U.S. FOOD & DRUG ADMINISTRATION

Propensity Score-Based Methods for Causal Inference and External Data Leveraging in Regulatory Clinical Studies

Lilly Yue, Ph.D.
Heng Li, Ph.D.
Division of Biostatistics, FDA/CDRH

June 22, 2023

Outline

- Propensity score-based methods for the regulatory clinical studies
- Application of propensity score-based methods for causal inference and external data leveraging in hybrid clinical studies
- Concluding Remarks

Propensity Score-Based Methods

- Traditional Propensity score (PS) methods - For causal inference
  - Matching on propensity scores
  - Stratification on propensity scores
  - Inverse probability of treatment weighting using propensity scores
- Propensity score-integrated approaches - For augmenting a prospective study with external data
  - Propensity score-integrated Power prior – Bayesian
  - Propensity score-integrated Composite likelihood – Frequentist
- External data – data external to traditional clinical study being planned, such as Real-world data or historical clinical study data

Adapting and Advancing Propensity Score Methodology in the Regulatory Settings

- Propensity score methodology - A ground-breaking statistical innovation for the *design* and *analysis* of observational studies, developed by Rosenbaum and Rubin in 1983.
- In 2002, it was adopted first by FDA/CDRH, for pre-market *confirmatory* non-randomized medical device studies (Yue, LQ, 2007, *JBS*).
- Around 2013, the 2-stage outcome-free PS study design framework was proposed to ensure the study integrity and transparency.
- Since 2018, the propensity score-integrated approaches, Bayesian and Frequentist, have been developed to augment a prospective study.
- It has been utilized for post-market safety evaluation of drugs and devices (Levenson and Yue, 2013).

Traditional Propensity Score Methodology

- Propensity score (PS): Conditional probability of receiving treatment A rather than treatment B, given a collection of observed baseline covariates.
- Replace the collection of confounding covariates with one scalar function of these covariates: the propensity score.
- Goal: Simultaneously balance many observed covariates between the two treatment groups, and then reduce bias in treatment comparison on outcomes.
- Propensity score is a balancing score - Conditional on the propensity score, the distribution of observed baseline covariates is the same between the two treatment groups.
- PS estimation: Statistical modeling of relationship between treatment group membership and covariates.

Limitations

- Propensity score methods can only adjust for observed confounding covariates and not for unobserved ones.
- Propensity score is seriously degraded when important variables influencing treatment assignment and outcome have not been collected or included in PS model.
  - E.g., variable in data source level - clinical practice, availability of adjunct therapy, data collection, region, time of data collection.
- Propensity score may not eliminate all selection bias.
- It may not work when there are significant amount of missing data in covariates.

6

When Does It Work Better?

Braitman and Rosenbaum (2002):

- When outcome event is rare.
- When there are a large number of subjects in each treatment group.
- When there are many covariates observed.

Braitman, L., Rosenbaum, P. R. (2002). Rare outcomes, common treatments: Analytic strategies using propensity scores. Ann. Intern. Med. 137:693–696.

7

Propensity Score-Integrated Approaches

- Propensity score-integrated Power prior – Bayesian
- Propensity score-integrated Composite likelihood – Frequentist

for augmenting a prospective study, single-arm or randomized controlled trial (RCT), with external data

Propensity Score Redefined

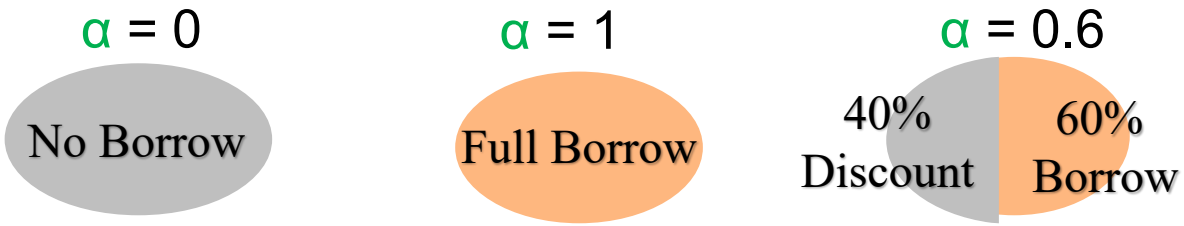
- Traditional propensity score: Conditional probability of receiving treatment A rather than treatment B, given a collection of observed baseline covariates.
  - Used to balance patient characteristics (covariates) between the two treatment groups.
  - For causal inference
- Re-defined propensity score: Conditional probability of being in the traditional study rather than in the external data source, given patient baseline covariates.
  - Used to balance patient characteristics (covariates) between the two patient groups: traditional study patients and external patients.
  - For augmenting a prospective study

