# Causal Inference Methods and Case Studies

STAT24630

Jingshu Wang

# Lecture 12

Topic: Propensity score estimation

- Observational study v.s. conditional randomized experiment
- Propensity score estimation
  - Find the sets of possible confounding covariates
  - Logistic regression
  - Construct propensity score strata
  - Assess covariates balancing
- Textbook Chapter 13

# Causal inference with observational data

- The core rationale is to conceptualize observational studies as conditional randomized experiments
  - Analyze observational data as if treatment has been randomly assigned conditional on measured pre-treatment covariates $\boldsymbol{X}_i$ (unconfoundedness: $W_i \perp \left(Y_i(0), Y_i(1)\right) \mid \boldsymbol{X}_i$ )

> Therefore "what randomized experiment are you trying to emulate?" is a key question for causal inference from observational data. For each causal effect that we wish to estimate using observational data, we can describe (i) the target trial that we would like to, but cannot, conduct, and (ii) how the observational data can be used to emulate that target trial.

-- *Causal Inference: What If* (Herman and Robins, 2020)

  - Not all observational data can be conceptualized as a conditional randomized experiment!

# Observational study V.S. conditional randomized experiments

1. Conditional randomized experiment: $W_i \perp \left(Y_i(0), Y_i(1)\right) \mid \boldsymbol{X}_i$ is a fact as we control treatment assignment mechanism

   Observational study: $W_i \perp \left(Y_i(0), Y_i(1)\right) \mid \boldsymbol{X}_i$ is an assumption. It is always possible that this assumption is violated.

2. Conditional randomized experiment: $e(\boldsymbol{X}_i) = P(W_i = 1 \mid \boldsymbol{X}_i)$ is known

   Observational study: $e(\boldsymbol{X}_i) = P(W_i = 1 \mid \boldsymbol{X}_i)$ needs to be estimated. Can introduce bias and suffer from estimation uncertainty

# Need to evaluate identifiability assumptions carefully

- SUTVA
  - Can any variable have a causal effect? Are there multiple versions of assignment? We need "sufficiently well-defined interventions"
    Example: effect of sex, heart transplant by different techniques
  - Interventions may not be well defined as the experiment is not really conducted

- Overlap
  $e(\boldsymbol{X}_i) = P(W_i = 1 | \boldsymbol{X}_i) \in (0,1)$ or $P(W_i = w | \boldsymbol{X}_i = \boldsymbol{x}) > 0$ for all $\boldsymbol{x}$ and $w$
  - Guaranteed by the nature of experiments
  - Not guaranteed in observational studies

- $L$ only contains pre-treatment covariates
- Unconfoundedness: $W_i \perp \left(Y_i(0), Y_i(1)\right) | \boldsymbol{X}_i$ is an untestable assumption!!

# Estimate ATE with observation data

- We can still use outcome regression, IPW and matching estimators
- For IPW and matching estimators, as the propensity scores are unknown, we need to estimate the propensity scores from data first
- Once we estimate the propensity scores, we can replace the true propensity scores by their estimates in IPW or matching
- We need good estimates of the true propensity scores → not an easy task!
- We will also discuss other estimators that are more robust to a poor estimate of the propensity scores: blocking, trimming, doubly robust estimator

# Propensity score estimation procedure

What is the criteria of a good estimated propensity score?

- Estimate $e(\boldsymbol{X}_i) = P(W_i = 1 | \boldsymbol{X}_i)$: a classification problem but not exactly a classification problem
  - The goal is not simply minimizing the mean square error or classification error
  - A good propensity score needs to achieve covariates balancing $W_i \perp \boldsymbol{X}_i \mid \hat{e}(\boldsymbol{X}_i)$
  - Even if $\hat{e}(\boldsymbol{X}_i)$ is NOT an accurate estimate of the true $e(\boldsymbol{X}_i)$, as long as it achieves covariates balancing, $\hat{e}(\boldsymbol{X}_i)$ is at least a balancing score which leads to unconfoundedness given $\hat{e}(\boldsymbol{X}_i)$

- Two stages to estimate the propensity score:
  1) Use an initial specified model, such as logistic regression, to obtain $\hat{e}(\boldsymbol{X}_i)$
  2) Check covariate balancing based on weights or matched sets defined by $\hat{e}(\boldsymbol{X}_i)$
  3) We can iterate back and forth between the above two stages, each time refining the specified model

