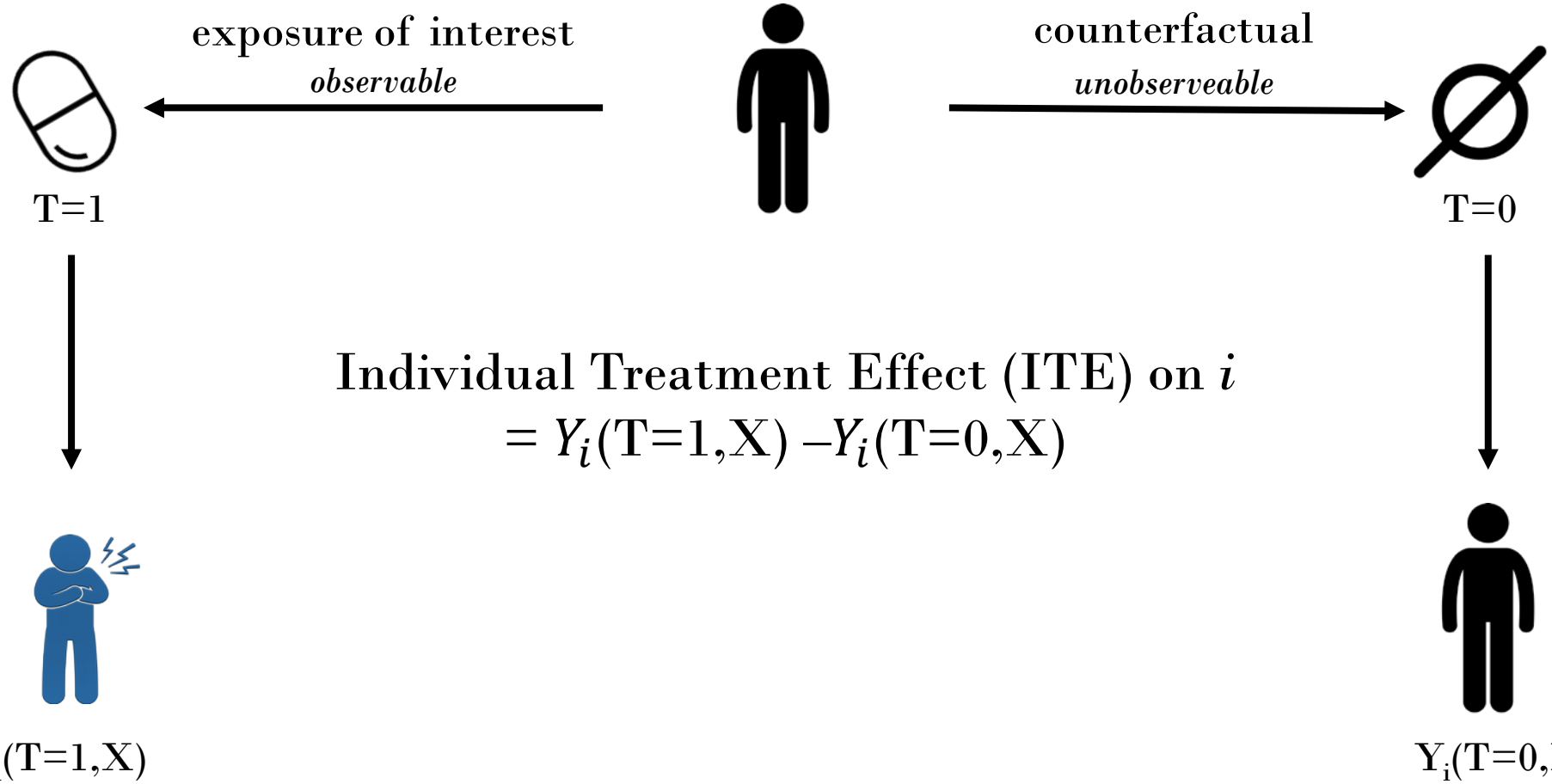
The Counterfactual χ-GAN

Amelia J Averitt, MPH MA MPhil

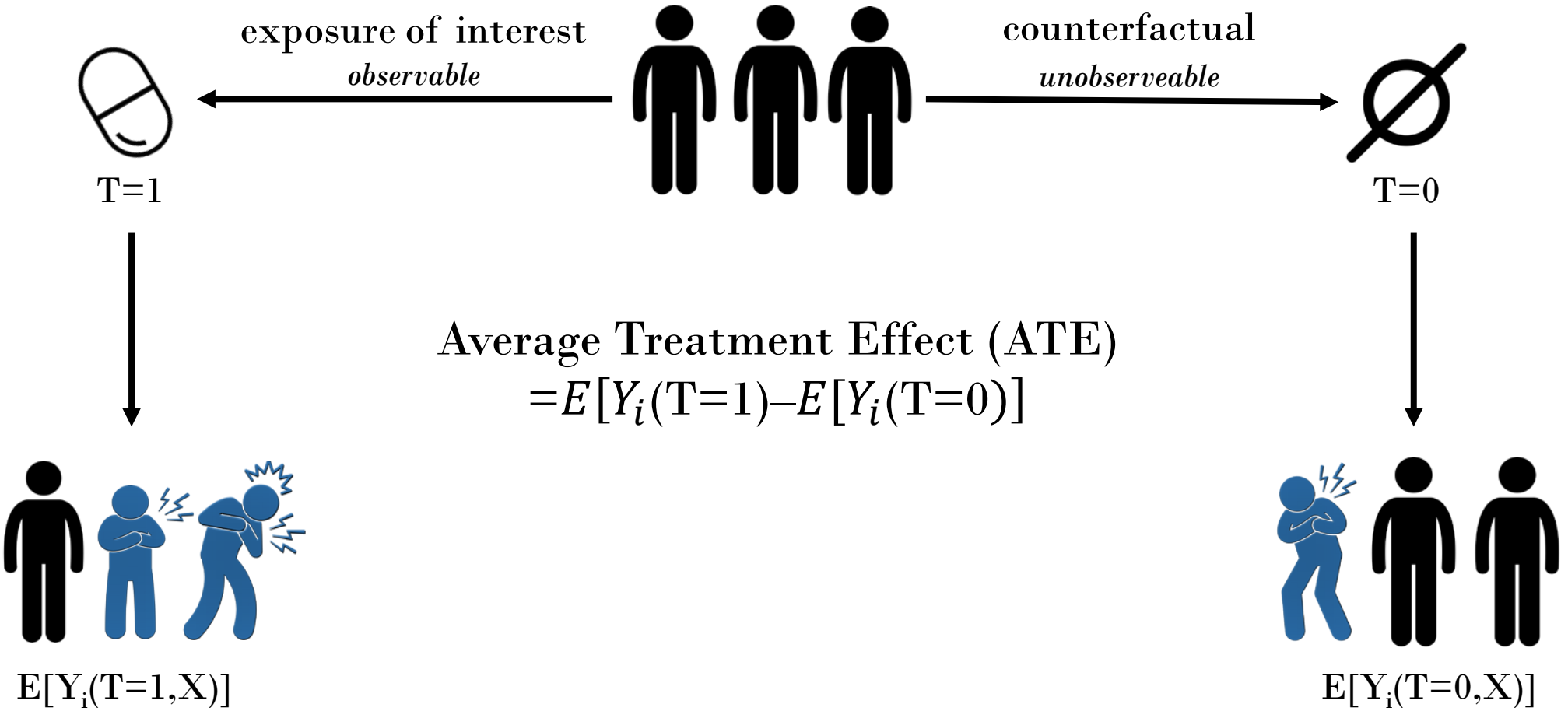
COLUMBIA | COLUMBIA UNIVERSITY DEPARTMENT OF BIOMEDICAL INFORMATICS

"Causal inference refers to the process of drawing a conclusion about cause and effect relationships" Vogt 2011

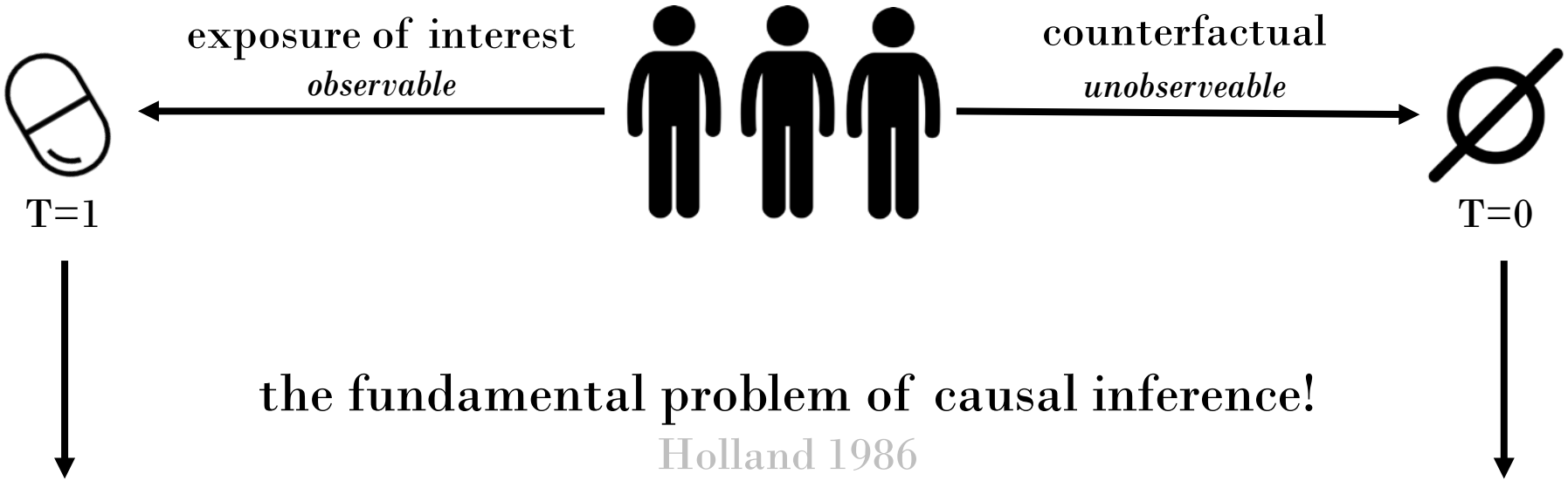
Titiunik 2007; Höfler 2005



Titiunik 2007; Höfler 2005



Titiunik 2007; Höfler 2005



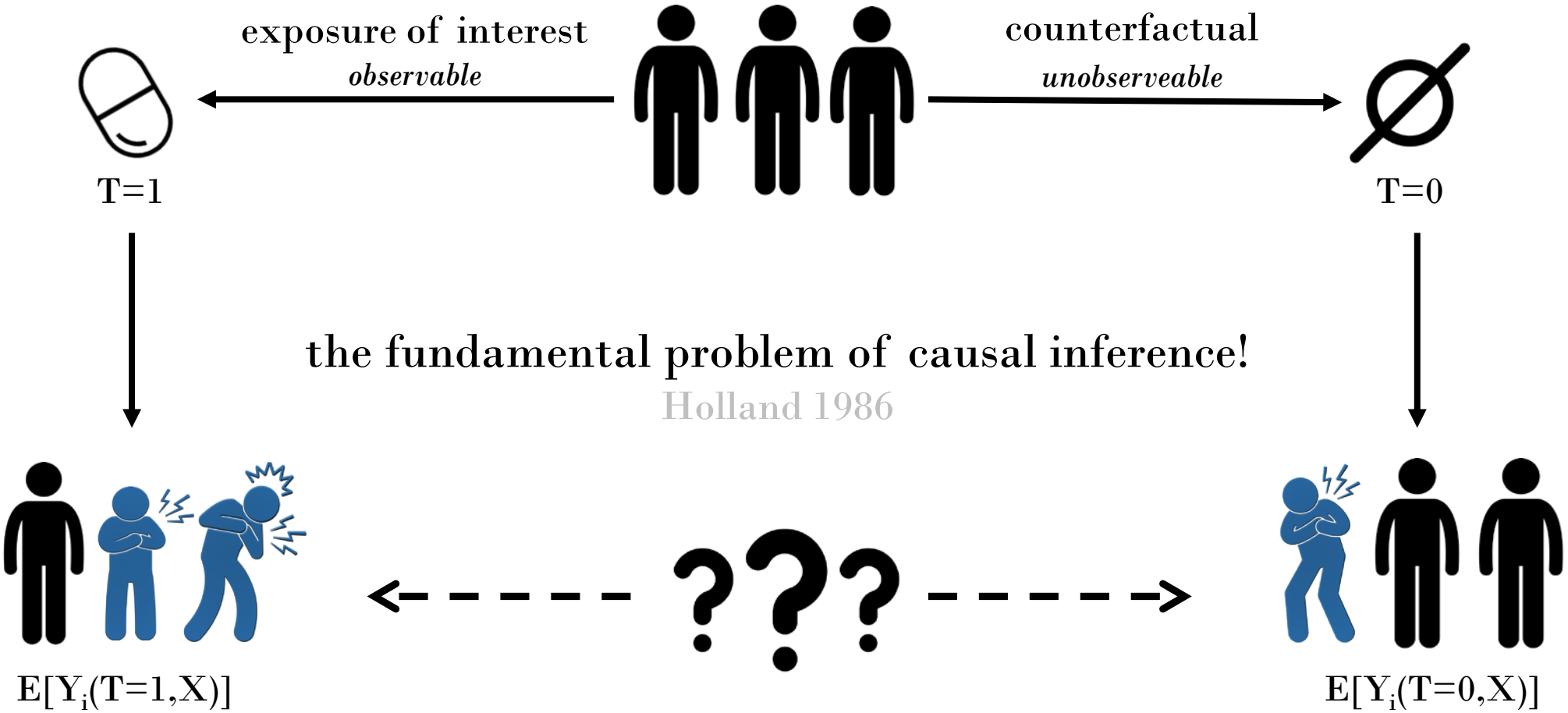
the fundamental problem of causal inference!