Power Prior (Bayesian)

- A power prior is constructed as

$$\pi(\boldsymbol{\theta}|D_0, \alpha) \propto [L(\boldsymbol{\theta}|D_0)]^{\alpha} \pi_0(\boldsymbol{\theta})$$

  - $\boldsymbol{\theta}$ **:** parameter of interest
  - $L(\boldsymbol{\theta}|D_0)$: likelihood function of the external data
  - $\pi_0(\boldsymbol{\theta})$: initial prior distribution for $\boldsymbol{\theta}$
  - α: *power prior parameter,* $0 \leq \boldsymbol{\alpha} \leq 1$
  - α controls how much external data to borrow



- Q: how and when to determine α for a prospective study?

Ref. Chen, M-H and Ibrahim, J.G., (2000) Power Prior Distribution for Regression Models. Statistical Science, 15(1): 46-60

Composite Likelihood (Frequentist)

- General form (weighted product of probability density functions):

$$L(\theta|Y) = \prod_i f(y_i|\theta)^{\lambda_i}$$

where λ_i is nonnegative weight to be chosen, and can be used to down-weight external patient info.

- Set:
  - $\lambda_i = 1$, if the patient *i* is from the traditional study
  - $0 < \lambda_i \leq 1$, if the patient *i* is from the external data source

Q: how and when to determine λ for a prospective study?

Ref. Lindsay, BG (1988). Composite likelihood method. *Contemporary mathematics*, 80(1): 221-239. Varin et al (2011). An overview of composite likelihood methods. *Statistica Sinica*, P5-42.

Propensity Score-Integrated Approaches for Augmenting

- Propensity score methodology
  - Study design
    - PS estimation
    - Using estimated PS, to
      - ✓ Select comparable patients from external data source
      - ✓ Determine how much info to borrow from each external patient, i.e., determine α in power prior or λ in composite likelihood.

- Power prior or composite likelihood method
  - Outcome analysis
  - Down-weight external data when needed.

Outline

- Propensity score-based methods for the regulatory clinical studies
- Application of propensity score-based methods for causal inference and external data leveraging in hybrid clinical studies
  - Hybrid study - data consist of two parts:
    - Data on patients prospectively enrolled in an investigational clinical study
    - External data
- Concluding Remarks

Three Types of Hybrid Study

Type 1: Non-randomized comparative study with external control group

to approximate a traditional randomized controlled trial (RCT)

Type 2: Single-arm hybrid study

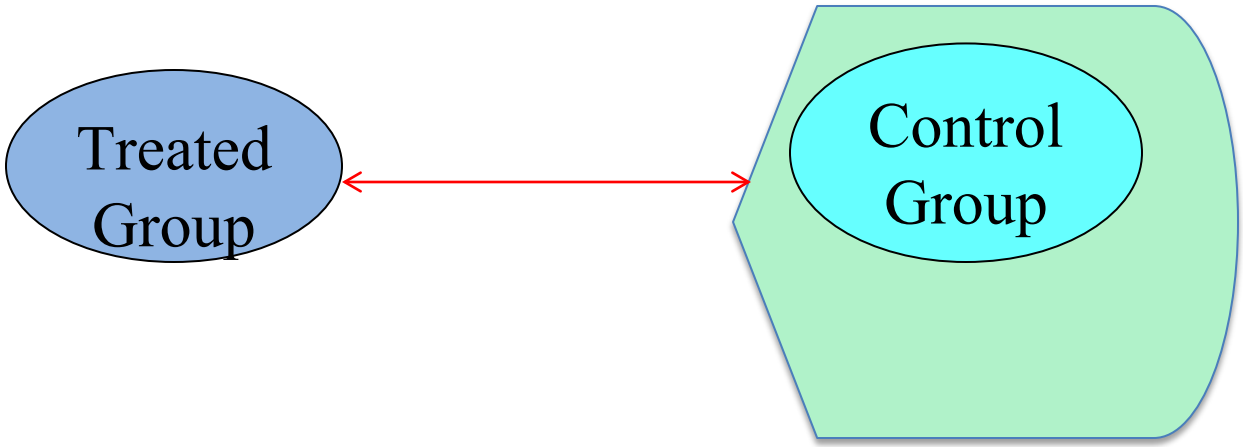
to approximate a traditional single-arm study

