Application of Bayesian borrowing methods in clinical trials for children with T2DM using simulation and case studies

Roberto Crackel, PhD,
Senior Mathematical Statistician, FDA/CDER/OB/DBII
Yoonhee Kim, PhD,
Lead Mathematical Statistician, FDA/CDER/OB/DBII

Disclaimer

The views expressed herein are those of the author and should not be construed as FDA's views or policies

Outline

- Background and Motivation
  - Sample Size Challenges in Pediatric Studies
- Bayesian Borrowing
  - Bayesian model with mixture priors
  - Effective Sample Size
  - Simulation study
    - Operating Characteristics
- Case Example
- Summary

Background / Motivation

Pediatrics with Type 2 Diabetes

- Significant increases from 34 per 100,000 in 2001 to 67 per 100,000 in 2017 (by 95%) in the number of youth living with type 2 diabetes were observed in pediatrics aged 10-19 years old in the United States*.

- Sample sizes in T2D pediatric study were calculated using assumptions of treatment effect size (Δ) and the standard deviation (SD) derived from T2D adult studies given the power and type 1 error control rates

* https://www.cdc.gov/media/releases/2021/p0824-youth-diabetes.html

Comparison of Assumed SD and Observed SD in Completed T2D Pediatric Studies

Study Drug	Assumed SD and Planned Study Power[a]	Observed SD[b] and Retrospective Power[c]
Sitagliptin	SD = 1.1% Power=82%	SD = 1.6% Retrospective Power=51%
Sitagliptin and Metformin	SD = 1.1% Power=86%	SD = 1.4% Retrospective Power=67%
Colesevelam	SD = 1.0% Power=80%	SD = 1.5% Retrospective Power=46%
Liraglutide	SD = 1.2% Power=80%	SD = 1.8% Retrospective Power=46%
Extended release Exenatide	SD = 1.0%, Power=74%	SD = 1.5% Retrospective Power=41%
Dulaglutide	SD = 1.4% Power=87%	SD = 1.6% Retrospective Power=78%
Empagliflozin	SD = 0.9% Power=85%	SD = 1.7% Retrospective Power=36%
Linagliptin	SD = 0.9% Power=85%	SD = 1.7% Retrospective Power=36%
Dapagliflozin	SD[d] = 1.7% Power=80%	SD = 1.8% Retrospective Power=75%
Saxagliptin	SD[d] = 1.7% Power=80%	SD = 1.5% Retrospective Power=Not available

a Parameters for power calculations were obtained from protocols or statistical analysis plans available on clinicaltrials.gov or from journal publications

b Observed standard deviation for the 6-month change from baseline in HbA1c in the pediatric study. Back calculated from the confidence intervals in the product label or on clinicaltrials.gov or publicly available reviews

c Retrospective power based on observed SD and original assumption of the treatment effect given the same α and sample size

- *Sources: study protocols available on clinicaltrials.gov; US product inserts (USPI)*

d SD updated after blinded review of ongoing study data

Bayesian Borrowing Methods in T2D Pediatric Studies

- Alternative method as part of a pediatric extrapolation approach when justified to balance the need of larger study and study feasibility
  - Publicly available review for Belimumab* as a precedent of using Bayesian Borrowing methods in systemic lupus erythematosus pediatric studies

- The purpose of this research is to explore the Bayesian informative priors and its properties and to study operating characteristics using simulation data and to apply these methods to a case example

* https://www.fda.gov/media/127912/download

Bayesian Borrowing

Assumptions / Considerations

- An informative Bayesian prior in the analysis
  - Requires assumptions on the similarity between pathophysiology and the mechanisms of action of the study drugs for T2DM in adults and pediatrics
  - Assumes treatment effects are similar between adults and pediatrics

- When selecting historical adult studies
  - Similar population and background therapies

The Bayesian Model

$$\hat{\delta}_P \sim N(\delta_P, SE_P^2)$$

$$\delta_P \sim \omega \cdot N\left(\hat{\delta}_A, SE_A^2\right) + (1-\omega) \cdot N(0, \sigma_0^2) \quad \text{[Mixture Prior]}$$

- $\hat{\delta}_P$ is the observed pediatric treatment effect and SE_P is the standard error of $\hat{\delta}_P$
- ω is the prior weight to be given to the information from the adult study
- $N\left(\hat{\delta}_A, SE_A^2\right)$ is the adult component where $\hat{\delta}_A$ is the observed adult treatment effect and SE_A is the standard error of $\hat{\delta}_A$
- $N(0, \sigma_0^2)$ is the vague component;
  - Center the mean at 0 (i.e., no pediatric treatment effect)
  - For σ_0^2, we consider:
    - Variance of the pediatric treatment effect based on 1 pediatric patient per treatment arm ($UISD_P^2$ ≣ Unit information variance)
    - In a 1:1 randomization ratio, then $\text{UISD}_\text{P} = \sqrt{SD_P^2 + SD_P^2} = \sqrt{2} * SD_P = \sqrt{n} * SE_P$
      - n is the common sample size per arm in the pediatric study
      - SD_P is the observed patient level pediatric standard deviation
    - For general a:b randomization ratio (Reference: Zhang, et al, Prior Effective Sample Size When Borrowing on the Treatment Effect Scale (2024))

