Bayes Bayesian Biostatistics Utrecht 2023

Systematic Value Exploration of Propensity Score and Borrowing Approaches

Anduena Rexhepi

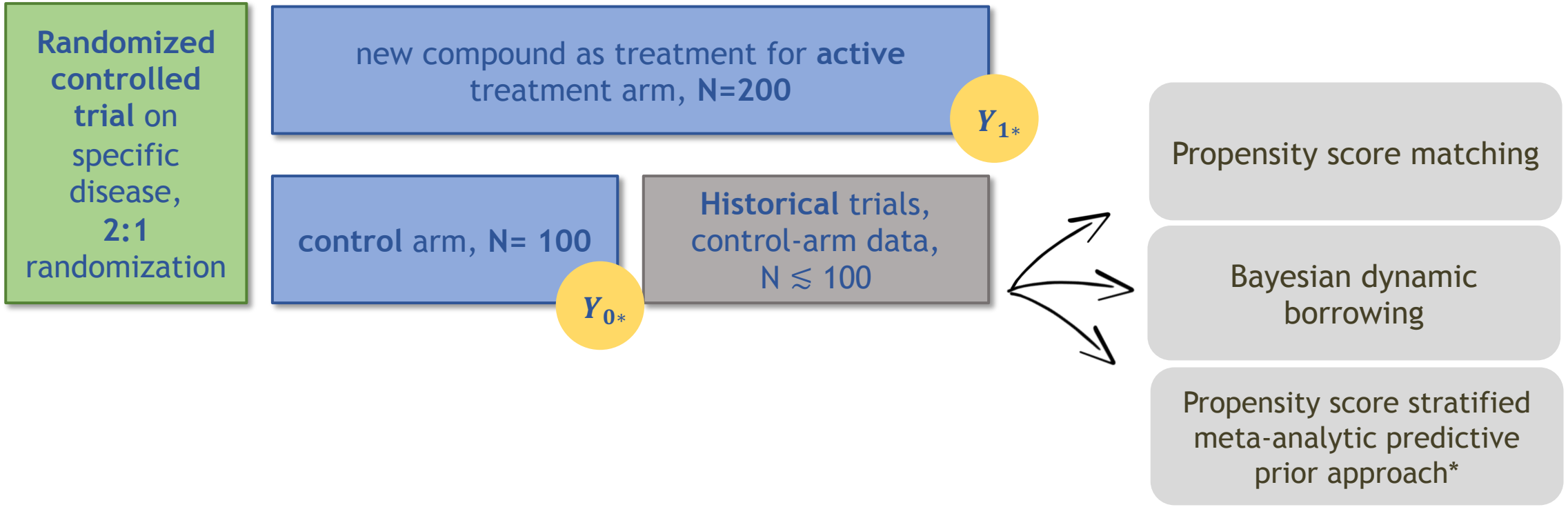
October 26, 2023

Boehringer Ingelheim

universität ulm

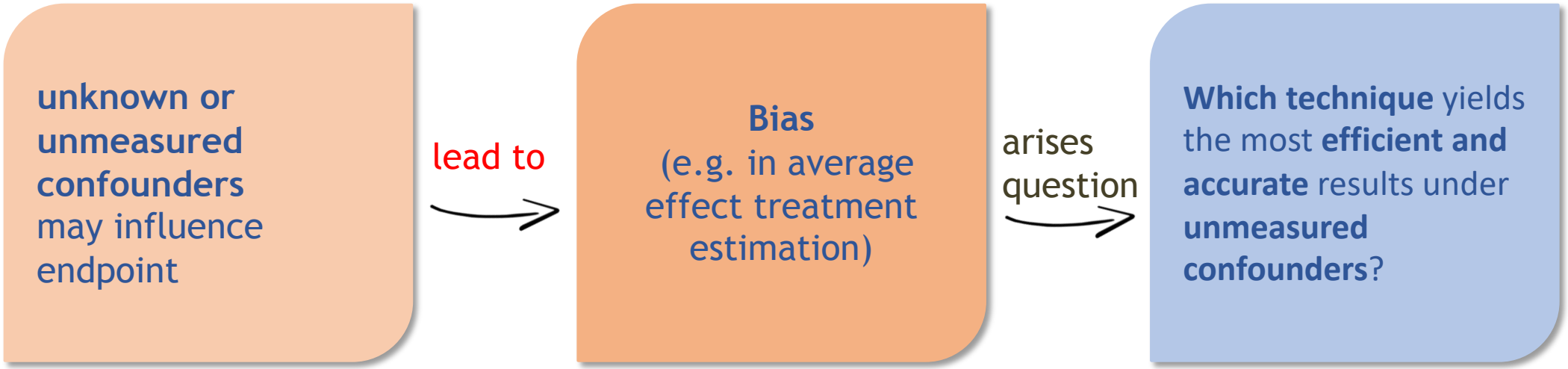
Introduction

Borrowing information from historical trials



* Zhu, A. Y., Roy, D., Zhu, Z., & Sailer, M. O. (2023). Propensity score stratified map prior and posterior inference for incorporating information across multiple po- tentially heterogeneous data sources. *Journal of Biopharmaceutical Statistics*, 1–15.

Question


unknown or unmeasured confounders may influence endpoint
lead to
Bias (e.g. in average effect treatment estimation)
arises question
Which technique yields the most **efficient and accurate** results under **unmeasured confounders**?

Structure

Explanation of **methods & operating characteristics**

Simulation studies

Real case study

Conclusion & outlook

Methods & operating characteristics

Bayesian Biostatistics Utrecht 2023

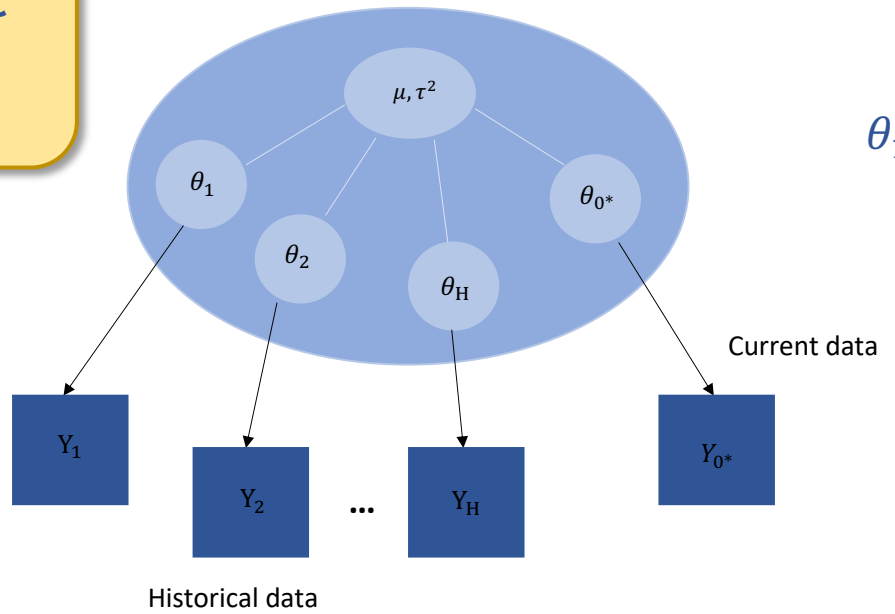
Propensity score matching

1. Estimate **propensity score** with a model involving **relevant** baseline characteristics X

   $e(X) := P(Z=1 \mid X)$ where $Z = \begin{cases} 1 & \text{if patient i from current trial} \\ 0 & \text{else} \end{cases}$.

2. **Match similar** historical control patients to current ones via **nearest neighbor matching algorithm** and **matching ratio** 1:1.

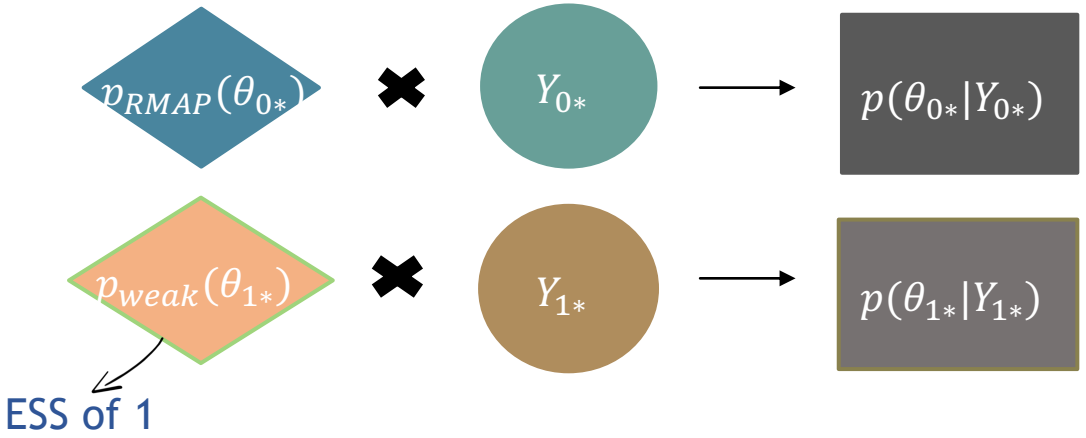
Bayesian dynamic borrowing



$$Y_h|\theta_h, \sigma_h^2 \sim N(\theta_h, \sigma_h^2) \quad h = 1, \dots, H$$
$$\theta_1, \dots, \theta_H, \theta_{0*}|\mu, \tau^2 \sim N(\mu, \tau^2) \quad h = 1, \dots, H$$
$$\mu \sim N(0, 100 \cdot \bar{\sigma}), \quad \mu \in R^{H+1}$$
$$\tau \sim HN\left(0, \frac{\bar{\sigma}}{c}\right), \; c \in \{16,8,4,2,1\}$$

c based on assumed level of **between-trial heterogeneity**

Similarity assumption may be violated when **prior-data conflict**



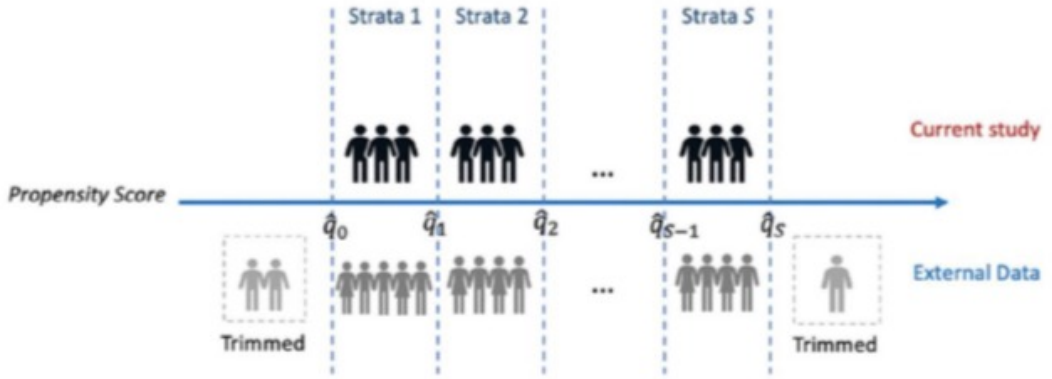
Increase standard deviation of components to reach target ESS (effective sample size).

Propensity score stratified meta-analytic predictive prior approach

1. Estimate **propensity score** with a model involving **relevant** baseline characteristics X

$$e(X) := P(Z=1 \mid X) \quad \text{where } Z = \begin{cases} 1 & \text{if patient i from current trial} \\ 0 & \text{else} \end{cases}.$$

2. **Stratification** + identification of **similar** historical patients (w.r.t to propensity score).



3. Apply **Bayesian analysis to each stratum separately** by additionally using **similarity measure** v_s and the **between-trial heterogeneity**.