Type 3. RCT with hybrid control group

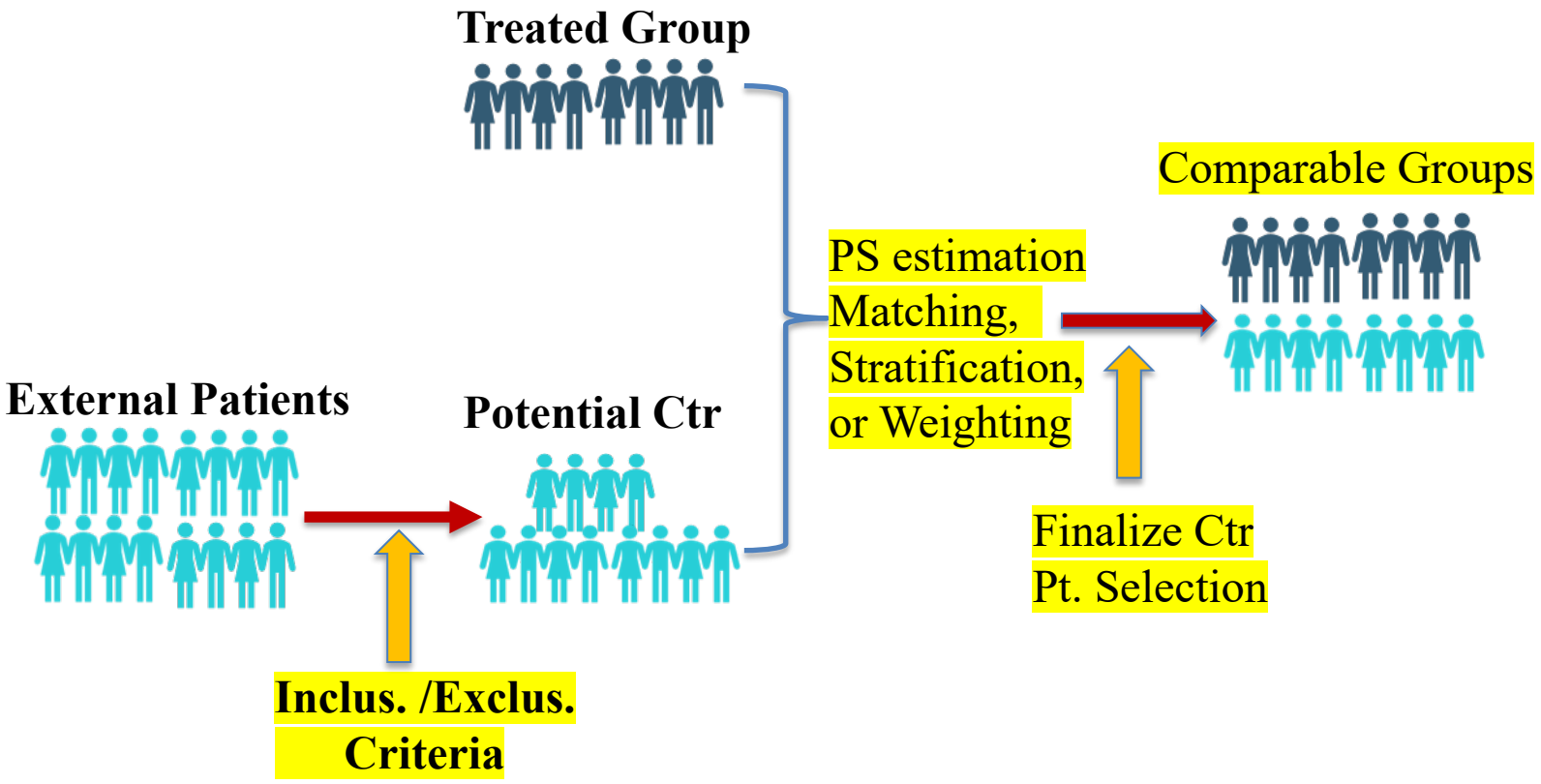
to approximate a traditional RCT

Type 1 Hybrid Study: Construct a Control Group

- Using PS, identify patients and construct a control group from external patients, based on patient baseline characteristics (not outcomes!) such that
  - the distribution of observed baseline covariates is similar between the treated and control patients
  - leading to comparable treatment groups in terms of baseline covariates.



