RCTrep: An R package for the validation of methods for treatment effect estimation using real-world data

PRESENTER: **Lingjie Shen**[1*]

INTRO:

- **Who cares?** – policy makers; regulators; real world evidence (RWE) evaluators.
- **Why?** There is an increasing attention for the leverage of large real-world data (RWD) in treatment effect estimation to drive fast and precise decision making.
- **Challenge:** Since we do not observe the true treatment effect for each individual- which is the fundamental problem of causal inference - validation of treatment effect estimation methods using RWD is challenging.
- **Aim:** In the absence of a ground truth, how can we validate different methods using RWD to select the most reasonable method for the data at hand, driving fast regulatory and clinical decision making?

METHODS:

- We identify under which conditions the estimate from randomized control trial (RCT) can be regarded as the ground truth for methods validation using RWD. We illustrate differences between RCT and RWD in Figure 1. We assume the RWD and RCT data are two random samples from a, potentially different, population, and hence allow for a fair comparison of estimates of treatment effect between two samples after population composition is controlled for.
- We consider a set of candidate treatment effect estimators $\mathcal{F} = \{f_1, \dots, f_m\}$, where $f(\boldsymbol{x})$: $\mathcal{X} \mapsto \mathbb{E}[Y(1) - Y(0) \mid \boldsymbol{X} = \boldsymbol{x}]$, $f(\boldsymbol{x})$ is an estimator of conditional average treatment effect of population with characteristics $\boldsymbol{X} = \boldsymbol{x}$. We select the best one using the following evaluation metric:

$$f^* = \text{argmin}_{f \in \mathcal{F}} \mathbb{L}(\hat{\tau}; f) = \text{argmin}_{f \in \mathcal{F}} \left(\hat{\tau} - \sum_{\boldsymbol{x}} w(\boldsymbol{x}) f(\boldsymbol{x}) \right)^2, \quad s.t.\ p(\boldsymbol{x}) = q(\boldsymbol{x}) w(\boldsymbol{x})$$

where $\hat{\tau}$ is an unbiased estimate of average treatment effect of a population that a RCT represents, $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ are the empirical density of $\boldsymbol{x}$ in RCT data and RWD, $w(\boldsymbol{x})$ is a weight for individuals in RWD with characteristics $\boldsymbol{X} = \boldsymbol{x}$.

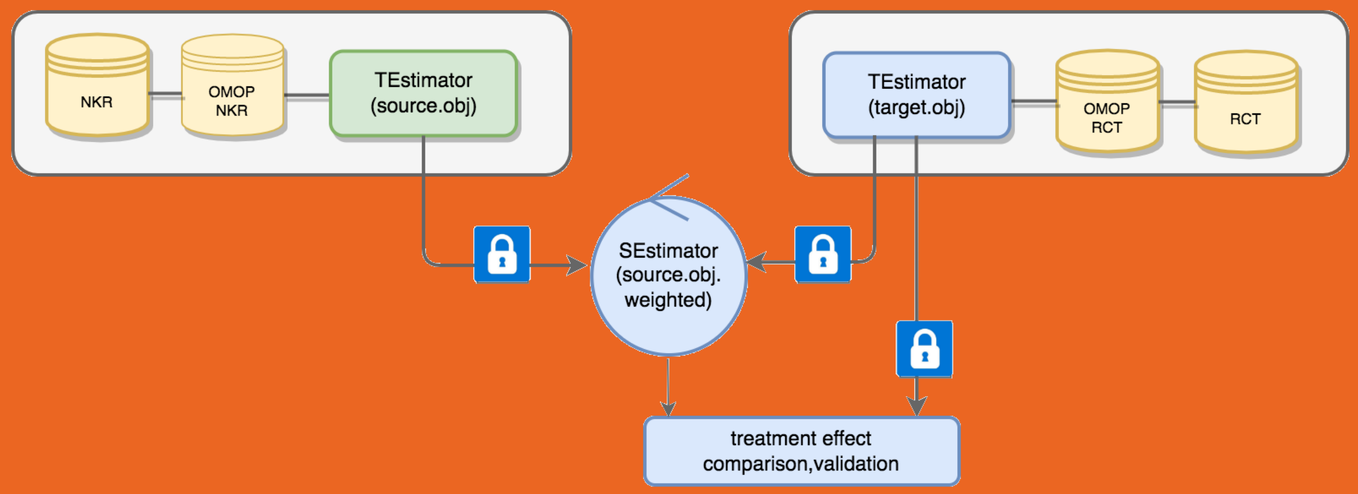

Figure 2: Diagram of RCTrep basic structure

- **TEstimator:** R6 class TEstimator is responsible for **estimating population- and subpopulation-level treatment effects**, and diagnosing assumptions.
- **SEstimator:** R6 class SEstimator is responsible for computing **weights**, so that the weighted covariates in **source.obj** and covariates in **target.obj** are **balanced**. **The two objects communicate within** the object of the class SEstimator, sharing either **unit-level data** or **aggregated data** for computing the weights.

