The Use of Historical Information in Clinical Trials

Berry Consultants
Statistical Innovation

Kert Viele
PSI Journal Club
July 17, 2014

Acknowledgements

- DIA BSWG coauthors on manuscript (appeared in Pharmaceutical Statistics)
  - Scott Berry, Beat Neuenschwander, Billy Amzal, Fang Chen, Nathan Enas, Brian Hobbs, Joe Ibrahim, Nelson Kinnersley, Stacy Lindborg, Sandrine Micallef, Satrajit Roychodhury, Laura Thompson
- Other helpful discussions
  - Jeff Wetherington, Linda Mundy

Introduction

- Many trials compare a novel treatment to a control arm.
- Control rarely exists in vacuum
  - many studies on control effectiveness
  - trials used for approval
  - trials post approval
  - etc.
- Prior to the study, we often believe we have a "good idea" of the control arm parameters.
- Can we use this information?
  - Can have several other talks on selecting appropriate historical studies…

Simple Example

- Dichotomous endpoint
- Current trial has 200 subjects on each of control and treatment arms
  - $Y_C \sim Bin(200, p_C)$ $Y_T \sim Bin(200, p_T)$
- Historical data available with 65/100 responses. ($Y_H \sim Bin(100, p_H)$)
- Primary Analysis
  - $H_0 : p_C = p_T$ versus $H_1 : p_C < p_T$

Berry Consultants
Statistical Innovation

Basic idea of historical borrowing

- Combine information from current and historical study
  - typically through informative prior, frequentist methods also possible
- Can be good, can be bad
  - if historical information is close to true current parameters, inferences much improved
  - sample size savings of 20% or more possible
  - if historical information is far from true current parameters, bias can occur
  - discrepancy between history/current we call drift
  - unfortunately, drift not known in advance

Berry Consultants
Statistical Innovation

Downweighting/Power priors

- Weight historical data relative to current
- Each historical subject "counts" as W current subjects
  - W=0 ignores historical data
  - W=1 corresponds to pooling
  - W<1 "downweights" historical subjects
  - W>1 "overweights" historical subjects (rarely done)
  - W=infinity is a single arm trial…

Berry Consultants
Statistical Innovation

Fixed Weights

- Place noninformative priors on p_C and p_T.
- Use likelihood for control arm of
  - $[p_C^{65} (1-p_C)^{35}]^W \quad [p_C^{YC} (1-p_C)^{(200-YC)}]$
  - W = weight of historical data
- Example W=0.2, the 65/100 acts like 13/20
- Consider W=0.0, 0.2, 0.4, 0.6, 0.8, 1.0
- Borrow "equivalent" of 100W subjects

Berry Consultants
Statistical Innovation

Fixed Weights

- Posterior mean for p_C is
  - $g(Y_C) = (65W+Y_C) / (100W + 200)$

$$= \left[\frac{100W}{100W+200}\right]\frac{65}{100} + \left[\frac{200}{100W+200}\right]\frac{Y_C}{200}$$

- Suppose you observed 146/200=73% responses on the control arm.
  - With W=0, posterior mean is 73%
  - With W=0.2, posterior mean is 72.27%
  - With W=1, posterior mean is 70.33% (pooled)
  - With W=100, posterior mean is 65.16%

Berry Consultants
Statistical Innovation

Fixed Weights
Single arm trials

- Note for large W
  - heavily weighted historical data acts as point prior at 0.65
  - current control data ignored
- Might as well not run control arm, save 200 subjects.
  - this becomes a single arm trial using OPC.
- So single arm trials act as the most extreme form of historical borrowing.

Operating Characteristics

- Obtain the posterior distribution of p_C.
- Posterior mean $g(Y_C)$ on previous slide
  - compute MSE of point estimate
  - $MSE(p_C) = E[(g(Y_C) - p_C)^2 \mid p_C]$
- Also perform hypothesis testing
  - Reject H_0 if $Pr(p_T > p_C) > 0.975$
  - compute type I error $Pr(\text{reject} \mid p_T = p_C)$
  - compute power $Pr(\text{reject} \mid p_T = p_C + 0.12)$
  - power and type I error also a function of p_C

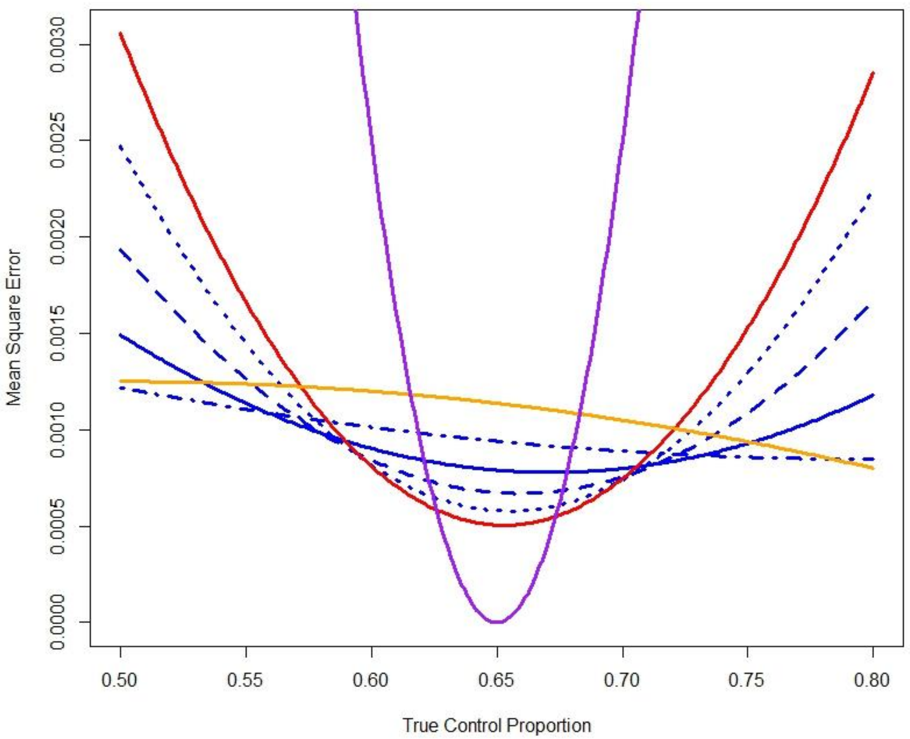
Berry Consultants
Statistical Innovation