External Control Group Construction



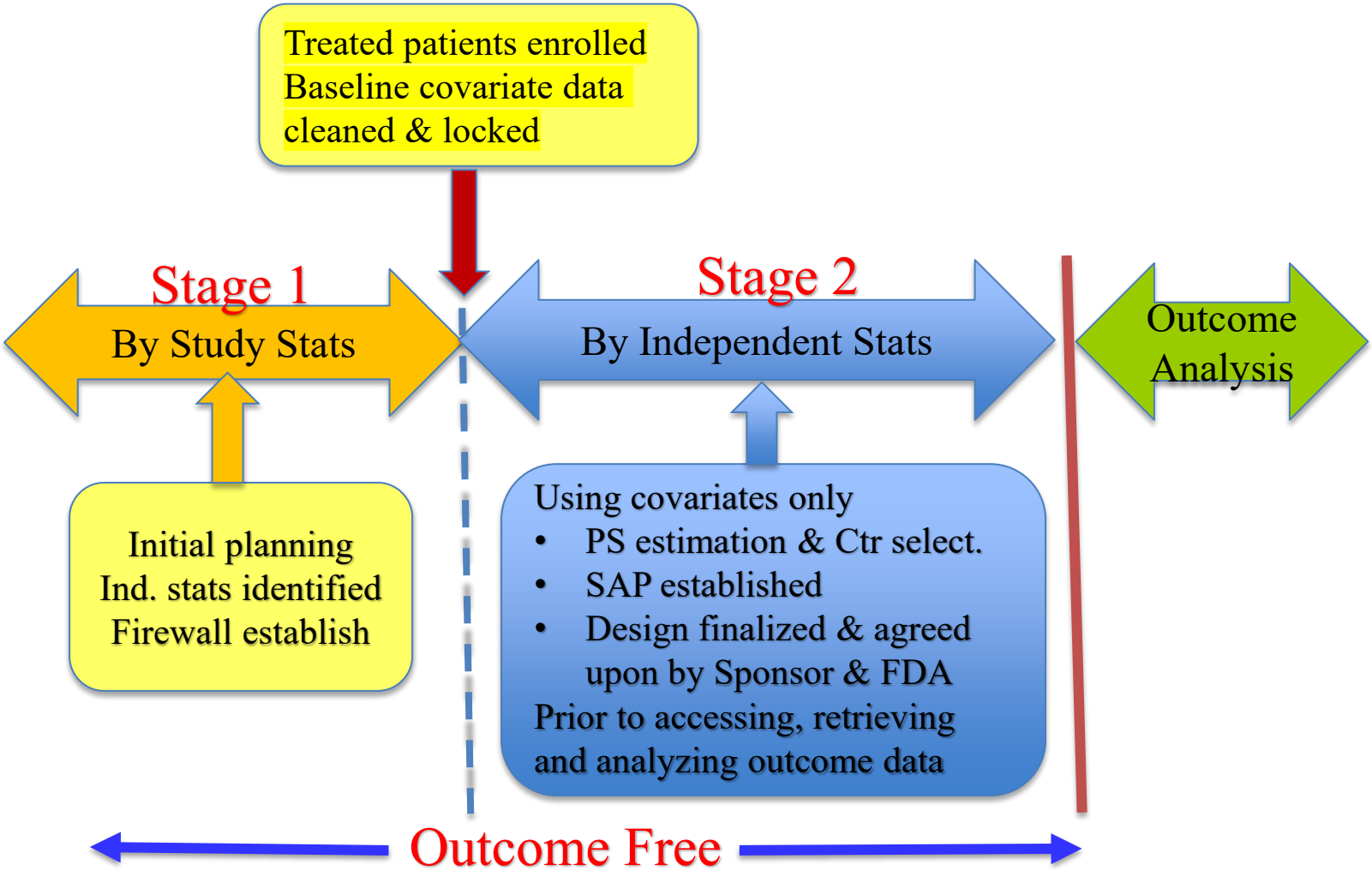
A Critical Question – Study Integrity

- Given that in many cases when designing a prospective study, data on external patients have already been collected and clinical outcomes are already available,
- And given the iterative nature of PS design,

Q: How to ensure the scientific validity of the study design and the interpretability of study results?

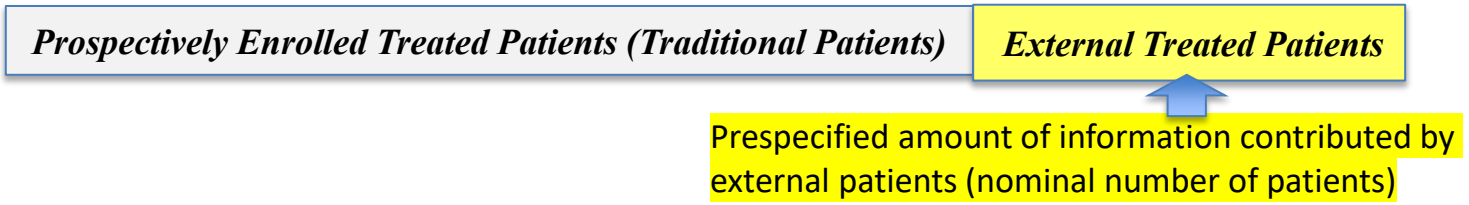