- During the whole process, we do not use the outcome data $Y_i^{\text{obs}}$

# The Barbiturate exposure data

- We aim to evaluate the effect of prenatal exposure to barbiturates
- The data set contains information on N = 7,943 men and women born between 1959 and 1961 in Copenhagen, Denmark.
- $N_t = 745$ men and women had been exposed in utero to substantial amounts of barbiturates due to maternal medical conditions. The comparison group consists of $N_c = 745$ individuals from the same birth cohort who were not exposed in utero to barbiturates.
- Outcome: barbiturate exposure on cognitive development in later years
- Treatment and control group can be systematically different: dataset contains 17 **pre-treatment** covariates that can potentially relate to both cognitive development and likelihood of being exposed to barbiturates

# The Barbiturate exposure data

**Table 13.1.** ***Summary Statistics Reinisch Data Set***

| Label | Variable Description | Controls ($N_c$ =7198) | | Treated ($N_t$ =745) | | t-Stat |
|---|---|---|---|---|---|---|
| | | Mean | (S.D.) | Mean | (S.D.) | Difference |
| sex | Sex of child (female is 0) | 0.51 | (0.50) | 0.50 | (0.50) | −0.3 |
| antih | Exposure to antihistamine | 0.10 | (0.30) | 0.17 | (0.37) | 4.5 |
| hormone | Exposure to hormone treatment | 0.01 | (0.10) | 0.03 | (0.16) | 2.5 |
| chemo | Exposure to chemotherapy agents | 0.08 | (0.27) | 0.11 | (0.32) | 2.5 |
| cage | Calendar time of birth | −0.00 | (1.01) | 0.03 | (0.97) | 0.7 |
| cigar | Mother smoked cigarettes | 0.54 | (0.50) | 0.48 | (0.50) | −3.0 |
| lgest | Length of gestation (10 ordered categories) | 5.24 | (1.16) | 5.23 | (0.98) | −0.3 |
| lmotage | Log of mother's age | −0.04 | (0.99) | 0.48 | (0.99) | 13.8 |
| lpbc415 | First pregnancy complication index | 0.00 | (0.99) | 0.05 | (1.04) | 1.2 |
| lpbc420 | Second pregnancy complication index | −0.12 | (0.96) | 1.17 | (0.56) | 55.2 |
| motht | Mother's height | 3.77 | (0.78) | 3.79 | (0.80) | 0.7 |
| motwt | Mother's weight | 3.91 | (1.20) | 4.01 | (1.22) | 2.0 |
| mbirth | Multiple births | 0.03 | (0.17) | 0.02 | (0.14) | −1.9 |
| psydrug | Exposure to psychotherapy drugs | 0.07 | (0.25) | 0.21 | (0.41) | 9.1 |
| respir | Respiratory illness | 0.03 | (0.18) | 0.04 | (0.19) | 0.7 |
| ses | Socioeconomic status (10 ordered categories) | −0.03 | (0.99) | 0.25 | (1.05) | 7.0 |
| sib | If sibling equal to 1, otherwise 0 | 0.55 | (0.50) | 0.52 | (0.50) | −1.6 |

# Logistic regression: specify a model to obtain $\hat{e}(\boldsymbol{X}_i)$

- Logistic regression is an extension of linear regression to regression binary response variable $W_i$ on the predictors $\widetilde{\boldsymbol{X}}_i$
  - Here, the predictors $\widetilde{\boldsymbol{X}}_i$ is not necessary the original set of pre-treatment covariates $\boldsymbol{X}_i$, we may drop some irrelevant covariates and add interaction terms
  - Logistic regression assumes the model

$$\pi_i = P\left(W_i = 1|\widetilde{\boldsymbol{X}}_i\right) = \frac{e^{\alpha+\boldsymbol{\beta}^T\widetilde{\boldsymbol{X}}_i}}{1+e^{\alpha+\boldsymbol{\beta}^T\widetilde{\boldsymbol{X}}_i}}$$

    or equivalently, $\text{logit}\left(P\left(W_i = 1|\widetilde{\boldsymbol{X}}_i\right)\right) = \alpha + \boldsymbol{\beta}^T\widetilde{\boldsymbol{X}}_i$
  - It also assumes that $W_i \sim \text{Bernoulli}(\pi_i)$
  - The log-likelihood function of the above model is

$$\sum_{i=1}^{N} W_i(\alpha + \boldsymbol{\beta}^T\widetilde{\boldsymbol{X}}_i) - \ln(1+\exp(\alpha + \boldsymbol{\beta}^T\widetilde{\boldsymbol{X}}_i))$$

  - We maximize the likelihood to obtain estimates $\hat{\alpha}$ and $\widehat{\boldsymbol{\beta}}$, and $\hat{e}(\boldsymbol{X}_i) = \frac{e^{\hat{\alpha}+\widehat{\boldsymbol{\beta}}^T\widetilde{\boldsymbol{X}}_i}}{1+e^{\hat{\alpha}+\widehat{\boldsymbol{\beta}}^T\widetilde{\boldsymbol{X}}_i}}$