Fixed Weights
(MSE as function of drift)

W=0 orange
W=1 red

single arm trial in purple

other weights in blue



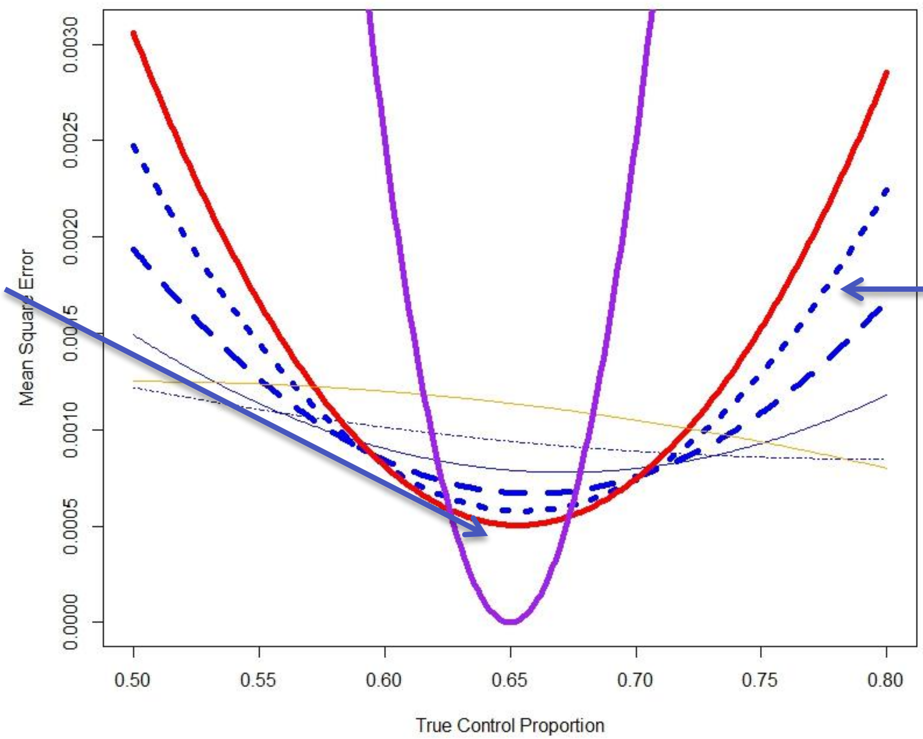
X-axis is p_C (current control Parameter)

Y-axis is MSE

If we knew drift, could select an ideal weight (of course we don't)

Fixed High Weights (MSE)

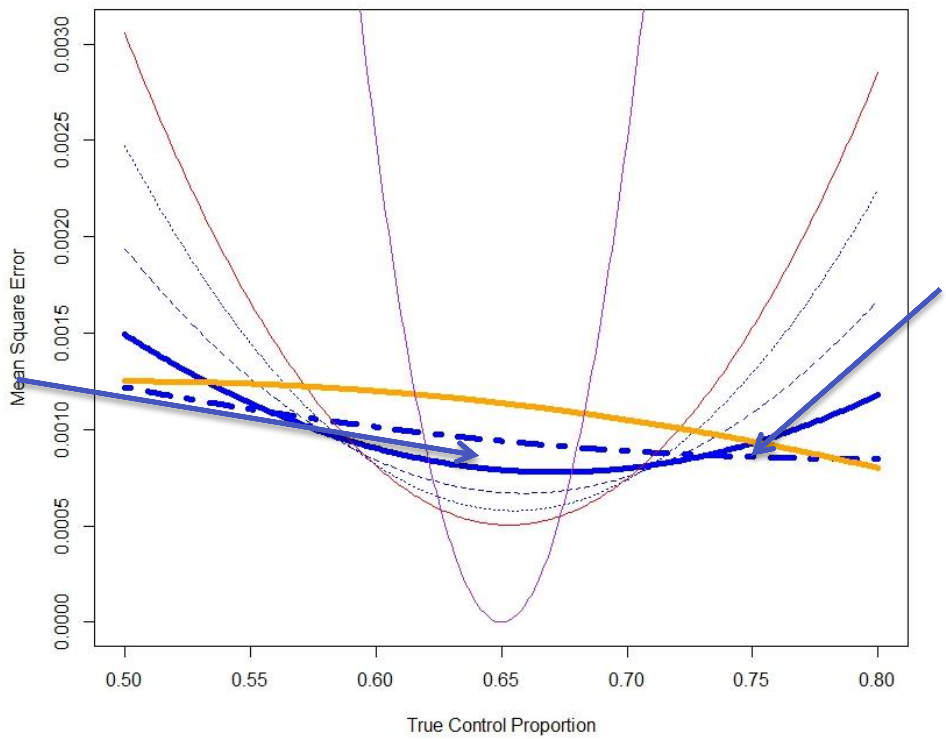
High Weight and No drift provides dramatic gains.



