Selecting Subpopulations for Causal Inference in Regression Discontinuity Studies

Fabrizia Mealli

University of Florence

European University Institute

fabrizia.mealli@unifi.it

Joint work with Laura Forastiere (Yale U.) and Alessandra Mattei (U. Florence)

ESEM 2022

Milan, August 24, 2022

Regression-discontinuity (RD) studies

- The RD design is a quasi-experimental design where the treatment status changes discontinuously according to some underlying variable (*forcing variable*) crossing a threshold
- Examples
  - ✓ Education: Treatment (e.g., scholarship offer, training course) assigned based on some test/qualifying score
  - ✓ Means-tested programs: Treatment (e.g., unemployment benefits, social benefits, university grants) assigned based on a wealth indicator (e.g., income)
  - ✓ Medicine and Public Health: Treatment (e.g., drug prescription, diet) assigned based on a biomarker (e.g., cholesterol, BMI)
  - ✓ Environmental studies (air quality regulations): Treatment (e.g., power plant closure, traffic restrictions) assigned based on pollutant levels (e.g., $PM_{2.5}$)
- Basic idea Units with values of the forcing variables in a neighborhood of the threshold but different levels should be as good as randomly assigned *(Thistlethwaite and Campbell, 1960)*
- Internal/external validity *(Angrist and Rokkanen, 2015)*

Motivating Study: The Brazil's Bolsa Familía (BF) Program

- Bolsa Família is a social welfare program of the Brazilian government, that started in 2003 and ended ded at the end of 2021
  - ✓ It has been replaced by a new welfare program, called AuxÃlio Brasil
- Objective: Reducing short-term poverty by direct cash transfers and fighting long-term poverty by increasing human capital among poor Brazilian people through conditional cash transfers
- Causal question: Assessing causal effects of the Bolsa Família program on health outcomes (here leprosy)
- We formally describe the Bolsa Família as a local randomized experiment
- Causal effects can be identified and estimated on a subpopulation where a local overlap assumption, a local SUTVA and a local ignorability assumption hold.
- Potential advantages of this framework over local regression methods based on continuity assumptions concern the definition of the causal estimands, the design and the analysis of the study, and the interpretation and generalizability of the results

Main Contributions

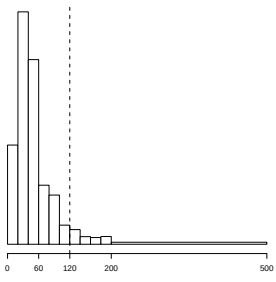
- Critical issue of local randomization approach is how to choose subpopulations
- We propose to use a Bayesian model-based finite mixture approach to clustering to classify observations into subpopulations where the RD assumptions hold and do not hold on the basis of the observed data
- This approach has important advantages
  - ✓ It allows to account for the uncertainty in the subpopulation membership, which is typically neglected (also in bandwidth selection);
  - ✓ It does not impose any constraint on the shape of the subpopulation;
  - ✓ It is scalable to high-dimensional settings;
  - ✓ It allows to account for rare outcomes;
  - ✓ It is robust to a certain degree of manipulation/selection of the running variable.

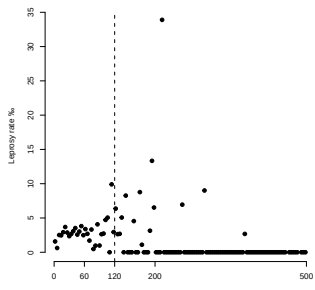
Main Contributions, cont' d

- The approach can be used as a design stage before the application of any type of analysis for any causal estimand
  - ✓ We can multiply impute subpopulation membership creating a set of complete membership datasets
  - ✓ Then for each complete membership dataset, we can estimate causal effects
  - ✓ Finally we can combine the complete-data inferences on the local causal effects to form one inference that properly reflects uncertainty on the subpopulation membership
- This will make any estimator more robust to deviations from the underlying assumptions as well as incorporate uncertainty e.g. of the bandwidth selection