Effective Sample Size (ESS)

- Effective Sample Size (ESS) tells us how much the prior information is worth in terms of sample size relative to a unit of information (e.g., 1 patient per arm)
- Expected local-information-ratio (ELIR) (Neuenschwander et al. 2020) :
  - The ELIR is defined as the mean of the prior information to the Fisher information ratio r(θ)

$$ESS_{ELIR(Prior)} = E_\theta\{r(\theta)\} = E_\theta\left\{\frac{i\big(p(\theta)\big)}{i_F(\theta)}\right\}$$

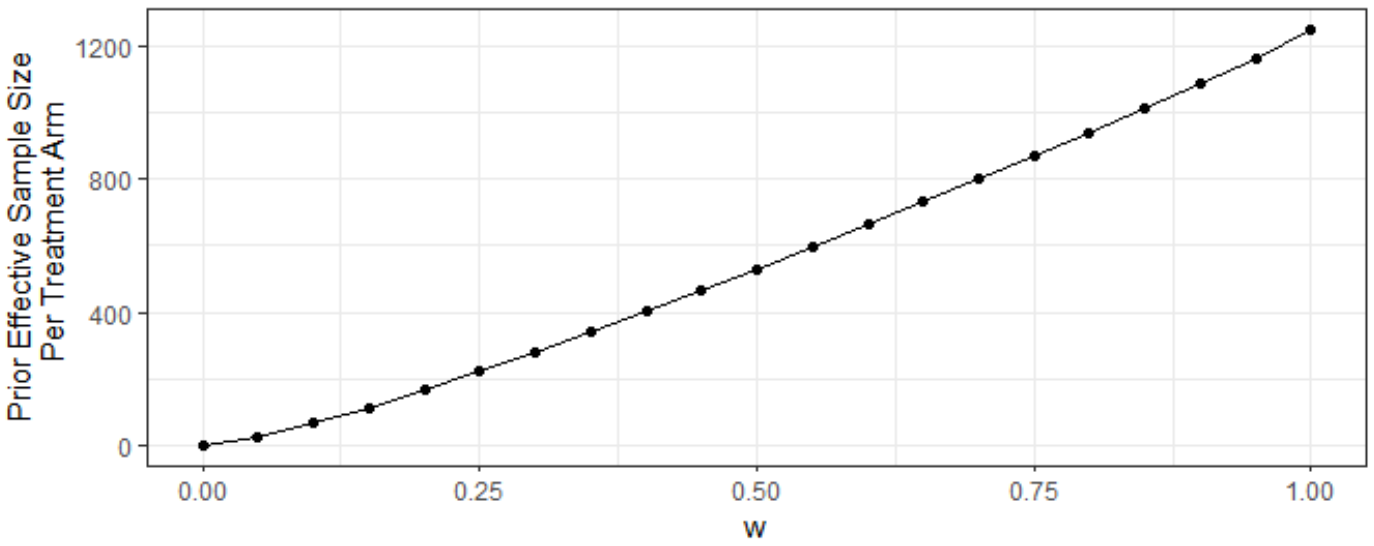
    - Predictive consistency
    - Reference: Neuenschwander, et al. (2020)
- A reference scale needs to be pre-specified which defines how many pediatric patients that $ESS_{ELIR(Prior)}$ is equivalent to
- Example: Under 1:1 randomization, if we take $UISD_P$ as the reference scale (1 patient per arm), then $ESS_{ELIR(Prior)}$= # of pairs of pediatric patients to be borrowed on the treatment and control arms
  - If $ESS_{ELIR(Prior)}$=50, the borrowed information equates to 50 pediatric patients added, in a pseudo manner, to **each** treatment and control arms

Simulation Study: Borrowing Decision

- **Objective:** Explore ESS and power for a "typical" T2DM pediatric studies
- **Example:** Currently enrolled:
  - N_P=120 (60 per treatment and control)
- **Assumptions for Sample Size Calculations**
  - N_P = 120; Δ_P = - 0.7% (assumed); SD_P = 1.5 (estimated)
  - Power = 72% (t-test)
- **Adult Information**
  - $\hat{\delta}_A$ = - 0.80%; N_A = 1000 (500 per arm); SE_A = 0.06
- **Mixture Prior**
  - $UISD_P=\sqrt{2} * 1.5 = 2.1213$, so $\sigma_0^2 = UISD_P^2 = 4.5$:

$$\omega * N(-0.8, 0.06^2) + (1-\omega) * N(0, 4.5)$$

Effective Sample Size



ω	$ESS_{ELIR(Prior)}$
0	1
0.05	23
0.10	65
0.20	167
0.30	281
0.40	404
0.50	531
0.60	664
0.70	800
0.80	942
0.90	1089
1	1250

Calculated using ess function in R with elir option

- Reference scale = $UISD_P = \sqrt{4.5} = 2.1213$
- ESS increases as the amount of borrowing increase
- 5% weight corresponds to adding 23 patients per arm
- ω=0 (no weight): $\left(\frac{2.1213}{2.1213}\right)^2 = 1$ pediatric patient per arm (Minimum)
- ω=1 (full weight): $\left(\frac{2.1213}{0.06}\right)^2 = 1250$ pediatric patients per arm (Maximum)