Average treatment effect

$$ATE = E[outcome_{treated}] - E[outcome_{control}]$$

Propensity score matching

$$\widehat{ATE}_{PSMATCH} := \frac{1}{|Y_{1*}|}\sum_{i \in Y_{1*}} outcome_i - \frac{1}{|Y_0|}\sum_{i \in Y_0} outcome_i$$

Bayesian dynamic borrowing

$$\widehat{ATE}_{BDB} := \sum_{k=1}^{K_1} w_{1,k} m_{1,k} - \sum_{k=1}^{K_0} w_{0,k} m_{0,k}$$

Propensity score stratified meta-analytic predictive prior approach

$$\widehat{ATE}_{PSMAP} := \frac{1}{\sum_{s=1}^{S}\omega_s}\sum_{s=1}^{S}\omega_s \cdot \widehat{ATE}^{(s)}$$

$$\widehat{ATE}^{(s)} := \sum_{k=1}^{K_1^{(s)}} w_{1,k}^{(s)} m_{1,k}^{(s)} - \sum_{k=1}^{K_0^{(s)}} w_{0,k}^{(s)} m_{0,k}^{(s)}$$

stratum weights $\omega_s = \frac{v_s}{\sum_{s=1}^{S} v_s}$; add up to 1;
reflect similarity between current and historical patients in stratum *s*

Confidence / Credible interval

Propensity score matching

$$CI_{PSMATCH} := \left[\widehat{ATE}_{PSMATCH} \pm t \cdot S \sqrt{\frac{1}{|Y_{1*}|} + \frac{1}{|Y_0|}}\right]$$

$(1 - \frac{\alpha}{2})$ quantile of t-distribution

pooled standard deviation

Bayesian dynamic borrowing

$$CrI_{BDB} := [\theta_{\frac{\alpha}{2}}, \theta_{1-\frac{\alpha}{2}}]$$

quantiles of posterior $p(\theta|y)$

Propensity score stratified meta-analytic predictive prior approach

$$CrI_{PSMAP} := [\delta_{\frac{\alpha}{2}}, \delta_{1-\frac{\alpha}{2}}]$$

quantiles of sample vector $\Delta := \begin{pmatrix} \sum_{s=1}^{S} \omega_s \Delta^{(s)}[1] \\ \vdots \\ \sum_{s=1}^{S} \omega_s \Delta^{(s)}[J] \end{pmatrix}$,

which contains the weighted sum of the difference between the posterior samples of $p(\theta_{1*}^{(s)}|Y_{1*})$ and $p\left(\theta_{0*}^{(s)}\middle|Y_{0*}\right)$ for each draw j

Operating characteristics

$$ATE = E[\widehat{ATE}] \; if \; unbiased$$
$$MSE = E\left[\left(\widehat{ATE} - ATE\right)^2\right]$$
$$Bias = E[\widehat{ATE}] - ATE$$
$$Var = E[\widehat{ATE}^2] - E[\widehat{ATE}]^2$$
$$Cov = E[1(ATE \in CI)]$$

In **simulation study: B replications** for specific scenario

$$\widehat{ATE} := \frac{1}{B}\sum_{b=1}^{B}\widehat{ATE}_b$$
$$\widehat{MSE} := \frac{1}{B}\sum_{b=1}^{B}(\widehat{ATE}_b - ATE)^2$$
$$\widehat{Bias} := \widehat{ATE} - \text{ATE}$$
$$\widehat{Var} := \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{ATE}_b - \widehat{ATE})^2$$
$$\widehat{Cov} := \frac{1}{B}\sum_{b=1}^{B} 1(ATE \in CI_b)$$

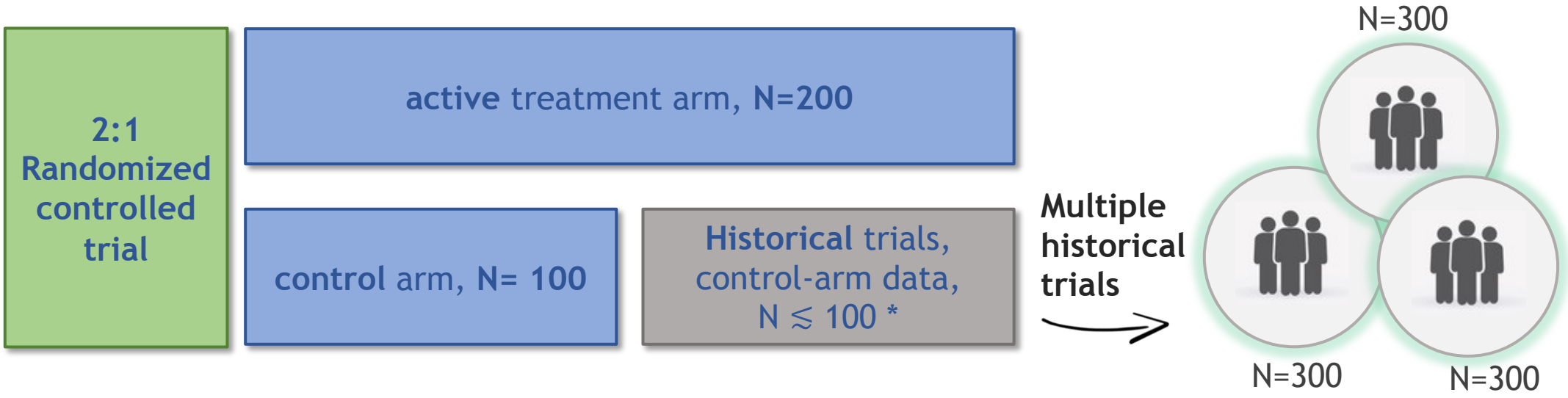
true ATE is predefined in simulation

$$= \begin{cases} 1 \; if \; ATE \in CI_b \\ 0 \; if \; ATE \notin CI_b \end{cases}$$

CI in coverage definition is to be replaced by CrI for Bayesian methods.

Simulation studies

Simulated trial design & data generation



blood pressure

predefined **ATE**

$$bp = 150 + 0.1 \cdot (AGE - 50) + 0.3 \cdot (BMI - 25) + (-2) \cdot T + \epsilon$$

unmeasured

treatment indicator

$\sim N(0, \sigma = 20)$

current patients

$AGE \sim N(50, \sigma = 10)$ and $BMI \sim N(25, \sigma = 3)$

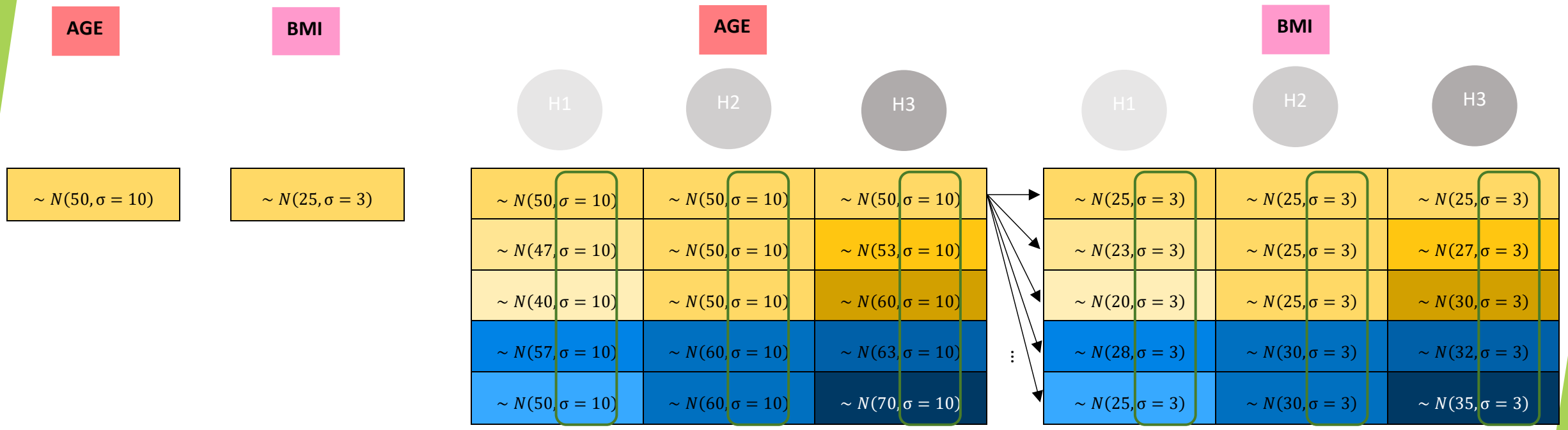
*ESS ≲ 100 or matched ≲ 100

Specifications in simulation

- Assumptions made in Bayesian methods:
  - **Moderate** level of between-trial heterogeneity
  - **80%** weight of historical information
- Each designed scenario run with $\boldsymbol{B = 1000}$ replications.

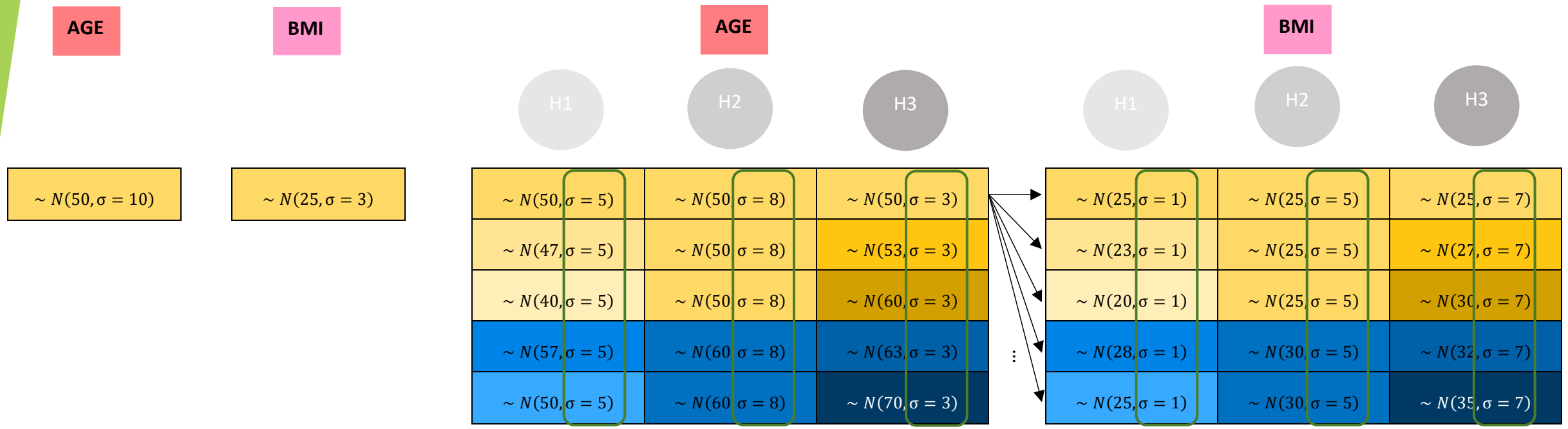
  → *High Performance Computing*

Different scenario designs – Case a



25 different scenarios

Different scenario designs – Case b



current patients

historical patients

25 different scenarios

Results

- Prior-data conflict → $\widehat{MSE}$ rise
- PS Methods less conditions for ‚almost unbiased' → $\emptyset\left(\mu_{BMI_{H_i}}\right) = 25 = \mu_{BMI_{Current}}$
- $\widehat{MSE}_{BDB}$ and $\widehat{Bias}_{BDB}$ invariant to covariate variability σ

- often PSMAP and BDB comparable except for strong prior-data conflict
- although $\widehat{MSE}_{PSMATCH}$ and $\widehat{MSE}_{PSMAP}$ similar behaviour, PS matching almost always most inefficient

Real case study

Real case study – schizophrenia*

Change in *positive & negative syndrome scale total score* from baseline at week 12

NCT01077700 | NCT01095562 | NCT01655680 | NCT01678755 — Abbvie sponsored trials

current control study (60) | historical control studies (≈ 200)

Inclusion criteria: between 20 and 55 years old patients

Baseline covariates: age, gender, race, height, weight, systolic & diastolic blood pressure, smoking status

Matched/prior ESS ~ 60; moderate level of between-trial heterogeneity; $w_{historical} = 80\ \%$

RESULTS

Method	$\widehat{ATE}$	CI / CrI	Interval length
PS Match	-4.08	[-5.37, -2.8]	2.57
Bayesian Borrowing	-3.79	[-5.07, -2.52]	2.55
PS-MAP	-4.2	[-5.39, -2.97]	2.42

*Data and access are provided by TransCelerate Platform and TransCelerate Biopharma and the participants.
Neither TransCelerate Biopharma nor any of the participants have contributed to, approved, or are in any way responsible for my research results.

Conclusion & outlook

Conclusion & outlook

Which technique yields the most **efficient and accurate** results under **unmeasured confounders**?

