Causal inference in the context of an error prone exposure: air pollution and mortality

Xiao Wu, Danielle Braun, Marianthi-Anna Kioumourtzoglou, Christine Choirat, Qian Di, Francesca Dominici

June 16, 2018

Introduction

- Long-term exposure to air pollution has consistently been associated with adverse health outcomes.
  - [Dockery et al., 1993, Pope III et al., 1995, Beelen et al., 2014, Di et al., 2017]

Introduction

- Long-term exposure to air pollution has consistently been associated with adverse health outcomes.
  - [Dockery et al., 1993, Pope III et al., 1995, Beelen et al., 2014, Di et al., 2017]
- Obtaining personal exposure information is not feasible.
  - Need to rely on air pollution concentrations predicted by spatio-temporal models (error-prone!).
  - Most epi studies have ignored exposure measurement error.

Introduction

- Long-term exposure to air pollution has consistently been associated with adverse health outcomes.
  - [Dockery et al., 1993, Pope III et al., 1995, Beelen et al., 2014, Di et al., 2017]
- Obtaining personal exposure information is not feasible.
  - Need to rely on air pollution concentrations predicted by spatio-temporal models (error-prone!).
  - Most epi studies have ignored exposure measurement error.
- Some studies have developed methodology to quantify and correct for error
  - E.g. [Alexeeff et al., 2015]
  - They do not use methods robust to model misspecification.
  - They do not use **Generalized Propensity Scores** to adjust for confounding.
    $\rightarrow$ focus of this work.

- **Goal:** To estimate the causal effect of long-term exposure to fine particles, $PM_{2.5}$ on health outcomes at the zip code level in the Medicare population in New England area.

Data

- Exposure: Annual averaged $PM_{2.5}$ concentrations aggregated at zip code level as categorical variable.
  - While $PM_{2.5}$ concentrations are continuous, our interest is in comparing exposure categories based on pre-specified $PM_{2.5}$ cutoffs.
  - Can help inform future policy decisions (E.g. $PM_{2.5}$ levels of 8, 10, 12 $\mu g/m^3$(NAAQS Table)).
- Outcome of Interest: Mortality, health outcome counts, at zip code level over a year.
- Study Population: Medicare participants in New England area (2000–2012).

Challenges

1. **Confounding:**
   - Observational studies, such as this one, have limitations due to lack of randomization.
     - Factors that vary and are associated both with $PM_{2.5}$ levels and mortality (e.g. SES-related factors) may confound exposure comparisons.
   - In causal inference, using propensity scores (the probability of a unit being assigned to a particular exposure given the pretreatment confounders) is a common practice.

Challenges

2. **Measurement Error:**
   - It is not feasible to measure the exact $PM_{2.5}$ exposure in every single zip code.
   - However, $PM_{2.5}$ concentrations can be predicted using spatio-temporal models [Di et al., 2016a, Di et al., 2016b, Di et al., 2017].
   - These predicted $PM_{2.5}$ concentrations are:
     - Error-prone.

Challenges

2. **Measurement Error:**
   - Internal Validation data
     - Measured $PM_{2.5}$ concentrations (gold-standard).
     - Predicted $PM_{2.5}$ using the spatio-temporal model (error-prone).

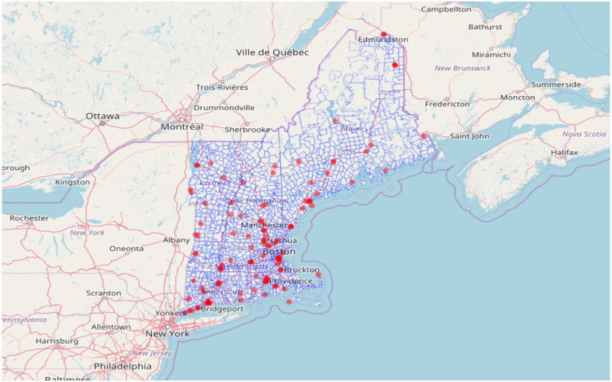


Figure: Locations of monitor stations in New England (in red). Zip code areas are drawn in blue.