Results

- Δ_P = - 0.7%, SD_P=1.5, N_P=120 (60 each arm), and prior: $\omega * N(-0.8, 0.06^2) + (1-\omega) * N(0, 4.5)$
- Reference scale = $UISD_P$ = 2.1213

ω	$ESS_{ELIR(Prior)}$	Psuedo N=120 + 2*$ESS_{ELIR(Prior)}$	Trial or Conditional Power (conditioned on adult data)
0	1	122	72%
0.05	23	166	75%
0.09	56	232	77%
0.095	**60**	**240**	**77%**
0.15	113	346	79%
0.16	124	368	80%

- No borrowing results in 72% conditional power ≈ frequentist power
- 5% weight (i.e., borrowing 23 pediatric patients per arm) corresponds to 75% conditional power
- 9.5% weight (i.e., borrowing 60 patients per arm or the currently enrolled amount) corresponds to 77% conditional power
  - Reasonable threshold for maximum borrowing allowed to declare a study win

Simulation: Operating Characteristics

- Conditional power and type 1 error, Mean Squared Error (MSE), Bias, Half-Width of 95% Credible Interval
- Define:
  - Conditional power = $\mathrm{P}\left[\mathrm{P}\left[(\delta_{\mathrm{P}} < 0)\middle|\hat{\delta}_{\mathrm{P}}\right] > 0.975 \middle| \Delta_{\mathrm{P}} < 0\right]$
  - Conditional type 1 error = $\mathrm{P}\left[\mathrm{P}\left[(\delta_{\mathrm{P}} < 0)\middle|\hat{\delta}_{\mathrm{P}}\right] > 0.975 \middle| \Delta_{\mathrm{P}} = 0\right]$

  ** Conditioned on observed adult data, meeting the 97.5% criteria, and underlying assumed pediatric treatment effect
- <u>Adult data</u>:

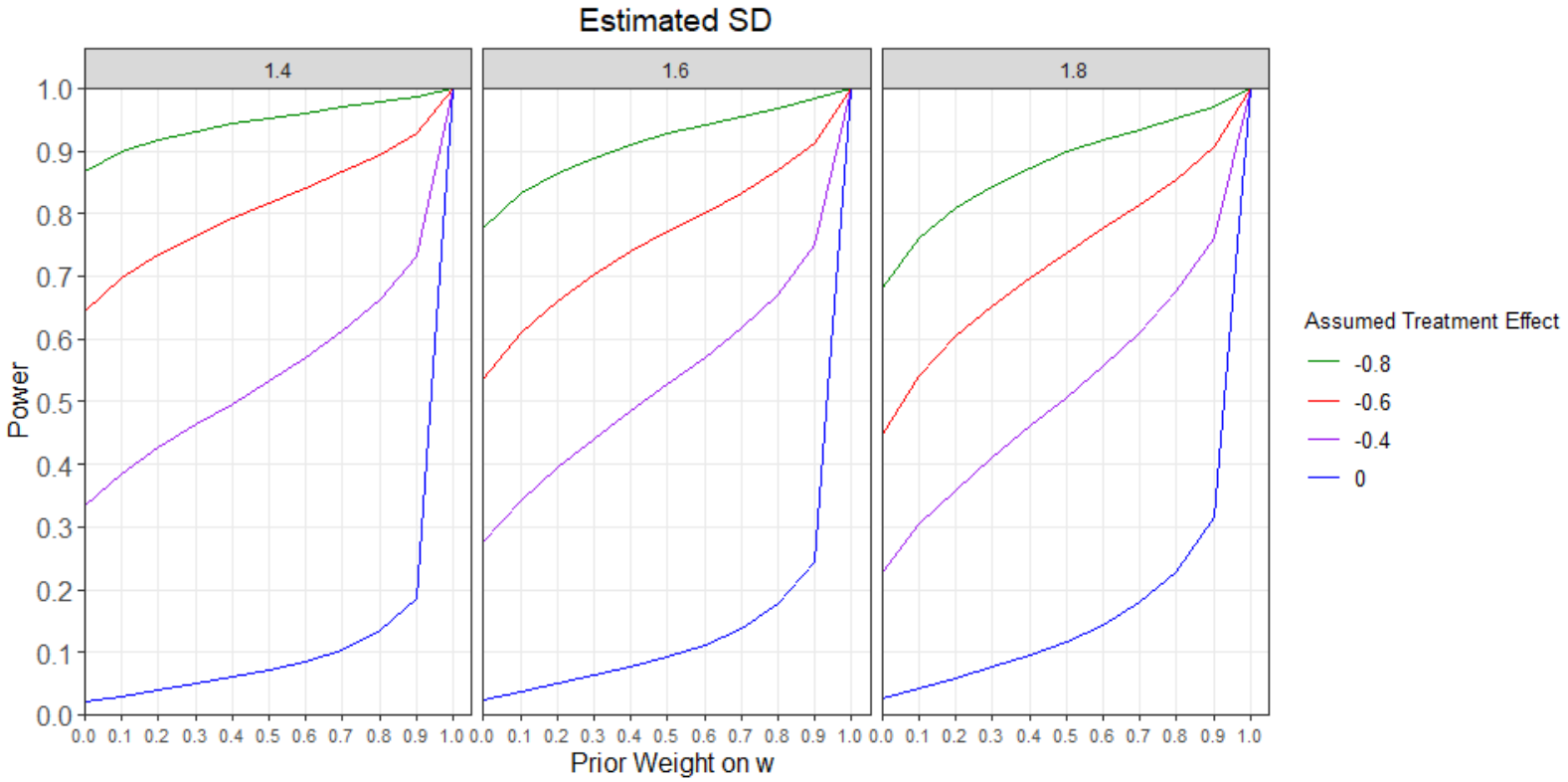
  $\hat{\delta}_A$ = - 0.80%, N_A=1000, SE_A=0.06
- <u>Mixture prior</u>:

  $\omega * N(-0.8, 0.06^2) + (1 - \omega) * N(0, UISD_P^2)$
- <u>Explore range</u>:

  Δ_P = 0, -0.4, -0.6, -0.8

  SD_P= 1.4, 1.6, 1.8
- 10,000 simulations per combination

Simulation: Operating Characteristics of Conditional Power



- Conditional power:
  - Largest when Δ_P = -0.8 = $\hat{\delta}_A$
  - Increases as borrowing increases
- Small Δ_P and large SD_P require more borrowing to achieve adequate conditional power
- For ω =0, the conditional power ≈ to the frequentist power of the stand- alone pediatric study
- For Δ_P = 0 and ω =0, the conditional type 1 error ≈ frequentist type 1 error = 0.05

Simulation: Operating Characteristics Findings

- When Δ_P = -0.8 = $\hat{\delta}_A$:
  - Little or no bias regardless of the amount of borrowing,
  - MSE and half-width of 95% Credible Interval is smallest compared to Δ_P = -0.6, -0.4, 0

- As the distance from Δ_P and $\hat{\delta}_A$ = -0.8 increases:
  - MSE and half-width of 95% Credible Interval increases

Case Example

Sitagliptin and Metformin

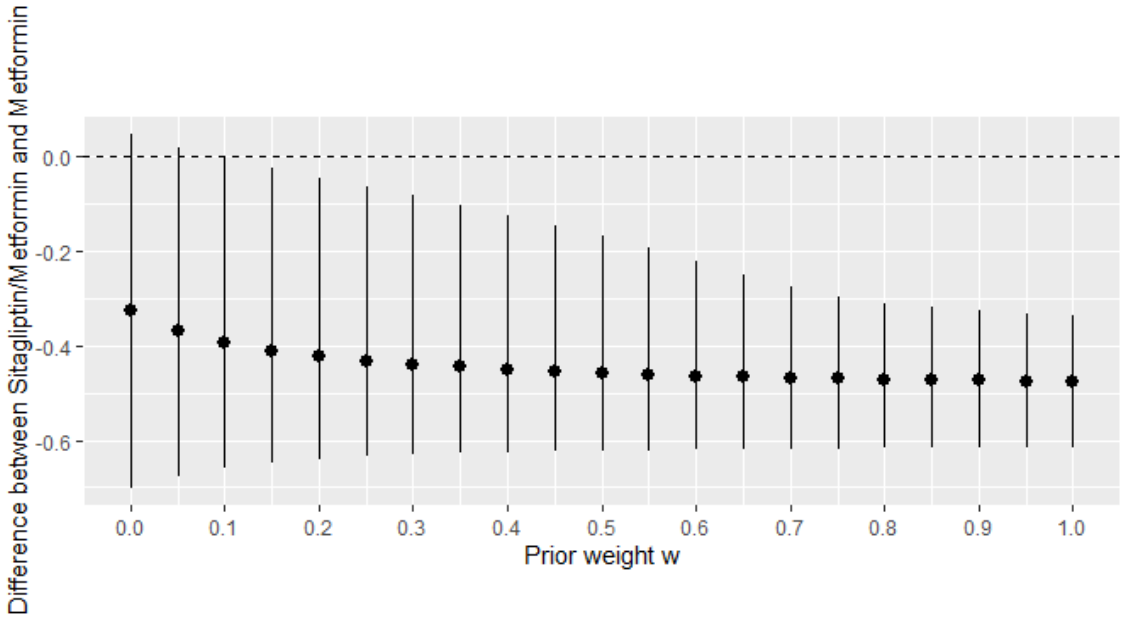
- Pediatric Information
  - N_P = 220 (107 to Sitagliptin and Metformin; 113 to Metformin)
  - $\hat{\delta}_P$ = - 0.33% and 95% CI: (-0.70, 0.05)
  - $SE_P \approx 0.1913$
  - SD_P = 1.4182

- Adult Information
  - N_A= 452 (223 to Sitagliptin and Metformin; 229 to Metformin)
  - $\hat{\delta}_A$ = - 0.50%
  - SE_A = 0.0765

- Mixture Prior: $\omega * N(-0.50, 0.0765^2) + (1-\omega) * N(0, 4.0226)$

$$\sigma_0^2 = UISD_P^2 = 2*1.4182^2 = 4.0226$$

Sitagliptin and Metformin Results: Unit Variance vs. Large Variance



Stand alone results:

$\hat{\delta}_P$ = - 0.33%

95% CI: (-0.70, 0.05)