BIAS All methods have strong requirements for **unbiased ATE estimates.**

MSE

- The PS Matching almost always most inefficient.
- Especially when stronger prior-data conflict **PSMAP most efficient.**

- Increasing number of replications, e.g. B=10000.
- Adjusting weight of robust prior component based on assumed level of prior-data conflict.
- Adjusting level of between-trial heterogeneity.
- Inclusion of more covariates.

Appendix

Bibliography

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when es- timating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, *10*(2), 150–161.

Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r using futures. *The R Journal*, *13*(2), 208–227. https://doi.org/10.32614/RJ- 2021-048

Bengtsson, H. (2022a). *Matrixstats: Functions that apply to rows and columns of matrices (and to vectors)* [R package version 0.63.0]. https://CRAN.R- project.org/package=matrixStats

Bengtsson, H. (2022b). *Matrixstats: Functions that apply to rows and columns of matrices (and to vectors)* [R package version 0.62.0]. https://CRAN.R- project.org/package=matrixStats

Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.

Czado, C., & Schmidt, T. (2011). *Mathematische statistik*. Springer-Verlag.

Falk, M., Hain, J., Marohn, F., Fischer, H., & Michel, R. (2014). *Statistik in theorie und praxis: Mit anwendungen in r*. Springer-Verlag.

Forster, M., & Sober, E. Aic scores as evidence: A bayesian interpretation. In: In *Philosophy of statistics*. Elsevier, 2011, pp. 535–549.

Gaujoux, R. (2023). *Dorng: Generic reproducible parallel backend for 'foreach' loops* [R package version 1.8.6]. https://CRAN.R-project.org/package=doRNG

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis (vol. 2).

Haig, G., Wang, D., Othman, A. A., & Zhao, J. (2016). The α7 nicotinic agonist abt- 126 in the treatment of cognitive impairment associated with schizophrenia in nonsmokers: Results from a randomized controlled phase 2b study. *Neuropsy- chopharmacology*, *41*(12), 2893–2902.

Haig, G. M., Bain, E., Robieson, W., Othman, A. A., Baker, J., & Lenz, R. A. (2014). A randomized trial of the efficacy and safety of the h3 antagonist abt-288 in cognitive impairment associated with schizophrenia. *Schizophrenia bulletin*, *40*(6), 1433–1442.

Haig, G. M., Wang, D., Zhao, J., Othman, A. A., & Bain, E. E. (2017). Efficacy and safety of the α7-nicotinic acetylcholine receptor agonist abt-126 in the treatment of cognitive impairment associated with schizophrenia: Results from a phase 2b randomized controlled study in smokers. *The Journal of Clinical Psychiatry*, *78*(3), 133.

Hashimoto, K. (2015). Targeting of α7 nicotinic acetylcholine receptors in the treat- ment of schizophrenia and the use of auditory sensory gating as a translational biomarker. *Current pharmaceutical design*, *21*(26), 3797–3806.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric prepro- cessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1–28. https://doi.org/10.18637/jss.v042.i08

Bibliography

Holton, G. A. (2013). *Value-at-risk: Theory and practice, second edition,* published online at https://www.value-at-risk.net.

James, G., Witten, D., Hastie, T., Tibshirani, R., James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). Statistical learning. *An introduction to statistical learning: with applications in R*, 15–57.

Kass, R. E., & Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the american statistical association*, *90*(431), 928–934.

Liu, M., Bunn, V., Hupf, B., Lin, J., & Lin, J. (2021). Propensity-score-based meta- analytic predictive prior for incorporating real-world and historical data. *Statistics in medicine*, *40*(22), 4794–4808.

Microsoft & Weston, S. (2022). *Foreach: Provides foreach looping construct* [R pack- age version 1.5.2]. https://CRAN.R-project.org/package=foreach

Neuenschwander, B., Capkun-Niggli, G., Branson, M., & Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials*, *7*(1), 5–18.

Neuenschwander, B., & Schmidli, H. Use of historical data. In: In *Bayesian methods in pharmaceutical research*. Chapman; Hall/CRC, 2020, pp. 111–137.

Neuenschwander, B., Weber, S., Schmidli, H., & O'Hagan, A. (2020). Predictively consistent prior effective sample sizes. *Biometrics*, *76*(2), 578–587.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project. org/

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project. org/

Ray, K., & Szabó, B. (2019). Debiased bayesian inference for average treatment effects. *Advances in Neural Information Processing Systems*, *32*.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Ross, S. (2010). *A first course in probability*. Pearson.

Röver, C. (2020). Bayesian random-effects meta-analysis using the bayesmeta R package. *Journal of Statistical Software*, *93*(6), 1–51. https://doi.org/10.18637/ jss.v093.i06

Sawamoto, R., Oba, K., & Matsuyama, Y. (2022). Bayesian adaptive randomization design incorporating propensity score-matched historical controls. *Pharmaceutical Statistics*, *21*(5), 1074–1089.

Schmidli, H., Gsteiger, S., & Neuenschwander, B. Incorporating historical data into randomized controlled trials. In: In *Handbook of statistical methods for randomized controlled trials*. Chapman; Hall/CRC, 2021, pp. 545–566.

Bibliography

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., & Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, *70*(4), 1023–1032.

Shotwell, M. (2022). *Sas7bdat: Sas7bdat reverse engineering documentation* [R package version 0.6]. https://CRAN.R-project.org/package=sas7bdat

Stan Development Team. (2022a). RStan: The R interface to Stan [R package version 2.21.7]. https://mc-stan.org/

Stan Development Team. (2022b). RStan: The R interface to Stan [R package version 2.21.5]. https://mc-stan.org/

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., et al. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), 1–26.

Weber, S. (2020). *Rbest: R bayesian evidence synthesis tools* [R package version 1.6- 1]. https://CRAN.R-project.org/package=RBesT

Weber, S., Li, Y., Seaman, J. W., Kakizume, T., & Schmidli, H. (2021). Applying meta-analytic-predictive priors with the R Bayesian evidence synthesis tools. *Journal of Statistical Software*, *100*(19), 1–32. https://doi.org/10.18637/jss. v100.i19

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2022a). *Dplyr: A grammar of data manipulation* [R package version 1.0.9]. https://CRAN.R-project.org/ package=dplyr

Wickham, H., Miller, E., & Smith, D. (2022b). *Haven: Import and export 'spss', 'stata' and 'sas' files* [R package version 2.5.0]. https://CRAN.R-project.org/ package=haven

Xie, Y. Knitr: A comprehensive tool for reproducible research in R (V. Stodden, F. Leisch, & R. D. Peng, Eds.) [ISBN 978-1466561595]. In: *Implementing reproducible computational research* (V. Stodden, F. Leisch, & R. D. Peng, Eds.). Ed. by Stodden, V., Leisch, F., & Peng, R. D. ISBN 978-1466561595. Chapman; Hall/CRC, 2014. http://www.crcpress.com/product/isbn/9781466561595

Zhao, Q.-Y., Luo, J.-C., Su, Y., Zhang, Y.-J., Tu, G.-W., & Luo, Z. (2021). Propensity score matching with r: Conventional methods and new features. *Annals of translational medicine*, *9*(9).

Zhu, A. Y., Roy, D., Zhu, Z., & Sailer, M. O. (2023). Propensity score stratified map prior and posterior inference for incorporating information across multiple po- tentially heterogeneous data sources. *Journal of Biopharmaceutical Statistics*, 1–15.

Zhu, H. (2021). *Kableextra: Construct complex table with 'kable' and pipe syntax* [R package version 1.3.4]. https://CRAN.R-project.org/package=kableExtra

SHT – Bias estimate

PS Model:
only AGE
adjustment

	$\hat{ATE}_{PSMATCH}$	$\hat{Bias}_{PSMATCH}$	$\hat{ATE}_{DBB}$	$\hat{Bias}_{DBB}$	$\hat{ATE}_{PSMAP}$	$\hat{Bias}_{PSMAP}$	AGE_H	BMI_H
1	-2.03	-0.03	-2.01	-0.01	-2.01	-0.01	N(50, σ=10)	N(25, σ=3)
2	-2.03	-0.03	-2.01	-0.01	-2.01	-0.01	N(50, σ=10)	N(25, σ=6)
3	-2.78	-0.78	-2.56	-0.56	-2.73	-0.73	N(50, σ=10)	N(30, σ=3)
4	-2.78	-0.78	-2.56	-0.56	-2.73	-0.73	N(50, σ=10)	N(30, σ=6)
5	-2.01	-0.01	-2.01	-0.01	-2.01	-0.01	N(50, σ=5)	N(25, σ=3)
6	-2.01	-0.01	-2.01	-0.01	-2.01	-0.01	N(50, σ=5)	N(25, σ=6)
7	-2.76	-0.76	-2.56	-0.56	-2.68	-0.68	N(50, σ=5)	N(30, σ=3)
8	-2.76	-0.76	-2.56	-0.56	-2.68	-0.68	N(50, σ=5)	N(30, σ=6)
9	-2.10	-0.10	-2.38	-0.38	-2.04	-0.04	N(60, σ=10)	N(25, σ=3)
10	-2.10	-0.10	-2.38	-0.38	-2.04	-0.04	N(60, σ=10)	N(25, σ=6)
11	-2.85	-0.85	-2.85	-0.85	-2.69	-0.69	N(60, σ=10)	N(30, σ=3)
12	-2.85	-0.85	-2.85	-0.85	-2.69	-0.69	N(60, σ=10)	N(30, σ=6)
13	-2.43	-0.43	-2.39	-0.39	-2.03	-0.03	N(60, σ=5)	N(25, σ=3)
14	-2.43	-0.43	-2.38	-0.38	-2.03	-0.03	N(60, σ=5)	N(25, σ=6)
15	-3.18	-1.18	-2.86	-0.86	-2.71	-0.71	N(60, σ=5)	N(30, σ=3)
16	-3.18	-1.18	-2.85	-0.85	-2.71	-0.71	N(60, σ=5)	N(30, σ=6)

almost **unbiased** when
$\mu_{BMI_{Historical}} = 25 = \mu_{BMI_{Current}}$
(except for scenarios 13+14)

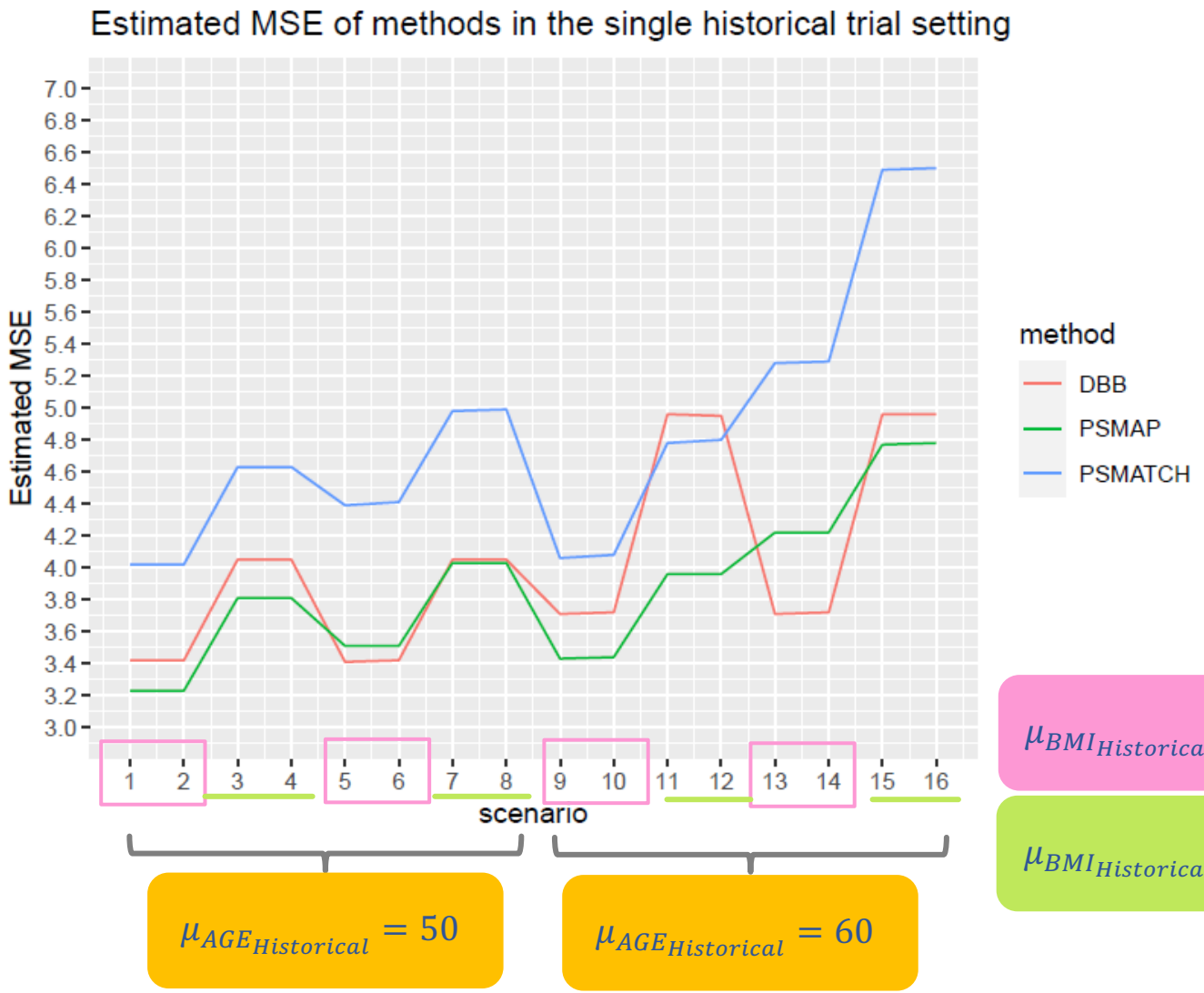
almost **unbiased** when
$\mu_{BMI_{Historical}} = 25 = \mu_{BMI_{Current}}$ **AND**
$\mu_{AGE_{Historical}} = 50 = \mu_{AGE_{Current}}$