Holland 1986

$E[Y_i(T=1,X)]$

$E[Y_i(T=0,X)]$

Titiunik 2007; Höfler 2005



the Rubin causal model Rubin 1974

In an cohort of treatment and comparator units, $Y_i(T=1)$ and $Y_i(T=0)$ are potential outcomes in that either of these two outcomes can be *potentially observed*.

Contrast the mean reported outcome in each arm

$$\hat{E}[Y|T=1] = \frac{1}{n}\sum_{i=1}^{n} Y_i(T=1, X=X_i)$$

$$\hat{E}[Y|T=0] = \frac{1}{n}\sum_{i=1}^{n} Y_i(T=0, X=X_i)$$

the Rubin causal model Rubin 1974

Approximate the counterfactual treatment effect when two assumptions are met.

the Rubin causal model Rubin 1974

Approximate the counterfactual treatment effect when two assumptions are met.

1. Stable Unit Treatment Value Assumption (SUTVA) Cox 1958; Rubin 1986
   - the [potential outcome of] one unit should be unaffected by the particular assignment of treatments to the other units. Laffers 2016

the Rubin causal model Rubin 1974

Approximate the counterfactual treatment effect when two assumptions are met.

1. Stable Unit Treatment Value Assumption (SUTVA) Cox 1958; Rubin 1986
   - the [potential outcome of] one unit should be unaffected by the particular assignment of treatments to the other units. Laffers 2016

2. Strong Ignorability/ Exchangeability Rosenbaum 1983a
   - $\big(Y_i(1), Y_i(0)\big) \perp T_i \mid X_i = x \quad \forall\, x$

violations of these assumptions
bias causal estimates

enforcing or correcting these assumptions
improves causal estimates

how do we uphold strong ignorability?

how do we uphold strong ignorability?

by enforcing feature balance

Imbens 2009; Morgan 2014; Ho 2007

$\tilde{F}(X|T=1) = \tilde{F}(X|T=0)$ where $\tilde{F}(\cdot)$ is the empirical distribution

how do we uphold strong ignorability?

by enforcing feature balance

Imbens 2009; Morgan 2014; Ho 2007

$\tilde{F}(X|T=1) = \tilde{F}(X|T=0)$ where $\tilde{F}(\cdot)$ is the empirical distribution

experimental data from an RCT!

how do we uphold strong ignorability?

by enforcing feature balance

Imbens 2009; Morgan 2014; Ho 2007

$\tilde{F}(X|T=1) = \tilde{F}(X|T=0)$ where $\tilde{F}(\cdot)$ is the empirical distribution

experimental data from an RCT!

theoretically guaranteed by randomization

Fisher 1926; Fisher 1935

how do we uphold strong ignorability?

by enforcing feature balance

Imbens 2009; Morgan 2014; Ho 2007

$\tilde{F}(X|T=1) = \tilde{F}(X|T=0)$ where $\tilde{F}(\cdot)$ is the empirical distribution

experimental data from an RCT!

expensive, unethical,
not representative,
poor generalizability,
& narrow scope

World Medical Association 1997; Rothman 2000;
DiMasi 2014; Gabler 2009; Longford 1999;
Kravitz 2004; Lachin 1988; Kernan 1999

or we could use observational data

data that is passively collected without any engineering adjustments Czitrom 1997

- electronic health record (EHR) data Murdoch 2013

or we could use observational data

data that is passively collected without any engineering adjustments Czitrom 1997

- electronic health record (EHR) data Murdoch 2013

suitable for studying rare outcomes

there's a lot of it Imai 2009

representative Concato 2004; Thadani 2006; Kleinberg 2011

inferences are more externally valid Steckler 2008; Rothwell 2006

observational data is non-randomized and requires manipulations to enforce strong ignorability

manipulations to enforce strong ignorability with observational data

1. matching
2. adjustment/ stratification
3. weighting

manipulations to enforce strong ignorability with observational data

1 matching

2 adjustment/ stratification

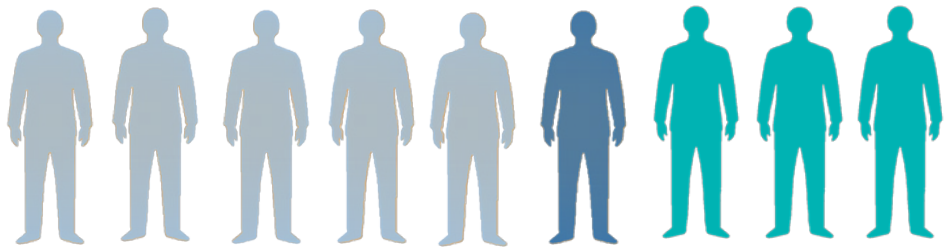
3 weighting

what is weighting?

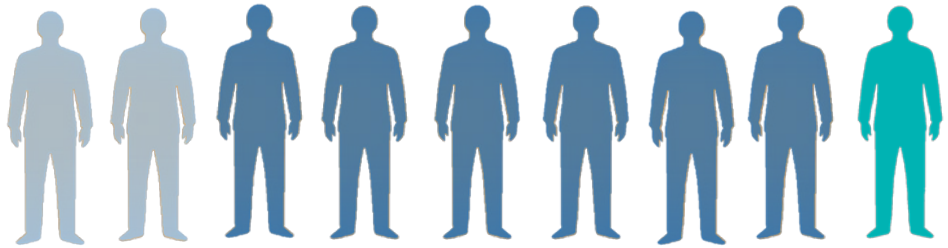
when the sample is not representative of the population, we can disproportionally consider units to make the sample *look* more like the population.

weighting and the counterfactual

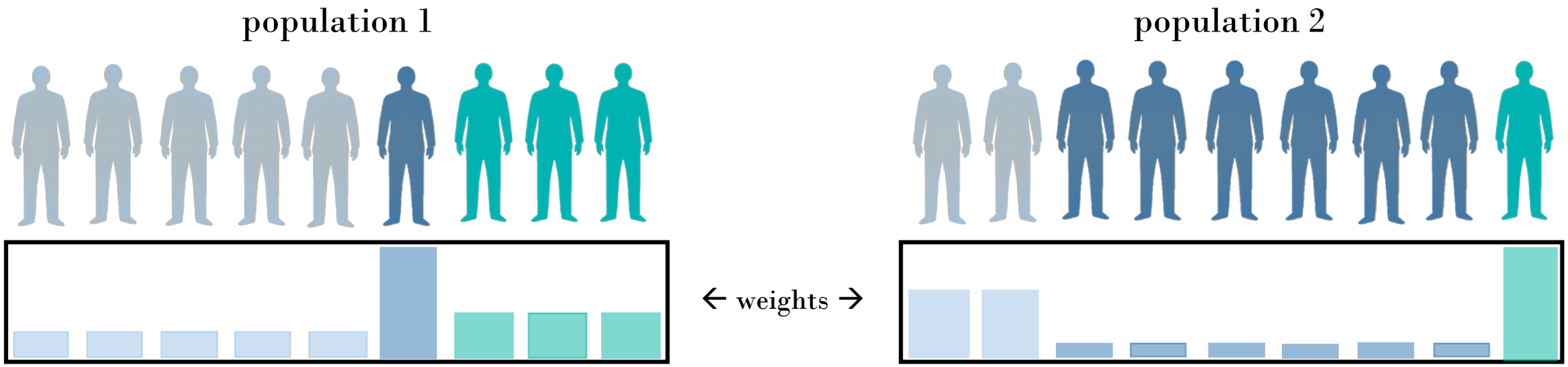
population 1



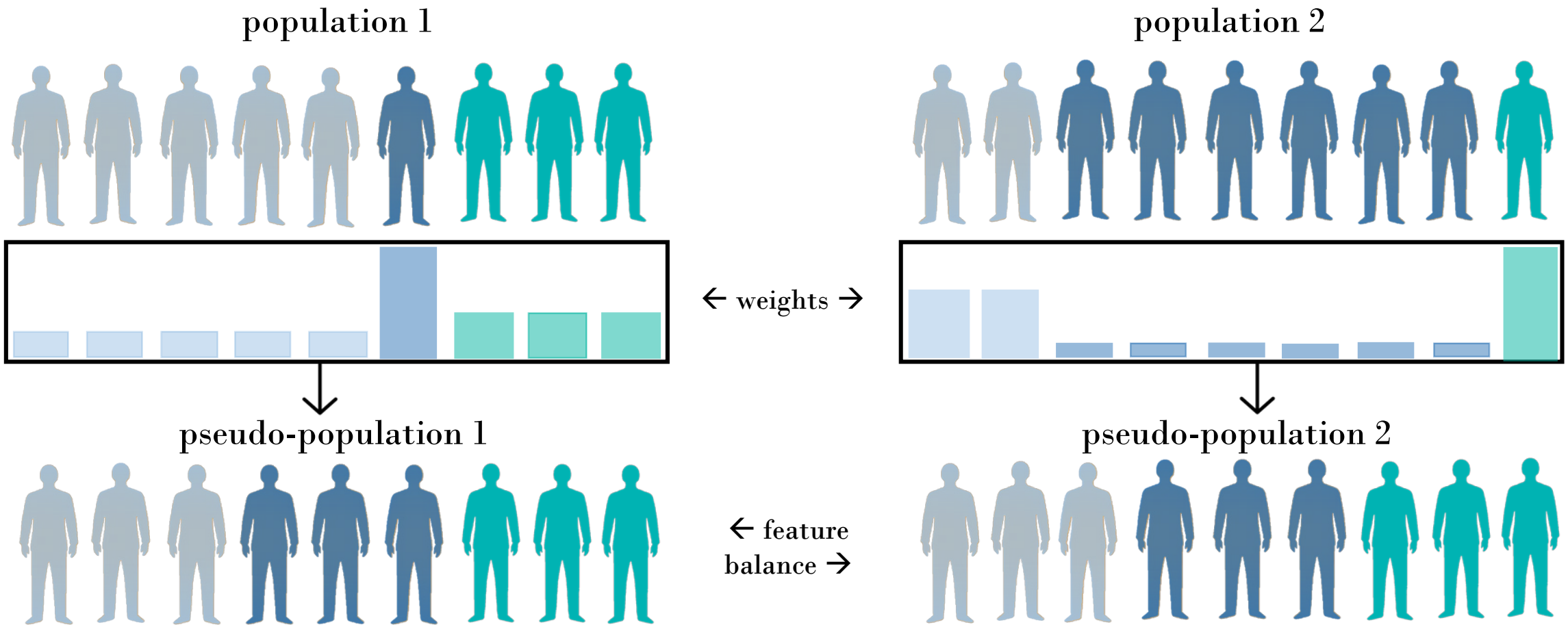
population 2



weighting and the counterfactual

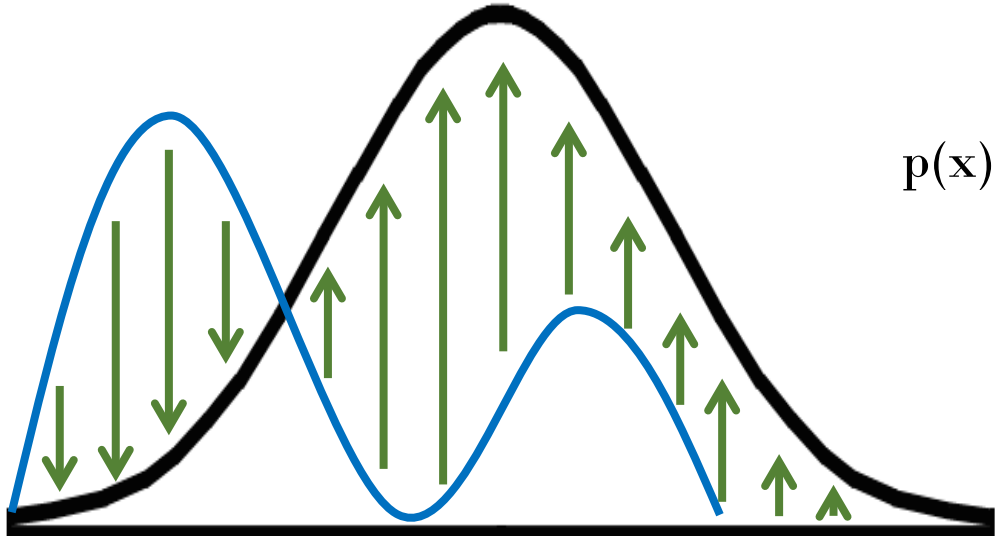


weighting and the counterfactual



this is related to importance sampling

Importance sampling draws samples from a *proposal distribution* and re-weights the distribution using importance weights so that the weighted distribution represents your target distribution.
Hammersley 1966



p(x) = target distribution = counterfactual distribution
q(x) = proposal distribution = real-world data
importance weights = feature balancing weights

a common method of weighting

a common method of weighting

Inverse Probability of Treatment Weighting (IPW). Units are weighted according to inverse of their probability of being assigned to the treatment conditional on their measured, baseline features, a metric often called the propensity score. Rubin 2015; Rosenblatt 1965; Rosenbaum 1984; Austin 2011

$$w = \frac{T}{P(T = 1|X)} + \frac{1 - T}{P(T = 1|X)}$$

a common method of weighting

Inverse Probability of Treatment Weighting (IPW). Units are weighted according to inverse of their probability of being assigned to the treatment conditional on their measured, baseline features, a metric often called the propensity score. Rubin 2015; Rosenblatt 1965; Rosenbaum 1984; Austin 2011

$$w = \frac{T}{P(T = 1|X)} + \frac{1 - T}{P(T = 1|X)}$$

- flexible and robust causal modeling under selection on observables Imai 2013

- model dependent!
- unstable weights/feature imbalance/bias if propensity scores very close to 0 or 1 King 2016

how can we learn balancing weights for causal inference such that the weights are more stable and model agnostic?

