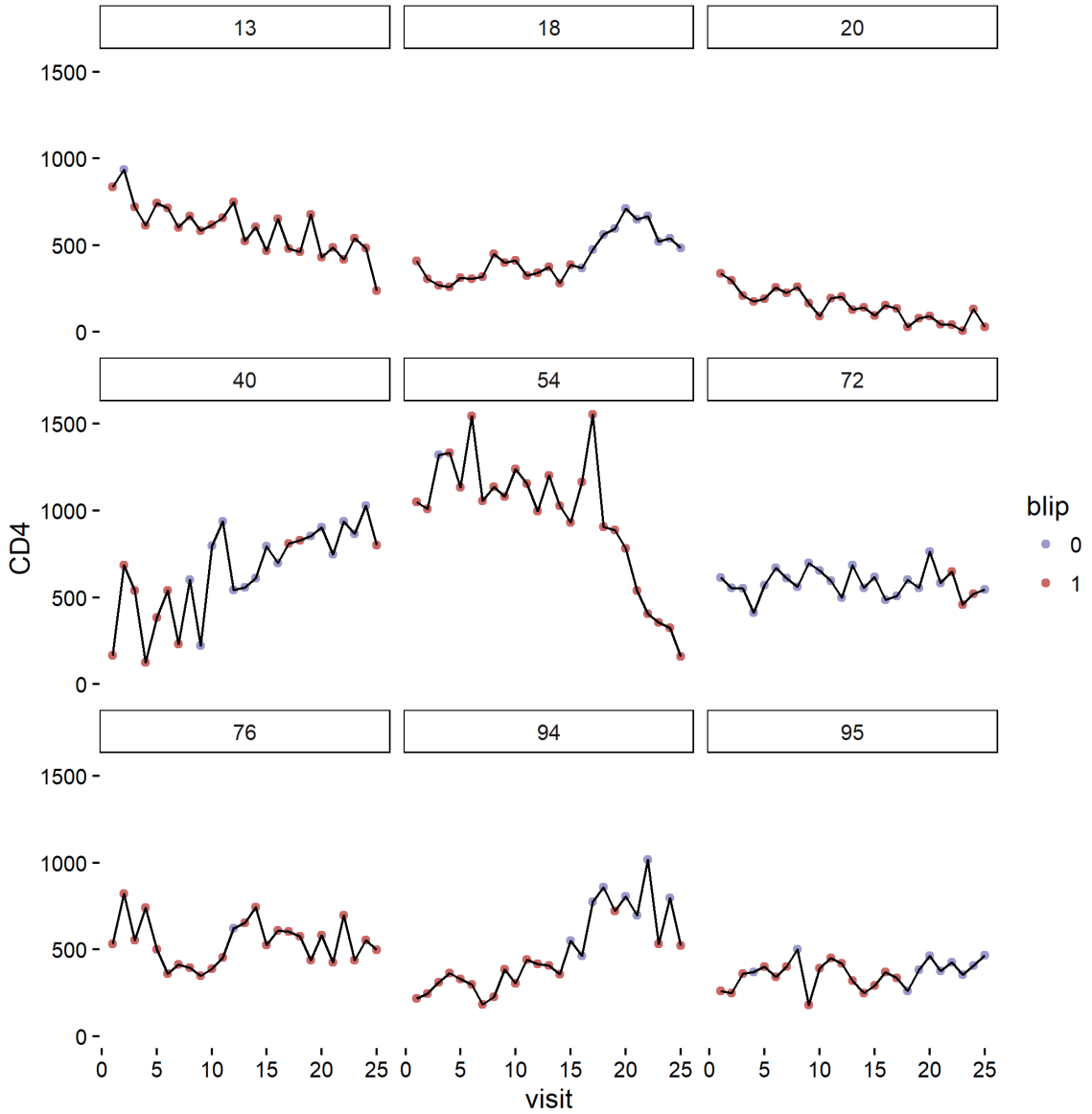
The Personalized and Population AUCs

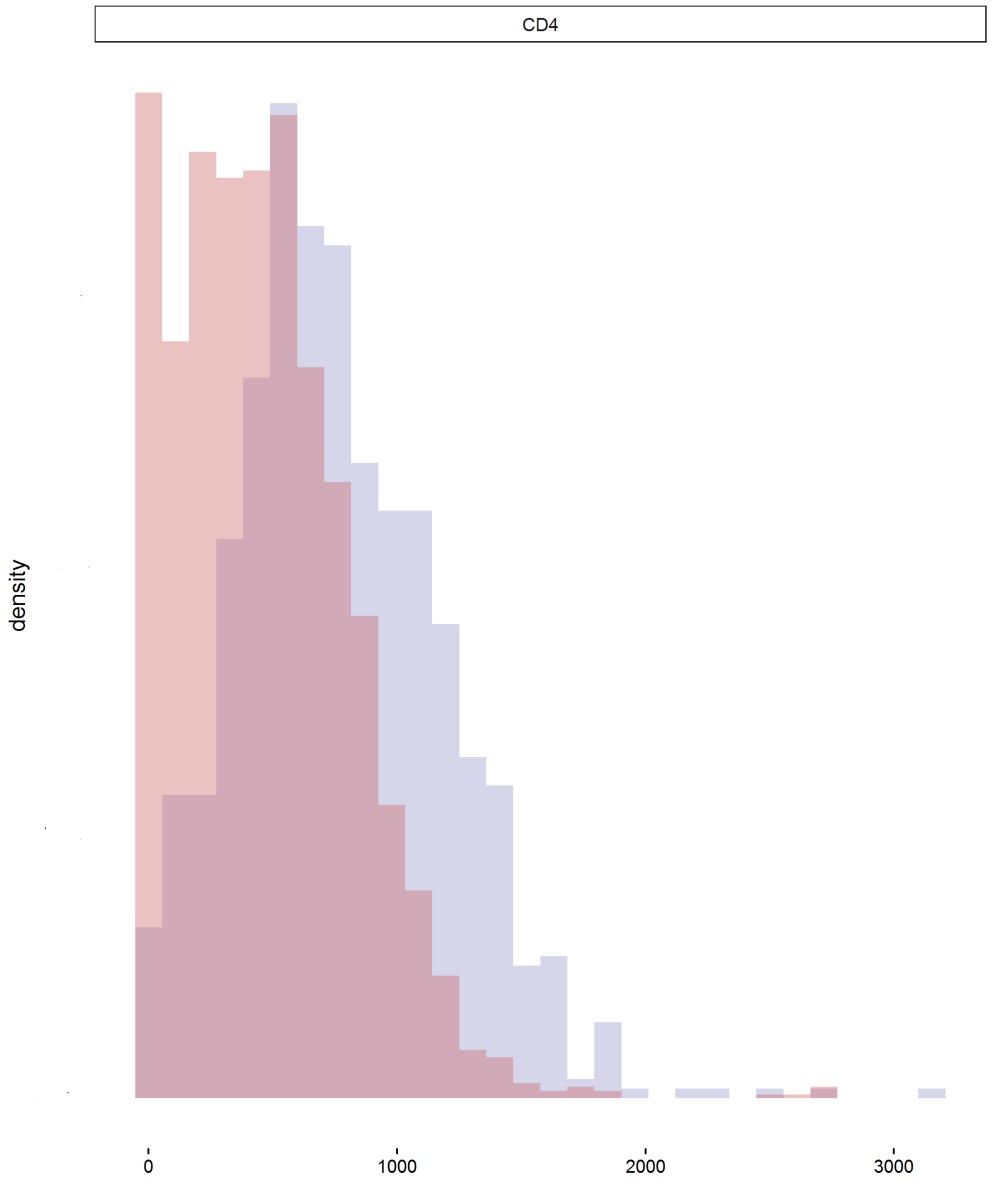
Outline

A motivating example

Data: The Yale Prospective Longitudinal HIV Cohort

Problem: Evaluate CD4 as a predictor of blip status




CD4
density
0
1000
2000
3000

	control	case
obs. # 1	X_1	
⋮	⋮	
obs. # k	X_k	
obs. # k+1		Y_{k+1}
⋮		⋮
obs. # N		Y_N

The AUC is the probability that an observation drawn from a negative/control/non-diseased subject is less than an independent observation from a positive/case/diseased subject.

$$\theta(P) = P(X < Y) = \mathrm{E}(F_X(Y))$$

$$\hat{\theta}(\vec{X}, \vec{Y}) = \frac{1}{k(N-k)} \sum_{i,j} \{X_i < Y_j\}$$

We wish to extend the AUC to clusters

- markers: longitudinal measurements of tumour antigens (CEA, CA15-3, TPS), response: progression/non-progression of breast cancer (Emir 2000)
- markers: how long an officer detains a suspect (clustered by officer), response: non-Black ("control") or Black ("case") suspect status (Ridgeway 2006)

	control	case
unit # 1	$X_1 = (X_{11}, \ldots, X_{1m_1})$	$Y_1 = (Y_{11}, \ldots, Y_{1n_1})$
$\vdots$	$\vdots$	$\vdots$
unit #N	$X_N = (X_{N1}, \ldots, X_{Nm_N})$	$Y_N = (Y_{N1}, \ldots, Y_{Nn_N})$