almost **unbiased** when
$\mu_{BMI_{Historical}} = 25 = \mu_{BMI_{Current}}$

PS – Propensity Score

SHT – MSE estimate

$$MSE = Var + Bias^2$$



- $\widehat{MSE}_{PSMATCH}$ and $\widehat{MSE}_{PSMAP}$ similar behaviour
- $\widehat{MSE}_{PSMATCH}$ almost everywhere highest
- σ_{BMI} *irrelevant*
- For $\widehat{MSE}_{DBB}$ also σ_{AGE} *irrelevant*
- when $\mu_{BMI_H} = 30$ → prior-data conflict → MSE estimates increase

a – Bias estimate

almost **unbiased** when

$$\emptyset\left(\mu_{BMI_{H_1}}, \mu_{BMI_{H_2}}, \mu_{BMI_{H_3}}\right) = 25 = \mu_{BMI_{Current}} \text{ AND}$$

$$\emptyset\left(\mu_{AGE_{H_1}}, \mu_{AGE_{H_2}}, \mu_{AGE_{H_3}}\right) = 50 = \mu_{AGE_{Current}}$$

$\hat{ATE}_{PSMATCH}$	$\hat{Bias}_{PSMATCH}$	$\hat{ATE}_{DBB}$	$\hat{Bias}_{DBB}$	$\hat{ATE}_{PSMAP}$	$\hat{Bias}_{PSMAP}$	AGE_{H_1}	AGE_{H_2}	AGE_{H_3}	BMI_{H_1}	BMI_{H_2}	BMI_{H_3}
-1.93	0.07	-1.92	0.08	-1.91	0.09				N(25, σ=3)	N(25, σ=3)	N(25, σ=3)
-1.93	0.07	-1.92	0.08	-1.91	0.09				N(23, σ=3)	N(25, σ=3)	N(27, σ=3)
-1.94	0.06	-1.91	0.09	-1.91	0.09	N(50, σ=10)	N(50, σ=10)	N(50, σ=10)	N(20, σ=3)	N(25, σ=3)	N(30, σ=3)
-2.68	-0.68	-2.57	-0.57	-2.65	-0.65				N(28, σ=3)	N(30, σ=3)	N(32, σ=3)
-2.69	-0.69	-2.53	-0.53	-2.65	-0.65				N(25, σ=3)	N(30, σ=3)	N(35, σ=3)
-1.96	0.04	-1.92	0.08	-1.9	0.1				N(25, σ=3)	N(25, σ=3)	N(25, σ=3)
-1.96	0.04	-1.92	0.08	-1.9	0.1				N(23, σ=3)	N(25, σ=3)	N(27, σ=3)
-1.96	0.04	-1.91	0.09	-1.9	0.1	N(47, σ=10)	N(50, σ=10)	N(53, σ=10)	N(20, σ=3)	N(25, σ=3)	N(30, σ=3)
-2.71	-0.71	-2.56	-0.56	-2.64	-0.64				N(28, σ=3)	N(30, σ=3)	N(32, σ=3)
-2.71	-0.71	-2.5	-0.5	-2.64	-0.64				N(25, σ=3)	N(30, σ=3)	N(35, σ=3)
-1.92	0.08	-1.91	0.09	-1.91	0.09				N(25, σ=3)	N(25, σ=3)	N(25, σ=3)
-1.92	0.08	-1.91	0.09	-1.91	0.09				N(23, σ=3)	N(25, σ=3)	N(27, σ=3)
-1.92	0.08	-1.9	0.1	-1.9	0.1	N(40, σ=10)	N(50, σ=10)	N(60, σ=10)	N(20, σ=3)	N(25, σ=3)	N(30, σ=3)
-2.67	-0.67	-2.52	-0.52	-2.64	-0.64				N(28, σ=3)	N(30, σ=3)	N(32, σ=3)
-2.67	-0.67	-2.41	-0.41	-2.63	-0.63				N(25, σ=3)	N(30, σ=3)	N(35, σ=3)
-2.05	-0.05	-2.37	-0.37	-1.94	0.06				N(25, σ=3)	N(25, σ=3)	N(25, σ=3)
-2.00	0.00	-2.36	-0.36	-1.91	0.09				N(23, σ=3)	N(25, σ=3)	N(27, σ=3)
-1.92	0.08	-2.31	-0.31	-1.87	0.13	N(57, σ=10)	N(60, σ=10)	N(63, σ=10)	N(20, σ=3)	N(25, σ=3)	N(30, σ=3)
-2.75	-0.75	-2.92	-0.92	-2.58	-0.58				N(28, σ=3)	N(30, σ=3)	N(32, σ=3)
-2.67	-0.67	-2.83	-0.83	-2.54	-0.54				N(25, σ=3)	N(30, σ=3)	N(35, σ=3)
-1.99	0.01	-2.35	-0.35	-1.94	0.06				N(25, σ=3)	N(25, σ=3)	N(25, σ=3)
-1.86	0.14	-2.32	-0.32	-1.86	0.14				N(23, σ=3)	N(25, σ=3)	N(27, σ=3)
-1.68	0.32	-2.24	-0.24	-1.73	0.27	N(50, σ=10)	N(60, σ=10)	N(70, σ=10)	N(20, σ=3)	N(25, σ=3)	N(30, σ=3)
-2.61	-0.61	-2.86	-0.86	-2.54	-0.54				N(28, σ=3)	N(30, σ=3)	N(32, σ=3)
-2.43	-0.43	-2.7	-0.7	-2.41	-0.41				N(25, σ=3)	N(30, σ=3)	N(35, σ=3)

PS Model: just AGE adjustment

almost **unbiased** when

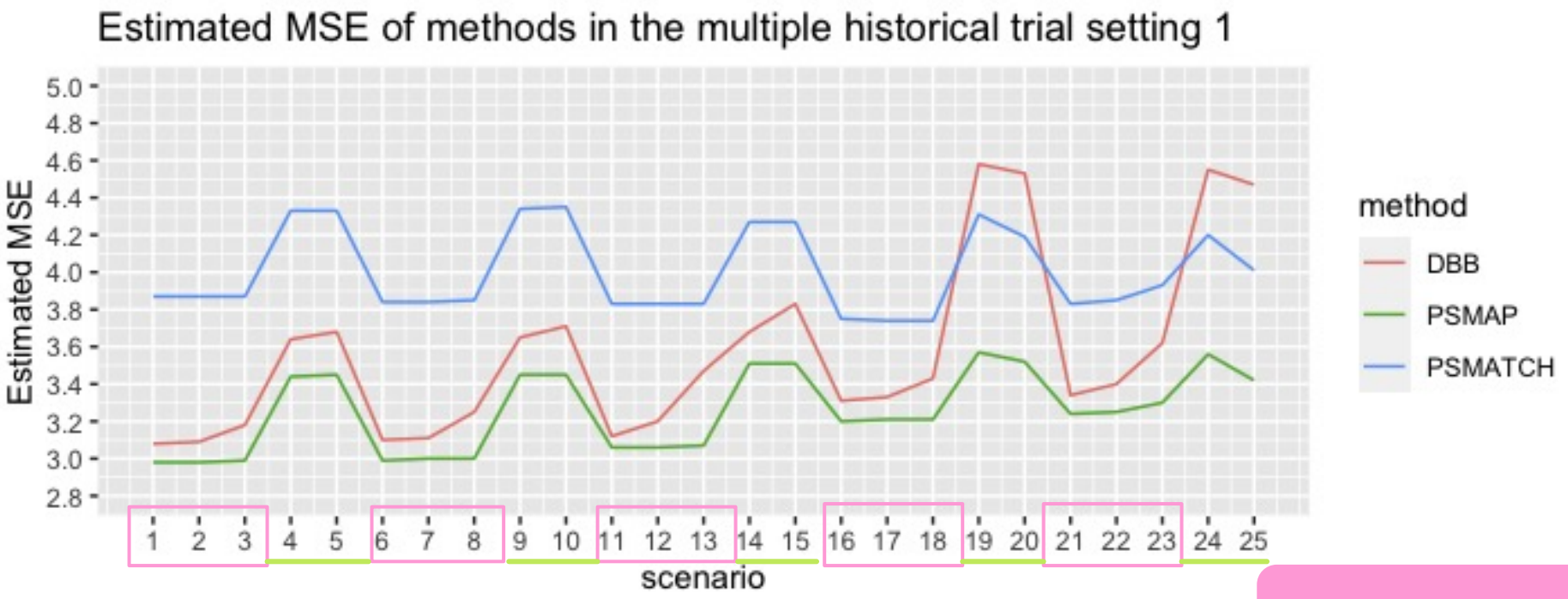
$$\emptyset\left(\mu_{BMI_{H_1}}, \mu_{BMI_{H_2}}, \mu_{BMI_{H_3}}\right) = 25 = \mu_{BMI_{Current}}$$

(except for scenario 23)

a – MSE estimate

$$MSE = Var + Bias^2$$

$$\sigma_{AGE_{H_i}} = 10 \; for \; i = 1,2,3$$
$$\sigma_{BMI_{H_i}} = 3 \; for \; i = 1,2,3$$



$\emptyset \; \mu_{BMI_{Historical}} = 25$

$\emptyset \; \mu_{BMI_{Historical}} = 30$

- $PSMAP$ most efficient method, $PSMATCH$ least efficient (except 19,20,24,25) but similar $\widehat{MSE}$ behaviour
- DBB comparable to $PSMAP$ (except for some scenarios)