# Selecting the covariates and interactions

- We can not include all 17 covariates and their 17*18/2 = 162 quadratic and interactions terms in the logistic regression, and want to select a subset of these terms

Step 1: select a subset of basic covariates based on scientific understanding

- Basic covariates: covariates that are a priori viewed as important for explaining the assignment and plausibly related to some outcome measures
- In the Barbiturate exposure data
  - lmotage: mother's age, which is plausibly related to cognitive outcomes for the child
  - ses: mother's socio-economic status, which is strongly related to the number of physician visits dur- ing pregnancies and thus exposes the mother to greater risk of barbiturate prescriptions
  - sex: sex of the child, which may be associated with measures of cognitive outcomes

| Variable | EST | (s. e.) | t-Stat |
|---|---|---|---|
| Intercept | −2.38 | (0.06) | −41.0 |
| sex | −0.01 | (0.08) | −0.2 |
| lmotage | 0.48 | (0.04) | 11.7 |
| ses | 0.10 | (0.04) | 2.6 |

# Selecting the covariates and interactions

## Step 2: add additional linear terms

- For each of the covariate not yet added, calculate the likelihood ratio statistics assessing the null hypothesis that the newly included covariate has a zero coefficient
- Add the covariate with the largest likelihood ratio statistics
- Stop if all likelihood ratio statistics of the remaining covariates are smaller to a cutoff (Say $C_L = 1$)
- Similar to forward stepwise regression

# Selecting the covariates and interactions

**Table 13.4.** ***Likelihood Ratio Statistics for Sequential Selection of Covariates to Enter Linearly; Barbiturate Data***

| Covariate | Step → | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| sex | – | – | – | – | – | – | – | – | – | – | – |
| antih | 17.5 | 0.5 | 1.6 | 1.3 | 2.1 | 1.8 | 1.6 | 1.6 | 1.7 | 1.3 | – |
| hormone | 3.9 | 0.3 | 0.7 | 0.7 | 0.4 | 0.8 | 0.7 | 0.7 | 0.7 | 0.8 | 0.9 |
| chemo | 10.0 | 36.6 | 41.9 | – | – | – | – | – | – | – | – |
| cage | 0.8 | 5.8 | 6.4 | 7.2 | 7.6 | 7.9 | – | – | – | – | – |
| cigar | 4.3 | 2.3 | 3.5 | 3.7 | 3.0 | 2.1 | 2.1 | 1.7 | 2.1 | – | – |
| lgest | 0.4 | 11.1 | 5.0 | 6.4 | 7.3 | 5.5 | 5.6 | – | – | – | – |
| lmotage | – | – | – | – | – | – | – | – | – | – | – |
| lpbc415 | 0.6 | 0.0 | 0.2 | 0.2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 |
| lpbc420 | 1308.0 | – | – | – | – | – | – | – | – | – | – |
| motht | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| motwt | 6.1 | 1.5 | 0.6 | 1.2 | 2.5 | 2.7 | 2.4 | 3.4 | – | – | – |
| mbirth | 4.6 | 66.1 | – | – | – | – | – | – | – | – | – |
| psydrug | 93.1 | 29.8 | 38.9 | 46.8 | – | – | – | – | – | – | – |
| respir | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| ses | – | – | – | – | – | – | – | – | – | – | – |
| sib | 21.0 | 13.8 | 12.5 | 15.0 | 15.7 | – | – | – | – | – | – |

# Selecting the covariates and interactions

## Step 3: add additional quadratic and interaction terms

- Say we now have $K_L$ linear covariates selected

- Quadratic and interaction terms are $K_L(K_L + 1)/2$
- Actual quadratic and interaction terms can be less as quadratic of binary covariate is itself

- Follow the same procedure in step 2 to add these terms sequentially
- There can be a different cutoff for the likelihood ratio statistics $C_Q$ (say $C_Q = 2.71$, corresponding to a 10% significance level)