A: Outcome data need to be kept out of sight during study design.

2-stage Outcome-Free Study Design Using PS



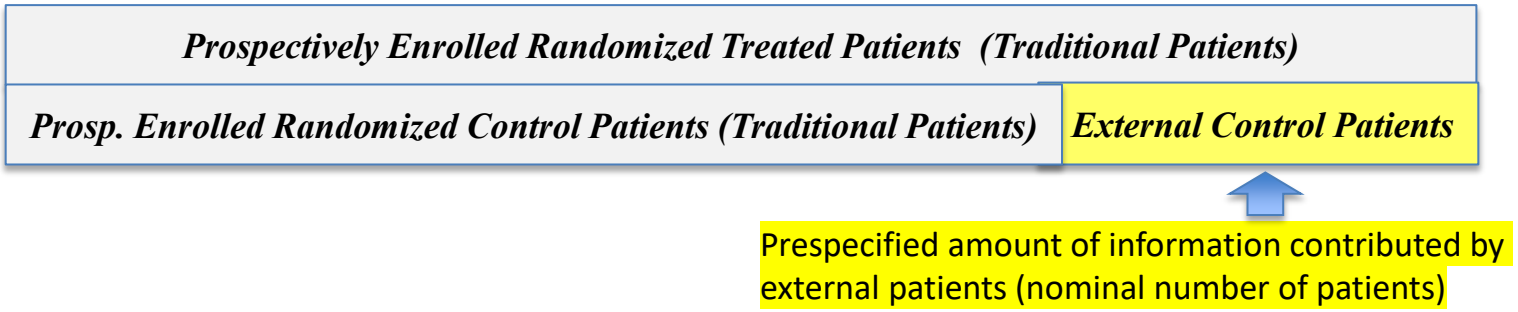
Data Leveraging: Type 2 and Type 3 Hybrid Studies

Type 2: Single-arm study leveraging external patients



to approximate a traditional single-arm study (target study)

Type 3: RCT leveraging external patients



to approximate a traditional RCT (target study)

Balancing Covariates for Data Leveraging

- What if external patients and traditional patients "look different"?