Previous Studies: Measurement Error in Causal Inference

- Lots of literature on methods to adjust for measurement error in confounders/outcomes.
  - [Stürmer et al., 2005, McCandless et al., 2012, McCaffrey et al., 2013, Webb-Vargas et al., 2015, Shu and Yi, 2017, Shu and Grace, 2018]

Previous Studies: Measurement Error in Causal Inference

- Lots of literature on methods to adjust for measurement error in confounders/outcomes.
  - [Stürmer et al., 2005, McCandless et al., 2012, McCaffrey et al., 2013, Webb-Vargas et al., 2015, Shu and Yi, 2017, Shu and Grace, 2018]
- Some literature on methods to adjust for measurement error in exposure.
  - [Babanezhad et al., 2010, Braun et al., 2017]
  - Focus on binary exposure.

Previous Studies: Measurement Error in Causal Inference

- Lots of literature on methods to adjust for measurement error in confounders/outcomes.
  - [Stürmer et al., 2005, McCandless et al., 2012, McCaffrey et al., 2013, Webb-Vargas et al., 2015, Shu and Yi, 2017, Shu and Grace, 2018]
- Some literature on methods to adjust for measurement error in exposure.
  - [Babanezhad et al., 2010, Braun et al., 2017]
  - Focus on binary exposure.
- No literature on methods to adjust for measurement error in a categorical exposure in causal settings.

Proposed Approach; RC-GPS

- Regression calibration (RC)-based adjustment for an error-prone exposure combined with generalized propensity scores (GPS) to adjust for confounding (RC-GPS).
  - **RC:** RC model is fit using continuous $PM_{2.5}$ regressing true $PM_{2.5}$ on error-prone $PM_{2.5}$ and additional predictive covariates.
  - **GPS:** Outcome analysis adjusting for confounding using GPS is conducted by transforming the continuous exposure into a categorical exposure.

Notations

- Main study: $(Y, \widehat{W_c}, \widehat{W}, \mathbf{C}, \mathbf{D})$.
- Internal validation study: $(X, \widehat{W}, \mathbf{D})$.
  - $\widehat{W}$ denote error-prone exposures.
  - $\widehat{W_c}$ is the category of $\widehat{W}$ determined by policy.
  - X denote true exposures.
  - $\mathbf{C}$ denote potential confounders measured without error associated with the true exposure and outcome.
  - $\mathbf{D}$ denote covariates measured without error are predictive of the true exposure.

RC: Assumptions

1. **Transportability:** need to assume that the relationship between X and W would be the same in locations where X is observed and in those that it is not.
2. **Non-differential measurement error:** the conditional distribution of outcome Y given $(W, X, \mathbf{D})$ depends only on $(X, \mathbf{D})$.
3. **Small measurement error:** $\text{tr}(\Sigma_{X|W,\mathbf{D}})$ is small.

GPS

- Follow framework described in [Yang et al., 2016, Imbens, 2000].
- **GPS:** the conditional probability of receiving each level of exposure given the confounders.
- Define $GPS(\mathbf{x}) = (GPS_{x_1}, GPS_{x_2}, ..., GPS_{x_n})$;

$$GPS_{x_i} = Pr(X_i = x_i | \mathbf{C_i} = \mathbf{c})$$

  for every possible exposures x_i (i=1, 2,..., n).
- Attention: GPS is a vector! The individual GPS_{x_i} are called GPS elements.

GPS

- [Yang et al., 2016, Imbens, 2000] propose to estimate the average of the potential outcomes separately for each exposure group.
  - Different from the binary exposure case, in which we construct groups based on similar propensity scores and compare observations within these groups.
  - Doing so is more challenging with multiple exposures.
    - Would require comparisons in groups constructed based on similar values of a vector of propensity scores.