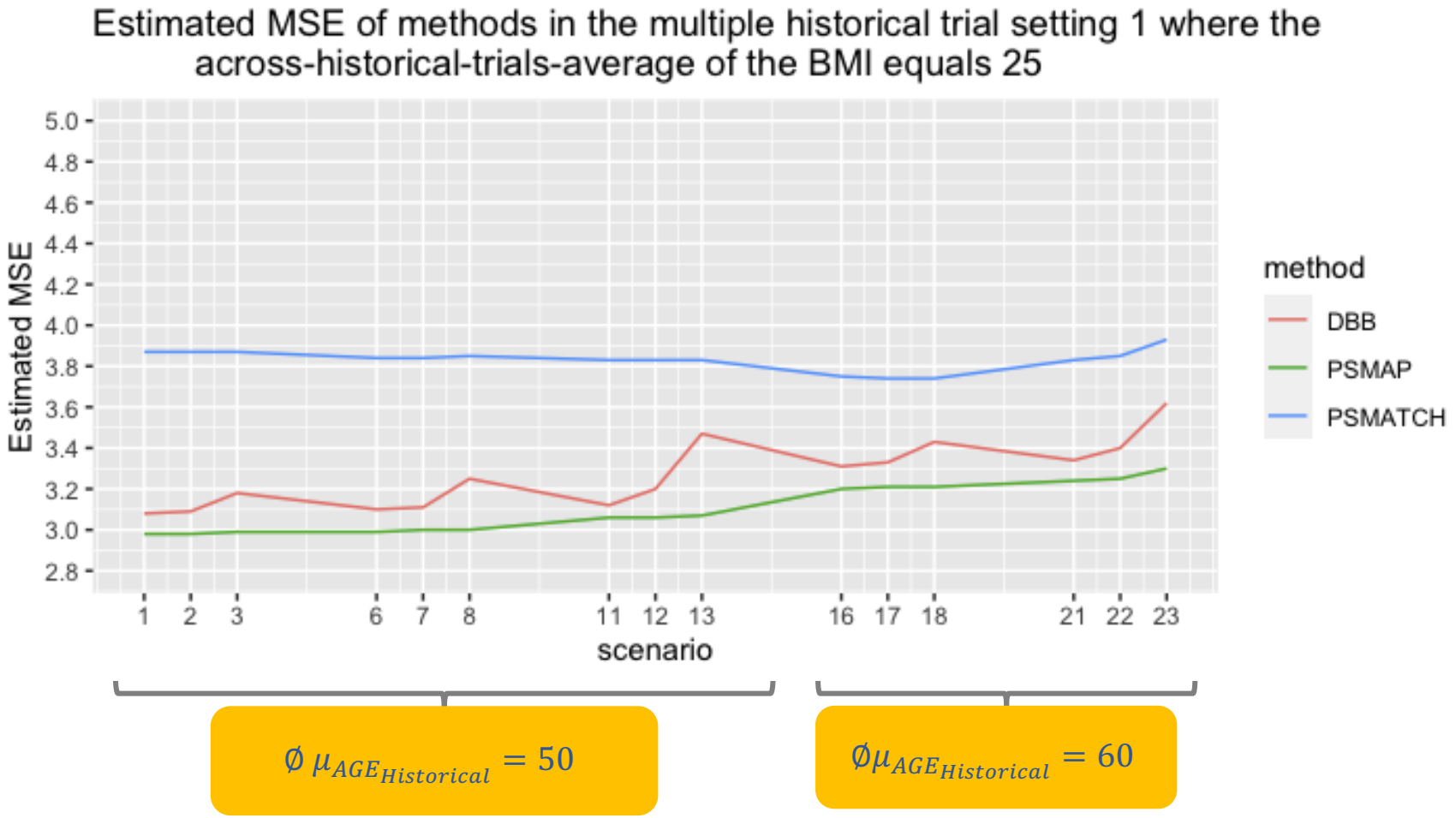
a – MSE estimate

$\emptyset\ \mu_{BMI_{Historical}} = 25$

$MSE = Var + Bias^2$

$\sigma_{AGE_{H_i}} = 10\ for\ i = 1,2,3$

$\sigma_{BMI_{H_i}} = 3\ for\ i = 1,2,3$



- BMI between-trial heterogeneity has influence on $\widehat{MSE}_{DBB}$
  → PS methods more stable w.r.t. that
- Increasing between-trial AGE heterogeneity
  → increase in $\widehat{MSE}_{PSMATCH}$ & decrease in $\widehat{MSE}_{PSMAP}$

PS – Propensity Score

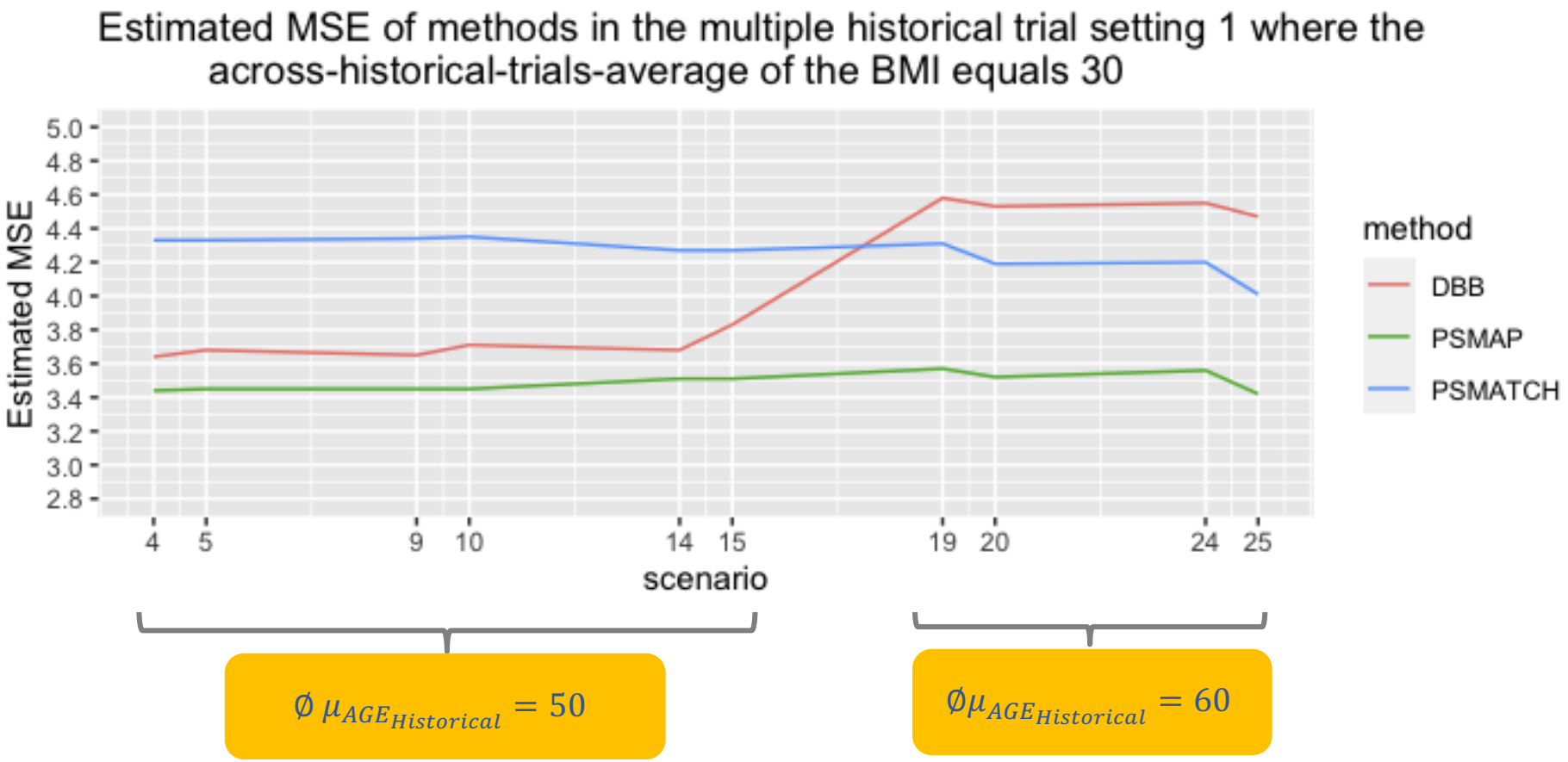
a – MSE estimate

$\emptyset\ \mu_{BMI_{Historical}} = 30$

$MSE = Var + Bias^2$

$\sigma_{AGE_{H_i}} = 10\ for\ i = 1,2,3$

$\sigma_{BMI_{H_i}} = 3\ for\ i = 1,2,3$



- MSE estimates higher than before
- $\widehat{MSE}_{DBB}$ increase in last four scenarios due to different $\emptyset\mu_{AGE_H}$ and $\emptyset\mu_{BMI_H}$
  → PS methods more stable w.r.t. that

b – Bias estimate

almost **unbiased** when

$$\emptyset\left(\mu_{BMI_{H_1}}, \mu_{BMI_{H_2}}, \mu_{BMI_{H_3}}\right) = 25 = \mu_{BMI_{Current}}$$

AND

$$\emptyset\left(\mu_{AGE_{H_1}}, \mu_{AGE_{H_2}}, \mu_{AGE_{H_3}}\right) = 50 = \mu_{AGE_{Current}}$$

$\widehat{Bias}_{DBB}$ very similar to case I
→ invariant to σ of covariates

	$\hat{ATE}_{PSMATCH}$	$\hat{Bias}_{PSMATCH}$	$\hat{ATE}_{DBB}$	$\hat{Bias}_{DBB}$	$\hat{ATE}_{PSMAP}$	$\hat{Bias}_{PSMAP}$
1	-1.90	0.10	-1.92	0.08	-1.9	0.1
2	-1.87	0.13	-1.92	0.08	-1.89	0.11
3	-1.83	0.17	-1.91	0.09	-1.86	0.14
4	-2.62	-0.62	-2.57	-0.57	-2.58	-0.58
5	-2.58	-0.58	-2.53	-0.53	-2.55	-0.55
6	-1.88	0.12	-1.92	0.08	-1.89	0.11
7	-1.86	0.14	-1.92	0.08	-1.88	0.12
8	-1.82	0.18	-1.9	0.1	-1.85	0.15
9	-2.61	-0.61	-2.56	-0.56	-2.59	-0.59
10	-2.57	-0.57	-2.5	-0.5	-2.55	-0.55
11	-1.89	0.11	-1.91	0.09	-1.9	0.1
12	-1.86	0.14	-1.91	0.09	-1.87	0.13
13	-1.83	0.17	-1.9	0.1	-1.83	0.17
14	-2.61	-0.61	-2.51	-0.51	-2.58	-0.58
15	-2.58	-0.58	-2.41	-0.41	-2.54	-0.54
16	-2.29	-0.29	-2.37	-0.37	-1.9	0.1
17	-2.20	-0.20	-2.36	-0.36	-1.85	0.15
18	-2.07	-0.07	-2.31	-0.31	-1.76	0.24
19	-2.95	-0.95	-2.92	-0.92	-2.46	-0.46
20	-2.82	-0.82	-2.83	-0.83	-2.37	-0.37
21	-2.04	-0.04	-2.35	-0.35	1.93	0.07
22	-1.88	0.12	-2.32	-0.32	-1.8	0.2
23	-1.65	0.35	-2.24	-0.24	-1.62	0.38
24	-2.63	-0.63	-2.85	-0.85	-2.47	-0.47
25	-2.40	-0.40	-2.7	-0.7	-2.29	-0.29

AGE_{H_1}	AGE_{H_2}	AGE_{H_3}	BMI_{H_1}	BMI_{H_2}	BMI_{H_3}
			N(25, σ=1)	N(25, σ=5)	N(25, σ=7)
			N(23, σ=1)	N(25, σ=5)	N(27, σ=7)
N(50, σ=5)	N(50, σ=8)	N(50, σ=3)	N(20, σ=1)	N(25, σ=5)	N(30, σ=7)
			N(28, σ=1)	N(30, σ=5)	N(32, σ=7)
			N(25, σ=1)	N(30, σ=5)	N(35, σ=7)
			N(25, σ=1)	N(25, σ=5)	N(25, σ=7)
			N(23, σ=1)	N(25, σ=5)	N(27, σ=7)
N(47, σ=5)	N(50, σ=8)	N(53, σ=3)	N(20, σ=1)	N(25, σ=5)	N(30, σ=7)
			N(28, σ=1)	N(30, σ=5)	N(32, σ=7)
			N(25, σ=1)	N(30, σ=5)	N(35, σ=7)
			N(25, σ=1)	N(25, σ=5)	N(25, σ=7)
			N(23, σ=1)	N(25, σ=5)	N(27, σ=7)
N(40, σ=5)	N(50, σ=8)	N(60, σ=3)	N(20, σ=1)	N(25, σ=5)	N(30, σ=7)
			N(28, σ=1)	N(30, σ=5)	N(32, σ=7)
			N(25, σ=1)	N(30, σ=5)	N(35, σ=7)
			N(25, σ=1)	N(25, σ=5)	N(25, σ=7)
			N(23, σ=1)	N(25, σ=5)	N(27, σ=7)
N(57, σ=5)	N(60, σ=8)	N(63, σ=3)	N(20, σ=1)	N(25, σ=5)	N(30, σ=7)
			N(28, σ=1)	N(30, σ=5)	N(32, σ=7)
			N(25, σ=1)	N(30, σ=5)	N(35, σ=7)
			N(25, σ=1)	N(25, σ=5)	N(25, σ=7)
			N(23, σ=1)	N(25, σ=5)	N(27, σ=7)
N(50, σ=5)	N(60, σ=8)	N(70, σ=3)	N(20, σ=1)	N(25, σ=5)	N(30, σ=7)
			N(28, σ=1)	N(30, σ=5)	N(32, σ=7)
			N(25, σ=1)	N(30, σ=5)	N(35, σ=7)