High drift and high weight produce biases and poor MSE

Fixed Low Weights (MSE)

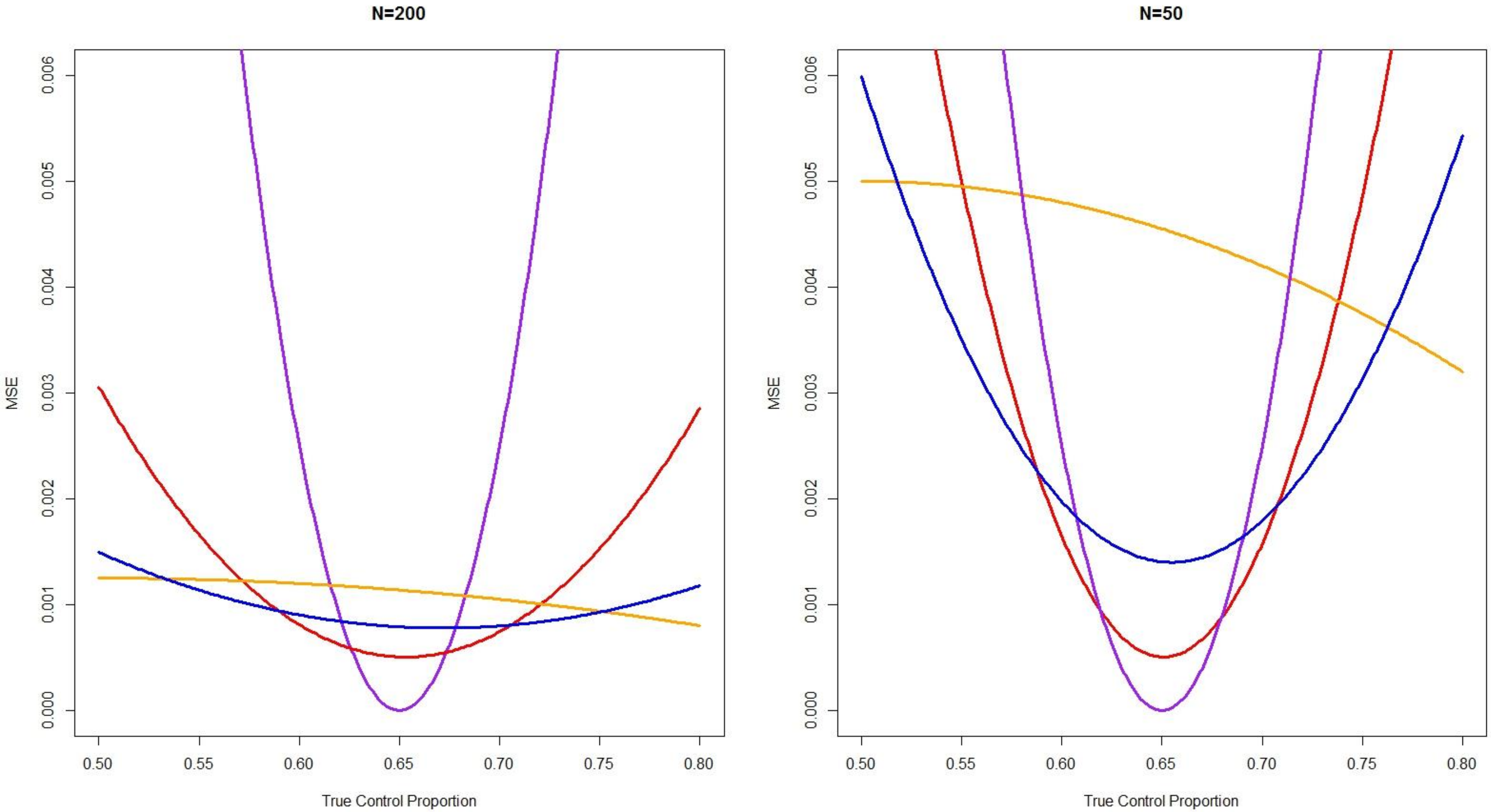
Low weights and low drift produce more modest gains (compared to ignoring history)