The BF study: BF Benefit Allocation Rule

- BF benefit allocation rule: A family must (1) meet eligibility criteria; and (2) apply for the Bolsa Família benefits
  - ✓ Data on families who applied for some welfare programs and registered in Cadastro Único in 2007-08 for the first time
- Eligibility: Per capita household income (forcing variable) falling below or above a pre-fixed threshold (120 Brazil Real $\simeq$ 36.5 USD per month)
- Here we focus on intention-to-treat effects of eligibility statuses, i.e, a sharp RD design with respect to income eligibility





RD Designs as Local Randomized Experiments

- Traditionally, RD designs are viewed as quasi-experimental designs with a non-probabilistic assignment mechanism
- RD designs as local randomized experiments (LREs) in a neighbourhood of the threshold *(e.g., Cattaneo et al., 2015; Li, Mattei and Mealli, 2015; Sales and Hansen, 2020; Mattei and Mealli, 2016)*
- Probabilistic formulation of the assignment mechanism underlying RD designs within the potential outcome approach *(Li, Mattei and Mealli, 2015; Mattei and Mealli, 2016)*
  - ✓ The forcing variable is viewed as a random variable
  - ✓ Local randomization: there exists at least a subpopulation, $\mathcal{U}_{s_0}$, around the threshold where the forcing variable, and therefore the treatment/eligibility status, can be seen as randomly assigned
  - ✓ Other assignment mechanisms are considered in *Branson and Mealli (2020)*
  - ✓ Focus on finite population causal effects for units in $\mathcal{U}_{s_0}$
  - ✓ $\mathcal{U}_{s_0}$ is chosen in such a way so that LR assumptions hold

Framing RDDs as local randomized experiments

(Rubin, 1974, 1978)

- $i =$ Unit/Family ($i = 1, \ldots, N$)
- $\mathbf{X}_i =$ Vector of covariates
- $Z_i =$ BF benefit eligibility status:

$$Z_i = z \in \{0, 1\} = \{\text{Ineligible}, \text{Eligible}\}$$

- $S_i =$ Per capita household income: The forcing variable

$$Z_i = \mathbf{1}\{S_i \leq s_0\} \qquad s_0 = 120 \text{ Brazil Real (threshold)}$$

- Assume that there is no interference
- $Y_i(s) =$ Potential outcomes for the indicator of the presence of at least a leprosy case (after 2009) in family i under a monthly per capita income equal to $s \in \mathbb{R}$

$$Y_i(s) = \begin{cases} 1 & \text{If there is at least a leprosy case in family } i \text{ given } s \\ 0 & \text{If there is no leprosy case in family } i \text{ given } s \end{cases}$$

Local Overlap, Local RD-SUTVA and Local Estimands

Assumption 1. Local Overlap. There exists a subset of units, $\mathcal{U}_{s_0}$, such that for each $i \in \mathcal{U}_{s_0}$, $\Pr(S_i \leq s_0) > \epsilon$ and $\Pr(S_i > s_0) > \epsilon$ for some sufficiently large $\epsilon > 0$

Assumption 2. Local RD-SUTVA. For each $i \in \mathcal{U}_{s_0}$, consider two eligibility statuses $z' = \mathbf{1}(S_i = s' \leq s_0)$ and $z'' = \mathbf{1}(S_i = s'' \leq s_0)$, with possibly $s' \neq s''$. If $z' = z''$ then $Y_i(s') = Y_i(s'')$

- ✓ Under Local RD-SUTVA for each $i \in \mathcal{U}_{s_0}$, there are only two potential outcomes for the indicator of the presence of at least a leprosy case: $Y_i(0)$ and $Y_i(1)$

Causal Estimand. Local relative risk

$$RR_{\mathcal{U}_{s_0}} \equiv \frac{Pr\{Y_i(1) = 1; i \in \mathcal{U}_{s_0}\}}{Pr\{Y_i(0) = 1; i \in \mathcal{U}_{s_0}\}}$$