- Units are IID, but a given unit's observations are dependent
- X and Y are vectors of control and case observations of lengths M and N
- the lengths M and N are random
- (X, Y, M, N) is a random vector with an unknown distribution

With (X_1, Y_1, M_1, N_1) and (X_2, Y_2, M_2, N_2), being two independent draws, we define the population AUC as

$$\theta_{12}(P) = \frac{1}{\mathrm{E}(M_1)\,\mathrm{E}(N_2)}\,\mathrm{E}\left(\sum_{i=1}^{M_1}\sum_{j=1}^{N_2}\{X_{1i} < Y_{2j}\}\right)$$

- ignoring the cluster structure, the probability that a randomly selected control observation is less than an independently selected case obesrvation
- this parameter/the model appears to be implied by what the applied researchers have in mind

- the medical field has lately focused on personalizing treatment
- "The individuality of the patient should be at the core of every treatment decision. One-size-fits-all approaches to treating medical conditions are inadequate; instead, treatments should be tailored to individuals . . ." (National Academy of Medicine, 2018)

Besides the population AUC

$$\theta_{12}(P) = \frac{1}{\mathrm{E}(M_1)\,\mathrm{E}(N_2)}\,\mathrm{E}\left(\sum_{i=1}^{M_1}\sum_{j=1}^{N_2}\{X_{1i} < Y_{2j}\}\right)$$

we define the personalized AUC as:

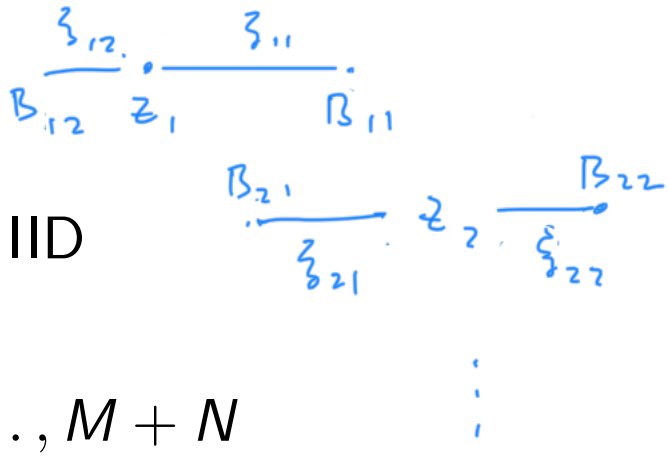
$$\theta_{11}(P) = \mathrm{E}\left(\frac{\sum_{i=1}^{M_1}\sum_{j=1}^{N_1}\{X_{1i} < Y_{1j}\}}{M_1 N_1}\right)$$

.

- Both the population and personalized AUC, like the usual AUC, are bounded between 0 and 1, 1/2 represents poor discrimination, and distance from 1/2 represents increasing discrimination.
- Differences:
  - the population AUC is the probability that a typical control observation in the population is less than a typical case observation
  - the personalized AUC is the AUC of a typical cluster