how can we learn balancing weights for causal inference such that the weights are more stable and model agnostic?

use a Generative Adversarial Network (GAN)!

why GANs?

implicit generative models only specify a stochastic procedure with which to generate data Mohamed 2017

- full distributional matching on feature
- less prone to instability originating from model specification

why GANs?

implicit generative models only specify a stochastic procedure with which to generate data Mohamed 2017

- full distributional matching on feature
- less prone to instability originating from model specification

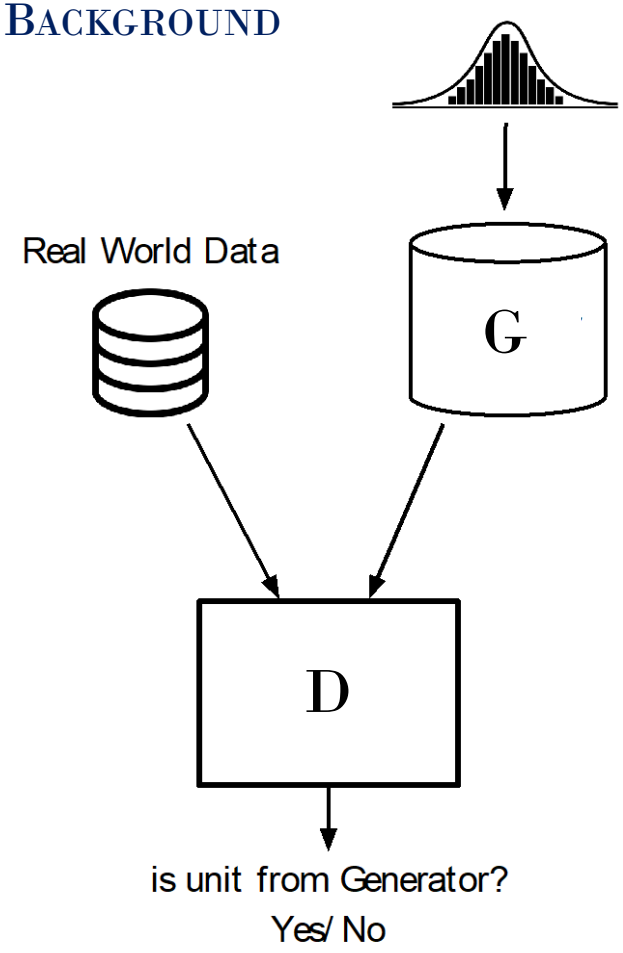
vs

prescribed models provide an explicit parametric specification for a distribution Diggle 1984

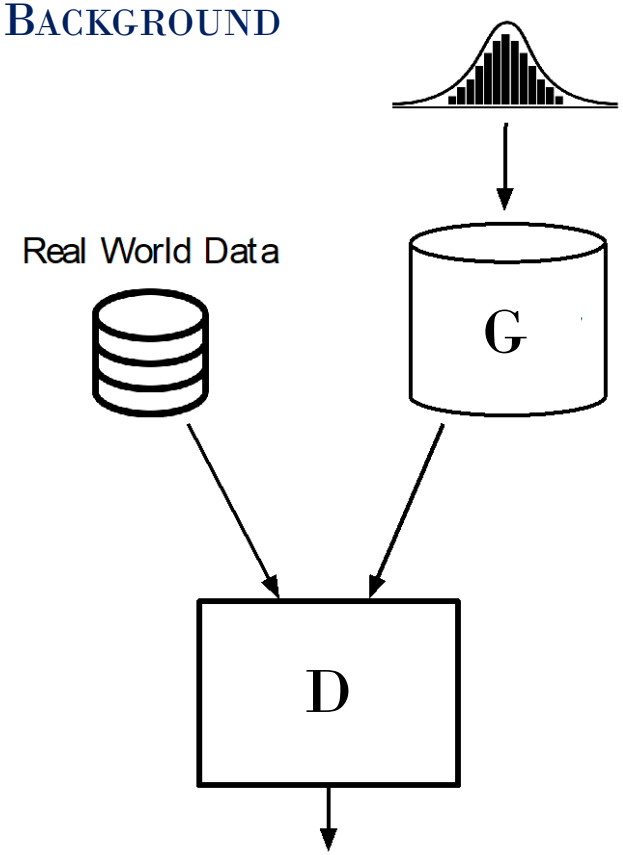
- often used to model propensity scores, etc that are used in weighting
- model dependence

vanilla GAN

vanilla GAN



vanilla GAN



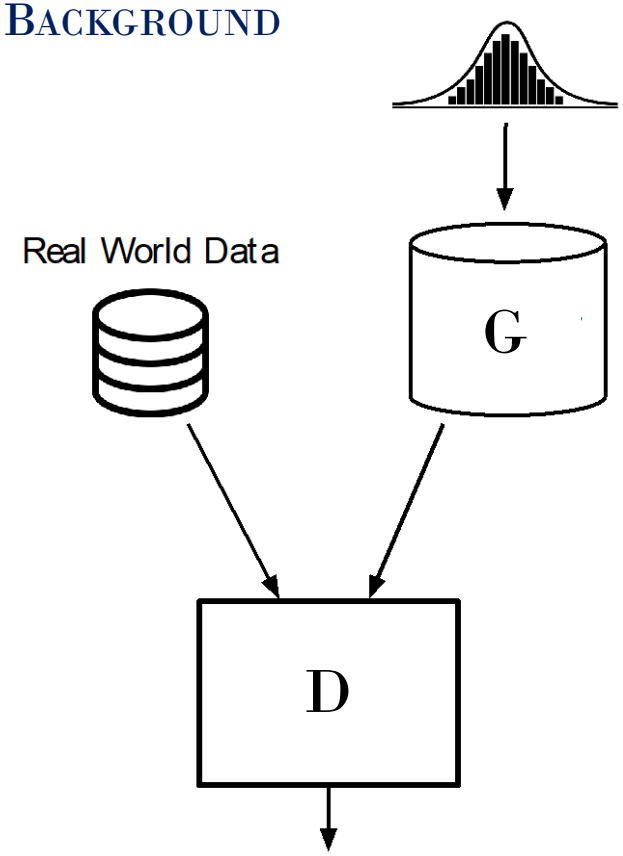
log probability of D predicting that "real" data is genuine

log probability of D predicting that "fake" data is not genuine

$$min_G max_D V(D,G) = E_{x \sim p_{data}}(x)[\log D(x)] + E_{z \sim p_z}(z)[\log(1 - D(G(z)))]$$

vanilla GAN

the optimal solution to this expression minimizes the Jensen-Shannon divergence.

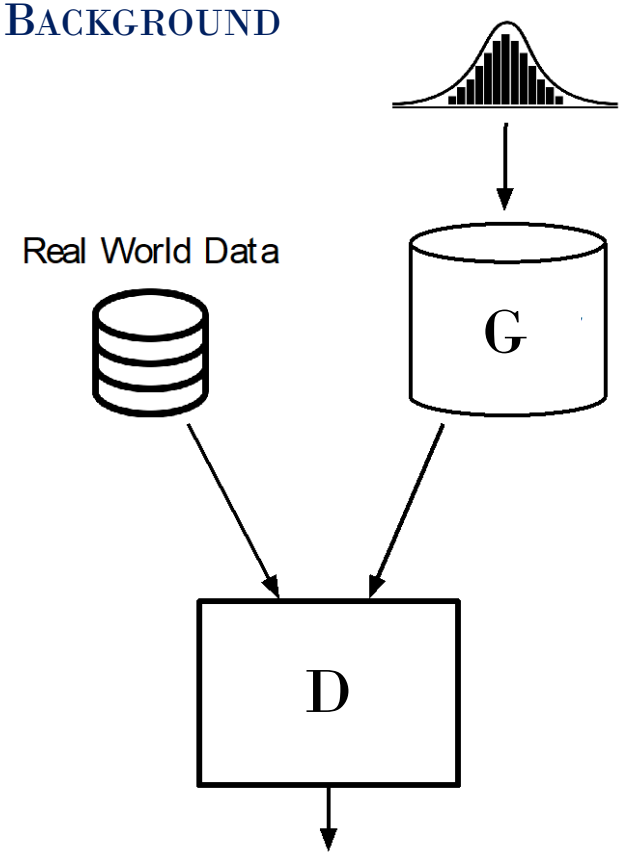


log probability of D predicting that "real" data is genuine

log probability of D predicting that "fake" data is not genuine

$$min_G max_D V(D,G) = E_{x \sim p_{data}}(x)[\log D(x)] + E_{z \sim p_z}(z)[\log(1 - D(G(z)))]$$

vanilla GAN



the optimal solution to this expression minimizes the Jensen-Shannon divergence.

this won't suffice for causal inference. we need something new!

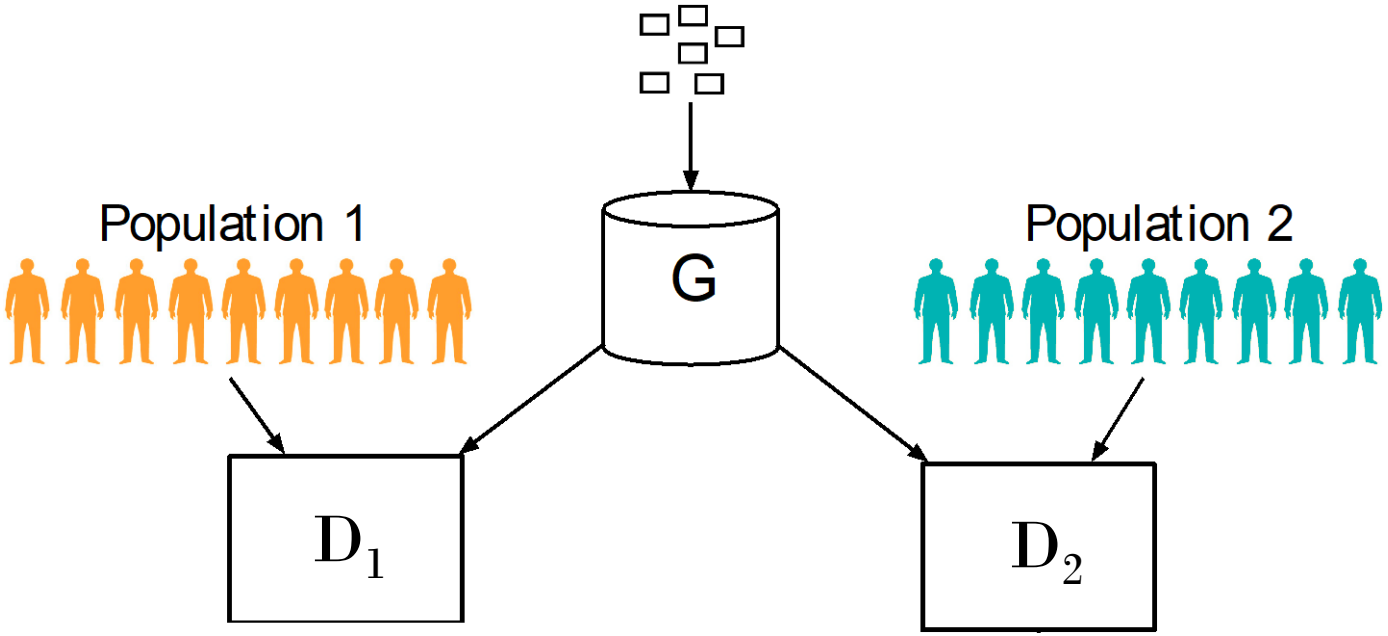
log probability of D predicting that "real" data is genuine

log probability of D predicting that "fake" data is not genuine

$$min_G max_D V(D,G) = E_{x \sim p_{data}}(x)[\log D(x)] + E_{z \sim p_z}(z)[\log(1 - D(G(z)))]$$