Probabilistic Treatment Assignment Mechanism for RD Designs

Assumption 3. Local Randomization (LR). For each $i \in \mathcal{U}_{s_0}$,

$$\Pr\left(S_i \mid Y_i(0), Y_i(1), \mathbf{X}_i\right) = \Pr\left(S_i\right)$$

✓ Under local randomization, for each $i \in \mathcal{U}_{s_0}$,

$$\Pr(Z_i = 1) = \Pr(S_i \leq s_0)$$

Assumption 3'. Local Unconfoundedness (LU).
For each $i \in \mathcal{U}_{s_0}$,

$$\Pr\left(S_i \mid Y_i(0), Y_i(1), \mathbf{X}_i\right) = \Pr\left(S_i \mid \mathbf{X}_i\right)$$

✓ Under local unconfoundedness, for each $i \in \mathcal{U}_{s_0}$,

$$\Pr(Z_i = 1 \mid \mathbf{X}_i) = \Pr(S_i \leq s_0 \mid \mathbf{X}_i)$$

Potential Advantages of RDDs as LREs

When assumptions are judged plausible, there are several advantages of RD designs as LREs:

- Sub-population treatment effects: Treatment effects are for subpopulation members rather than local average treatment effects for those at the cutoff only; results are more easily generalizable
- Focus on different estimands (e.g., RR for rare events)
- No modelling assumptions: Modeling assumptions on the relationship between the running variable and the outcome can be avoided, causal effect not necessarily additive
- Ignorable treatment assignment: The treatment assignment mechanism is allowed to be random rather than deterministic as in typical RD analyses; finite population inference can be used
- Discrete running variables can be easily accounted for
- Visualize covariate balancing (e.g., appealing for doctors)

Selection of Subpopulations $\mathcal{U}_{s_0}$: State of the Art

- Local randomization based methods; falsification tests *(Cattaneo et al., 2015; Li, Mattei, Mealli, 2015; Licari and Mattei, 2020)*
  - ✓ Assume LR and select subpopulations where pre-treatment variables are well balanced in the two subsamples defined by the assignment
  - ✓ Randomization or model-based Bayesian tests, possibly with adjustment for multiplicities
  - ✓ These methods usually rely on assumptions on the shape of the subpopulations and are not immediately applicable when LU rather than LR is assumed
- Local unconfoundedness based methods
  - ✓ Assume LU and construct a subpopulation conditioning on observables and the discontinuity using penalized matching methods *(Keele et al., 2015)*
  - ✓ Assume LU and other designs using covariates *(Branson and Mealli, 2020)*
- LR and LU based methods do not scale up well, assume a shape, and do not directly account for the uncertainty about a selected subpopulation
- Recently, Ricciardi et al. (2021) propose to include in the RDD analysis units who belong to balanced and homogeneous clusters defined using a Dirichlet process mixture model

Selection of Subpopulations $\mathcal{U}_{s_0}$: Our Proposal

- The problem of selecting suitable subpopulations, $\mathcal{U}_{s_0}$, as a clustering problem
- Sample units in a RD study come from three subpopulations:

$$\mathcal{U}_{s_0}^{-} = \{i \notin \mathcal{U}_{s_0} : S_i < s_0\} \qquad \mathcal{U}_{s_0} = \{i : S_i \in \mathcal{I}_{s_0}\} \qquad \mathcal{U}_{s_0}^{+} = \{i \notin \mathcal{U}_{s_0} : S_i > s_0\}$$

  where $\mathcal{I}_{s_0}$ is a neighborhood around s_0

- Crucial issue: We have some information on each subpopulation but we do not know which subpopulation each unit belongs to
- What do we know about the three subpopulations?
  - ✓ Each unit belongs to only one of the 3 subpopulations
  - ✓ For units who belong to $\mathcal{U}_{s_0}$ the RD assumptions hold
  - ✓ For units who belong to either $\mathcal{U}_{s_0}^{-}$ or $\mathcal{U}_{s_0}^{+}$ some RD assumptions may fail to hold
- Idea: Use clustering methods to ascertain, on the basis of the information we have, which subpopulation each unit belongs to