GPS; Notations and Assumptions

- Let X_i denote the exposure for individual i, $X_i \in \mathbb{X}$.
- Let $Y_i(x)$ denote the potential outcome for an exposure x.
- Let $Y_i^{obs} = Y_i(X_i)$.
- Let $I_i(x) = 1$ if $X_i = x$ and 0 otherwise.
- Assumptions;
  1. SUTVA: $X = x$ implies $Y = Y(x)$.
  2. Overlap; $P(x|\mathbf{c}) > 0$ for all $x, \mathbf{c}$.
  3. Weak Unconfoundedness; $I_i(x) \perp\!\!\!\perp Y_i(x) | P(x|\mathbf{C}_i)$ for all $x \in \mathbb{X}$.

GPS; Average Causal Effects

- Under weak unconfoundedness;
  $E[Y_i(x') - Y_i(x)]$
  $= E[E[Y_i^{obs}|X_i = x', P(x'|\mathbf{C}_i)]] - E[E[Y_i^{obs}|X_i = x, P(x|\mathbf{C}_i)]]$
- For example, the average exposure effects with respect to two exposures, e.g. 1 and 2, can be expressed as,

$$
\begin{aligned}
& E[Y(1) - Y(2)] \\
&= E[Y(1)] - E[Y(2)] \\
&= E[E[Y^{obs}|X = 1, GPS_1]] - E[E[Y^{obs}|X = 2, GPS_2]]
\end{aligned}
$$

  - This allows us to estimate $E[E[Y^{obs}|X = 1, GPS_1]]$ and $E[E[Y^{obs}|X = 2, GPS_2]]$ separately.
- Consider three different GPS implementations; 1) subclassification, 2) IPTW, 3) matching.

RC-GPS: ME Adj

Stage 1: Measurement Error Correction

1. Fit a RC model in the validation study. More specifically, fit $E(X|W, \mathbf{D}) = \gamma_0 + \gamma_1 W + \boldsymbol{\gamma}_2^T \mathbf{D}$ to obtain estimated $\boldsymbol{\gamma}$, i.e. $\hat{\gamma}$ in the validation study.

RC-GPS: ME Adj

Stage 1: Measurement Error Correction

1. Fit a RC model in the validation study. More specifically, fit $E(X|W, \mathbf{D}) = \gamma_0 + \gamma_1 W + \boldsymbol{\gamma_2}^T \mathbf{D}$ to obtain estimated $\boldsymbol{\gamma}$, i.e. $\hat{\gamma}$ in the validation study.
2. Under the transportability assumption, estimate $\hat{X} = \hat{\gamma}_0 + \hat{\gamma}_1 W + \hat{\boldsymbol{\gamma}}_{\mathbf{2}}{}^T \mathbf{D}$ in the main study.

RC-GPS: ME Adj

Stage 1: Measurement Error Correction

1. Fit a RC model in the validation study. More specifically, fit $E(X|W, \mathbf{D}) = \gamma_0 + \gamma_1 W + \boldsymbol{\gamma}_\mathbf{2}^T \mathbf{D}$ to obtain estimated $\boldsymbol{\gamma}$, i.e. $\hat{\gamma}$ in the validation study.
2. Under the transportability assumption, estimate $\hat{X} = \hat{\gamma}_0 + \hat{\gamma}_1 W + \hat{\boldsymbol{\gamma}}_\mathbf{2}{}^T \mathbf{D}$ in the main study.
3. Based on pre-defined categories, transform $\hat{X}$ into $\hat{X}_c \in \mathbb{X}_c = \{1, 2, ..., n\}$, a categorical variable.