Low weight and High drift produce gains over broader area

More potential benefit with smaller sample sizes

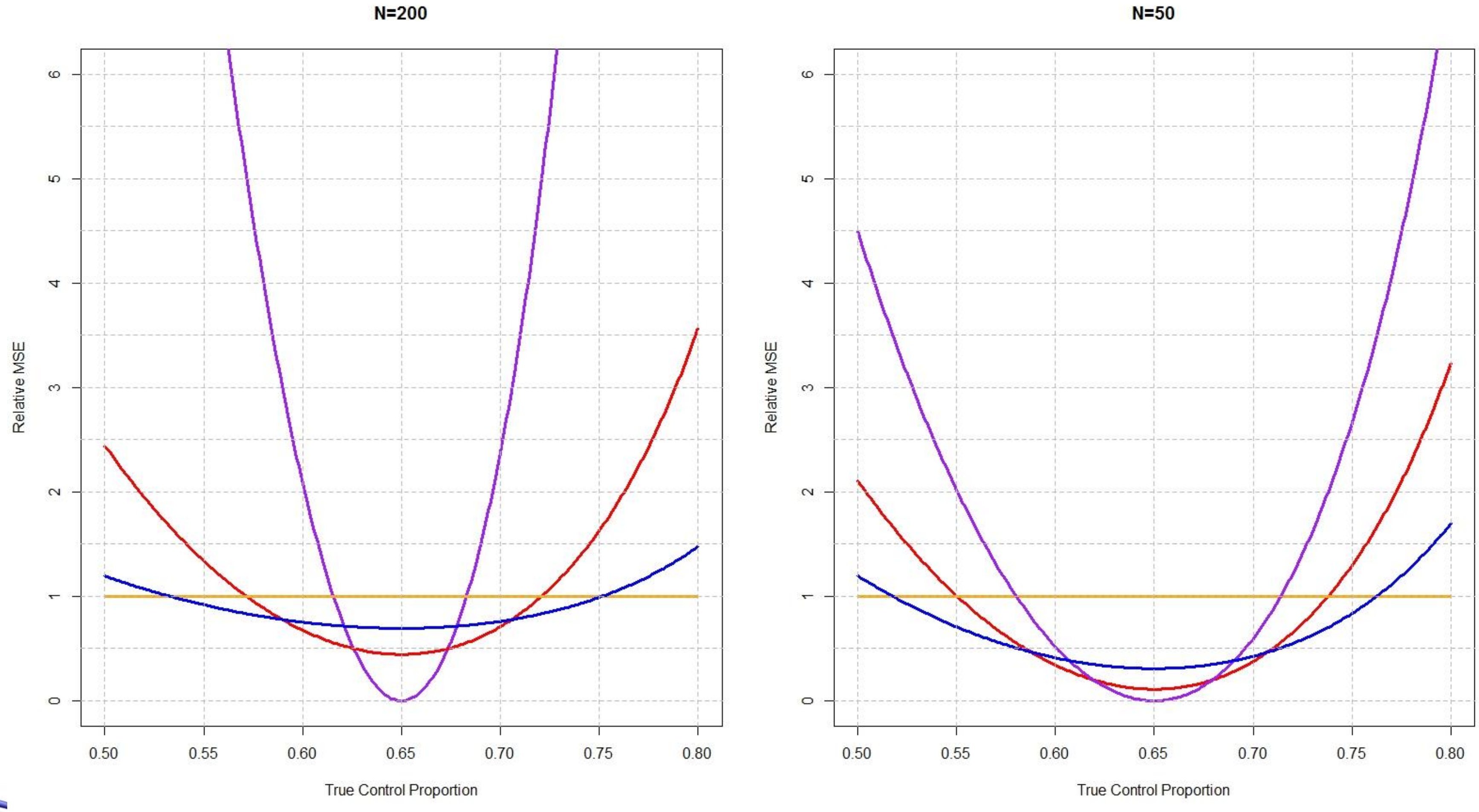
purple = single arm, red=pooled, blue=downweight 0.4, orange=ignore history



Berry Consultants
Statistical Innovation

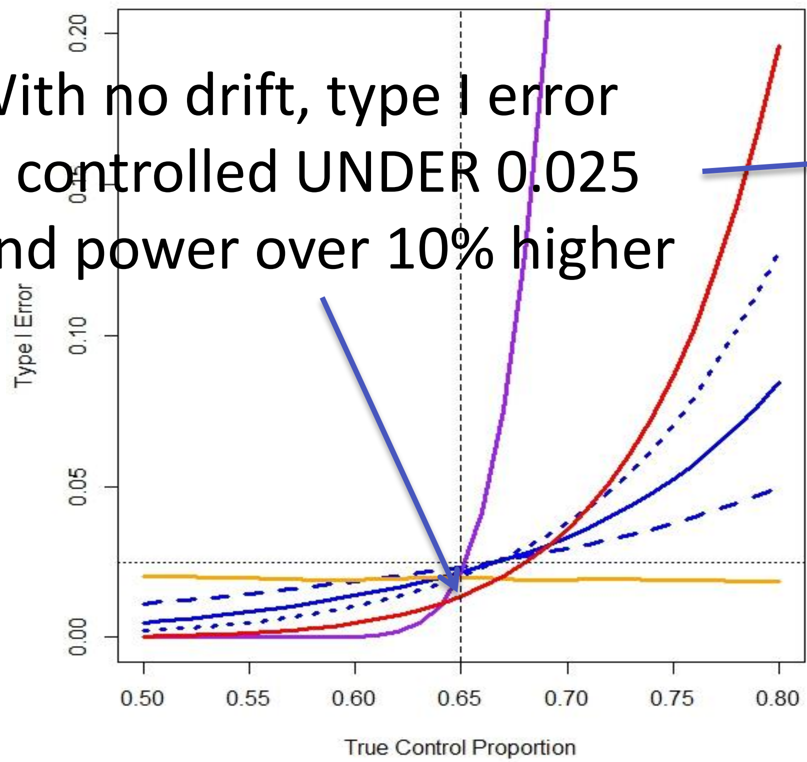
Different sample sizes MSE relative to ignoring history