Selection of Subpopulations $\mathcal{U}_{s_0}$: A Finite Mixture Model Approach

- A finite mixture model-based approach *(e.g., McLachlan and Basford, 1988; Titterington, Smith, and Markov, 1985)*

$$
\begin{aligned}
&p\left(S_i, \{Y_i(s)\}_{s\in\mathbb{R}_+} \mid \mathbf{X}_i; \boldsymbol{\theta}\right) = \\
&\quad \pi_i(\mathcal{U}_{s_0}^-)\, p(S_i \mid \mathbf{X}_i; i \in \mathcal{U}_{s_0}^-; \boldsymbol{\eta}^-)\, p(\{Y_i(s)\}_{s\in\mathbb{R}_+} \mid S_i, \mathbf{X}_i; i \in \mathcal{U}_{s_0}^-; \boldsymbol{\gamma}^-) + \\
&\quad \pi_i(\mathcal{U}_{s_0})\, p(S_i \mid \mathbf{X}_i, i \in \mathcal{U}_{s_0}; \boldsymbol{\eta})\, p(Y_i(0), Y_i(1) \mid \mathbf{X}_i, i \in \mathcal{U}_{s_0}; \boldsymbol{\gamma}) + \\
&\quad \pi_i(\mathcal{U}_{s_0}^+)\, p(S_i \mid \mathbf{X}_i, i \in \mathcal{U}_{s_0}^+; \boldsymbol{\eta}^+)\, p(\{Y_i(s)\}_{s\in\mathbb{R}_+} \mid S_i, \mathbf{X}_i, i \in \mathcal{U}_{s_0}^+; \boldsymbol{\gamma}^+)
\end{aligned}
$$

where

$$
\pi_i(\mathcal{U}_{s_0}^-) = Pr(i \in \mathcal{U}_{s_0}^- \mid \mathbf{X}_i; \boldsymbol{\alpha}) \geq 0 \quad \pi_i(\mathcal{U}_{s_0}^+) = Pr(i \in \mathcal{U}_{s_0}^+ \mid \mathbf{X}_i; \boldsymbol{\alpha}) \geq 0
$$
$$
\pi_i(\mathcal{U}_{s_0}) = Pr(i \in \mathcal{U}_{s_0} \mid \mathbf{X}_i; \boldsymbol{\alpha}) \geq 0
$$

are the mixing probabilities, with $\pi_i(\mathcal{U}_{s_0}^-) + \pi_i(\mathcal{U}_{s_0}) + \pi_i(\mathcal{U}_{s_0}^+) = 1$, $(\boldsymbol{\eta}^-, \boldsymbol{\gamma}^-)$, $(\boldsymbol{\eta}, \boldsymbol{\gamma})$ and $(\boldsymbol{\eta}^+, \boldsymbol{\gamma}^+)$ are parameter vectors defining each mixture component, and $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\eta}^-, \boldsymbol{\gamma}^-, \boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\eta}^+, \boldsymbol{\gamma}^+)$ is the complete set of parameters specifying the mixture

- Units with close values of S may belong to different subpopulations

The BF Study: Summary Statistics

- Population of $N = 152\,602$ families who registered in Cadastro Único in 2007-08 for the first time

$$\text{Eligible } (S \leq 120) = 138\,220 \quad \text{Ineligible } (S > 120) = 14\,382$$

- Forcing variable: Per capita household income (S)

Statistics	All	$S \leq 120$	$S > 120$
Min	0.0	0.0	120.2
Median	46.7	40.0	190.0
Max	500.0	120.0	500.0

- Outcome variable: Leprosy rate in 2009 (Y) ‰

All	$S \leq 120$	$S > 120$
2.78	2.80	2.57

The BF Study: Background Characteristics

Household structure

Covariate	Eligible	Ineligible
Mean age	21.89	45.52
Min age	10.32	36.55
Household size	3.02	2.11
No children	1.38	0.46
No adults	1.60	1.05
Children not at school	0.04	0.01
Presence of weak people	0.23	0.12

Household head's characteristics

Covariate	Eligible	Ineligible
Male	0.87	0.76
Race: Hispanic	0.88	0.83
Primary/Middle Education	0.47	0.61
Occupation: Unemployed	0.49	0.25