Outline

binary response model



individual effects $\vec{\xi} = (\xi_1, \ldots, \xi_{M+N})$ IID

cluster effects $Z \perp\!\!\!\perp (\xi_1, \ldots, \xi_{M+N})$

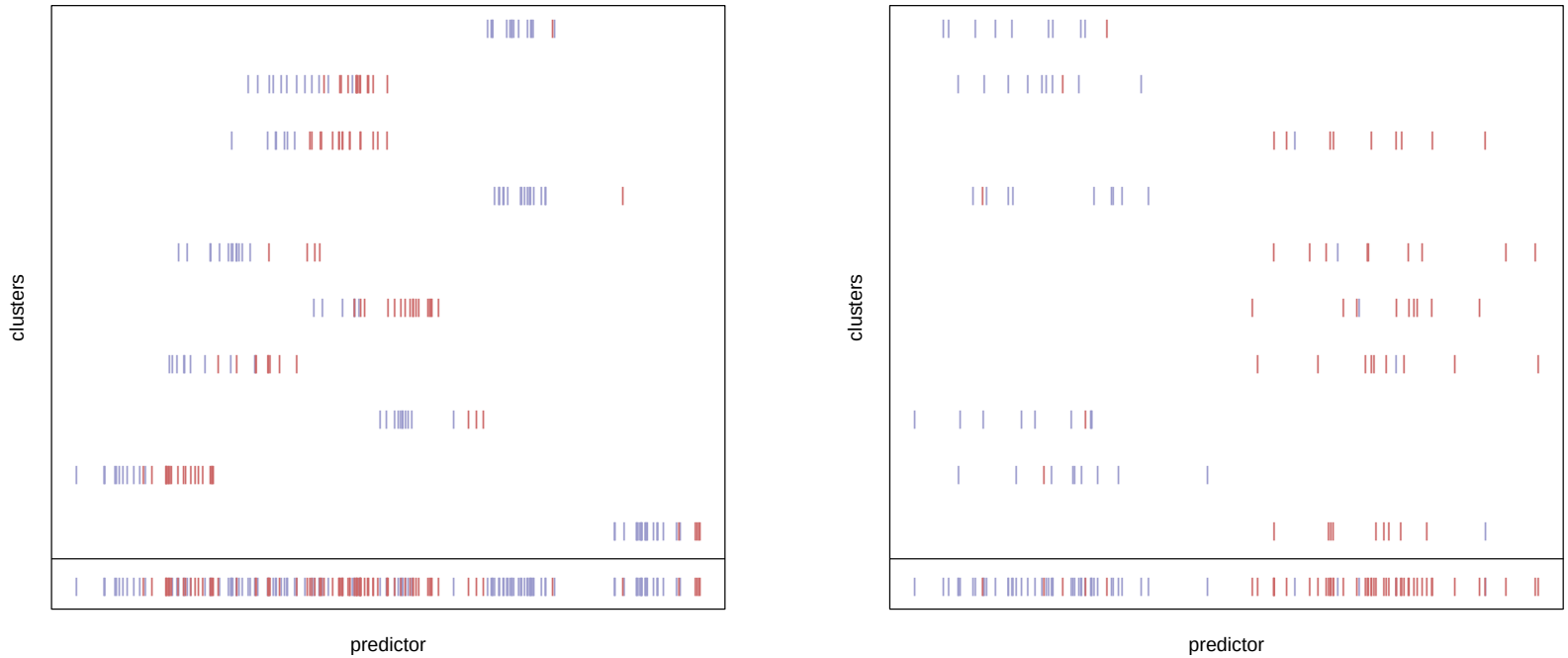
markers $B_i = Z + \xi_i, i = 1, \ldots, M + N$

case status indicators

$$D_i \mid \vec{Z}, \vec{\xi} \sim \text{bernoulli with parameter } \sigma(\beta_0 Z + \beta_1 \xi_i)$$

$$M = \sum_{i=1}^{M+N} (1 - D_i), \qquad N = \sum_{i=1}^{M+N} D_i.$$

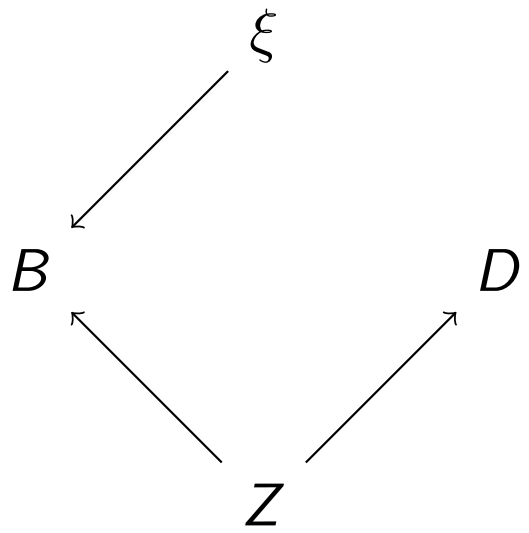
The control and case observations in a cluster, X_i and Y_i, are B_i such that $D_i = 0$ and $D_i = 1$



Left: informative personalized AUC, uninformative population AUC
Right: informative population AUC, uninformative personalized AUC