purple = single arm, red=pooled, blue=downweight 0.4, orange=ignore history



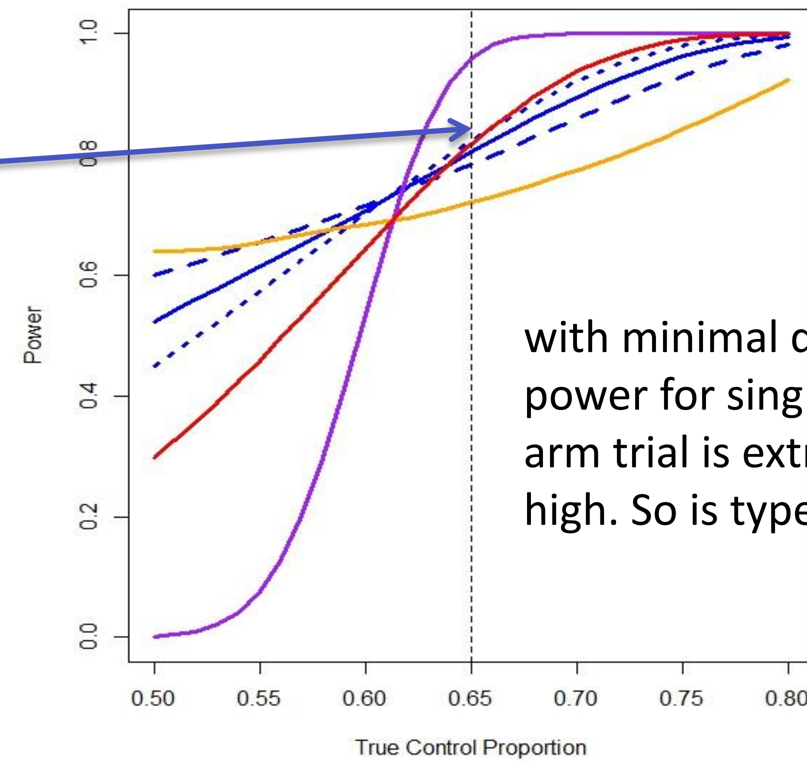
Fixed Weights (Testing)

Type I error



With no drift, type I error is controlled UNDER 0.025 and power over 10% higher

Power (for 0.12 gain)

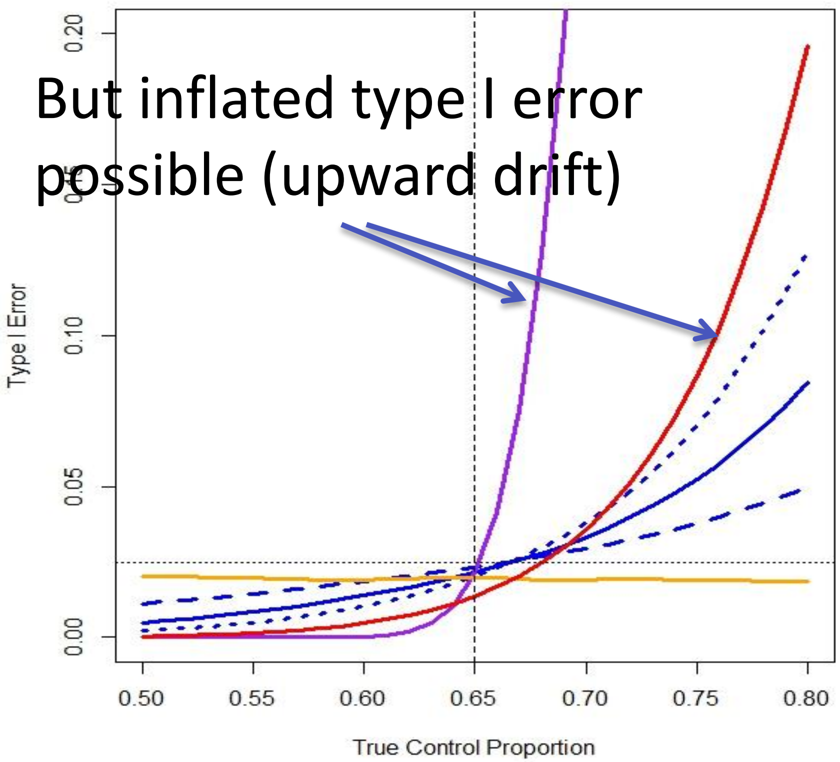


with minimal drift power for single arm trial is extremely high. So is type I error