Living and economic conditions

Covariate	Eligible	Ineligible
Rural (Urban)	0.40	0.23
Apartment (Other)	0.95	0.97
Home ownership: Homeowner	0.58	0.75
No rooms pc	1.57	2.92
House of bricks/row dirt	0.91	0.96
Water treatment	0.78	0.86
Water supply	0.63	0.79
Lighting	0.79	0.91
Bathroom fixture	0.62	0.49
Waste treatment	0.63	0.80
Zero PC expenditure	0.22	0.17
PC expenditure	2.89	4.06
Other programs	0.06	0.10

BF Study: Mixture-Model Specification

- Model for the mixing probabilities: conditional probit

$$\pi_i(\mathcal{U}_{s_0}^-) = \Pr(G_i^*(-) \leq 0) \qquad \pi_i(\mathcal{U}_{s_0}^+) = \Pr(G_i^*(-) > 0 \text{ and } G_i^*(+) \leq 0)$$
$$\pi_i(\mathcal{U}_{s_0}) = 1 - \pi_i(\mathcal{U}_{s_0}^-) - \pi_i(\mathcal{U}_{s_0}^+)$$

  where $G_i^*(-) = \alpha_0^- + \mathbf{X}_i'\boldsymbol{\alpha}_X^- + \epsilon_i^-$ and $G_i^*(+) = \alpha_0^+ + \mathbf{X}_i'\boldsymbol{\alpha}_X^+ + \epsilon_i^+$, with $\epsilon_i^- \sim N(0,1)$ and $\epsilon_i^+ \sim N(0,1)$, independently

- Models for the forcing variable (per capita household income): Log-normal models

$$\begin{aligned}
\log(S_i) \mid \mathbf{X}_i, i \in \mathcal{U}_{s_0}^- &\sim N\left(\beta_0^- + \mathbf{X}_i'\boldsymbol{\beta}_X^-; \sigma_-^2\right) \\
\log(S_i) \mid \mathbf{X}_i, i \in \mathcal{U}_{s_0}^+ &\sim N\left(\beta_0^+ + \mathbf{X}_i'\boldsymbol{\beta}_X^+; \sigma_+^2\right) \\
\log(S_i) \mid \mathbf{X}_i, i \in \mathcal{U}_{s_0} &\sim N\left(\beta_0 + \mathbf{X}_i'\boldsymbol{\beta}_X; \sigma^2\right)
\end{aligned}$$

- Models for the outcome (probit link):

$$\begin{aligned}
\Pr(Y_i(s) = 1 \mid \mathbf{X}_i, i \in \mathcal{U}_{s_0}^-) &= \Phi\left(\gamma_0^- + \log(s)\gamma_1^- + \mathbf{X}_i'\boldsymbol{\gamma}_X^-\right) \\
\Pr(Y_i(s) = 1 \mid \mathbf{X}_i, i \in \mathcal{U}_{s_0}^+) &= \Phi\left(\gamma_0^+ + \log(s)\gamma_1^+ + \mathbf{X}_i'\boldsymbol{\gamma}_X^+\right) \\
\Pr(Y_i(z) = 1 | \mathbf{X}_i, i \in \mathcal{U}_{s_0}) &= \Phi\left(\gamma_{0,z} + \mathbf{X}_i'\boldsymbol{\gamma}_X\right) \qquad z = 0, 1
\end{aligned}$$

BF Study: Bayesian Inference

- We assume that parameters are a priori independent
- We use weakly informative priors
  - ✓ Multivariate normal priors for the coefficients (mean vector $= \mathbf{0}$; covariance-variance matrix $= \sigma^2 \cdot \mathbb{I} = 100 \cdot \mathbb{I}$)
  - ✓ Scaled inverse-χ^2 priors for the variances (3 df; scale parameter $= 1/3$)
- Finite sample estimands
- MCMC algorithm: For $\ell = 1 \ldots, L$
  - ✓ Impute missing subpopulation membership for each unit using a data augmentation step
  - ✓ Update the model parameters using Gibbs sampling
  - ✓ For each unit i in $\mathcal{U}_{s_0}$, draw the missing potential outcome, $Y_i^{mis} = Z_i Y_i(0) + (1 - Z_i) Y_i(1)$ from its posterior predictive distribution and calculate

$$RR^{\ell}_{\mathcal{U}_{s_0}} = \frac{\sum_{i:i\in\mathcal{U}_{s_0}} [Z_i Y_i^{obs} + (1 - Z_i) Y_i^{\ell}(1)] \Big/ N^{\ell}_{\mathcal{U}_{s_0}}}{\sum_{i:i\in\mathcal{U}_{s_0}} [(1 - Z_i) Y_i^{obs} + Z_i Y_i^{\ell}(0)] \Big/ N^{\ell}_{\mathcal{U}_{s_0}}}$$