- 10% borrowing required to exclude 0 from the 95% credible interval
  - ESS=31 (per arm) [Reference scale = $\sqrt{2} * 1.4182 = 2.0056$]
  - 2*ESS < N_P = 220
  - Posterior Mean and 95% Credible Interval:

    -0.393 (-0.659, -0.003)

Sitagliptin and Metformin Results: Borrowing required to tip results

Variance of vague component	Borrowing required to tip results	$ESS_{ELIR(Prior)}$	Mean	Median	Cri95L	Cri95U	Posterior probability of efficacy	Updated weight
1	19%	66	-0.392	-0.428	-0.648	-0.003	0.976	0.465
Unit Variance (4.0226)	10%	31	-0.393	-0.427	-0.659	-0.003	0.976	0.439
10	7%	22	-0.396	-0.430	-0.660	-0.006	0.977	0.453
100	2%	6	-0.391	-0.424	-0.665	-0.001	0.975	0.414
1000	0.7%	2	-0.394	-0.428	-0.663	-0.004	0.976	0.435
10000	0.2%	1	-0.390	-0.424	-0.666	-0.0001	0.975	0.409

- The smaller the variance of the vague component, the more borrowing and larger $ESS_{ELIR(Prior)}$ is required to tip results
- Regardless of the variance of the vague component, estimates of summary statistics are similar
- Small variances should still be considered to incorporate uncertainty

Summary

- Larger observed SD in pediatric studies have challenged study feasibility
- Bayesian borrowing methods from the adult population reduces the number of pediatric patients that need to be enrolled to maintain adequate study power
- Effective Sample Size (e.g., $ESS_{ELIR(Prior)}$) tells us how much the prior information is worth in terms of sample size relative to a unit of information (e.g., 1 pediatric patient per arm)
- Small pediatric treatment effects and/or large SD requires more borrowing to achieve adequate study power
- Difference in magnitude between the assumed pediatric and estimated adult treatment effects, and/or large SD impact operating characteristics
- In our case study, as the variance of the vague component increases, less borrowing (and smaller ESS) is required to tip results, but yields similar estimates of summary statistics as smaller variances
- Small variances should still be considered to incorporate uncertainty

References

- Pediatrics living with T2DM: https://www.cdc.gov/media/releases/2021/p0824-youth-diabetes.html

- Belimumab Review: https://www.fda.gov/media/127912/download

- Neuenschwander, B., Weber, S., Schmidli, H., & O'Hagan, A. (2020). Predictively consistent prior effective sample sizes. *Biometrics*, *76*(2), 578-587: https://onlinelibrary.wiley.com/doi/10.1111/biom.13252

- Zhang, H., Anderson, K. M., Zimmer, Z., Golm, G., Sapre, A., & Ibrahim, J. G. (2024). Prior Effective Sample Size When Borrowing on the Treatment Effect Scale. *arXiv preprint arXiv:2404.13366*: [2404.13366] Prior Effective Sample Size When Borrowing on the Treatment Effect Scale (arxiv.org)

Acknowledgements

- ORISE Intern: Yuanyuan Ji (University of Nebraska)

- FDA/CDER/OTS/OB/DBII: James Travis, Yoonhee Kim, Yun Wang, Wenda Tu, Pablo Bonangelino, Sunghee Kim, Mark Rothmann