RC-GPS: Confounding Adj

Stage 2A: Design Phase with GPS

4. After obtaining $\hat{X}_c$ in the main study: Estimate GPS using a GLM relating $\hat{X}_c$ to $\mathbf{C}$.

RC-GPS: Confounding Adj

Stage 2A: Design Phase with GPS

4. After obtaining $\hat{X}_c$ in the main study: Estimate GPS using a GLM relating $\hat{X}_c$ to $\mathbf{C}$.

Stage 2B: Analysis Phase with GPS

5. Estimate $\hat{E}[Y(x)]$ for each exposure category $x \in \mathbb{X}_c = \{1, 2, ..., n\}$ after adjusting for confounding using GPS subclassfication, IPTW or matching methods.
6. Estimate the ATE as the contrast of $\hat{E}[Y(x)]$ and $\hat{E}[Y(x')]$ between any two exposure categories x, x'.
7. Estimate the variance of the ATE using bootstrap to jointly account for the variability in the estimation of RC parameters γ, GPS parameters $\boldsymbol{\eta}$, and outcome model parameters $\boldsymbol{\beta}$.

Simulations

- Simulation design:
  1. Sample sizes; main study $n_m = 2000$, internal validation study $n_v = 500$. Conducted 1000 iterations.
  2. Simulation strategy:
     $[W|\boldsymbol{C}, \boldsymbol{\tau}]$, $[X|W, \boldsymbol{D}, \gamma]$, $[Y|X, \boldsymbol{C}, \beta]$, follow linear models.
     $\boldsymbol{C} = (C_1, C_2, ..., C_6)$, which include a combination of continuous and categorical covariates.
     $\boldsymbol{D} = (D_1, D_2, D_3)$, which are three continuous covariates.
  3. Correlation between X and W: 0.82 (close to reality).
  4. Correlations between X and $\boldsymbol{C}$: ≈ 0.20.
  5. Assume three exposure levels for X, W.
     - Choose cut-off points such that equal proportion of observations in each exposure level.
     - Cut-off points should be pre-specified.

Simulation Results: Subclassification

- Results based on GPS with subclassification, in which we classify subjects into ten subclasses by deciles based on each GPS element.

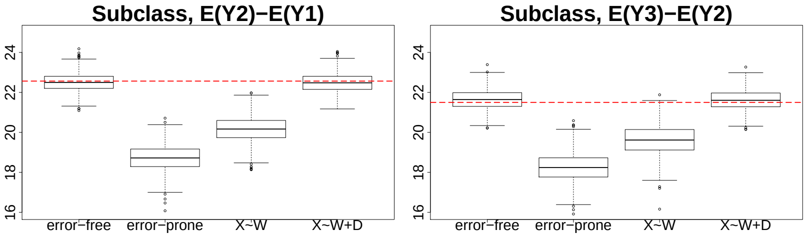
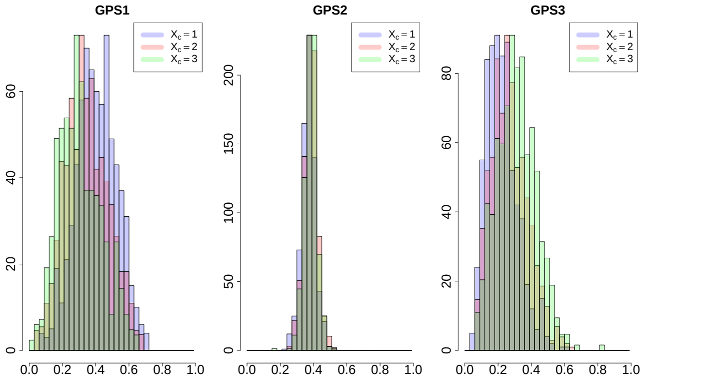


Figure: Subclassification: The red dashed line represents the true ATE.

- Results based on GPS with IPTW and matching are similar.

Simulation Results: Overlap

- When correlation between X and $\mathbf{C}$ is not too high ($<= 0.40$ in absolute), there is good overlap. Otherwise, we use a technique called trimming to improve overlap [Crump et al., 2009].