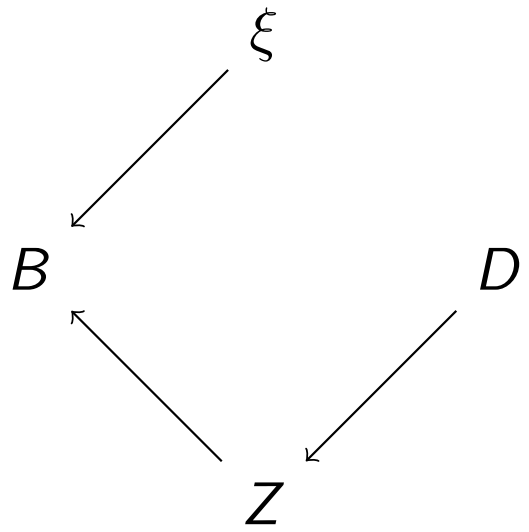
For $\beta_0 > 0$ and $\beta_1 = 0$, the population AUC is $> 1/2$ whereas the personalized AUC is $= 1/2$

cluster effect but no individual effect, Ex. #1



- the cluster effect Z represents a genuine signal of disease status D, such as viral load wrt HIV status
- ξ represents non-systematic measurement error on instruments measuring Z
- the population AUC better matches expectations of an AUC measurement than the personalized AUC, as the biomarker $B = Z + \xi$ isn't completely uninformative

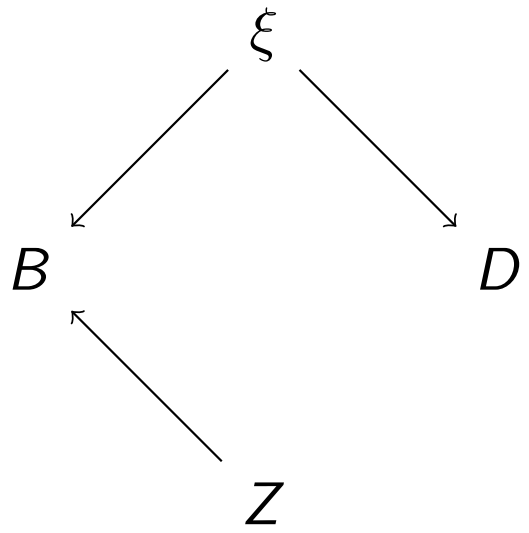
cluster effect but no individual effect, Ex. #2



- The cluster effect Z is a subject's dose of a possibly ineffective drug, and larger doses are administered to sicker patients.
- ξ represents non-systematic measurement error on instruments measuring Z as before
- the association between the marker and disease status implied by the population AUC is spurious

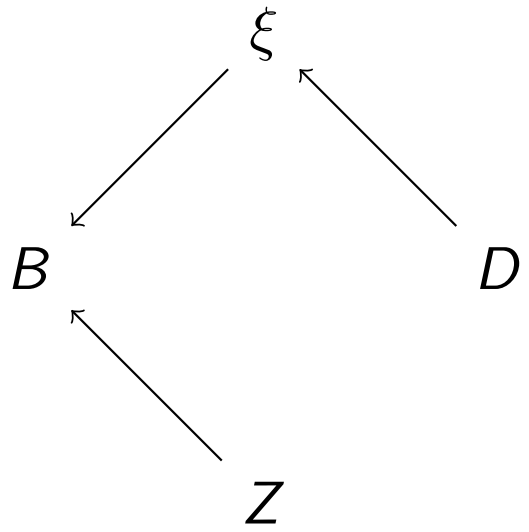
For $\beta_0 = 0$ and $\beta_1 > 0$, the population AUC is $\approx 1/2$ and the individual AUC is $> 1/2$

individual effect but no cluster effect, Ex. #1



- The markers B are measurements on a patient
- D indicates the presence of a disease with little or no causal dependence on a baseline measure Z but is indicated by the deviations ξ from the baseline
- Here the personalized AUC probably carries the correct interpretation

individual effect but no cluster effect, Ex. #2



- cluster effect Z is independent of D again
- ξ is the dose of a drug taken
- sicker individuals, those for which D is more likely to be 1, take higher doses ξ of the drug.
- personalized AUC represents a spurious association

- the marker is not a cause of case/disease status in any of the interpretations given above
- in each case, either the marker and disease status are both downstream effects of Z and $\vec{\xi}$, or the marker is an effect of the status