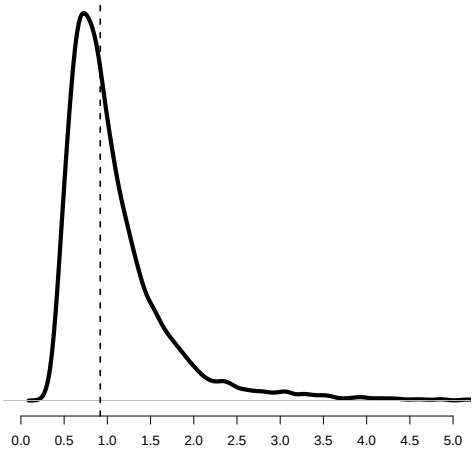
where $Y_i^{obs} = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$ and $N^{\ell}_{\mathcal{U}_{s_0}}$ is the number of units in $\mathcal{U}_{s_0}$

Posterior distributions of the mixing probabilities

Estimand	Median	2.5%	97.5%
$\pi(\mathcal{U}_{s_0}^-)$	0.433	0.430	0.435
$\pi(\mathcal{U}_{s_0}^+)$	0.074	0.073	0.075
$\pi(\mathcal{U}_{s_0})$	0.493	0.491	0.496
$N_{\mathcal{U}_{s_0}}$	75 273	74 893	75 653
$\sum_{i \in \mathcal{U}_{s_0}} (1 - Z_i)$	3 076	2 989	3 166
$\sum_{i \in \mathcal{U}_{s_0}} Z_i$	72 198	71 844	72 542

No assumption on the shape of the subpopulations; units with similar realized values of S may belong to different subpopulations

Posterior Distribution of $RR_{\mathcal{U}_{s_0}}$ (Finite Sample Causal Effect)



Median	2.5%	97.5%
0.769	0.455	2.933

Width of the 95% PCI $= 2.478$

$\Pr(RR_{\mathcal{U}_{s_0}} < 1) = 0.574$

Nice interpretation of our approach: the causal effect estimate is an average over the distribution of the proportion of units in $\mathcal{U}_{s_0}$ - cfr: worst-case bounds w.r.t. proportion of contamination in Bonvini, Kennedy (2020)

Summary statistics of the posterior distributions of $RR_{\mathcal{U}_{s_0}}$ for specific MSE-optimal subpopulations

Subpopulation: $\mathcal{U}_{s_0}$ ($N_{\mathcal{U}_{s_0}} = \sum_{i \in \mathcal{U}_{s_0}} (1 - Z_i) + \sum_{i \in \mathcal{U}_{s_0}} Z_i$)	Median	2.5%	97.5%	(width)	$Pr\left(RR_{\mathcal{U}_{s_0}} < 1\right)$
Uniform kernel ($p = 1$)					
$[70.2; 176.9] = [120 - 49.9; 120 + 56.9]$ $(26\,893 = 6\,077 + 20\,816)$	1.238	0.762	2.125	(1.363)	0.189
Triangular kernel ($p = 1$)					
$[39.0; 172.0] = [120 - 81.0; 120 + 52.0]$ $(83\,295 = 5\,128 + 78\,167)$	1.256	0.751	2.250	(1.499)	0.199
Uniform kernel ($p = 2$)					
$[0.0; 193.5] = [120 - 120.0; 120 + 73.5]$ $(145\,612 = 7\,392 + 138\,220)$	1.420	0.892	2.356	(1.464)	0.067
Triangular kernel ($p = 2$)					
$[0.0; 209.7] = [120 - 120.0; 120 + 89.7]$ $(146\,173 = 7\,953 + 138\,220)$	1.267	0.822	2.035	(1.213)	0.150