# Final model for the estimated propensity score

| Variable | EST | (s.e.) | t-Stat |
|---|---|---|---|
| Intercept | −5.67 | (0.23) | −24.4 |
| Linear terms | | | |
| sex | 0.12 | (0.09) | 1.3 |
| lmotage | 0.52 | (0.11) | 4.7 |
| ses | 0.06 | (0.09) | 0.6 |
| lpbc420 | 2.37 | (0.36) | 6.6 |
| mbirth | −2.11 | (0.36) | −5.9 |
| chemo | −3.51 | (0.67) | −5.2 |
| psydrug | −3.37 | (0.55) | −6.1 |
| sib | −0.24 | (0.22) | −1.1 |
| cage | −0.56 | (0.26) | −2.2 |
| lgest | 0.57 | (0.23) | 2.5 |
| motwt | 0.49 | (0.17) | 2.9 |
| cigar | −0.15 | (0.10) | −1.5 |
| antih | 0.17 | (0.13) | 1.3 |
| Second-order terms | | | |
| lpbc420 × sib | 0.60 | (0.19) | 3.1 |
| motwt× motwt | −0.10 | (0.02) | −4.5 |
| lpbc420 × psydrug | 1.88 | (0.39) | 4.8 |
| ses× sib | −0.22 | (0.10) | −2.2 |
| cage× antih | −0.39 | (0.14) | −2.8 |
| lpbc420× chemo | 1.97 | (0.49) | 4.0 |
| lpbc420 × lpbc420 | −0.46 | (0.14) | −3.3 |
| cage × lgest | 0.15 | (0.05) | 3.0 |
| lmotage × lpbc420 | −0.24 | (0.10) | −2.5 |
| mbirth× cage | −0.88 | (0.39) | −2.3 |
| lgest × lgest | −0.04 | (0.02) | −2.0 |
| ses× cigar | 0.20 | (0.09) | 2.2 |
| lpbc420× motwt | 0.15 | (0.07) | 2.0 |
| chemo × psydrug | −0.93 | (0.46) | −2.0 |
| lmotage× ses | 0.10 | (0.05) | 1.9 |
| cage × cage | −0.10 | (0.05) | −1.8 |

# Construct propensity score strata

- At the second stage, we need to evaluate covariates balancing based on $\hat{e}(\boldsymbol{X}_i)$

$$W_i \perp \boldsymbol{X}_i \mid \hat{e}(\boldsymbol{X}_i)$$

- Ideally, we want to stratify samples into blocks so that **each block has the exact same value of $\hat{e}(\boldsymbol{X}_i)$**, and assess whether $W_i \perp \boldsymbol{X}_i$ within each block.

- In practice, we need to coarsen $\hat{e}(\boldsymbol{X}_i)$ into discrete values

- Define a set of boundary points: $0 = b_0 < b_1 < \cdots < b_J = 1$
- Define block indicators

$$B_i(j) = \begin{cases} 1 & \text{if } b_{j-1} \leq \hat{e}(X_i) < b_j, \\ 0 & \text{otherwise,} \end{cases}$$

- We then assess: $W_i \perp \boldsymbol{X}_i \mid B_i(1), \cdots, B_i(J))$

# Find boundary points

How to find the boundary points $b_0 < b_1 < \cdots < b_J$?

- Intuitively, we want to make sure that $\hat{e}(\boldsymbol{X}_i)$ within each block / strata are close enough to each other
- Practically, we can check if $W_i \perp \hat{e}(\boldsymbol{X}_i)$ within each block
- We need estimated propensity score to be balanced within each strata, so that discretizing $\hat{e}(\boldsymbol{X}_i)$ do not introduce an extra bias

- Steps:
  1. Preprocessing: remove units if their estimated propensity score is too large or too small
     - Define $\underline{e}_{\mathrm{t}} = \min_{i:W_i=1} \hat{e}(X_i)$ , remove a control unit $i$ if $\hat{e}(\boldsymbol{X}_i) < \underline{e}_t$
     - Define $\bar{e}_{\mathrm{c}} = \max_{i:W_i=0} \hat{e}(X_i)$ , remove a treated unit $i$ if $\hat{e}(\boldsymbol{X}_i) < \bar{e}_c$
     - Ensure that there are both enough treated and control units within each strata

# Find boundary points

- Steps:
  1. Preprocessing: remove units if their estimated propensity score is too large or too small
  2. Sequential block splitting
     - Start with a single block $J = 1$ with $b_0 = \underline{e}_t$ and $b_1 = \bar{e}_c$
     - Define linearized propensity score

$$\hat{l}(\boldsymbol{X}_i) = \ln\left(\frac{\hat{e}(\boldsymbol{X}_i)}{1 - \hat{e}(\boldsymbol{X}_i)}\right)$$

     - For each of the current blocks, we assess whether we need to further split it into two
       - For block $j$, need to evaluate whether $W_i \perp \hat{l}(\boldsymbol{X}_i)$ within the block
       - Define the two-sample test statistics (assume equal variance of the two groups)

$$t_j = \frac{\bar{\ell}_{\mathrm{t}}(j) - \bar{\ell}_{\mathrm{c}}(j)}{\sqrt{s_\ell^2(j) \cdot (1/N_{\mathrm{c}}(j) + 1/N_{\mathrm{t}}(j))}}$$