Simulation Results: Covariate Balance

- Covariate balance is evaluated via the absolute standardized bias (ASB) [Harder et al., 2010] across all confounders between subpopulations with $X_c = x$ and subpopulations with $X_c \neq x$.

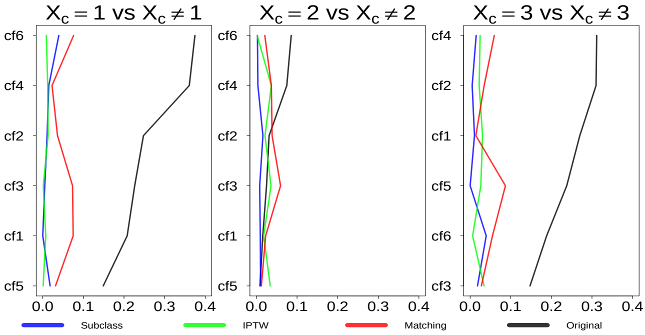


Figure: Absolute Standardized Bias (ASB). All three GPS implementations perform similarly and all improve confounder balance substantially.

Data Application: Background

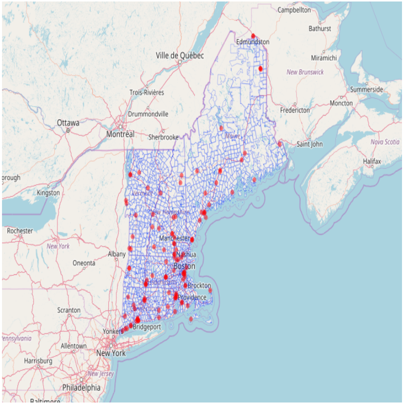


Figure: Locations of monitor stations in New England (in red). Zip code areas are drawn in blue.

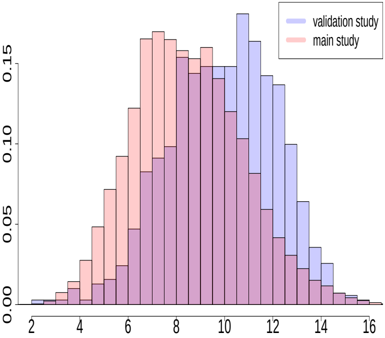


Figure: The distribution of annual mean predicted $PM_{2.5}$ exposures in the main and the validation study across 13 years (2000-2012).

Data Application: Main Results

Table: ATE of long-term $PM_{2.5}$ exposure on mortality measured by incidence rate ratios (IRRs). Error-prone implements GPS approaches to adjust confounding based on error-prone exposures. RC-GPS is based on the proposed approach adjusting for measurement error by RC model and adjusting confounding using GPS approaches based on corrected exposures. All 95% confidence intervals were obtained by bootstrap.

Results for Exposure Levels $PM_{2.5} \leq 8$ $\mu g/m^3$ vs. $8 < PM_{2.5} \leq 10$ $\mu g/m^3$			
	ATE [95% CI]		
	Subclassification	**IPTW**	**Matching**
GPS, Error-prone	1.013 [0.999, 1.029]	1.031 [1.021, 1.042]	1.020 [1.004, 1.036]
RC-GPS	1.025 [1.006, 1.045]	1.022 [1.007, 1.038]	1.028 [1.012, 1.045]
Results for Exposure Levels $PM_{2.5} \leq 8$ $\mu g/m^3$ vs. $PM_{2.5} > 10$ $\mu g/m^3$			
	ATE [95% CI]		
	Subclassification	**IPTW**	**Matching**
GPS, Error-prone	1.015 [0.993, 1.037]	1.050 [1.032, 1.068]	1.018 [0.996, 1.040]
RC-GPS	1.035 [0.999, 1.072]	1.030 [1.005, 1.056]	1.035 [1.015, 1.055]