Concluding Remarks

- Features of the model-based Bayesian mixture approach to the selection of subpopulations, $\mathcal{U}_{s_0}$, in RD designs
  - ✓ It explicitly accounts for the uncertainty about $\mathcal{U}_{s_0}$ membership
  - ✓ It imposes no constraint on the shape of $\mathcal{U}_{s_0}$
  - ✓ It provides inference on causal effects in a unique Bayesian framework
  - ✓ It works with different assignment mechanisms
  - ✓ It can be easily extended to fuzzy RDDs
- Alternative approaches could use our mixture approach just as a tool to select suitable subpopulations
  - ✓ Multiply impute sub-population membership creating a set of complete membership datasets
  - ✓ For each complete membership dataset, use units belonging to $\mathcal{U}_{s_0}$ to draw inference on the causal effects of interest using a proper mode of causal inference
  - ✓ Combine the complete-data inferences on the local causal effects to form one inference that properly reflects missing $\mathcal{U}_{s_0}$-membership uncertainty (and possibly sampling variability)
- Results may form the basis to conduct a sensitivity analysis

References

Bonvini M., Kennedy E.H. (2019). Sensitivity analysis via the proportion of unmeasured confounding. arXiv 1912.02793.

Branson Z., Mealli F. (2020). The local randomization framework for regression discontinuity designs: A review and some extensions. arXiv 1810.02761

Calonico S., Cattaneo M.D., Farrell M.H. (2020). Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs. Econometrics Journal 23(2) 192-210.

Cattaneo M.D., Frandsen B.R., Titiunik R. (2015). Randomization Inference in the Regression Discontinuity Design: An application to Party Advantages in the U.S. Senate. Journal of Causal Inference 3, 1-24.

Diebolt J., Christian P.R. (1994). Estimation of Finite Mixture Distributions through Bayesian Sampling. JRSS-B 56, 363-375.

Imbens G.W., Lemieux T. (2008). Regression Discontinuity Designs: A Guide to Practice. Journal of Econometrics 142, 615-635.

Eckles D., Ignatiadis N., Wager S., Wu H. (2020). Noise-Induced Randomization in Regression Discontinuity Designs. ArXiv working paper N. 2004.09458

Imbens G.W., Kalyanaraman K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. Review of Economic Studies 79, 933-959.

Keele, L.J., Titiunik, R. and Zubizarreta, J. R. (2015) Enhancing a Geographic Regression Discontinuity Design Through Matching to Estimate the Effect of Ballot Initiatives on Voter Turnout. JRSS-A 178, 223-239.

Lee D.S., Lemieux T. (2010). Regression Discontinuity Designs in Economics. Journal of Economic Literature 485, 281-355.

Li F., Mattei A., Mealli F. (2015). Bayesian Inference for Regression Discontinuity Designs with Application to the Evaluation of Italian University Grants. AOAS 9, 1906-1931.

Licari F., Mattei A. (2020). Assessing causal effects of extra compulsory learning on college students' academic performances. Journal of the Royal Statistical Society: Series A. To appear.

Mattei A., Mealli F. (2016). Regression Discontinuity Designs as Local Randomized. Observational Studies 2, 156-172.

McLachlan G., Basford K. (1988). Mixture Models: Inference and Application to Clustering. New York: Marcel Dekker.

Sales A., Hansen B.B. (2020). Limitless Regression Discontinuity. Journal of Educational and Behavioral Statistics, 45(2):143-174.

Ricciardi F., Liverani S. Baio G. (2021). Dirichlet Process Mixture Models for Regression Discontinuity Designs. ArXiv working paper N. 2003.11862

Richardson S., Green P.J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components. JRSS-B 59, 731-792.

Rubin D.B. (1974). Estimating Causal Effects of Treatments in Randomized and nonrandomized studies. Journal of Educational Psychology, 66: 688-701.

Rubin D.B. (1978). Bayesian Inference for Causal Effects. The Annals of Statistics 6, 34-58.

Thistlethwaite D., Campbell D. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment. Journal of Educational Psychology 51, 309-317.

Titterington D., Smith A., Markov U. (1985). Statistical Analysis of Finite Mixture Distributions. Chichester, U.K.: John Wiley & Sons.