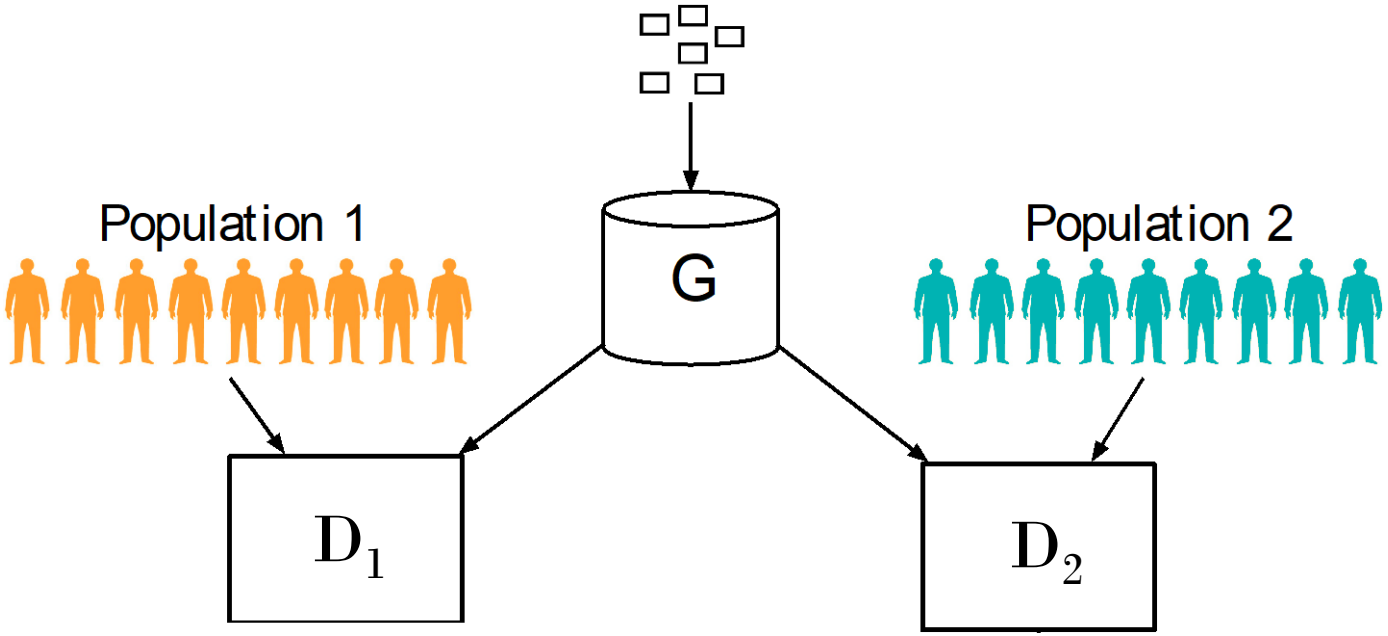
the Counterfactual χ-GAN

The Model



the Counterfactual χ-GAN

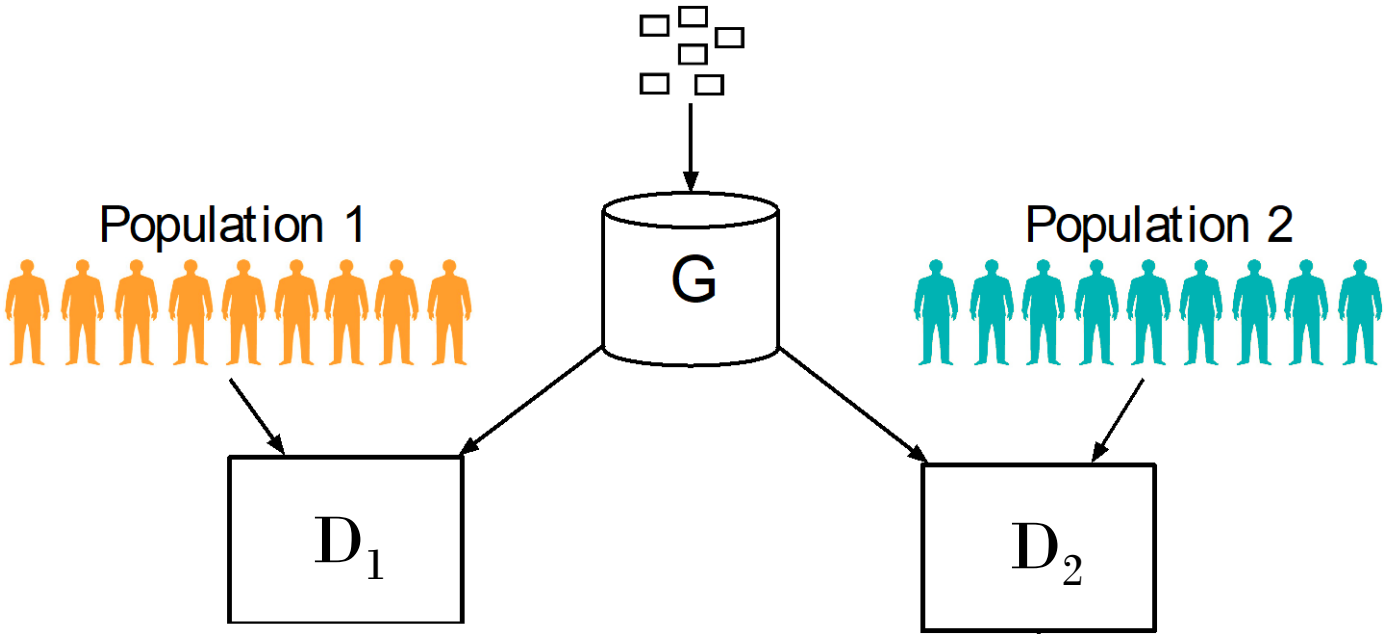
difference #1:
two GANs joined
at the generator



the Counterfactual χ-GAN

difference #1:
two GANs joined
at the generator

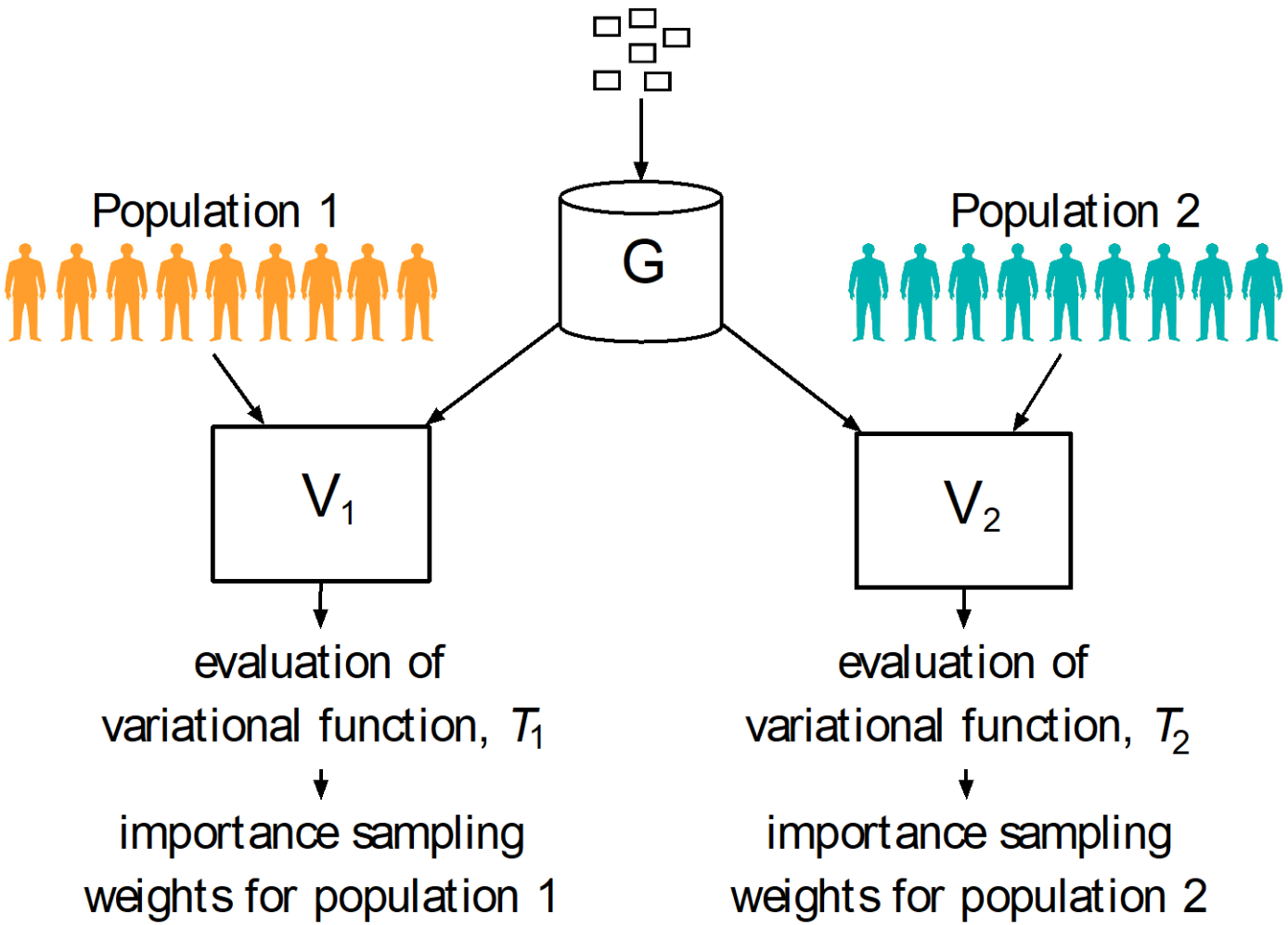
difference #2:
minimize the χ
divergence



the Counterfactual χ-GAN

difference #1:
two GANs joined
at the generator

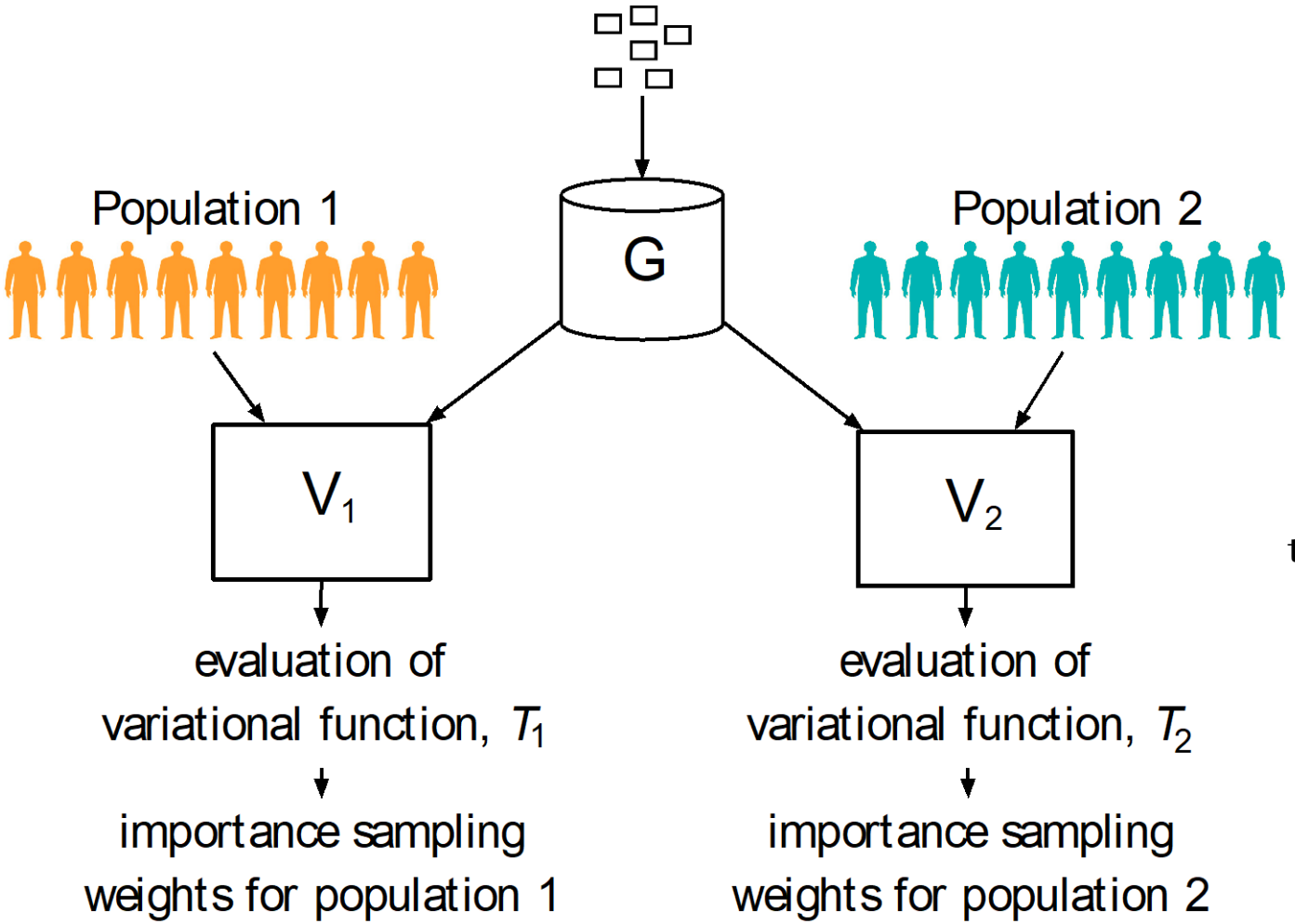
difference #2:
minimize the χ
divergence
Nowozin 2016



the Counterfactual χ-GAN

difference #1:
two GANs joined
at the generator

difference #2:
minimize the χ
divergence
Nowozin 2016



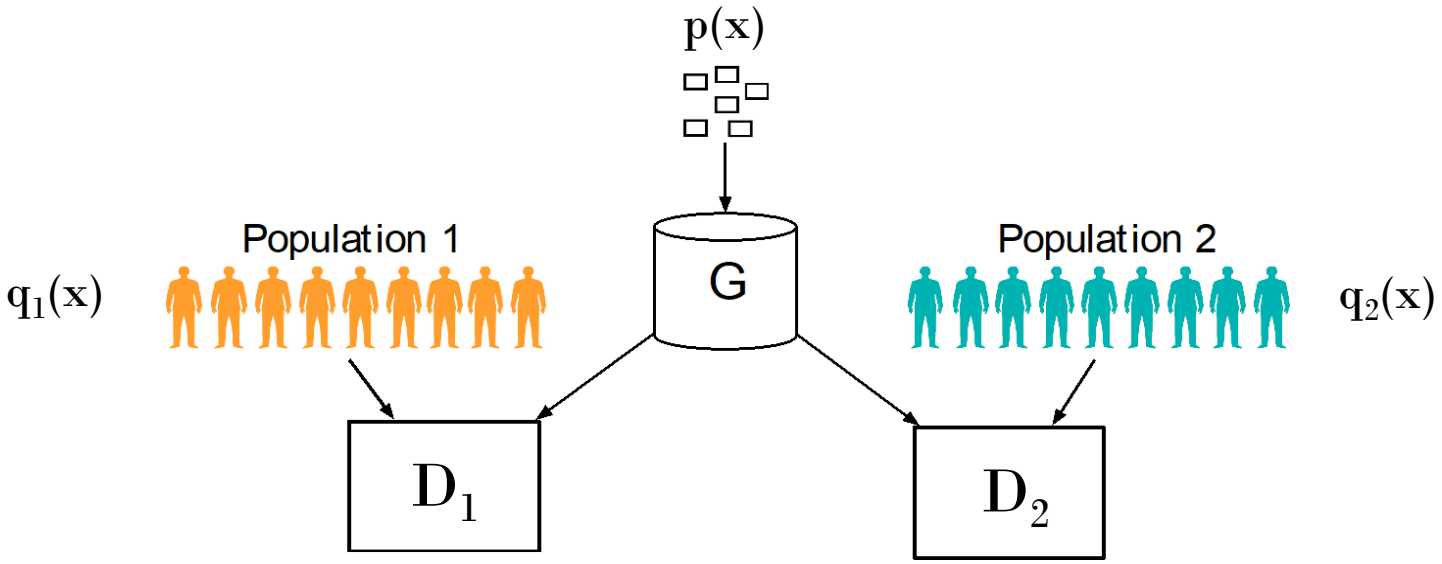
learns feature-
balancing weights
through an adversarial
training process.

