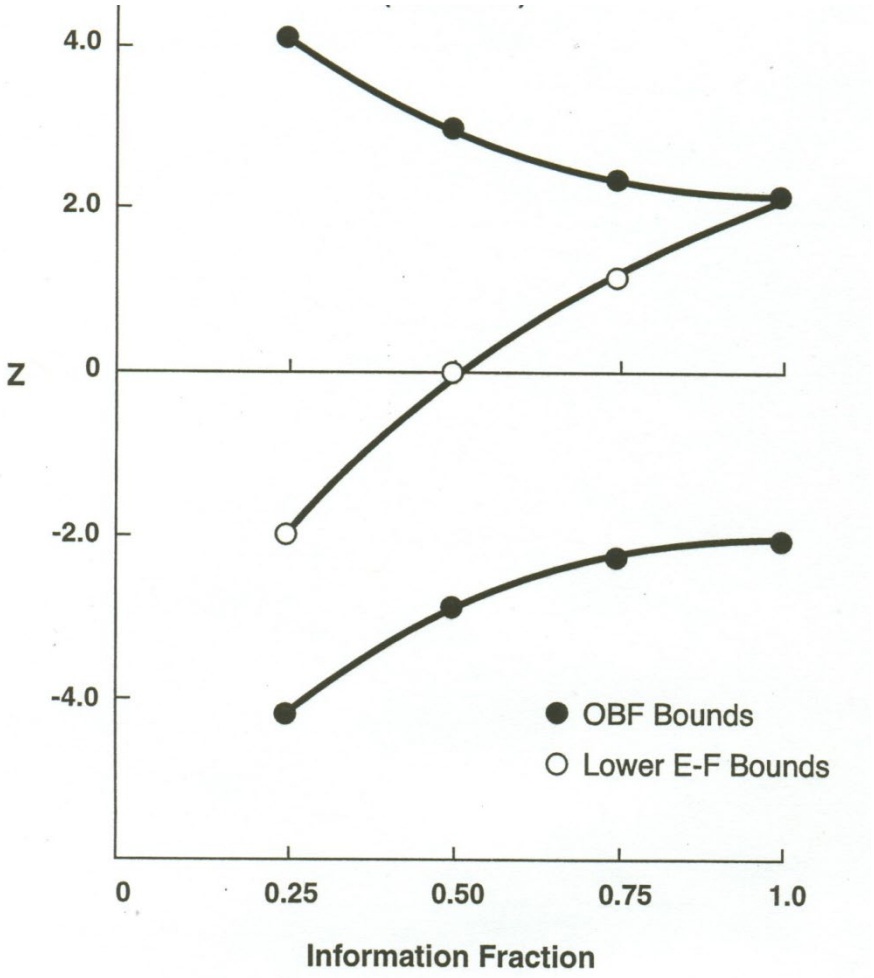


B2: Methods for Assessing Negative Trends

- "Demonstrating" Harm
  - An asymmetric sequential boundary for new intervention; not as demanding as for benefit
  - Symmetric boundaries such as group sequential boundaries for established interventions, which is better?
- Futility: little chance to demonstrate an effect
  - Beta Spending Functions (Emerson-Fleming, 1989)
  - Stochastic Curtailment (Halperin & Ware, 1974)
  - Conditional Power (Lan and Wittes, *Biometrics*, 1988).
  - Sequential Probability Ratio Test (SPRT) /Triangular Boundaries (Likelihood)
  - Predictive Power (Bayesian)

Group Sequential Boundaries For Negative Trends (E&F, 1989)

Upper bound for benefit 1α = 0.025



B.1: Curtailment

- Group sequential methods assess the current data without reference to future data
- Curtailment assesses the current trend with reference to potential future treatment effect – not possible to have a significant beneficial effect
- Ref: Halperin & Ware, 1974
- Very Conservative Method

B.2: Conditional Power

- Conditional power or stochastic curtailment assesses the current trend with reference to potential future treatment effect – probability of a trend reversal
- In the CDP, Canner via simulations calculated the chance of recovering from a null or negative trend to have a statistically significant result at the end

Conditional Power

- Also called “Stochastic Curtailed Sampling”
- Likelihood of a Trend Reversal
- Canner CDP Calculations
- Lan, Simon & Halperin.