       - Need to split Block $j$ into two blocks if $|t_j| > t_{\max} = 1.96$
       - Define the two sub-blocks: find the median of $\hat{e}(\boldsymbol{X}_i)$ within block $j$ as $b_j'$
         - Sub-block 1: all units with $\hat{e}(\boldsymbol{X}_i) < b_j'$; sub-block 2: all units with $\hat{e}(\boldsymbol{X}_i) \geq b_j'$
  3. Stopping rule: stop if every block either does not need to split or has a small enough size (too small to split)

# Construct blocks for Barbiturate exposure data

- Proprocessing: removed 2737 controls and 3 treated units

| Step | Block | Lower Bound | Upper Bound | Width | # Controls | # Treated | t-Stat | Median $\hat{e}(\boldsymbol{X}_i)$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.00 | 0.94 | 0.94 | 4462 | 742 | 36.3 | 0.06 |
| 2 | **1** | 0.00 | 0.06 | 0.06 | 2540 | 61 | 3.2 | 0.02 |
| | **2** | 0.06 | 0.94 | 0.88 | 1922 | 681 | 23.7 | 0.20 |
| 3 | **1** | 0.00 | 0.02 | 0.01 | 1280 | 20 | 2.2 | 0.01 |
| | **2** | 0.02 | 0.06 | 0.05 | 1260 | 41 | 0.5 | |
| | **3** | 0.06 | 0.20 | 0.14 | 1163 | 138 | 3.9 | 0.11 |
| | **4** | 0.20 | 0.94 | 0.74 | 759 | 543 | 10.9 | 0.37 |
| 4 | **1** | 0.00 | 0.01 | 0.00 | 644 | 6 | −0.0 | |
| | **2** | 0.01 | 0.02 | 0.01 | 636 | 14 | 1.7 | |
| | 3 | 0.02 | 0.06 | 0.05 | 1260 | 41 | 0.5 | |
| | **4** | 0.06 | 0.11 | 0.05 | 604 | 46 | −0.3 | |
| | **5** | 0.11 | 0.20 | 0.09 | 559 | 92 | 1.0 | |
| | **6** | 0.20 | 0.37 | 0.17 | 458 | 192 | 1.2 | |
| | **7** | 0.37 | 0.94 | 0.57 | 301 | 351 | 5.6 | 0.5 |

$t_{\max} = 2$

# Construct blocks for Barbiturate exposure data

| Step | Block | Lower Bound | Upper Bound | Width | # Controls | # Treated | t-Stat |
|---|---|---|---|---|---|---|---|
| 5 | 1 | 0.00 | 0.01 | 0.00 | 644 | 6 | −0.0 |
| | 2 | 0.01 | 0.02 | 0.01 | 636 | 14 | 1.7 |
| | 3 | 0.02 | 0.06 | 0.05 | 1260 | 41 | 0.5 |
| | 4 | 0.06 | 0.11 | 0.05 | 604 | 46 | −0.3 |
| | 5 | 0.11 | 0.20 | 0.09 | 559 | 92 | 1.0 |
| | 6 | 0.20 | 0.37 | 0.17 | 458 | 192 | 1.2 |
| | **7** | 0.37 | 0.50 | 0.13 | 181 | 144 | 2.5 |
| | **8** | 0.50 | 0.94 | 0.44 | 120 | 207 | 2.3 |
| 6 | 1 | 0.00 | 0.01 | 0.00 | 644 | 6 | −0.0 |
| | 2 | 0.01 | 0.02 | 0.01 | 636 | 14 | 1.7 |
| | 3 | 0.02 | 0.06 | 0.05 | 1260 | 41 | 0.5 |
| | 4 | 0.06 | 0.11 | 0.05 | 604 | 46 | −0.3 |
| | 5 | 0.11 | 0.20 | 0.09 | 559 | 92 | 1.0 |
| | 6 | 0.20 | 0.37 | 0.17 | 458 | 192 | 1.2 |
| | **7** | 0.37 | 0.42 | 0.05 | 101 | 61 | 0.3 |
| | **8** | 0.42 | 0.50 | 0.08 | 80 | 83 | 0.7 |
| | **9** | 0.50 | 0.61 | 0.11 | 73 | 90 | 0.8 |
| | **10** | 0.61 | 0.94 | 0.34 | 47 | 117 | −0.3 |

# Assess covariates balancing given the blocks

- Within each block, we test for the null hypothesis

$$\mathbb{E}\left[X_i | W_i = 1, B_i(j) = 1\right] = \mathbb{E}\left[X_i | W_i = 0, B_i(j) = 1\right]$$