importance sampling & the Counterfactual χ-GAN

Under importance sampling, those ranges with in which the ratio of $p(x)/q(x)$

- high → will have high importance weights, contribute more to expectations
- low → will have very small importance weights, contribute negligibly to expectations

the target distribution, $p(x)$ functions as the generator, which embodies the *overlapping* portions of the empirical distributions, $q(x)$ of the treatment arms.



importance sampling & the Counterfactual χ-GAN

$$\chi = \int q(x)\left[\frac{p(x)^2}{q(x)^2} - 1\right]dx$$

Dieng 2016

importance sampling & the Counterfactual χ-GAN

$$\chi = \int q(x)\left[\frac{p(x)^2}{q(x)^2} - 1\right]dx$$

Dieng 2016

$$\sigma_q^s = \frac{\mu_q^2}{n}\left[\int q(x)\left(\frac{p(x)}{q(x)}\right)^2 - 1\right]dx$$

importance sampling & the Counterfactual χ-GAN

$$\chi = \int q(x)\left[\frac{p(x)^2}{q(x)^2} - 1\right]dx$$

Dieng 2016

$$\sigma_q^s = \frac{\mu_q^2}{n}\left[\int q(x)\left(\frac{p(x)}{q(x)}\right)^2 - 1\right]dx$$

$$cGAN = \arg\min_{p} \sum_{t=1}^{T} \chi\left(p_G(x) \,\|\, p_{data_t}(x_t)\right)$$

change the objective function to minimize this divergence.

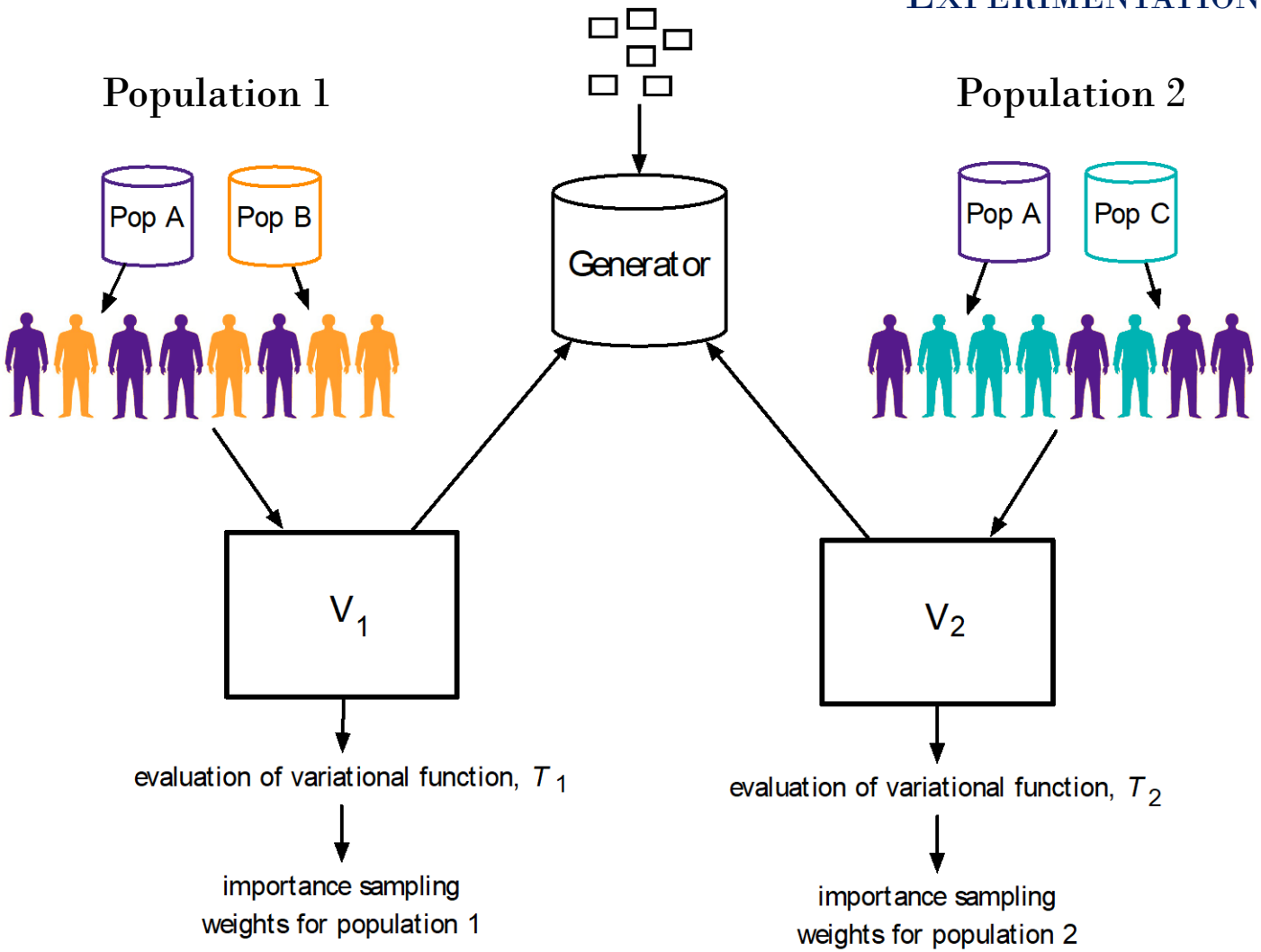
equivalent to minimizing the variance.

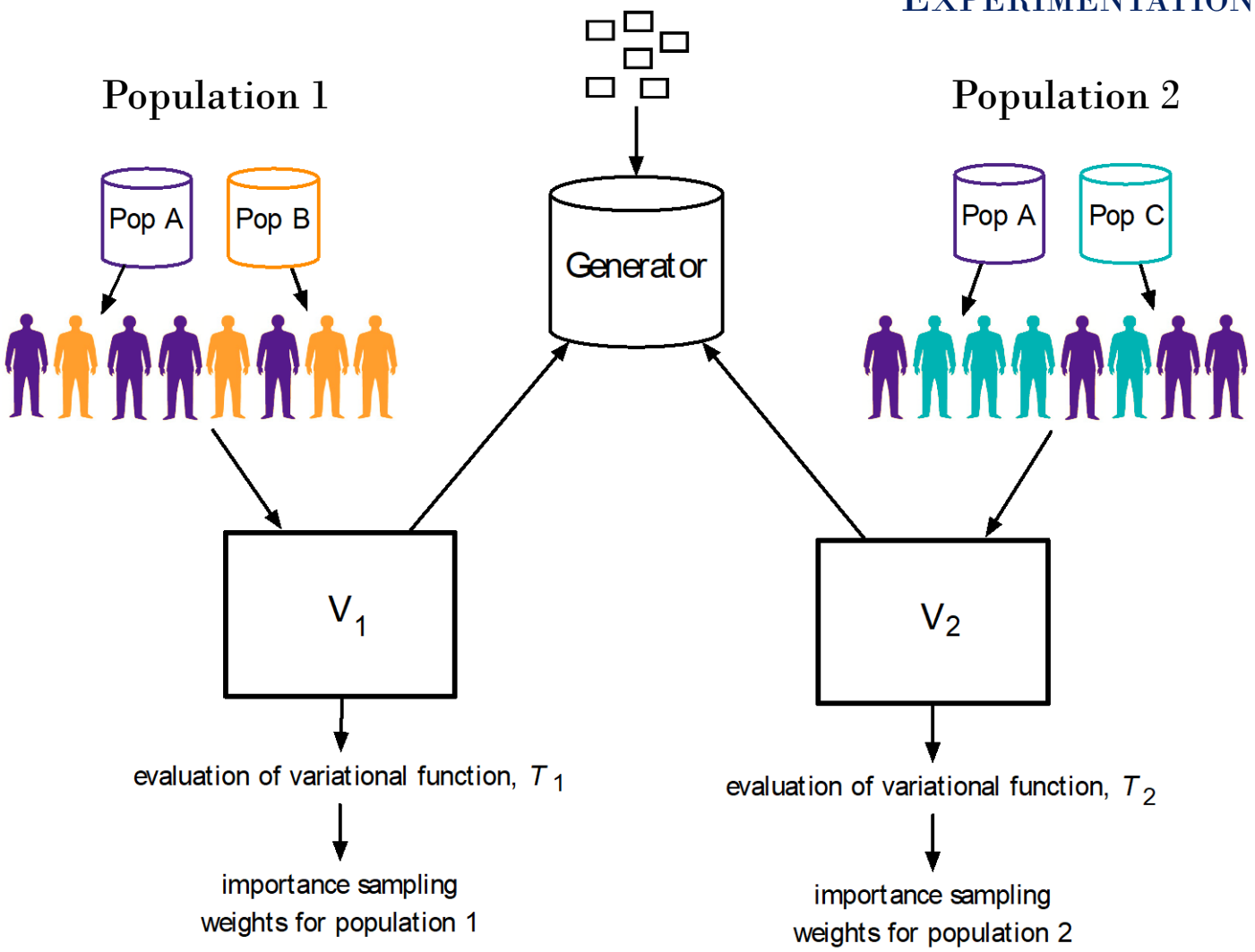
G = generator distribution

p_{data_t} = unweighted empirical treatment group, T

1. simulation
2. application to clinical data

1. simulation
2. application to clinical data


Population 1
Population 2
Pop A
Pop B
Generator
Pop A
Pop C
V_1
V_2
evaluation of variational function, T_1
evaluation of variational function, T_2
importance sampling
weights for population 1
importance sampling
weights for population 2



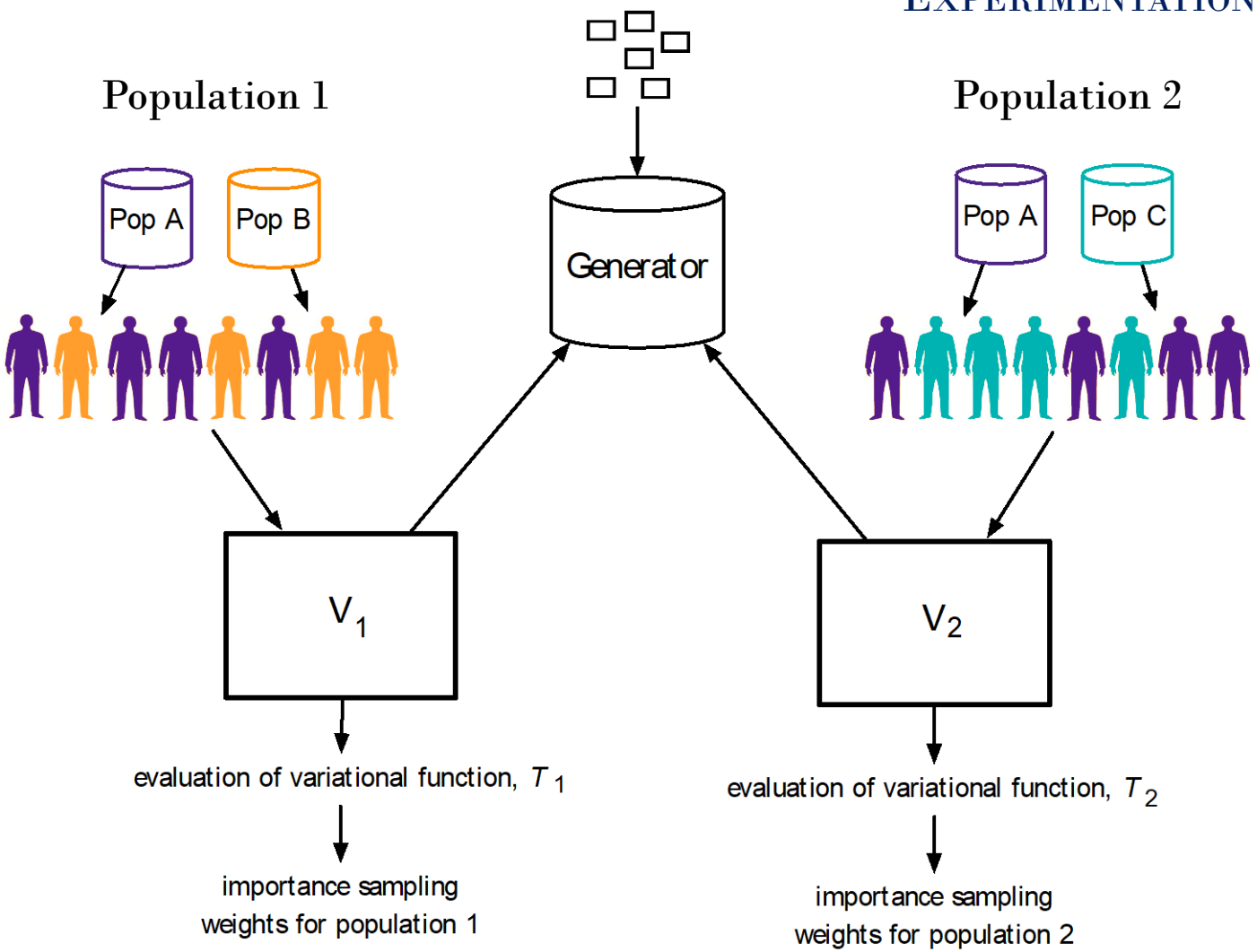
3 subpopulations A, B, C were drawn from randomly generated multivariate normal distributions.