PS Model:
just AGE adjustment

almost **unbiased** when

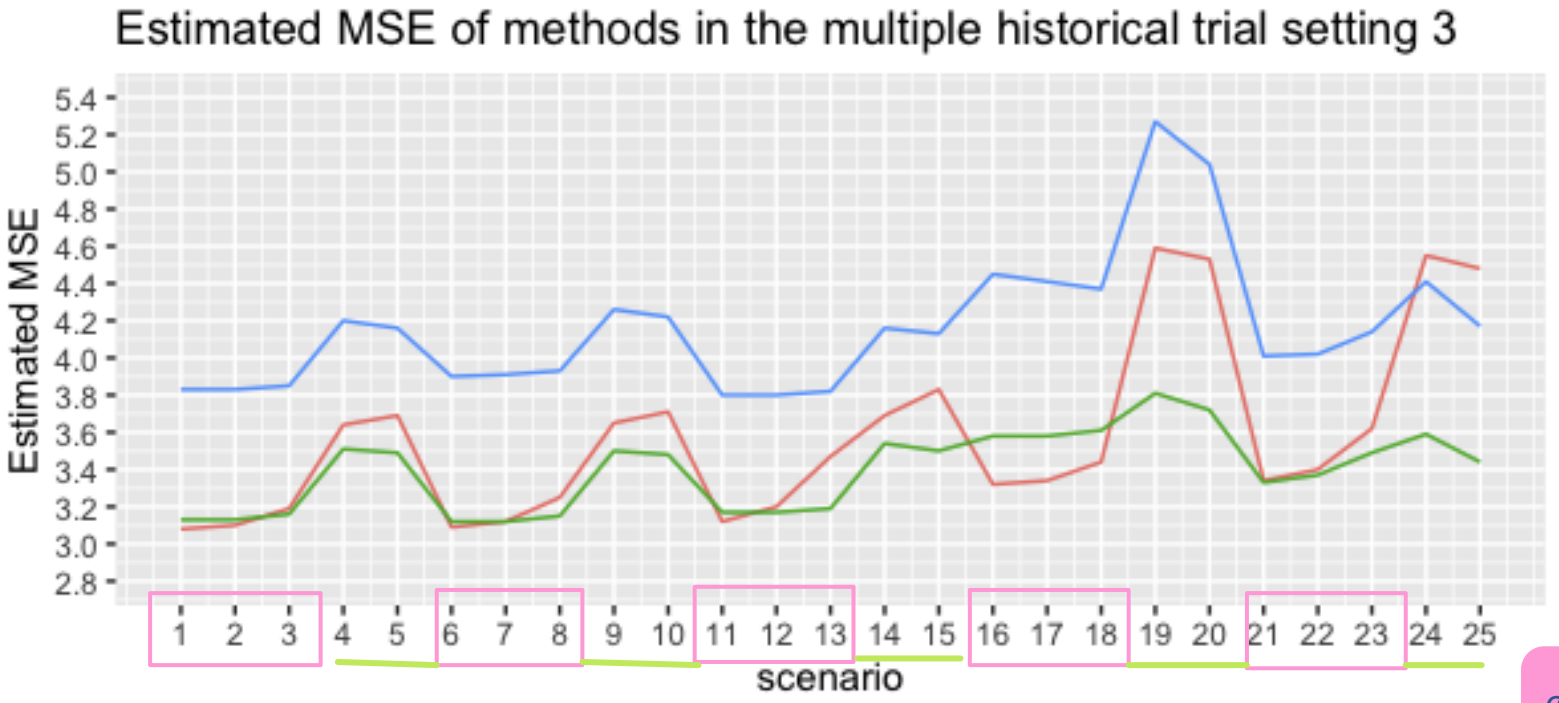
$$\emptyset\left(\mu_{BMI_{H_1}}, \mu_{BMI_{H_2}}, \mu_{BMI_{H_3}}\right) = 25 = \mu_{BMI_{Current}}$$

(except for scenarios 16+23 / 18+23)

b- MSE estimate

$$MSE = Var + Bias^2$$

$$\sigma_{AGE_{H_1}} = 5,\ \sigma_{AGE_{H_2}} = 8,\ \sigma_{AGE_{H_3}} = 3$$
$$\sigma_{BMI_{H_1}} = 1,\ \sigma_{BMI_{H_2}} = 5,\ \sigma_{BMI_{H_3}} = 7$$



- $\widehat{MSE}_{DBB}$ very similar to case I
  → invariant to σ of covariates
- again $PSMATCH$ & $PSMAP$ similar $\widehat{MSE}$ behaviour, but $\widehat{MSE}_{PSMATCH}$ almost everywhere highest

$\emptyset\ \mu_{BMI_{Historical}} = 25$

$\emptyset\ \mu_{BMI_{Historical}} = 30$

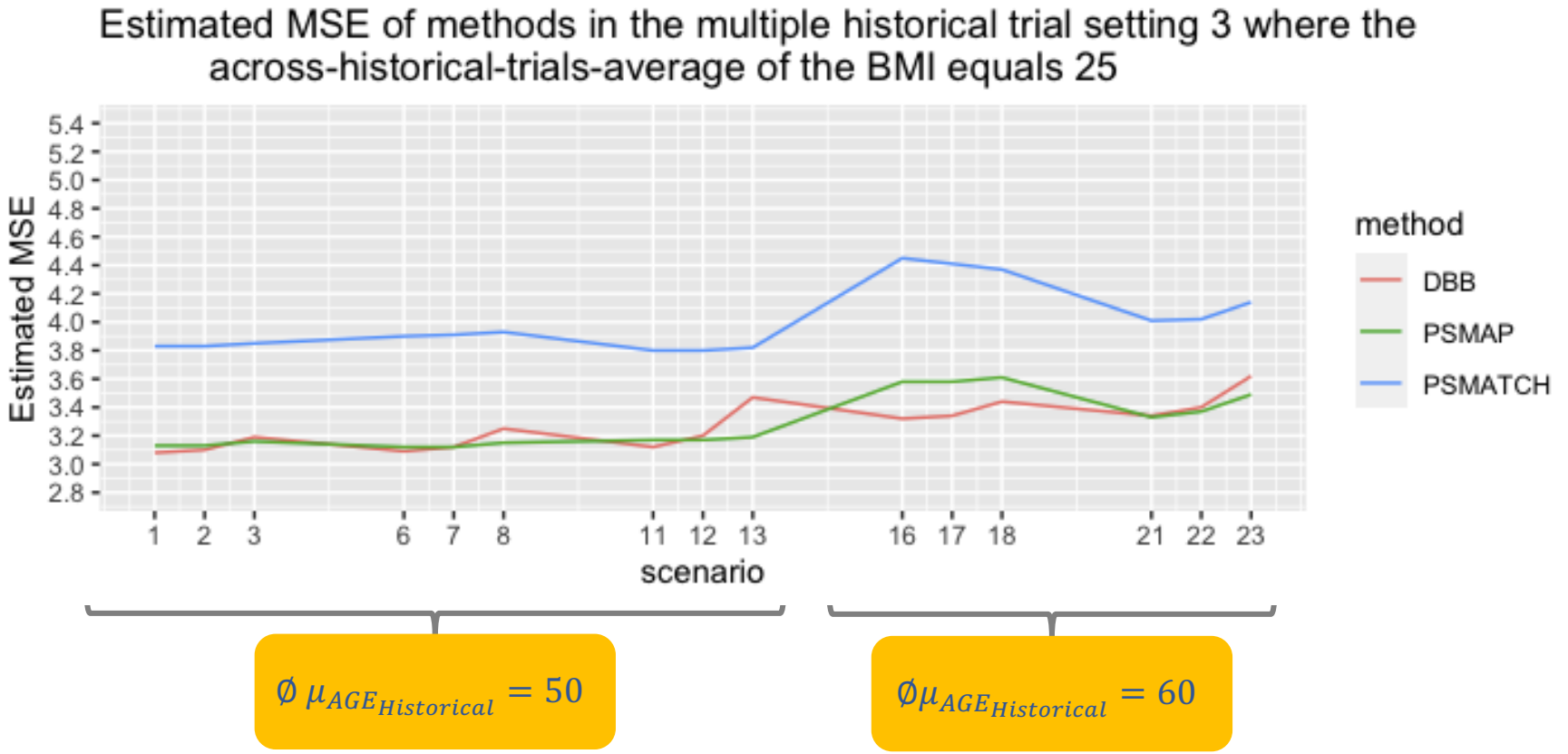
b- MSE estimate

$\emptyset\ \mu_{BMI_{Historical}} = 25$

$MSE = Var + Bias^2$

$\sigma_{AGE_{H_1}} = 5,\ \sigma_{AGE_{H_2}} = 8,\ \sigma_{AGE_{H_3}} = 3$

$\sigma_{BMI_{H_1}} = 1,\ \sigma_{BMI_{H_2}} = 5,\ \sigma_{BMI_{H_3}} = 7$



- $\widehat{MSE}$ of DBB and $PSMAP$ comparable
- when $\emptyset\mu_{AGE_H}$=60 the MSE estimates increase