Data Application: Covariate Balance

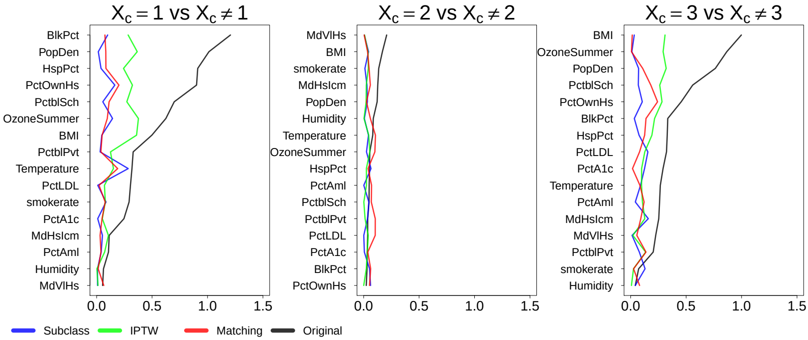


Figure: Absolute Standardized Bias (ASB). Each panel represents the absolute standardized differences for each confounders, between subpopulation with $X_c = x$ and subpopulation with $X_c \neq x$ in original data (black) and after GPS implementations (colored). All three GPS implementations improve the covariates balances for most of confounders.

Conclusions

- We develop an innovative approach to adjust for measurement error in causal inference setting.
  - Account for confounding using GPS.
- In simulations, we are able to fully adjust for the measurement error.
- In data application, we can detect causal effects of long-term exposure to $PM_{2.5}$ on all-cause mortality.
- Working on extending method to outcome analysis with continuous exposure.
  - Use continuous version of GPS.
  - Answer different policy question.
- Plan to control covariate balance using optimization methods.

Code

- Code to create the data set and implement the analysis described in the paper is available at: `https://github.com/wxwx1993/RC-GPS`
- Data files used to conduct the analysis are available on RCE: `/shared_space/ci3_nsaph/XiaoWu/RC_GPS`

References I

Alexeeff, S., Carroll, R., and Coull, B. (2015).
Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures.
Biostatistics, Under Revision.

Babanezhad, M., Vansteelandt, S., and Goetghebeur, E. (2010).
Comparison of causal effect estimators under exposure misclassification.
Journal of Statistical Planning and Inference, 140(5):1306–1319.

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z. J., Weinmayr, G., Hoffmann, B., Wolf, K., Samoli, E., Fischer, P., Nieuwenhuijsen, M., et al. (2014).
Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 european cohorts within the multicentre escape project.
The Lancet, 383(9919):785–795.

Braun, D., Gorfine, M., Parmigiani, G., Dominici, F., and Zigler, C. (2017).
Adjustment for mismeasured exposure using validation data and propensity scores.
Biostatistics, Accepted.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009).
Dealing with limited overlap in estimation of average treatment effects.
Biometrika, 96(1):187–199.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., and Schwartz, J. (2016a).
Assessing $PM_{2.5}$ Exposures with High Spatiotemporal Resolution across the Continental United States.
Environmental science & technology, 50(9):4712–4721.

References II

Di, Q., Koutrakis, P., and Schwartz, J. (2016b).
A hybrid prediction model for $PM_{2.5}$ mass and components using a chemical transport model and land use regression.
Atmospheric Environment, 131:390–399.

Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., Dominici, F., and Schwartz, J. D. (2017).
Air pollution and mortality in the medicare population.
New England Journal of Medicine, 376(26):2513–2522.

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris Jr, B. G., and Speizer, F. E. (1993).
An association between air pollution and mortality in six us cities.
New England journal of medicine, 329(24):1753–1759.

Harder, V. S., Stuart, E. A., and Anthony, J. C. (2010).
Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research.
Psychological methods, 15(3):234.

Imbens, G. (2000).
The role of the propensity score in estimating dose-response functions.

McCaffrey, D. F., Lockwood, J. R., and Setodji, C. M. (2013).
Inverse probability weighing with error-prone covariates.
Biometrika, 100(3):671–680.