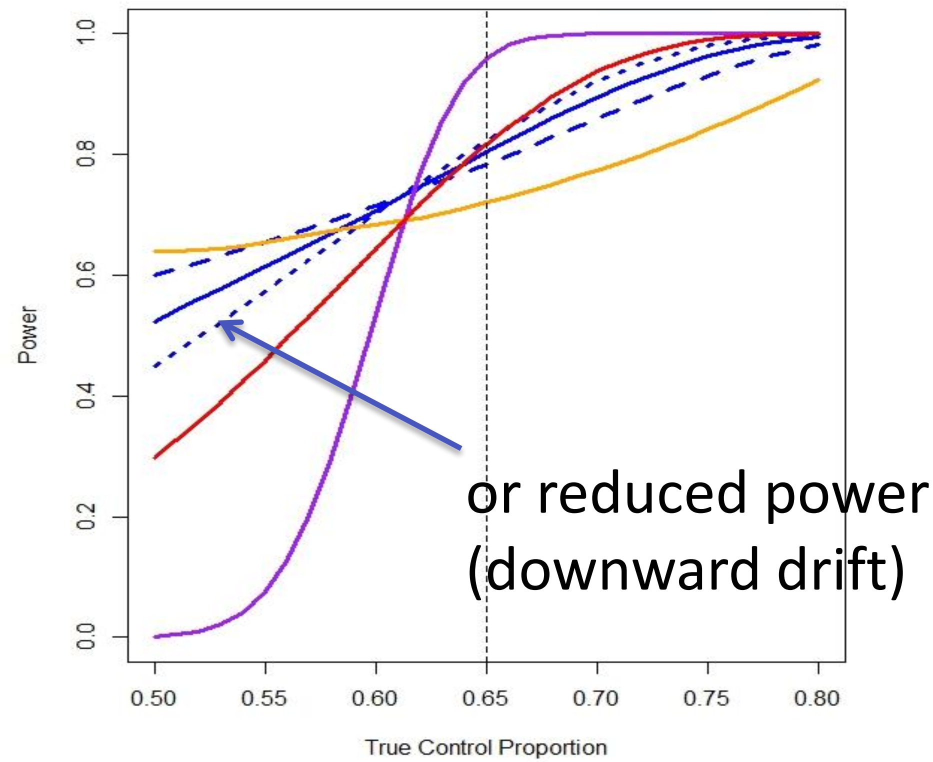
Berry Consultants
Statistical Innovation

Fixed Weights (Testing)

Type I error



Power (for 0.12 gain)



Dynamic Borrowing

- Desired weight depends on unknown drift
  - small drift = large weight
  - large drift = small weight
- The data itself provides information on drift

- Dynamic borrowing = amount of weight depends on agreement between historical and current data.

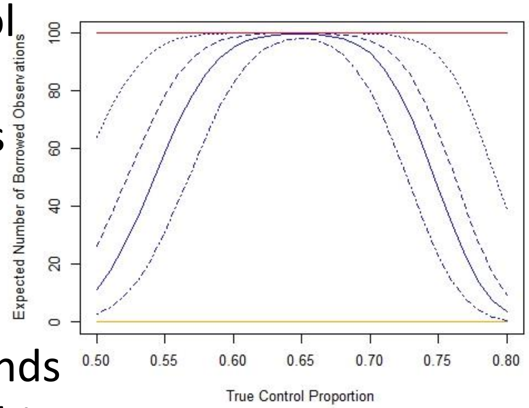
Test then pool

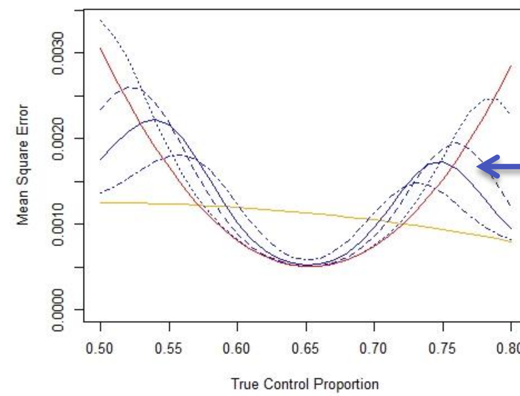
- Simple dynamic borrowing method
  - Perform hypothesis test of $H_0 : p_c = p_h$ versus $H_1 : p_c \neq p_h$ with size α (perhaps not 0.025)
  - If H_0 not rejected, pool (use W=1)
  - If H_0 rejected, ignore history (use W=0)
- Certainly "all or nothing" but incorporates current control data into decision on weight.

Test then pool

Red = always pool

Blue corresponds to α=0.20, 0.10, 0.05, 0.01

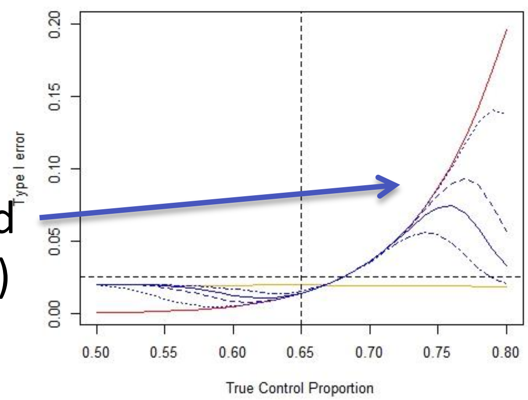


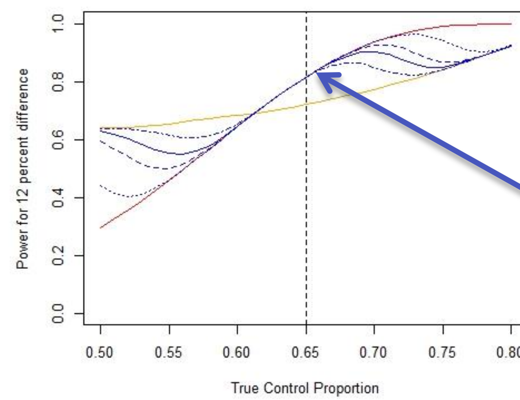


MSE increase bounded

Orange corresponds to always ignore history

Type 1 error inflation bounded (still can be large)





Retains power increase for minimal drift. Similar to ignoring history for large drift.