  Looking different means:

  The distribution of patient characteristics among the external patients

  $\neq$

  The distribution of patient characteristics among the traditional patients

- Some statistical adjustment is needed for a hybrid study to approximate the target study well ---- covariate balance.

Redefining Propensity Score for a Hybrid Study

- Propensity score (PS) (modified definition): Conditional probability of being a traditional patient rather than an external patient, given a collection of observed baseline covariates (applicable to data leveraging)
- Propensity score stratification with 5 strata
- Yellow---external patients, grey---traditional patients
- With equal number of traditional patients in each stratum
- Using PS quintiles among traditional patients as boundaries for PS strata

- 2-stage outcome-free propensity score design framework

Summary of Propensity Score Design

Determination of α or λ for a type 3 hybrid study

	PS Stratum					
	1	2	3	4	5	Total
Traditnl. Pts (n)	54	53	53	53	54	267
Treated (n)	41	28	39	36	39	183
Control (n)	13	25	14	17	25	84
External Pts (n)	**332**	270	233	201	156	1192
Leveraged Info. (n)	**19**	17	17	16	18	87
α or λ	**0.06**	**0.06**	**0.08**	**0.08**	**0.11**	

Concluding Remarks

- Novel statistical methods play a critical role in leveraging external data to support regulatory decisions
- Propensity score-based methods can be applied to the design and analysis of all three types of hybrid study
- Propensity score-integrated approaches have drawn a great deal of attention
- Software for implementation is available

References

- Li, H., Yue, L.Q. (2023) Propensity Score-Based Methods for Causal Inference and External data Leveraging in Regulatory Settings: From Basic Ideas to Implementation. Pharmaceutical Statistics. https://onlinelibrary.wiley.com/doi/10.1002/pst.2294
- Li, H. and Yue, L.Q. (2023) Clinical Studies Leveraging Real-World Data Using Propensity Score-Based Methods. In W. He, Y. Fang & H. Wang ed. Real-World Evidence in Medical Product Development. Springer.
- Austin, P. (2011). An introduction to propensity score methods for reducing the effect of confounding in observational studies, *Multivariate Behavioral Research,* 46:399-424
- Rosenbaum, PR, Rubin DB (1984). Reducing bias in observational studies using subclassification on the propensity score. *JASA*, 79:516-524.
- Rubin, DB (1997). Estimating casual effects from large data sets using propensity scores. *Ann Intern Med*, 127:757-763.
- Rubin, D.B. (2001). Using propensity score to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology,* **2**, 169-188.
- Rubin, D.B. (2007). The design versus the analysis of observational studies for causal effects: Parallel with the design of randomized trials. *Statistics in medicine*, **26**: 20-36.
- Rubin, D.B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2 (3), 808-840.

References (cont.)

- Yue, LQ., Lu, N. and Xu, Y. (2014). Designing pre-market observational comparative studies using existing data as controls: challenges and opportunities. *Journal of Biopharmaceutical Statistics* 24:994-1010.
- Yue, Q.L., Campbell, G., Lu, N., Xu, Y., & Zuckerman, B. (2016) Utilizing national and international registries to enhance pre-market medical device regulatory evaluation. *Journal of Biopharmaceutical Statistics,* 26 (6), 1136–1145.
- Li, H., Mukhi, V., Lu, N., Xu, Y. & Yue, Q.L. (2016). A note on good practice of objective propensity score design for premarket nonrandomized medical device studies with an example. Statistics in Biopharmaceutical Research, 8(3), 282-286
- Xu, Y., Lu, N., Yue, L.Q., Tiwari, R. (2019). A Study Design for Augmenting the Control Group in a Randomized Controlled Trial: A Quality Process for Interaction Among Stakeholders. *Therapeutic Innovation & Regulatory Science*.
- Imbens, G., Rubin, D.B. (2015) Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, New York.

References (cont.)