McCandless, L. C., Richardson, S., and Best, N. (2012).
Adjustment for missing confounders using external validation data and propensity scores.
Journal of the American Statistical Association, 107(497):40–51.

References III

Pope III, C. A., Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E., and Heath Jr, C. W. (1995).
Particulate air pollution as a predictor of mortality in a prospective study of us adults.
American journal of respiratory and critical care medicine, 151(3_pt_1):669–674.

Shu, D. and Grace, Y. Y. (2018).
Estimation of causal effect measures in the presence of measurement error in confounders.
Statistics in Biosciences, pages 1–22.

Shu, D. and Yi, G. Y. (2017).
Causal inference with measurement error in outcomes: Bias analysis and estimation methods.
Statistical methods in medical research, page 0962280217743777.

Stürmer, T., Schneeweiss, S., Avorn, J., and Glynn, R. J. (2005).
Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration.
American journal of epidemiology, 162(3):279–289.

Webb-Vargas, Y., Rudolph, K. E., Lenis, D., Murakami, P., and Stuart, E. A. (2015).
An imputation-based solution to using mismeasured covariates in propensity score analysis.
Statistical methods in medical research, page 0962280215588771.

Yang, S., Imbens, G. W., Cui, Z., Faries, D. E., and Kadziola, Z. (2016).
Propensity score matching and subclassification in observational studies with multi-level treatments.
Biometrics, 72(4):1055–1065.

GPS Subclassification

- Consider classifying individuals into k groups based on each GPS element.
- To estimate $E[Y_j(x)]$:
  1. Estimate the average value of $Y_i(x)$ in subclass k.

$$\hat{\mu}_{k,x} = \frac{1}{N_{k,x}} \sum_{j: q^{p(x|\mathbf{c})}_{x,k-1} \leq p(x|\mathbf{C}_j) < q^{p(x|\mathbf{c})}_{x,k}, X_j = x} Y^{obs}_j$$

     where $q^{p(x|\mathbf{c})}_{x,k}$ is the value of GPS_x in k-th quantile.
  2. Estimate overall average:

$$\begin{aligned}\hat{E}[Y_j(x)] &= \hat{E}\left[E[Y^{obs}_j | X_{c,j} = x, p(x|\mathbf{C}_j)]\right] \\ &= \sum_{k=1}^{K} \frac{N_k}{N} \hat{\mu}_{k,x}\end{aligned}$$

     where N_k is the number of individuals with the x-th GPS element falling into the interval $[q^{p(x|\mathbf{c})}_{x,k-1},\ q^{p(x|\mathbf{c})}_{x,k})$.

GPS IPTW

- Similar to IPTW on PS, use inverse of GPS to weigh observations [Imbens, 2000].
- The average exposure effects with respect to the different level of exposures, can be expressed as,

$$\hat{E}[Y_j(x)] = \hat{E}[\frac{Y_j^{obs} I_j(x)}{p(x|\mathbf{C}_j)}].$$

GPS Matching

- Define a matching function [Yang et al., 2016];

$$m_{gps}(x,p) = argmin_{j:X_j=x}||p(x|\mathbf{C}_j) - p||.$$

  - Match individuals with $X_j = x$ based on $p(x|\mathbf{C}_j)$.
  - That is, for an exposure x and for each element $p \in GPS_x$, we find the jth observation which minimizes the matching function.
  - This jth observation has $X_j = x$ and is matched based on $p(x|C_j)$.
- Impute $Y_i(x)$ as: $\hat{Y}_i(x) = Y^{obs}_{m_{gps}(x,p(x|\mathbf{C}_i))}$.
- The overall average of $Y_j(x)$ can be expressed as;

$$\hat{E}[Y_j(x)] = \frac{1}{N}\sum_{i=1}^{N} Y^{obs}_{m_{gps(x,p(x|\mathbf{C}_i))}}.$$