- FDA/CDER/OND/OCHEN/DDLO: Lisa Yanoff, Patrick Archdeacon

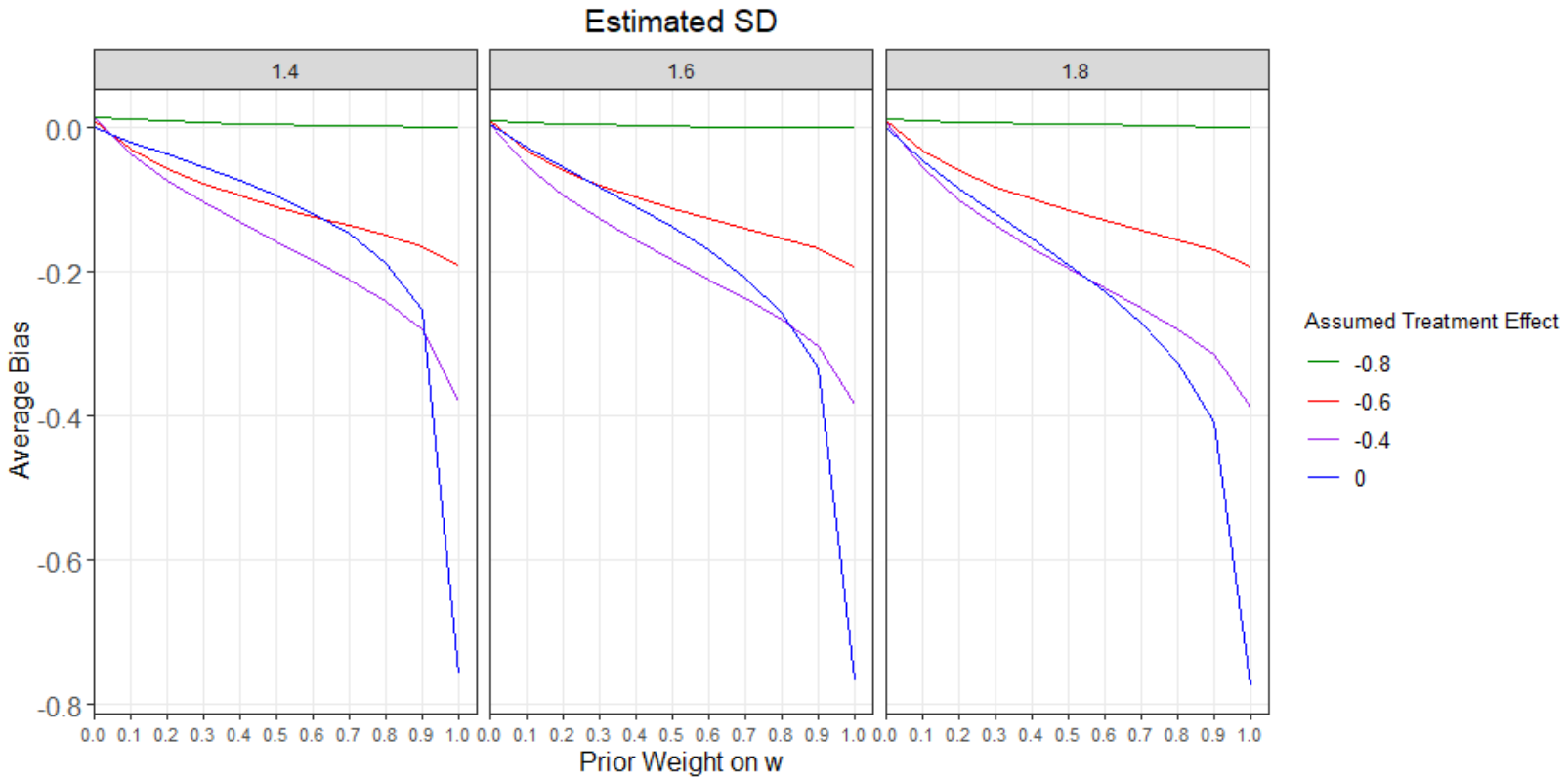
Back Up

Challenges with Larger SD Assumptions in T2D Pediatric Studies

SD	Total sample size needed to achieve 80% power with Δ = - 0.7%
0.9	52
1.0	64
1.1	78
1.2	94
1.3	110
1.4	126
1.5	144
1.6	164
1.7	186
1.8	208
1.9	232
2.0	256

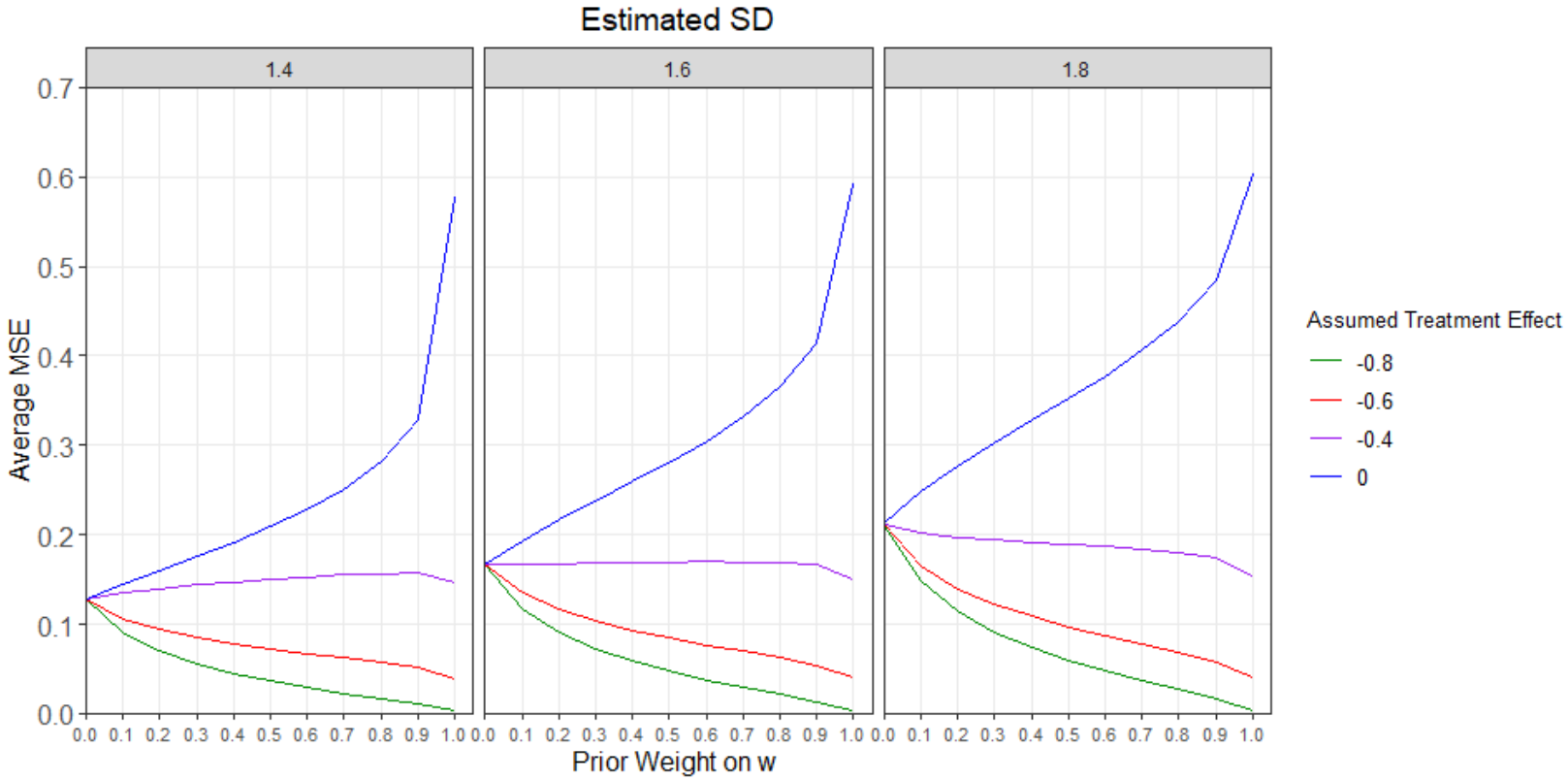
- As the SD increases by 50%, the sample size needed to preserve power increases by 125%
- Study feasibility issue with increasing the sample size in pediatric studies due to difficulties of recruitment