- Augmenting a single-arm investigational study with external data (type 2 hybrid studies)

  - Wang, C., Li, H., Chen, W., Lu, N., Tiwari, R., Xu, Y., Yue, L.Q. (2019). Propensity Score-Integrated Power Prior Approach for Incorporating Real-World Evidence in Single-Arm Clinical Studies. *JBS,* 29 (5),731-748.
  - Wang, C., Lu, N., Chen, W-C., Li, H., Tiwari, R., Xu, Y., Yue, L.Q. (2020). Propensity Score-Integrated Composite Likelihood Approach for Incorporating Real-World Evidence in Single-Arm Clinical Studies. *JBS,* 30(3), 495-507.
  - Lu, N., Wang, C., Chen, W-C., Li, H., Song, C., Tiwari, R., Xu, Y., and Yue, L.Q. (2022), Leverage Multiple Real-World Data Sources in Single-Arm Medical Device Clinical Studies, *Journal of Biopharmaceutical Statistics,* 32:1, 107-123, https://www.tandfonline.com/doi/full/10.1080/10543406.2021.1897994., *Published online: April 2021*
  - Chen, W-C., Lu, N., Wang, C., Li, H., Song, C., Tiwari, R., Xu, Y., Yue, LQ. (2022) Propensity Score-Integrated Approach to Survival Analysis: Leveraging External Evidence in Single-Arm Studies, *Journal of Biopharmaceutical Statistics,* VOL. 32, NO. 3, 400–413, https://doi.org/10.1080/10543406.2022.2080701

References (cont.)

- Augmenting the control and/or treated arm of an RCT with external data (type 3 hybrid studies)

  - Chen, W., Wang, C., Li, H., Lu, N. Tiwari, R., Xu, Y., Yue, L.Q.(2020). Propensity Score-Integrated Composite Likelihood Approach for Augmenting the Control Arm of a Randomized Controlled Trial by Incorporating Real-World Data. *JBS,* 30(3).
  - Lu, N., Wang, C., Chen, W-C., Li, H., Song, C., Tiwari, R., Xu, Y., **Yue, L.Q**. (2022). Propensity Score-Integrated Power Prior Approach for Augmenting the Control Arm of a Randomized Controlled Trial by Incorporating Multiple External Data Sources, *J of Biopharmaceutical Statistics,* 32:1, 158-169, *https://doi.org/10.1080/10543406.2021.1998098, Publication online: Nov 2021*
  - Li, H., Chen, WC., Wang, C. Lu, N., Song, C., Tiwari, R., Xu, Y., Yue, LQ. (2021). Augmenting Both Arms of a Randomized Controlled Trial Using External Data: An Application of the Propensity Score-Integrated Approaches. *Stat Biosci*. https://doi.org/10.1007/s12561-021-09315-5

References (cont.)

- Augmenting diagnostic studies with low prevalence disease
  - ✓ Chen, W-C., Li, H., Wang, C., Lu, N., Song, C., Tiwari, R., Xu, Y., Yue, L.Q. (2021), Evaluation of Diagnostic Tests for Low Prevalence Diseases: A Statistical Approach for Leveraging Real-World Data to Accelerate the Study, *JBS,* 31(1), 1-16. https://www.tandfonline.com/doi/full/10.1080//10543406.2021.1877724
  - ✓ Song, C., Li, H., Chen, W-C., Lu, N., Tiwari, R., Wang, C., Xu, Y., **Yue, LQ** (***Corresponding author)*** (2023). Principled Leveraging of External Data in the Evaluation of Diagnostic Devices via the Propensity Score-Integrated Composite Likelihood Approach, *Pharmaceutical Statistics,* DOI: 10.1002/pst.2295

- Mitigating study power loss caused by clinical trial disruptions due to the COVID-19 pandemic
  - ✓ Li, H., Chen, W.-C., Lu, N., Song, C., Wang, C., Tiwari, R., Xu, Y., and Yue, L.Q**.** (2022), Mitigating Study Power Loss Caused by Clinical Trial Disruptions Due to the COVID-19 Pandemic: Leveraging External Data via Propensity Score-Integrated Approaches, *Statistics in Biopharmaceutical Research*, VOL. 14, NO. 1, 87–91, https://www.tandfonline.com/doi/full/10.1080/19466315.2020.1860813. *Published online: 26 Jan 2021.*

2
9

Thank You!