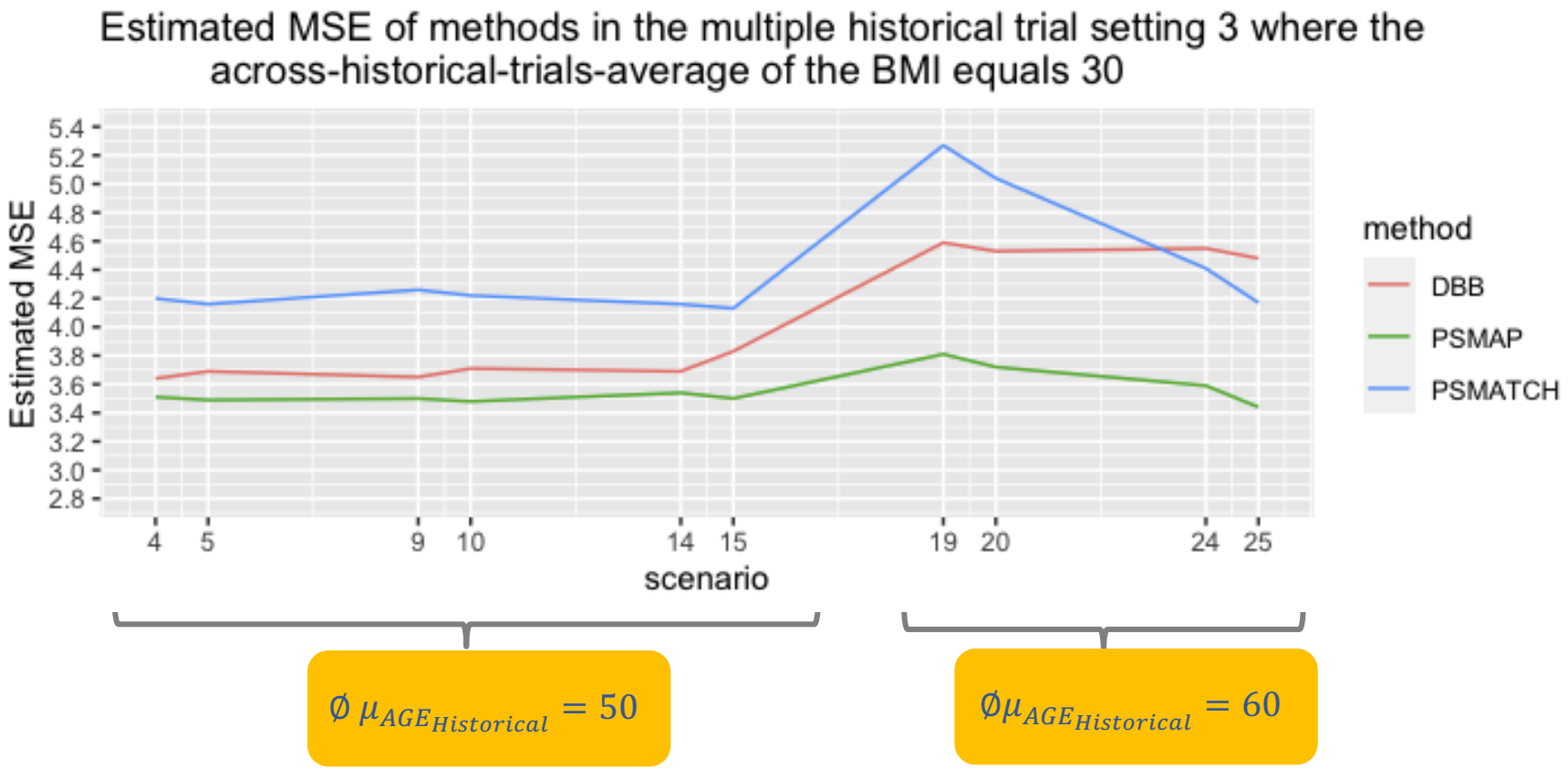
b- MSE estimate

$\emptyset\ \mu_{BMI_{Historical}} = 30$

$MSE = Var + Bias^2$

$\sigma_{AGE_{H_1}} = 5,\ \sigma_{AGE_{H_2}} = 8,\ \sigma_{AGE_{H_3}} = 3$

$\sigma_{BMI_{H_1}} = 1,\ \sigma_{BMI_{H_2}} = 5,\ \sigma_{BMI_{H_3}} = 7$

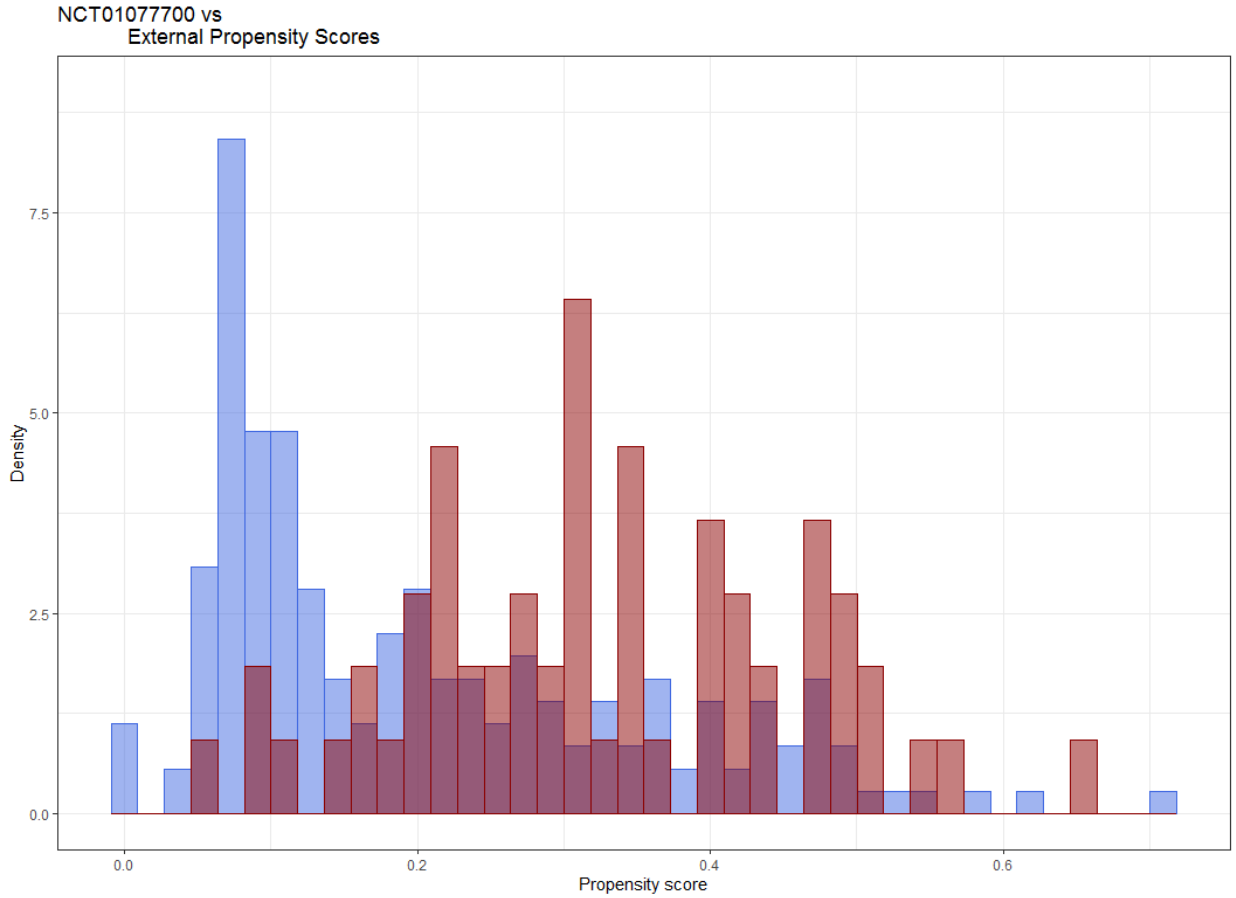


$\emptyset\ \mu_{AGE_{Historical}} = 50$

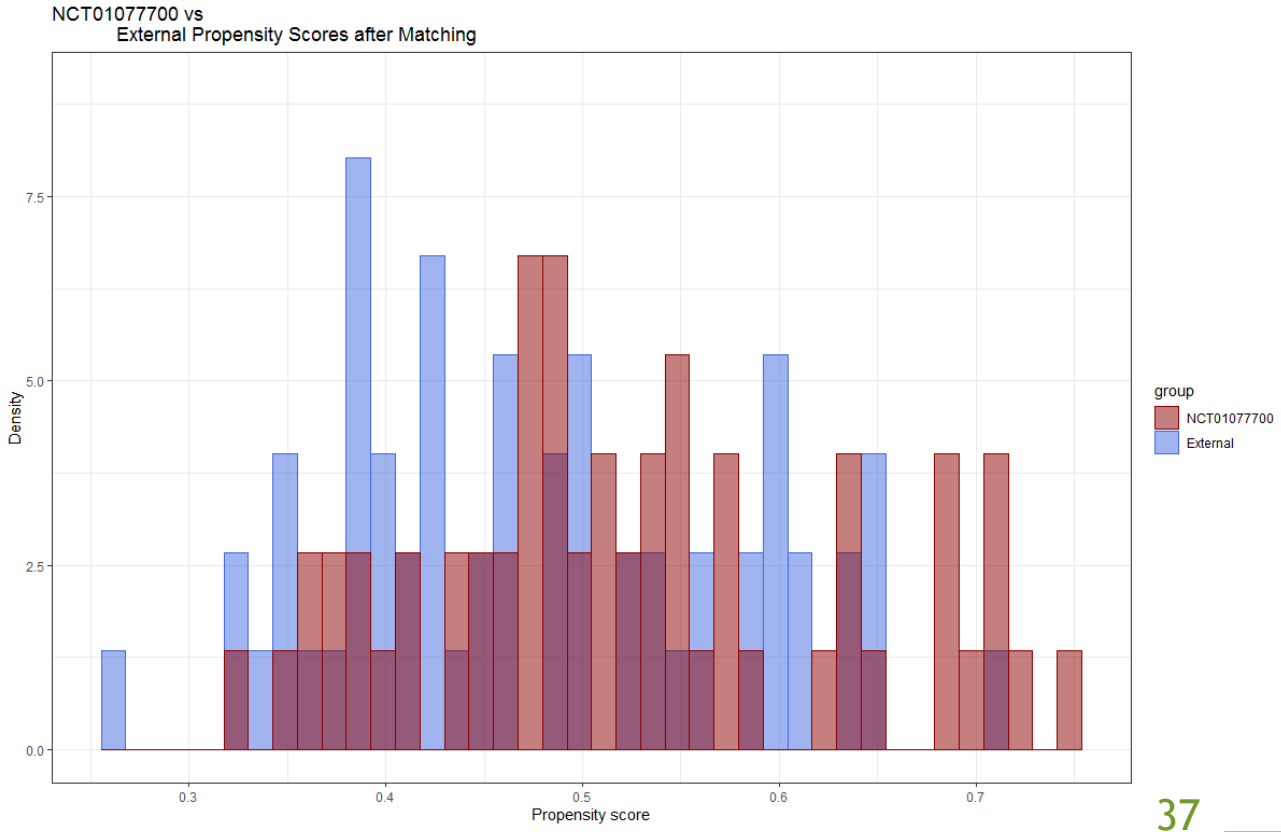
$\emptyset \mu_{AGE_{Historical}} = 60$

- MSE estimates higher than before
- from 19/20 to 24/25: $\widehat{MSE}_{DBB}$ no decrease compared to $\widehat{MSE}_{PSMATCH}$ & $\widehat{MSE}_{PSMAP}$

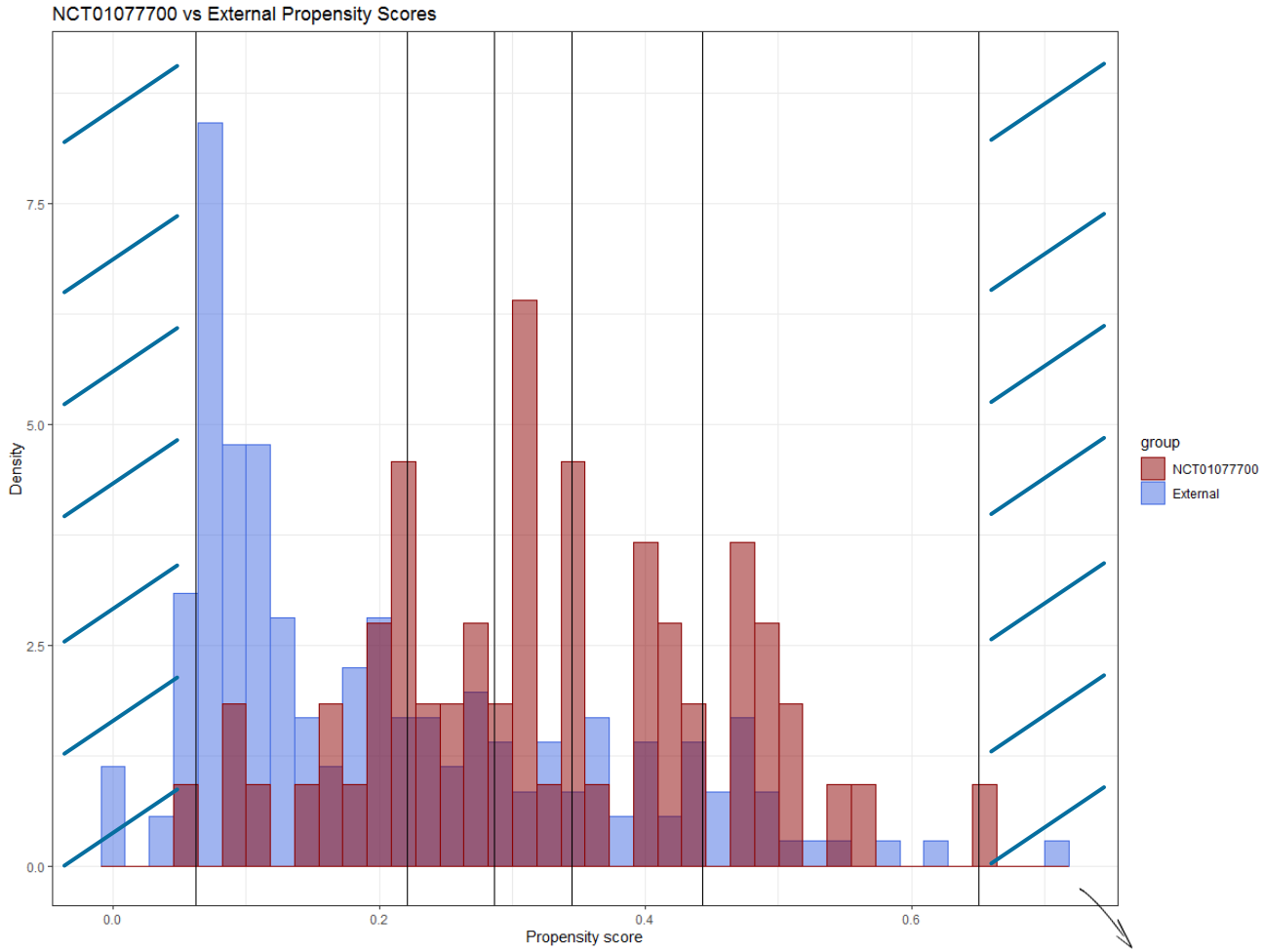
Schizophrenia



PS distribution after 1:1 PS Matching (60:60)