Outline

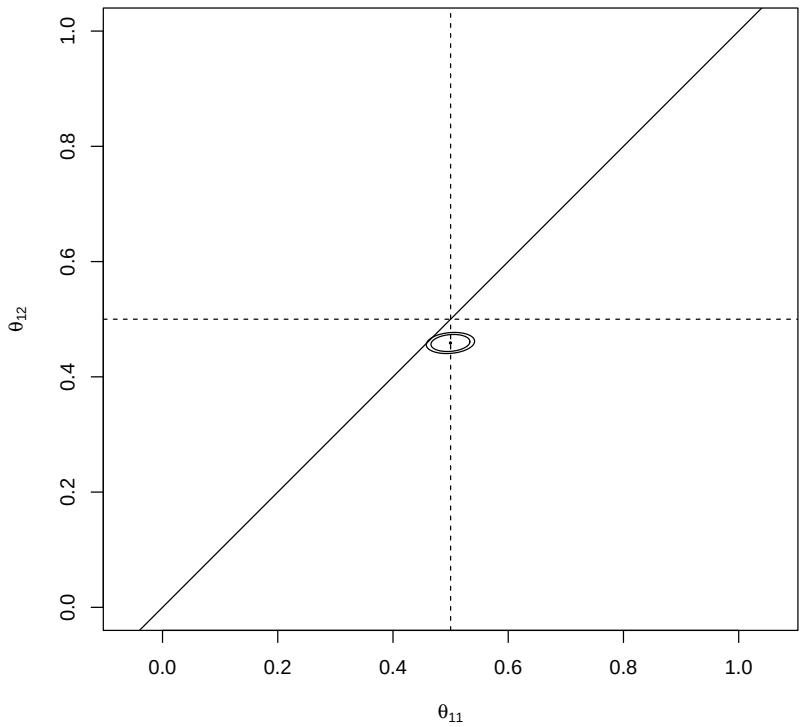
- The data consists of Terry stops in New York City and Boston.
- The analysis here focuses on the relationship between the duration of the stop and race of the suspect.
- We cluster the stops according to precinct, in the case of NYC, and according to the officer conducting the stop, in the case of Boston.

	NYC			Boston		
group	mean duration (SD)	count	freq.	mean duration (SD)	count	freq.
Asian	14.24 (21.16)	1139	0.02	25.00 (24.22)	53	0.01
Black Hispanic	11.01 (17.12)	4675	0.09	15.28 (18.73)	391	0.06
Black non-Hispanic	10.99 (16.78)	31588	0.58	19.06 (28.93)	3448	0.55
White Hispanic	11.21 (15.15)	11486	0.21	15.63 (15.96)	578	0.09
White non-Hispanic	12.85 (16.18)	4854	0.09	21.74 (33.01)	1760	0.28
other	11.84 (17.70)	261	0.00	20.89 (23.90)	93	0.01

Table: Summary estimates on the duration of Terry stops by racial group.

$\theta_{12} < \theta_{11}$

- With Black race as the binary classification, the AUC analysis looks for a difference in location between the distribution of stop durations of non-Black ("control") and Black ("case") suspects.
- For the NYC data, the population AUC estimate is $\hat{\theta}_{12} = 0.46$ with 95% CI 0.45—0.47, significantly different from the null value of $1/2$. The personalized AUC estimate is $\hat{\theta}_{11} = 0.50$ with a 95% CI 0.47—0.53.
- A test of equality $H_0 : \theta_{12} = \theta_{11}$ against $\theta_{12} < \theta_{11}$ returns a p-value of .05%. The Boston data is similar.


1.0
0.8
0.6
0.4
0.2
0.0
θ_{12}
0.0
0.2
0.4
0.6
0.8
1.0
θ_{11}

$\theta_{11} < \theta_{12}$

- Look at duration of stops between non-White ("con- trol") or White ("case") suspect status
- For the Boston data, the personalized AUC, 0.46 [0.40, 0.53], is more informative than the population AUC, 0.52 [0.48, 0.55],
- the test of equality versus $\theta_{11} < \theta_{12}$ returning a p-value of 2.5%.

No significant difference between θ_{12} and θ_{11}.

- duration of the stop between non-Hispanic ("control") and Hispanic ("case") suspects: For both the NYC and Boston data, neither the population AUC nor personalized AUC is significantly different from the null value $1/2$, and the test of equality of the two AUCs fails to reject.
- For the Boston data, whether one takes the case status to be non-Hispanic Black or non-Hispanic White, the two AUCs are statistically indistinguishable from each other and each is indistinguishable from the null value $1/2$.

Haben Michael, Lu Tian. The Population and Personalized AUCs. Forthcoming. Available at:
`https://haben-michael.github.io/`