- 5 discrete / 5 continuous features

2 populations (Pop1 & Pop2) are mixtures of subpopulations

- Pop 1 = subpop A & subpop B
- Pop 2 = subpop A & subpop C
- 4000 per arm, 2000 per subpopulation

Experimentation



3 subpopulations A, B, C were drawn from randomly generated multivariate normal distributions.

- 5 discrete / 5 continuous features

2 populations (Pop1 & Pop2) are mixtures of subpopulations

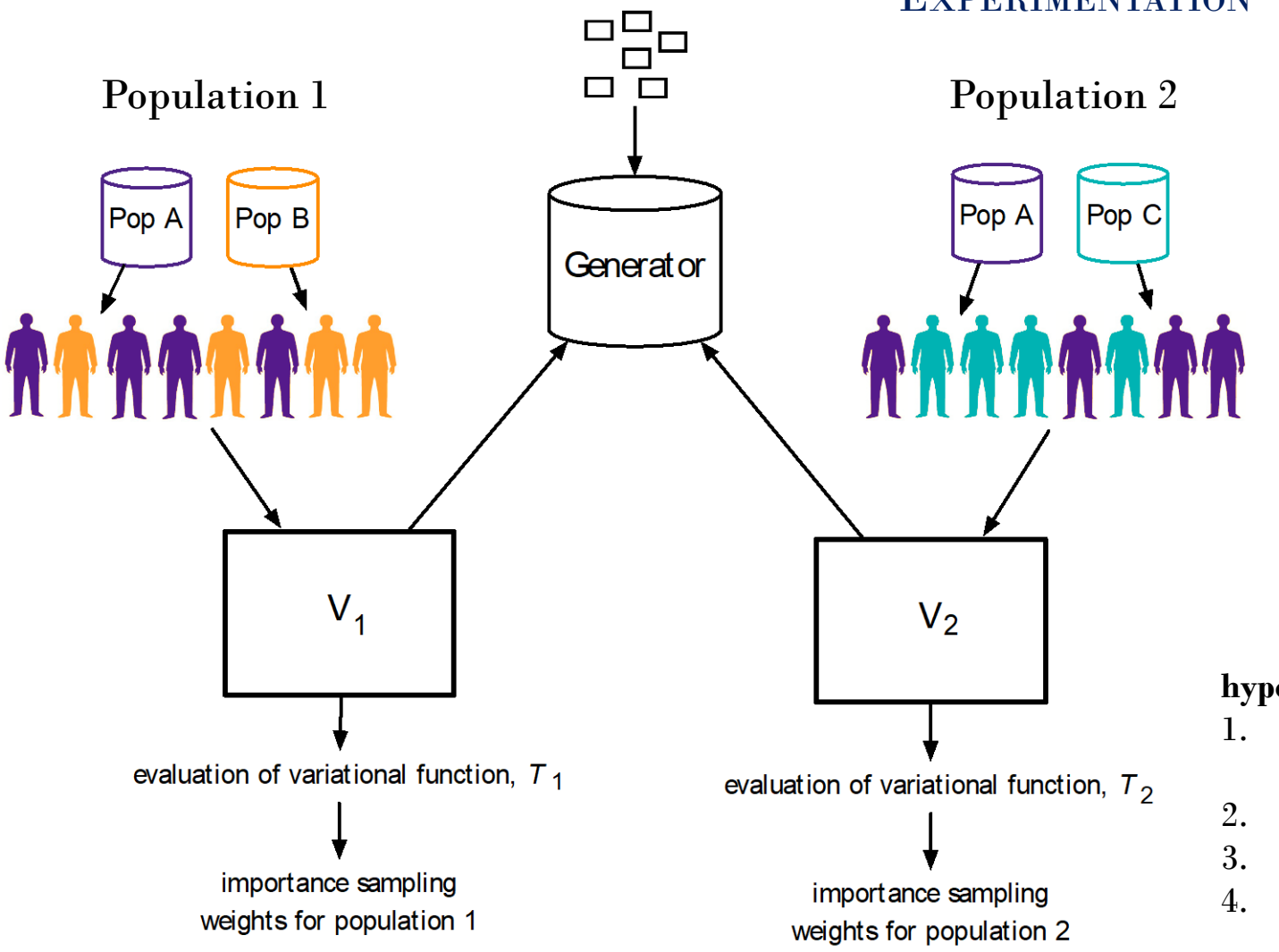
- Pop 1 = subpop A & subpop B
- Pop 2 = subpop A & subpop C
- 4000 per arm, 2000 per subpopulation

1 outcome. conditional on subpop

- Pop 1A ~ Gaussian (60, 1)
- Pop 1B ~ Gaussian (40, 1)
- Pop 2A ~ Gaussian (-10, 1)
- Pop 2C ~ Gaussian (10, 1)

$ATE_{mixture}$ = Pop1 – Pop2 = 50
$ATE_{overlap}$ = Pop 1A – Pop2A = 70

Experimentation



3 subpopulations A, B, C were drawn from randomly generated multivariate normal distributions.

- 5 discrete / 5 continuous features

2 populations (Pop1 & Pop2) are mixtures of subpopulations

- Pop 1 = subpop A & subpop B
- Pop 2 = subpop A & subpop C
- 4000 per arm, 2000 per subpopulation

1 outcome. conditional on subpop

- Pop 1A ~ Gaussian (60, 1)
- Pop 1B ~ Gaussian (40, 1)
- Pop 2A ~ Gaussian (-10, 1)
- Pop 2C ~ Gaussian (10, 1)

$ATE_{mixture}$ = Pop1 – Pop2 = 50

$ATE_{overlap}$ = Pop 1A – Pop2A = 70

hypotheses

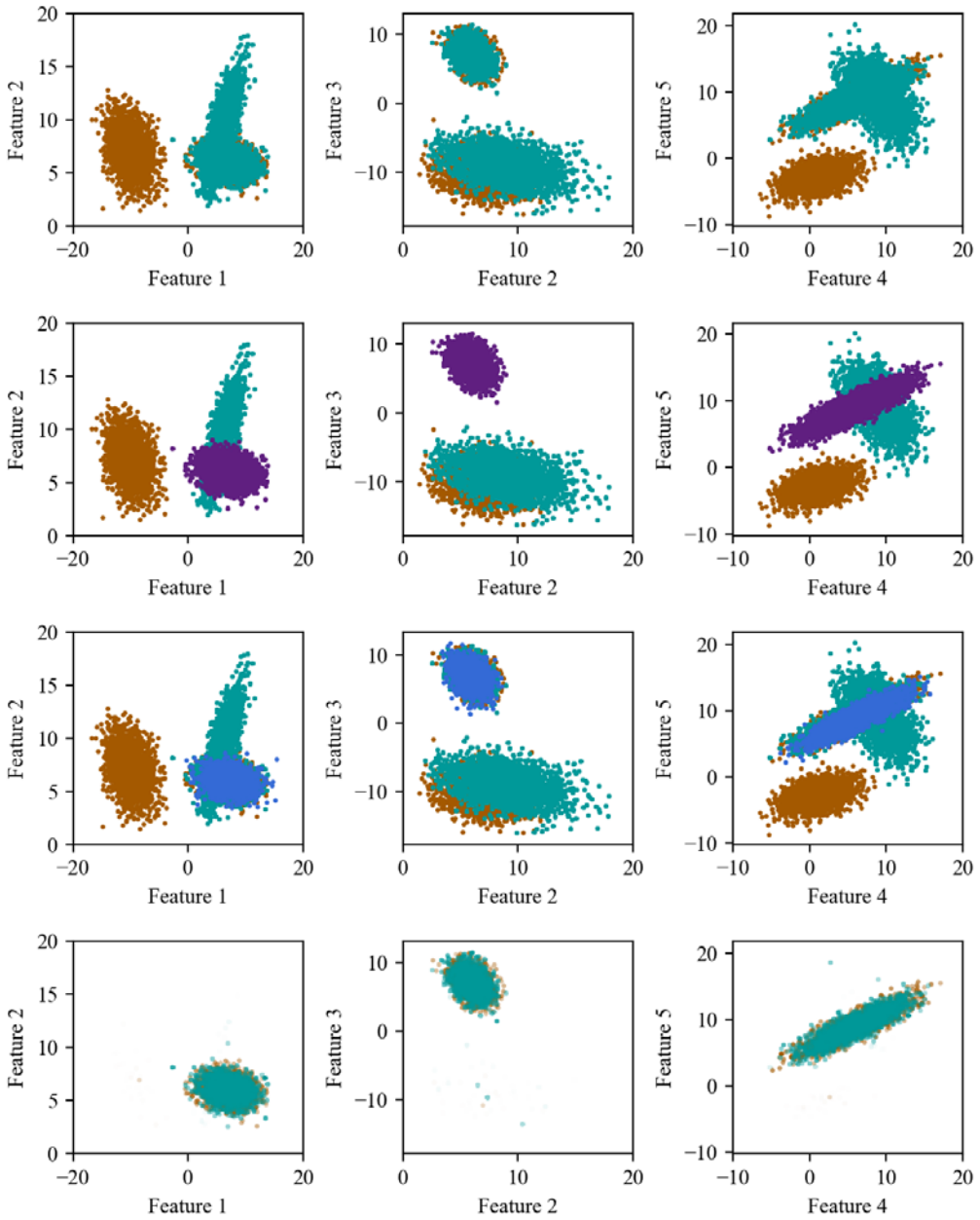
1. weights from overlapping population (Pop 1A/Pop 2A) will be high; weights from Pop 1B/Pop 2C will be low
2. weighting will make features more similar between Pops
3. cGAN-weighted ATE will be less biased than comparators
4. effective sample size of cGAN will be more reasonable than comparators