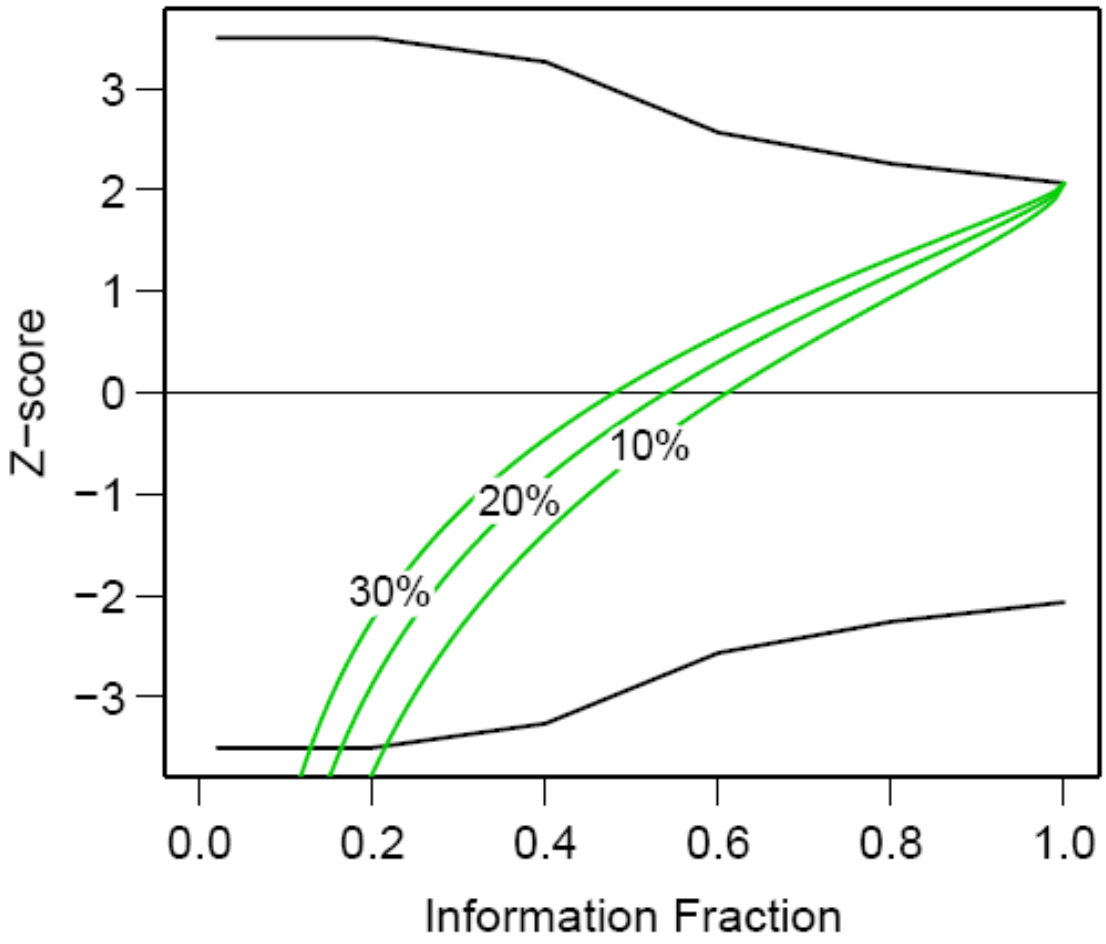
  Communications in Statistics-C 1:207-219, 1982
- Lan and Wittes, *Biometrics*, 1988.

Increase in but a Bound for Type I & Type II Error

$$\frac{\alpha}{\gamma_0} \qquad \frac{\beta}{\gamma_A}$$

Conditional Power Boundaries

(DeMets, CT, 2006)



C: Bayesian Methods for Interim Monitoring

- Cornfield began to think about Bayesian methods for RCTs, including interim analyses for the CDP
  - JASA, 1966 & Am J Epi, 1976
- Developed the concept of Relative Betting Odds (RBOs) Biometrics, 1969
- The odds in favor of the null hypothesis of no drug-placebo difference relative to a specified set of alternatives; that is,
- Ratio of likelihood of observed data under the null/ weighted mean of likelihood of observed data under a range of alternative hypotheses

Cornfield RBO's

- Used during review of 5mg estrogen arm: CDP (1970) JAMA "Modifications"
- Method "ahead of its time"
  - Except for Cornfield, others had trouble understanding how to interpret the RBO
- In two arm, treatment vs plbo, RBO more conservative than repeated testing
- Now, many Bayesian based methods in use – an extensive literature
  - Eg: Predictive power

Summary

- CDP researchers initially had few methods available for interim analyses
- Had to develop methods
- Relied on a number of statistical perspectives to develop them
- Had to rely on them to terminate 3 of the 5 arms
- Set a statistical monitoring framework for the rest of us to follow