Simulation: Operating Characteristics of Bias



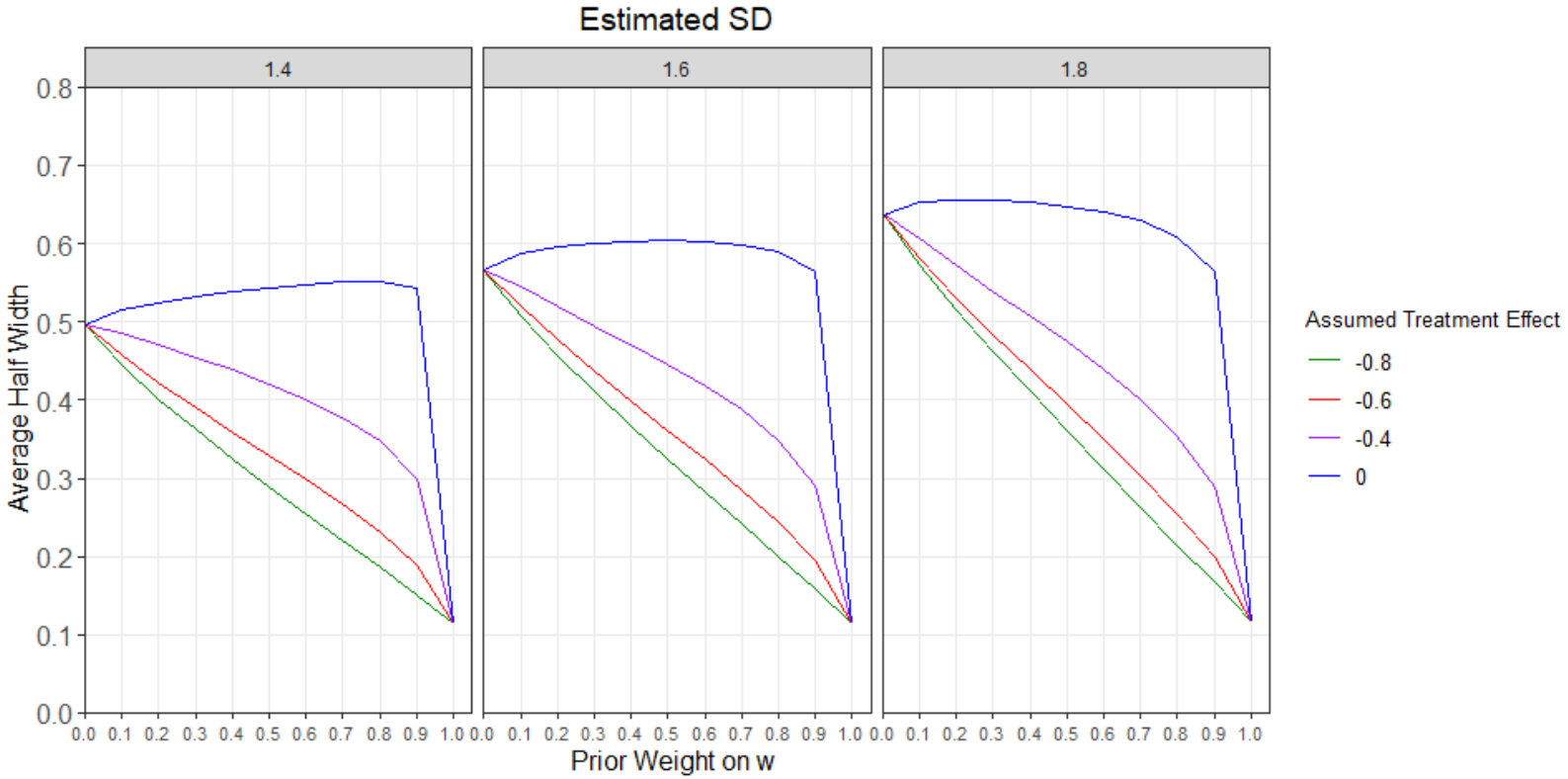
- For Δ_P = -0.8 = $\hat{\delta}_A$, there is little or no bias regardless of the amount of borrowing
- For $\Delta_P \neq$ -0.8 , the bias increases as the amount of borrowing increases