● Pop1 + Pop2

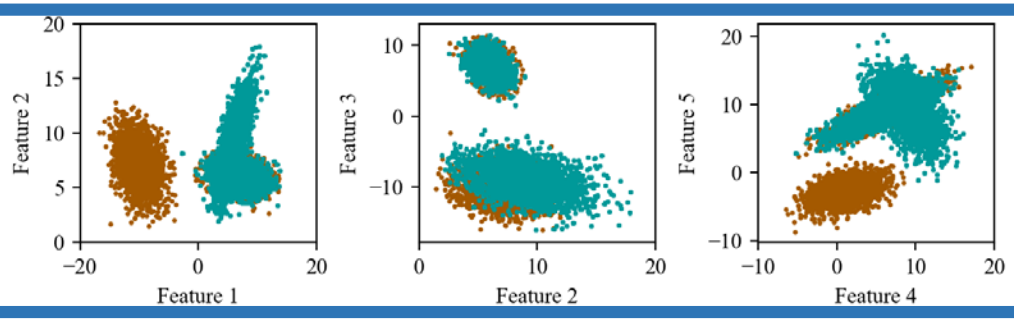
● Pop1A ● Pop1B + Pop2A + Pop2C

● Pop1 + Pop2 ● Gen

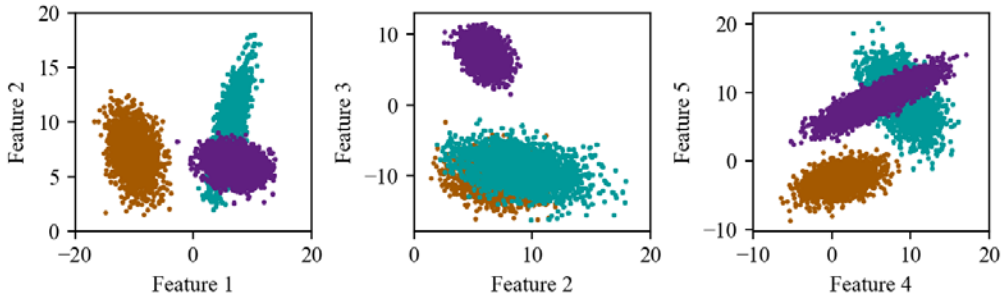
● Pop1 + Pop2



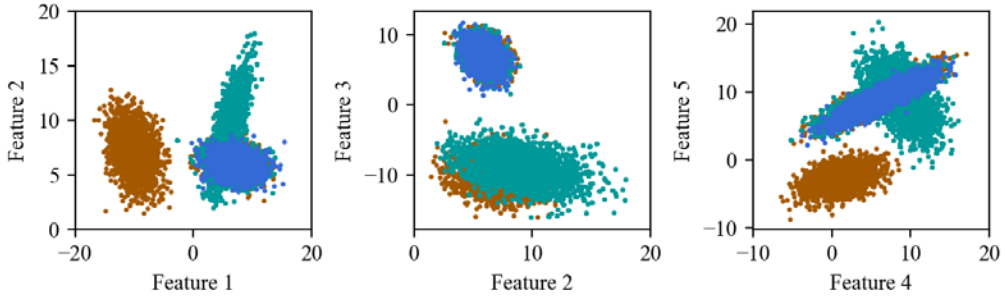
● Pop1 + Pop2



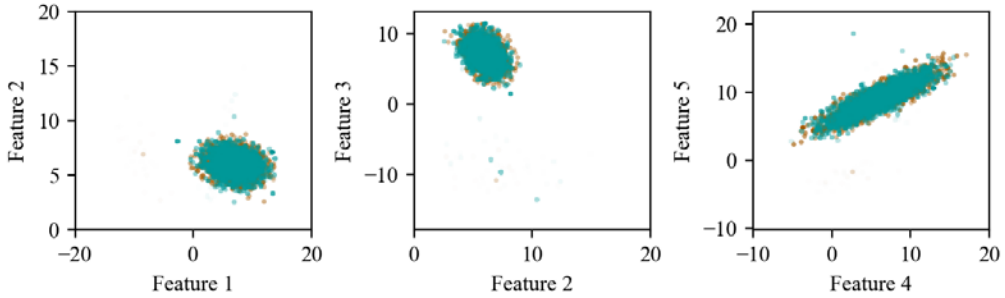
● Pop1A ● Pop1B + Pop2A + Pop2C



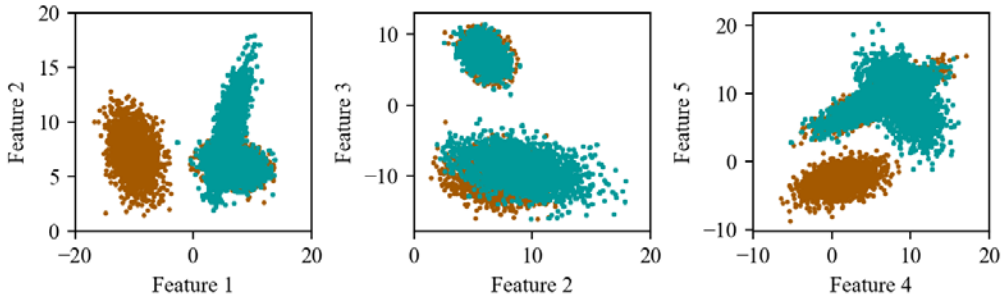
● Pop1 + Pop2 ● Gen



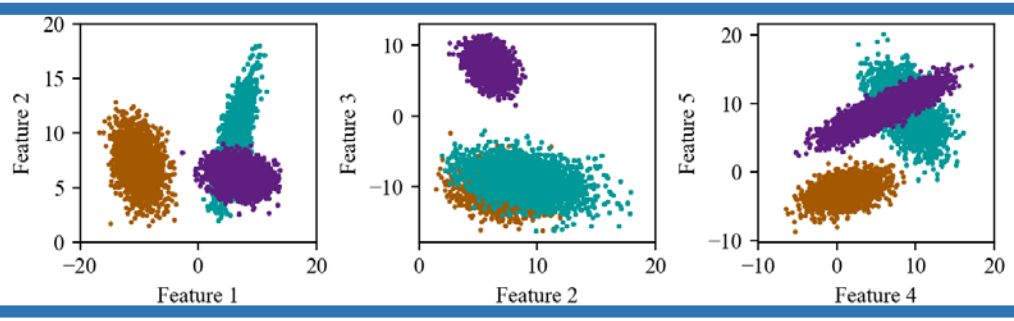
● Pop1 + Pop2



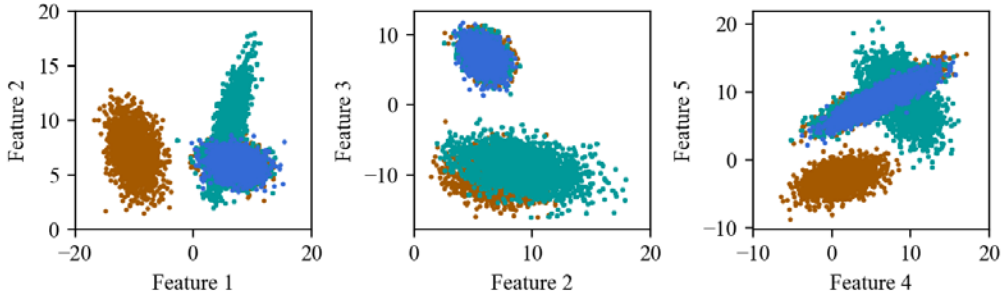
● Pop1 + Pop2



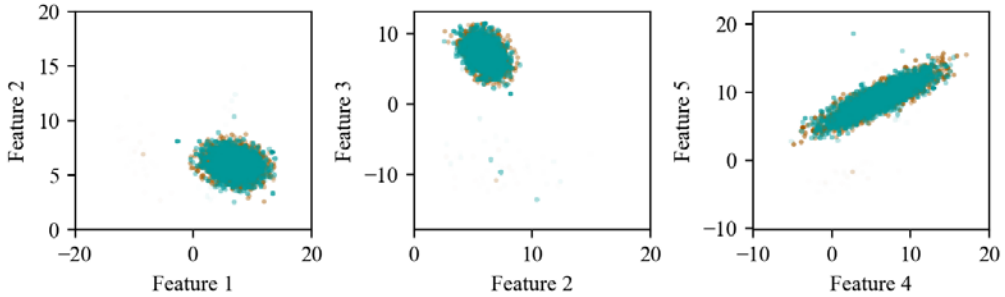
● Pop1A ● Pop1B + Pop2A + Pop2C



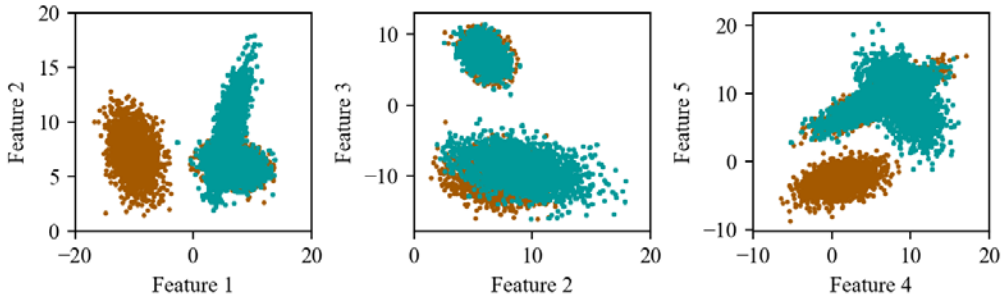
● Pop1 + Pop2 ● Gen



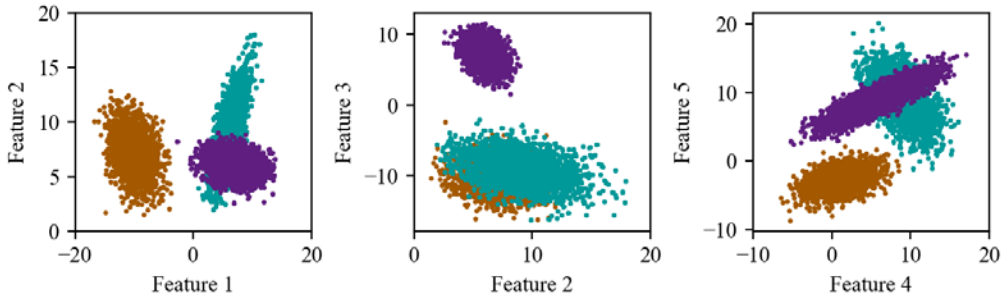
● Pop1 + Pop2



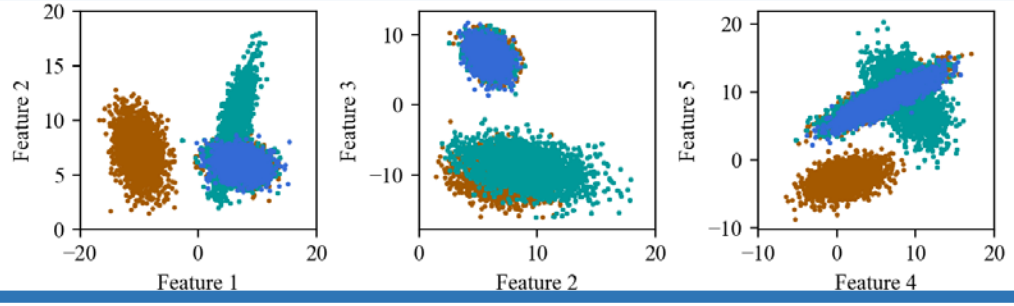
● Pop1 + Pop2



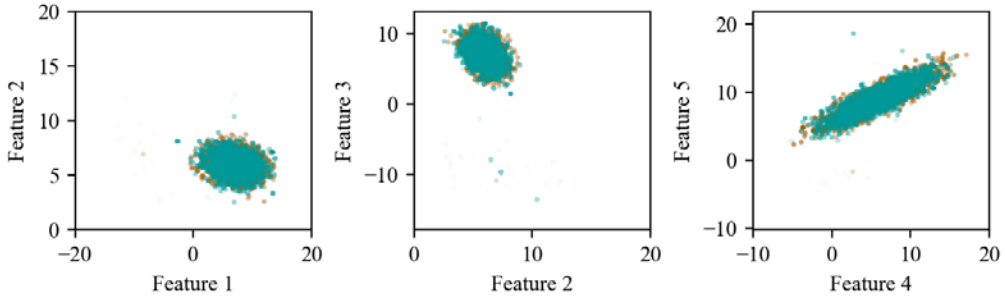