- For each covariate $k$, construct t-statistics within block $j$
  - Sample mean difference and its estimated squared standard error (assume equal variance)

$$\hat{\tau}_k^X(j) = \overline{X}_{\mathrm{t},k}(j) - \overline{X}_{\mathrm{c},k}(j) \qquad \hat{\mathbb{V}}_k^X(j) = s_k^2(j) \cdot \left(\frac{1}{N_{\mathrm{c}}(j)} + \frac{1}{N_{\mathrm{t}}(j)}\right)$$

  - Within-block t-statistics: $z_k(j) = \dfrac{\hat{\tau}_k^X(j)}{\sqrt{\hat{\mathbb{V}}_k^X(j)}}$

- Overall t-statistics averaged across blocks $\hat{\tau}_k^X = \sum_{j=1}^{J} \dfrac{N_{\mathrm{c}}(j) + N_{\mathrm{t}}(j)}{N} \cdot \hat{\tau}_k^X(j), \;\; \hat{\mathbb{V}}_k^X = \sum_{j=1}^{J} \left(\dfrac{N_{\mathrm{c}}(j) + N_{\mathrm{t}}(j)}{N}\right)^2 \cdot \hat{\mathbb{V}}_k^X(j)$

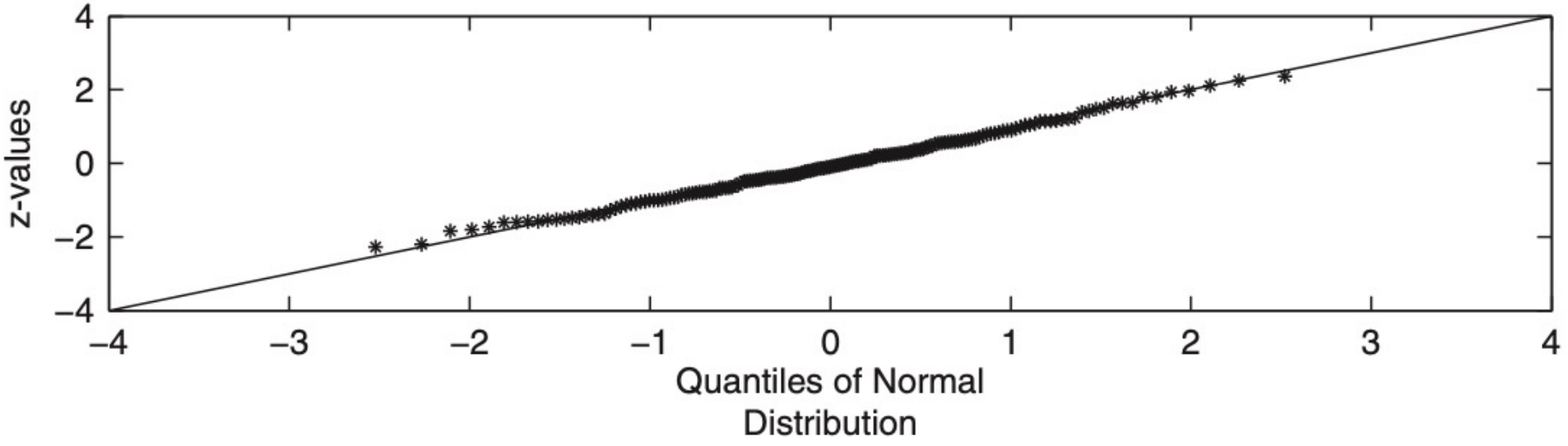
$$z_k = \frac{\hat{\tau}_k^X}{\sqrt{\hat{\mathbb{V}}_k^X}}$$