Simulation: Operating Characteristics of MSE



- Increases as Δ_P separates from $\hat{\delta}_A$= -0.8
- For Δ_P = 0, MSE increases as the borrowing increases
- For $\Delta_P \neq 0$, MSE tends to decrease as the borrowing increases

Simulation: Operating Characteristics of Half-Width of 95% Credible Interval



- Increases as Δ_P separates from $\hat{\delta}_A$= -0.8

- Converges for all Δ_P as borrowing approaches 100%

Sitagliptin Monotherapy

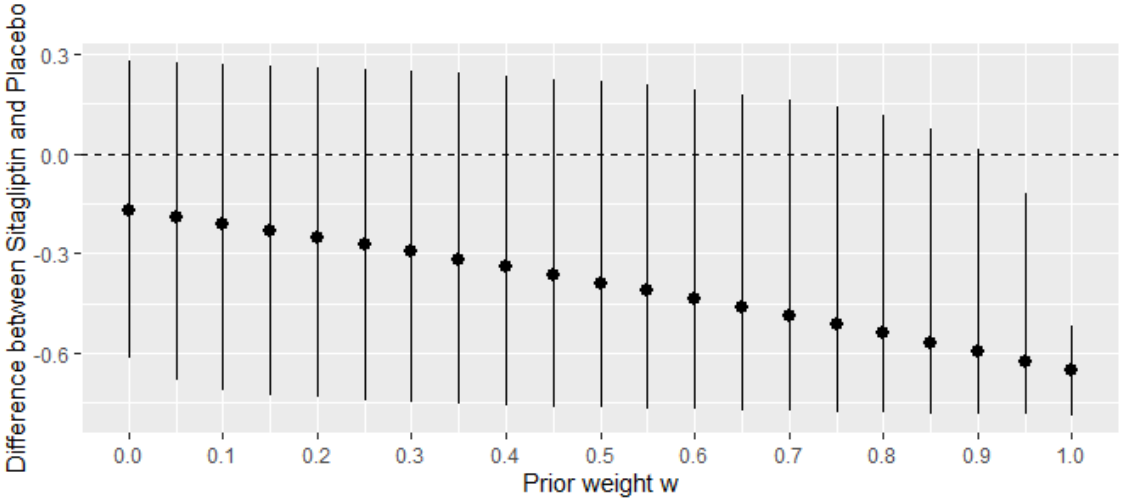
- Pediatric Information
  - N_P=190 (95 to sitagliptin; 95 to placebo)
  - $\hat{\delta}_P$ = - 0.17% and 95% CI: (-0.62, 0.28)
  - $SE_P \approx 0.2296$
  - $SD_P = 1.5824$

- Adult Information
  - 2 relevant studies
    - Summary statistics obtained from the label
  - Meta-analysis performed
    - N_A = 769 (422 to sitagliptin; 347 to placebo)
    - $\hat{\delta}_A$ = - 0.70%
    - SE_A = 0.0714

- Mixture Prior: $\omega * N(-0.70, 0.0714^2) + (1-\omega) * N(0, 5.008)$

  - $\sigma_0^2 = UISD_P^2 = 2*1.5824^2 = 5.008$

Sitagliptin Monotherapy Results: Unit Variance vs. Large Variance



Stand alone results:
$\hat{\delta}_P$ = - 0.17%
95% CI: (-0.62, 0.28)

- Mixture Prior

$$\omega * N(-0.70, 0.0714^2) + (1-\omega) * N(0, 5.008)$$

- 91% borrowing required to exclude 0 from the 95% credible interval
  - ESS=865 (per arm) [Reference scale = $\sqrt{2} * 1.5824 = 2.2379$]
  - 2*ESS > N_P = 190
  - Mean and 95% Credible Interval:
    -0.601 (-0.784, -0.002)

Sitagliptin Monotherapy Results: Borrowing required to tip results

Variance of vague component	Borrowing required to tip results	$ESS_{ELIR(Prior)}$	Mean	Median	Cri95L	Cri95U	Posterior probability of efficacy	Updated weight
1	96%	918	-0.605	-0.644	-0.784	-0.013	0.977	0.901
Unit Variance (5.008)	91%	865	-0.601	-0.643	-0.784	-0.002	0.975	0.893
10	88%	836	-0.602	-0.644	-0.784	-0.006	0.976	0.895
100	70%	665	-0.603	-0.644	-0.784	-0.007	0.976	0.895
1000	42%	399	-0.602	-0.643	-0.784	-0.004	0.976	0.894
10000	19%	180	-0.603	-0.644	-0.784	-0.008	0.976	0.896

- The smaller the variance of the vague component, the more borrowing and larger $ESS_{ELIR(Prior)}$ is required to tip results
- Regardless of the variance of the vague component, summary statistics are similar


FDA
U.S. FOOD & DRUG
ADMINISTRATION