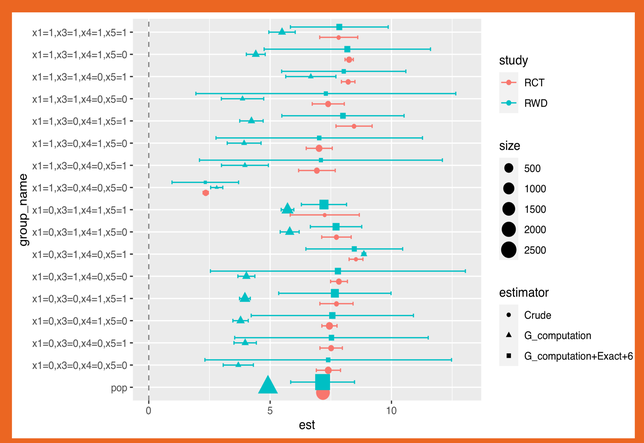

Figure 3: A working example of RCTrep. We use the **G-computation method** to adjust the **treatment assignment mechanism,** and use **exact matching** to adjust the **sampling mechanism.** Results show that **only** correcting for **both mechanisms** can allow for **comparison** of treatment effect estimation between **RWD** and **RCT** data.

Paper

Package

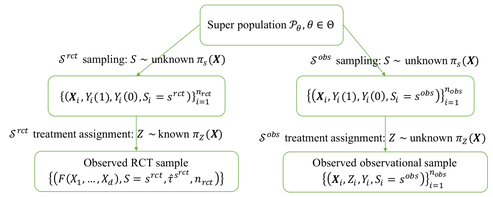

Figure 1: The mechanisms of RWD and RCT data generations

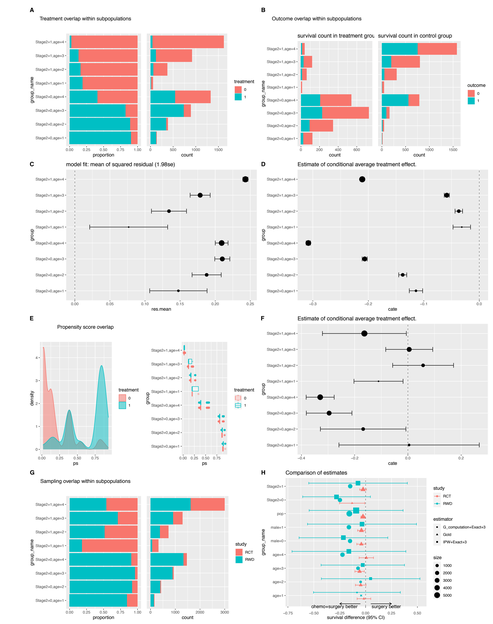

Figure 4: Estimates comparison between NKR and QUASAR trial using RCTrep. Subfigure (a)(b) diagnoses overlap of treatment within subgroups in NKR data and survival in treatment and control groups in NKR data. Figure (c)(d) diagnoses G-computation model fit and estimates of treatment effect in subgroups. Figure (e) (f) diagnoses propensity score overlap between treatment and control groups and estimates of treatment effect using inverse propensity score weighting. Figure (d)(h) diagnoses covariates balance between KNR and QUASAR trial and comparison of estimates from QUASAR and estimates from KNR with and without weighting.

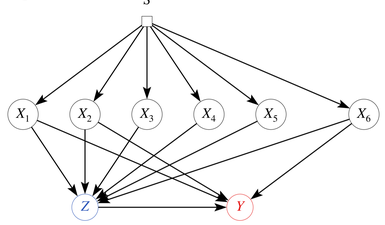

Figure 5: Illustration of adjustment sets in TEstimator and SEstimator. S is an indicator of selection into RCT and Z is an indicator of selection into treatment group.

Lingjie Shen[1*],
Gijs Geleijnse[2],
Maurits Kaptein[3]

[1]Department of Methodology and Statistics, Tilburg University,
[2]IKNL,
[3]Jheronimus Academy of Data Science,
[*]L.Shen@uvt.nl