# Covariate balancing for Barbiturate exposure data

| | Within Blocks | | | | | | | | | | Overall | | 1-Block |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | t-Test | F-Test (z-Value) | t-Test |
| Covariate | | | | | | | | | | | | | |
| sex | −0.05 | −2.27 | 1.97 | 0.81 | 0.89 | −1.28 | 0.04 | −0.39 | −1.42 | 1.14 | 0.13 | 1.22 | −0.73 |
| antih | −0.67 | −0.47 | 0.67 | 0.03 | 0.37 | −0.25 | 0.38 | −0.53 | −0.11 | 0.27 | −0.17 | −2.88 | 3.21 |
| hormone | −0.14 | −0.42 | −0.65 | −1.00 | 0.25 | 0.71 | −0.22 | −1.05 | −1.10 | 0.21 | −0.99 | −0.66 | 1.66 |
| chemo | 0.55 | −0.39 | −0.78 | −0.75 | −1.17 | 1.47 | −0.94 | 0.61 | 0.66 | 0.29 | −0.27 | −0.61 | 1.76 |
| cage | −1.41 | −0.29 | −1.04 | −0.46 | 2.11 | 0.28 | 0.20 | 0.46 | −1.48 | −0.74 | −1.38 | 0.34 | 1.15 |
| cigar | −0.37 | 0.55 | 0.58 | 1.50 | 0.31 | −0.93 | 0.21 | −0.99 | 0.25 | −0.39 | 0.52 | −1.17 | −3.13 |
| lgest | 0.90 | 0.58 | −0.07 | −0.82 | 0.79 | −0.36 | 0.05 | −0.33 | −1.14 | 1.21 | 0.71 | −1.48 | 0.12 |
| lmotage | −2.20 | −1.37 | 0.56 | 1.64 | 0.95 | 0.60 | −0.96 | −1.73 | −1.47 | 0.36 | −1.26 | 1.45 | 8.56 |
| lpbc415 | −0.48 | −1.84 | −1.00 | −0.34 | 0.59 | 0.44 | −0.20 | −0.16 | 1.07 | −0.10 | −1.49 | −0.82 | 0.75 |
| lpbc420 | 1.04 | 0.84 | −0.67 | −0.86 | −1.61 | 1.80 | −0.39 | 1.62 | 1.14 | −1.80 | 0.51 | 0.59 | 32.04 |
| motht | −0.84 | 0.45 | −0.67 | 0.75 | 0.64 | 0.09 | 0.30 | −1.37 | −0.60 | −0.13 | −0.50 | −1.37 | 0.90 |
| motwt | 1.23 | 1.14 | 0.12 | −1.23 | −0.05 | −0.45 | −0.32 | 1.94 | −0.01 | −0.47 | 1.08 | −0.18 | 1.44 |
| mbirth | −0.44 | −0.80 | −1.54 | −0.37 | 1.80 | 0.20 | 0.00 | 2.25 | −1.58 | −1.60 | −1.28 | 1.00 | −2.93 |
| psydrug | −0.66 | −1.01 | 1.05 | −0.15 | −0.78 | 0.06 | −0.18 | 0.08 | 0.09 | 0.89 | −0.29 | −1.40 | 6.32 |
| respir | −0.49 | 0.53 | −0.21 | 0.98 | 1.38 | 0.24 | −0.78 | −1.51 | 0.22 | −0.28 | 0.24 | −0.49 | 0.19 |
| ses | −0.60 | −0.31 | −0.74 | 1.16 | 0.82 | −0.08 | −0.03 | −0.82 | −0.91 | 0.36 | −0.56 | −1.37 | 5.19 |
| sib | 1.42 | 2.37 | −1.09 | −1.58 | −1.53 | 0.11 | 0.63 | 1.63 | 1.19 | 0.23 | 0.98 | 1.64 | 1.48 |

# Evaluate covariate balancing within the blocks

- We want to know if there are any imbalancing within any strata
- For 170 t-statsitics for 10 blocks and 17 covariates, plot QQ plot to assess whether these statistics follow $\mathcal{N}(0,1)$



# What if we adjust for all covariates but no interactions or quadratic forms?

- In practice, what is often done is that we run a logistic regression on all available covariates without adding any interactions and quadratic terms
- This is equivalent to setting $C_L = 0, C_Q = \infty$