Schizophrenia



PS-MAP

- Overlapping coefficients v_s: 0.6, 0.66, 0.76, 0.71, 0.8
- Strata weights ω_s: 0.17, 0.19, 0.21, 0.2, 0.23

Current	12	12	12	12	12
Historical	107	19	15	2	17

Trimming region
≈ 20 trimmed

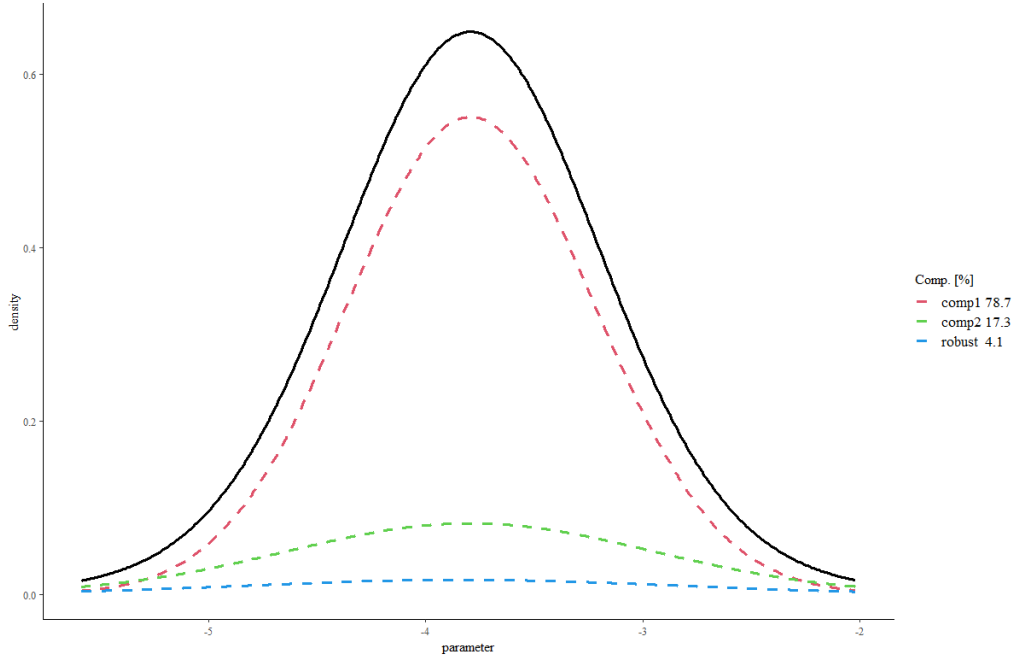
Mixture Distribution

Priors $p_{RMAP}(\theta_{0*})$ and $p_{weak}(\theta_{1*})$ are approximated via **mixture distributions**

$$p_{RMAP}(\theta_{0*}) \approx \hat{p}(\theta_{0*} \mid Y_{0*}) = \sum_{k=1}^{K_0} w_{0,k} f_{0,k}(m_{0,k}), \mathrm{K}_0 \in N$$

$$p_{weak}(\theta_{1*}) \approx \hat{p}(\theta_{1*} \mid Y_{1*}) = \sum_{k=1}^{K_1} w_{1,k} f_{1,k}(m_{1,k}), \mathrm{K}_1 \in N$$

Mixture distributions



Increase standard deviation of mixture components to reach target ESS.

Effective Sample Size (ESS)

ESS represents the **amount of information** the **MAP prior** contains

Expected Local Information Ratio ESS

$$ESS_{ELIR} = E_\theta\big(r(\theta)\big) = E_\theta\left[\frac{i(p(\theta))}{i_F(\theta)}\right] < min\left(\left(\frac{\sigma}{\tau}\right)^2, N\right)$$

prior information
If prior density $p_k(\theta) \sim N(\theta|m, s^2)$,
then $i\big(p_k(\theta)\big) = \frac{1}{s^2}$

Fisher information
If sampling distribution $f(Y|\theta) \sim N(Y|\theta, \sigma^2)$,
then $i_F(\theta) = \frac{1}{\sigma^2}$

within-trial to between-trial variance ratio; $\in \{256,64,16,4,1\}$ depending on assumed between-trial heterogeneity

number of available historical patients

superior to other ESS versions

Predictive consistency: expected posterior ESS = prior ESS + sample size N

Neuenschwander, B., Weber, S., Schmidli, H., & O'Hagan, A. (2020). Predictively consistent prior effective sample sizes. *Biometrics*, *76*(2), 578–587.
Weber, S., Li, Y., Seaman, J. W., Kakizume, T., & Schmidli, H. (2021). Applying meta-analytic-predictive priors with the R Bayesian evidence synthesis tools. *Journal of Statistical Software*, *100*(19), 1–32. https://doi.org/10.18637/jss. v100.i19
Neuenschwander, B., Capkun-Niggli, G., Branson, M., & Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials*, *7*(1), 5–18.

Logistic Regression

$X = (X_1, \ldots, X_d)$ denotes the d baseline characteristics of patients

Propensity Score

$$e(X) := P(Z = 1|X) := \frac{exp^{\beta_0+\beta_1 X_1+\cdots+\beta_d X_d}}{1+exp^{\beta_0+\beta_1 X_1+\cdots+\beta_d X_d}}$$

estimate $e(X)$ by

$$\hat{e}(X) = \frac{exp^{\hat{\beta}_0+\hat{\beta}_1 X_1+\cdots+\hat{\beta}_d X_d}}{1+exp^{\hat{\beta}_0+\hat{\beta}_1 X_1+\cdots+\hat{\beta}_d X_d}}$$

logit of $e(X) := log\left(\frac{e(X)}{1-e(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_d X_d$ is the linear relationship of covariates

to get estimates $\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_d$ maximize $\prod_{i:z_i=1} \hat{e}(X_i) \prod_{j:z_j=0} (1 - \hat{e}(X_j))$ via **Maximum-Likelihood approach**

James, G., Witten, D., Hastie, T., Tibshirani, R., James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). Statistical learning. *An introduction to statistical learning: with applications in R*, 15–57.

Nearest Neighbor Matching Algorithm

Data $\mathbf{D} \coloneqq \{\boldsymbol{Y_{1*}}, \boldsymbol{Y_{0*}}, \boldsymbol{Y_{hist}}\}$ where Y_{1*} contains current treated patients, Y_{0*} current control patients and Y_{hist} historical control patients

→ use data D to **estimate PSs** $\hat{\boldsymbol{e}}(\boldsymbol{X_i})$ for all patients in D

Goal: match patients $\in Y_{hist}$ to Y_{0*}

Define **distance** $d_{ij} \coloneqq \left|\underbrace{logit\left(\hat{e}(X_i)\right)}_{\coloneqq log\left(\frac{\hat{e}(X_i)}{1-\hat{e}(X_i)}\right)} - logit\left(\hat{e}(X_j)\right)\right|$ $i \in Y_{0*}, j \in Y_{hist}$ and **caliper width** $c \coloneqq 0.2 \cdot \sqrt{\frac{\sigma^2_{Y_{0*}} + \sigma^2_{Y_{hist}}}{2}}$

Order the logit of the PS in Y_{0*} in descending order

Zhao, Q.-Y., Luo, J.-C., Su, Y., Zhang, Y.-J., Tu, G.-W., & Luo, Z.(2021).Propensity score matching with r:Conventional methods and new features. *Annals of translational medicine*,*9*(9).

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when es- timating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, *10*(2), 150–161.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Nearest Neighbor Matching Algorithm

Algorithm 1 Nearest Neighbor Matching

Define vector *matched* with length $|Y_{0*}|$
Set caliper c
for $i = 1 \rightarrow |Y_{0*}|$ **do**
    Set distance $d_{i,opt} = \infty$ and $matched[i] = NA$
    for $j \in 1 \rightarrow |Y_{hist}|$ **do**
        Calculate d_{ij}
        if $d_{ij} \leq c_{opt}$ **then**
            if $d_{ij} \leq d_{i,opt}$ **then** $d_{i,opt} = d_{ij}$ and matched$[i] = j$
            else$d_{i,opt} = d_{i,opt}$
            end if
        else$d_{i,opt} = d_{i,opt}$
        end if
    end for
    if $d_{i,opt} < \infty$ **then** $matched[i] = j$, $Y_{hist} = Y_{hist} \setminus \{j\}$, update c
    end if
end for
return $matched$

1:1 matching without replacement with caliper

Zhao, Q.-Y., Luo, J.-C., Su, Y., Zhang, Y.-J., Tu, G.-W., & Luo, Z.(2021).Propensity score matching with r:Conventional methods and new features. *Annals of translational medicine*,*9*(9).
Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when es- timating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, *10*(2), 150–161.
Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1).

MCMC Algorithm

An MCMC algorithm is used within RBesT::gMAP() and by extracting infomation from historical information a MCMC sample is returned

Let $\theta^1, \theta^2, \ldots$ be independent sequences drawn by starting at θ^0,
θ^t denotes a previous sequence
Transition distribution

$$T_t(\theta^t|\theta^{t-1})$$

based on

is constructed such that **Markov Chain** converges to posterior distribution $p(\theta|Y)$

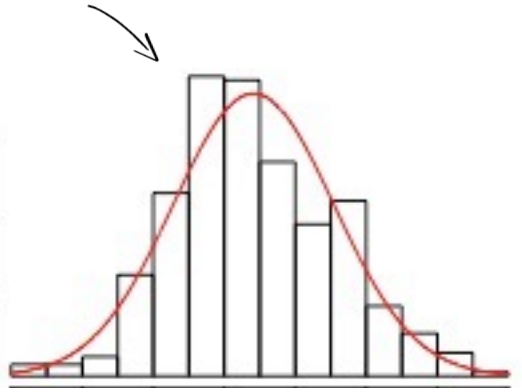
Stan software uses an **MCMC algorithm** & samples via a **Hamiltonian Monte Carlo (HMC) simulation algorithm** given a Bayesian model

MCMC – Markov Chain Monte Carlo

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis (vol. 2).

EM Algorithm

After RBesT::gMAP() returns a MCMC sample, the **EM algorithm** within RBesT::automixfit() **fits** a **parametric distribution.**



It is given observed data Y, parameters w, a and b and the density

number of observations

number of components

$$p(Y|w,a,b) = \sum_{i=1}^{N}\sum_{k=1}^{K} w_k p_k(Y_i|a_k,b_k).$$

component weights

k^{th} component

Numerically problematic to maximize $log\ p(Y|w,a,b)$ via Maximum-Likelihood approach
→ extension to $(Y, \widetilde{Y})$

This leads to $\boldsymbol{E_{\widetilde{Y}|Y,w^{(n)},a^{(n)},b^{(n)}}\ \log\left(p\left(Y,\widetilde{Y}\middle|w,a,b\right)\right) =: Q\left(w,a,b\middle|w^{(n)},a^{(n)},b^{(n)}\right) =: Q_n(w,a,b)}$

EM – Expectation Maximization

Weber, S., Li, Y., Seaman, J. W., Kakizume, T., & Schmidli, H. (2021). Applying meta-analytic-predictive priors with the R Bayesian evidence synthesis tools. *Journal of Statistical Software*, *100*(19), 1–32. https://doi.org/10.18637/jss. v100.i19

EM Algorithm

After initial guess of parameters, steps [E] and [M]

[E]: Calculate $\boldsymbol{Q_n(w, a, b)}$
[M]: find the **parameters** $\boldsymbol{w^{(n+1)}}$, $\boldsymbol{a^{(n+1)}}$ **and** $\boldsymbol{b^{(n+1)}}$ that maximize $Q_n(w, a, b)$
such that $\sum_{k=1}^{K} w_k^{n+1} = 1$
(→ Lagrange technique)

are repeated until $Q^{(n)}(w, a, b)$ converges to a maximum.

EM – Expectation Maximization

Weber, S., Li, Y., Seaman, J. W., Kakizume, T., & Schmidli, H. (2021). Applying meta-analytic-predictive priors with the R Bayesian evidence synthesis tools. *Journal of Statistical Software*, *100*(19), 1–32. https://doi.org/10.18637/jss. v100.i19