Berry Consultants
Statistical Innovation

Hierarchical Models (not the only way)

- In general, let p_C be current control rate
- $p_1, \ldots, p_H$ are true rates from historical studies
- $Y_0 \sim Bin(n_0, p_C)$ [current data]
- $Y_h \sim Bin(n_h, p_h)$ [historical data]
- $logit(p_C), \ldots, logit(p_H) \sim N(\mu, \tau)$
- $\mu \sim N(\mu_0, \tau_0)$, $\tau \sim \pi(\tau)$

Hierarchical Models

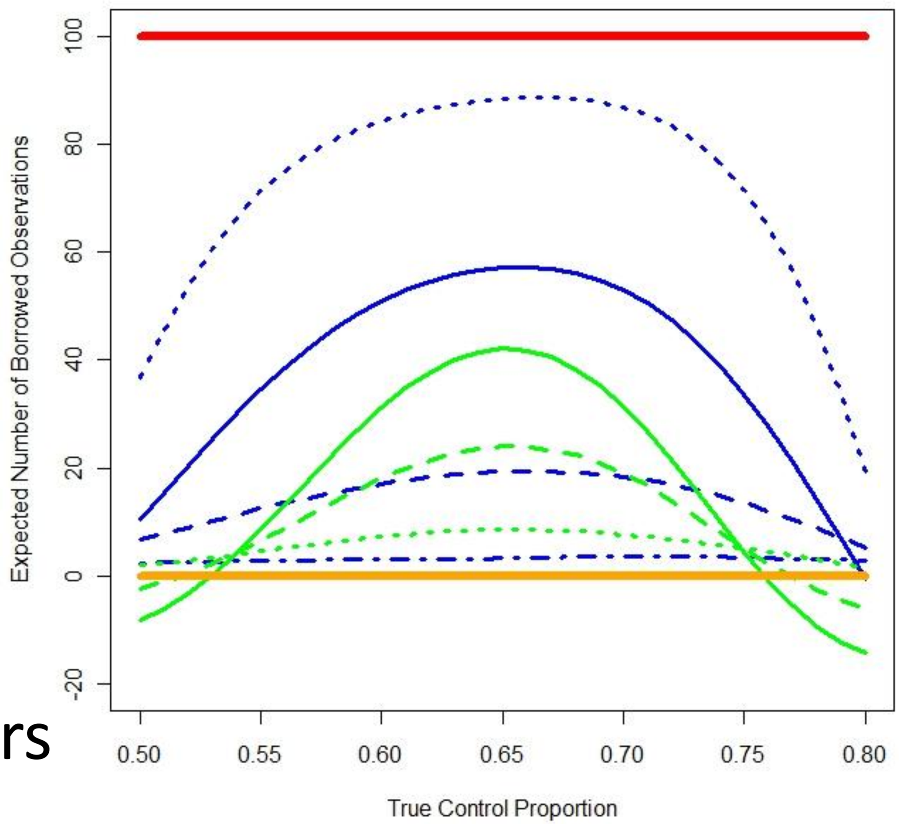
- $\text{logit}(p_C),\ldots,\text{logit}(p_H) \sim N(\mu,\tau)$
- τ measures across study variation
- A fixed τ corresponds to a specific weight
- We use an IGamma prior, here we obtained good operating characteristics.
  - other prior structures available
- Creates dynamic borrowing
  - generally lower τ when current data agrees with history, and thus higher weight
  - generally larger τ when current data disagrees with history, and thus lower weight.

Berry Consultants
Statistical Innovation

Hierarchical Models (expected borrowing behavior)