- Within block statistics have heavier tailed than standard Gaussian

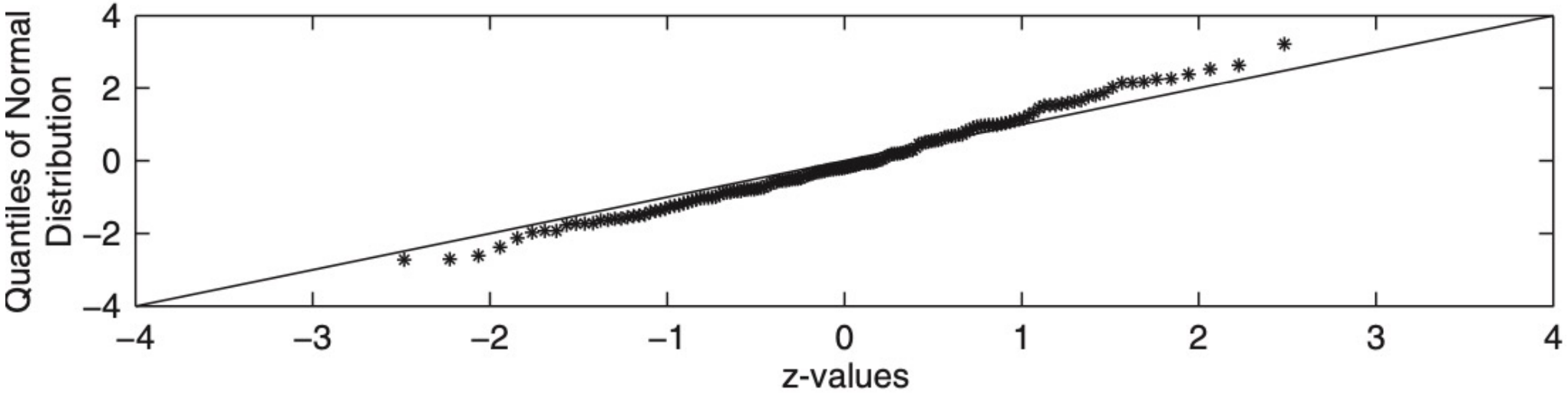


**Table 13.9.** *z-Values for Balancing Tests: Simple Linear Propensity Score Specification; Barbiturate Data*

| | Within Blocks | | | | | | | | | Overall | | 1-Block |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | t-Test | F-Test (z-Value) | t-Test |
| Covariate | | | | | | | | | | | | |
| sex | 1.68 | 0.41 | −0.39 | 0.09 | −0.25 | −0.51 | 0.78 | −0.63 | −0.20 | 1.47 | −1.16 | −0.87 |
| antih | −0.98 | 1.75 | 0.17 | 0.29 | −1.11 | 0.60 | −0.51 | −0.07 | 0.68 | −0.18 | −0.54 | 3.43 |
| hormone | −0.34 | −0.75 | −0.45 | 1.23 | −1.38 | 0.73 | 1.23 | 0.22 | −0.54 | −0.58 | −0.16 | 1.78 |
| chemo | −1.00 | −2.37 | −0.37 | −0.90 | −1.44 | −1.22 | 2.36 | 1.88 | 0.51 | −2.03 | 2.41 | −0.02 |
| cage | −2.54 | 0.38 | −1.40 | 1.08 | 0.60 | −0.71 | 1.76 | −0.59 | −0.07 | −2.07 | 1.11 | 0.86 |
| cigar | −0.41 | 0.61 | −0.36 | 0.95 | 2.21 | −1.16 | −0.87 | −1.59 | 0.67 | 0.04 | 0.70 | −2.96 |
| lgest | −0.06 | −0.81 | 1.06 | 1.88 | −0.63 | 1.18 | −0.92 | −1.86 | 1.19 | −0.01 | 0.80 | −0.31 |
| lmotage | 0.50 | 1.66 | 1.86 | 1.30 | 2.04 | −0.10 | −1.34 | −2.57 | −0.63 | 1.58 | 2.26 | 10.74 |
| lpbc415 | −1.10 | −1.10 | −1.53 | 0.42 | 0.91 | 0.46 | 0.40 | 0.48 | −0.03 | −1.34 | −0.58 | 0.98 |
| lpbc420 | 1.69 | −1.93 | 0.73 | −1.97 | −1.93 | 0.17 | 2.63 | 2.52 | 1.82 | 0.77 | 3.09 | 36.35 |
| motht | −1.94 | 0.61 | 0.19 | −0.27 | 1.02 | −0.48 | −0.15 | 0.27 | −0.59 | −1.35 | −0.70 | 0.57 |
| motwt | −0.92 | 0.34 | −0.70 | −1.59 | −0.94 | 0.30 | 0.06 | −0.07 | 1.43 | −1.01 | −0.29 | 1.31 |
| mbirth | −0.65 | −0.91 | 2.95 | −1.22 | −1.22 | 3.24 | 1.35 | −0.85 | −1.65 | −0.62 | 2.33 | −3.26 |
| psydrug | −0.25 | −1.37 | −0.02 | −0.72 | −1.50 | −1.94 | 0.63 | 0.45 | 2.76 | −1.30 | 3.09 | 7.20 |
| respir | −0.63 | −0.60 | 1.97 | −1.00 | 1.27 | 0.49 | 0.08 | −0.39 | −0.59 | −0.30 | 0.05 | 0.19 |
| ses | −0.30 | 1.62 | 1.52 | 0.03 | 0.87 | −0.12 | −1.92 | −1.40 | 1.14 | 0.63 | 0.97 | 5.61 |
| sib | −2.24 | −1.00 | −2.24 | −1.67 | −2.80 | 0.25 | 1.58 | 2.21 | 2.18 | −2.93 | 3.09 | −0.78 |