Y-axis shows expected number of borrowed subjects

Different curves are different priors

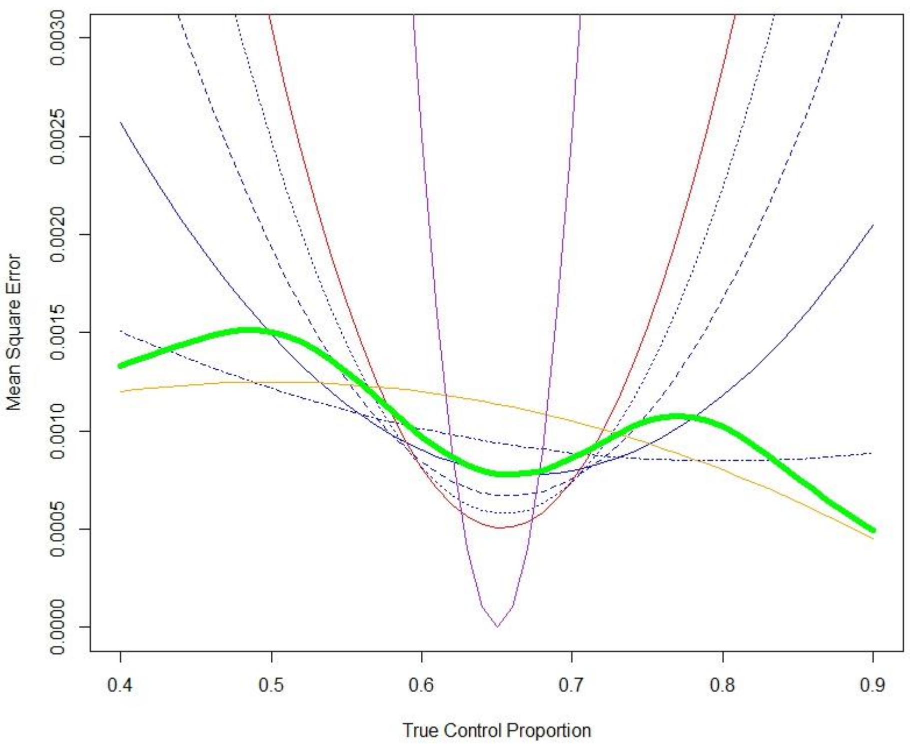


Dynamic Borrowing - E[borrow] greatest for low drift

Berry Consultants
Statistical Innovation

MSE for dynamic borrowing

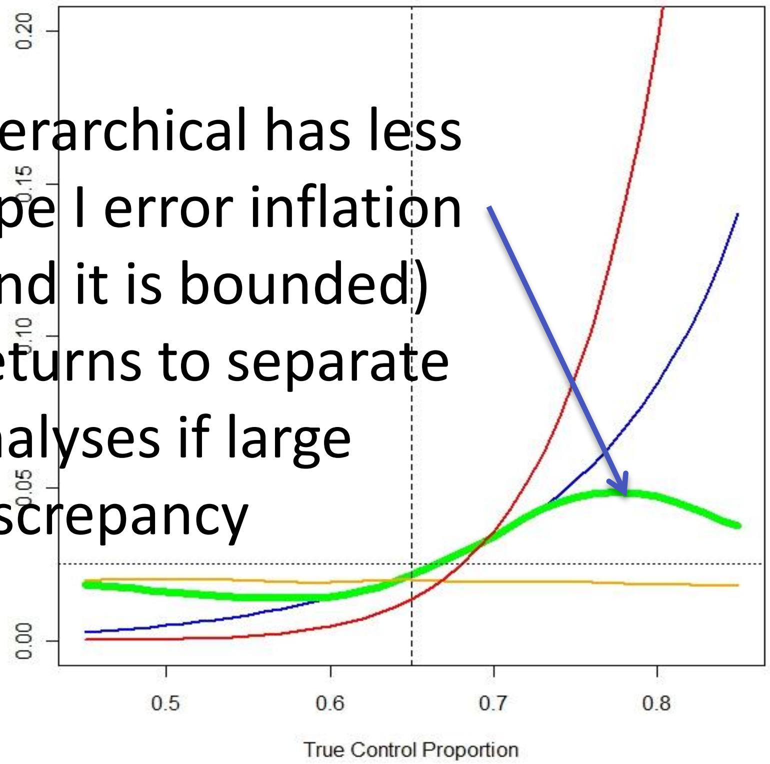
Green curve shows MSE for posterior mean using dynamic borrowing



Note inflation of MSE is bounded over ignoring history

Hierarchical Models

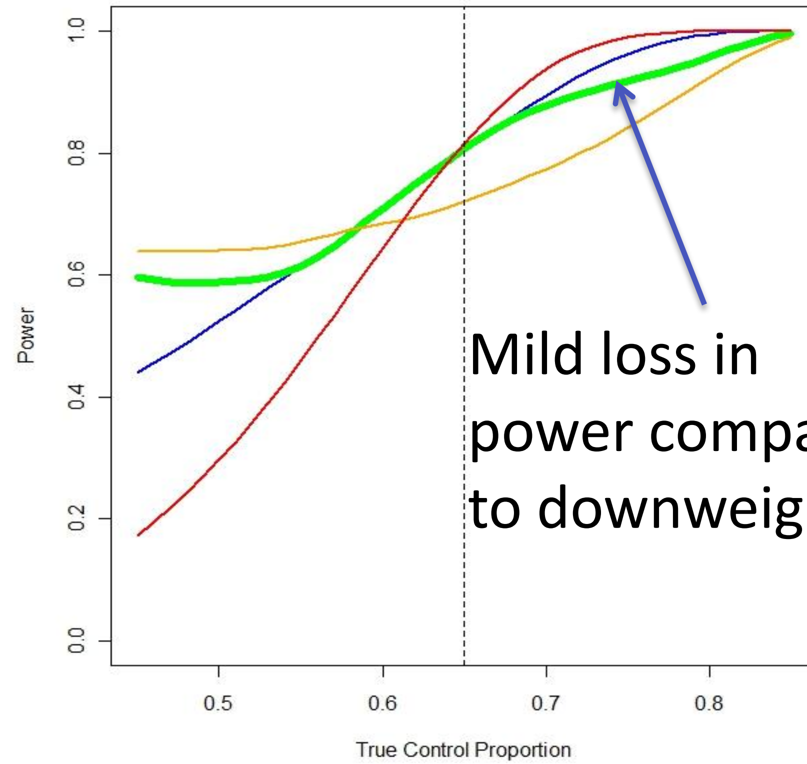
Type 1 Error



Hierarchical has less type I error inflation (and it is bounded)

Returns to separate analyses if large discrepancy

Power (0.12 gain)



Mild loss in power compared to downweighting

Berry Consultants
Statistical Innovation

Comments on type I error

- Usual definition of type I error conditions on historical data
  - $\alpha(p_C)$ = Pr(success | $p_C=p_T$, Y_H)
  - regardless of borrowing method type I error is inflated for SOME p_C (those with large drift)
- You could argue strictly this precludes historical borrowing.
  - However, this can lead to some unintuitive decisions.
  - In some situations we may be confident about likely range of drift (antibiotics)

Berry Consultants
Statistical Innovation

Summary

- Historical borrowing
  - may improve point estimates
  - may reduce type I error
  - may increase power
  - can result in substantial sample size savings
- There will be situations where historical borrowing is NOT beneficial
  - large expected drift, or high variation in drift