● Pop1A ● Pop1B + Pop2A + Pop2C



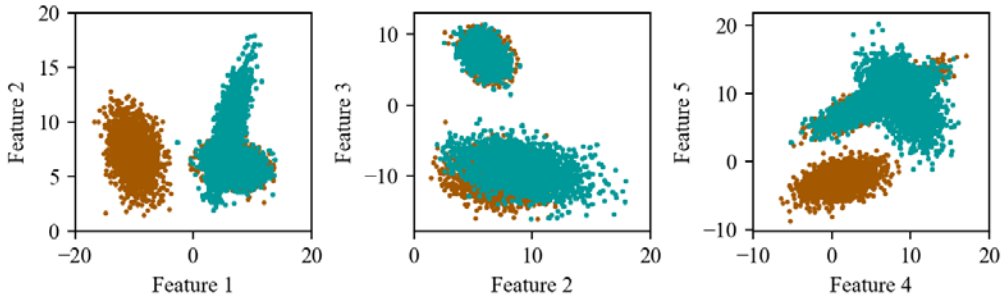
● Pop1 + Pop2 ● Gen



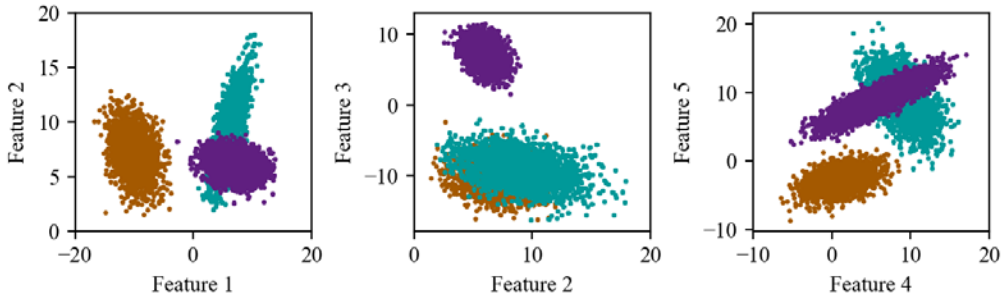
● Pop1 + Pop2



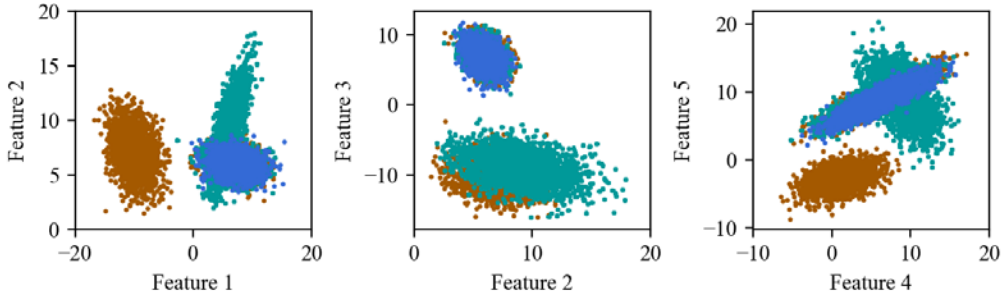
● Pop1 + Pop2



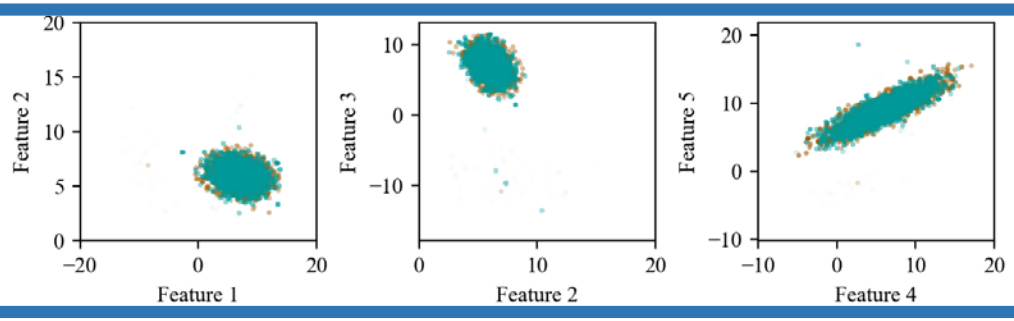
● Pop1A ● Pop1B + Pop2A + Pop2C



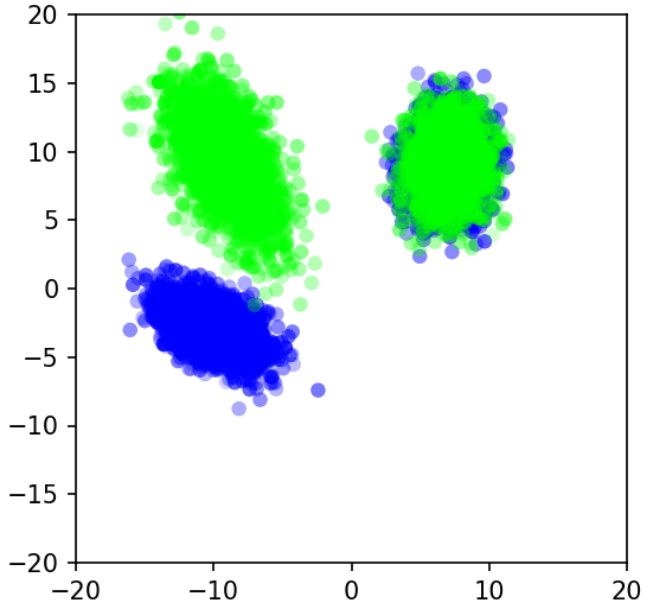
● Pop1 + Pop2 ● Gen

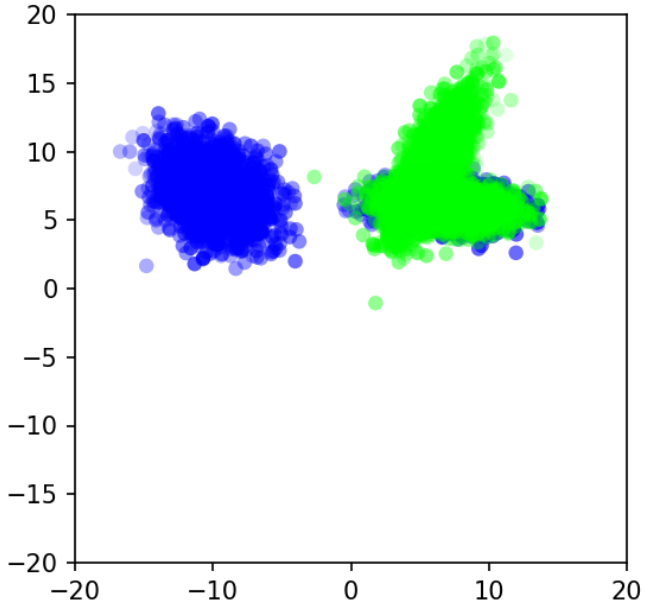


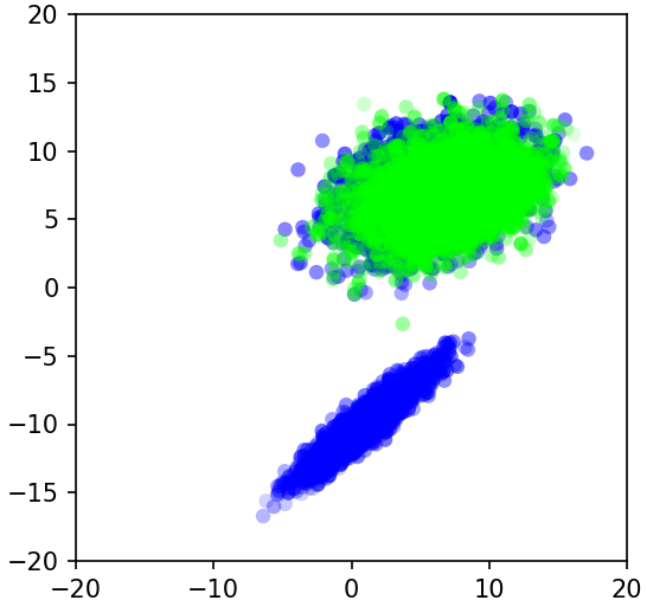
● Pop1 + Pop2



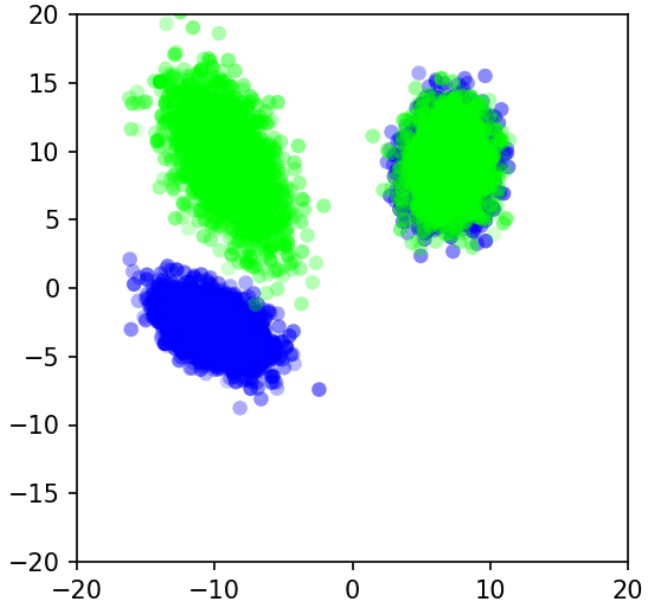
Experimentation

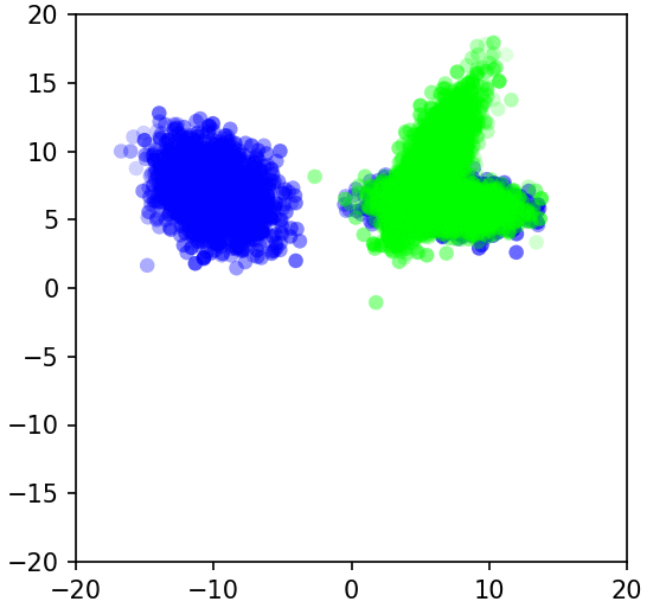


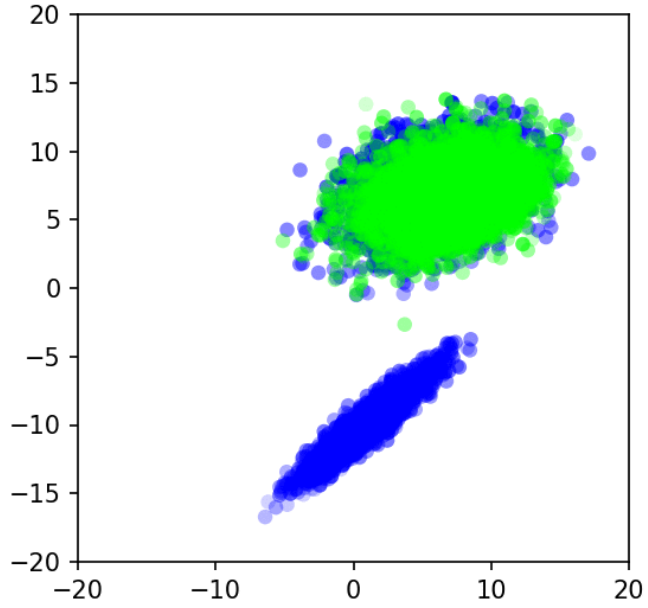




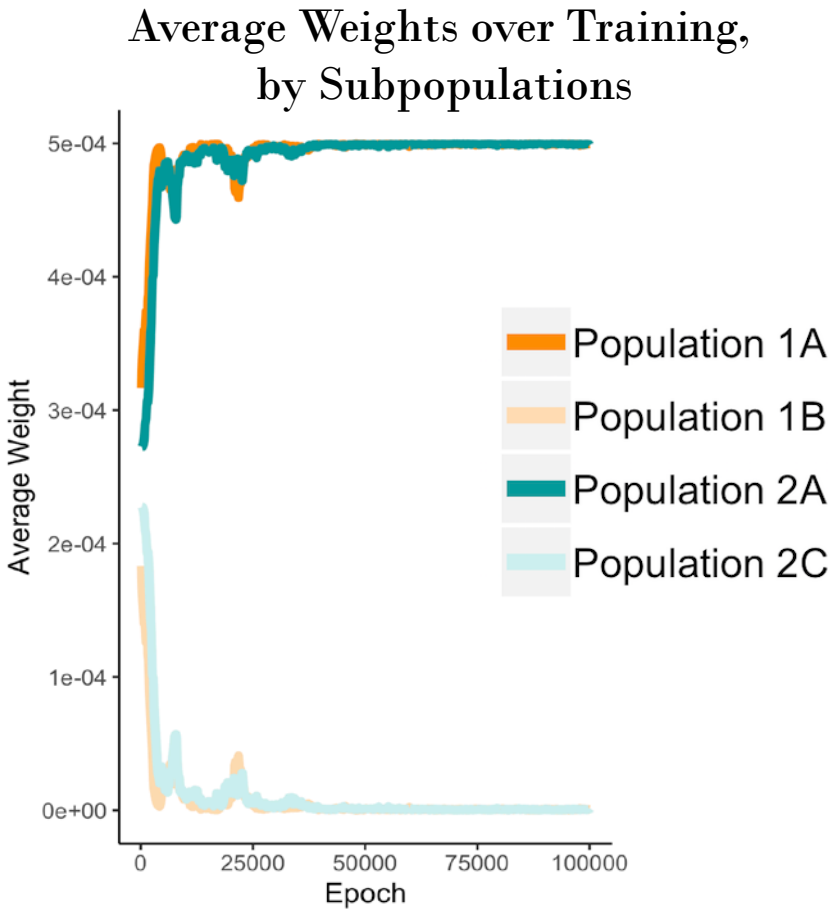
Experimentation





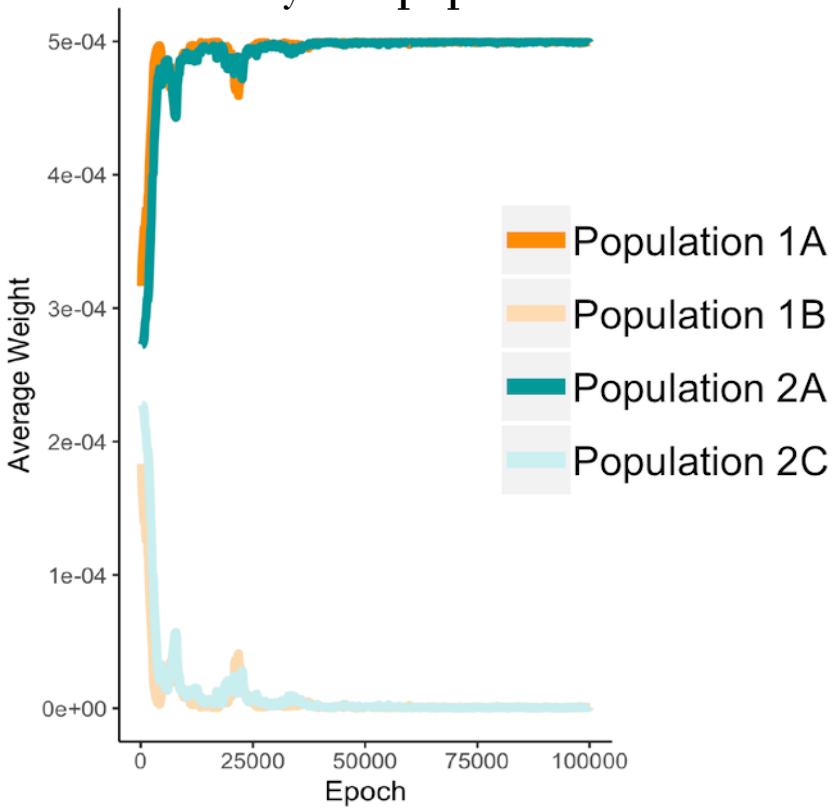


HYPOTHESIS 1: weights from Pop 1A/Pop 2A will be high; weights from Pop 1B/Pop 2C will be low

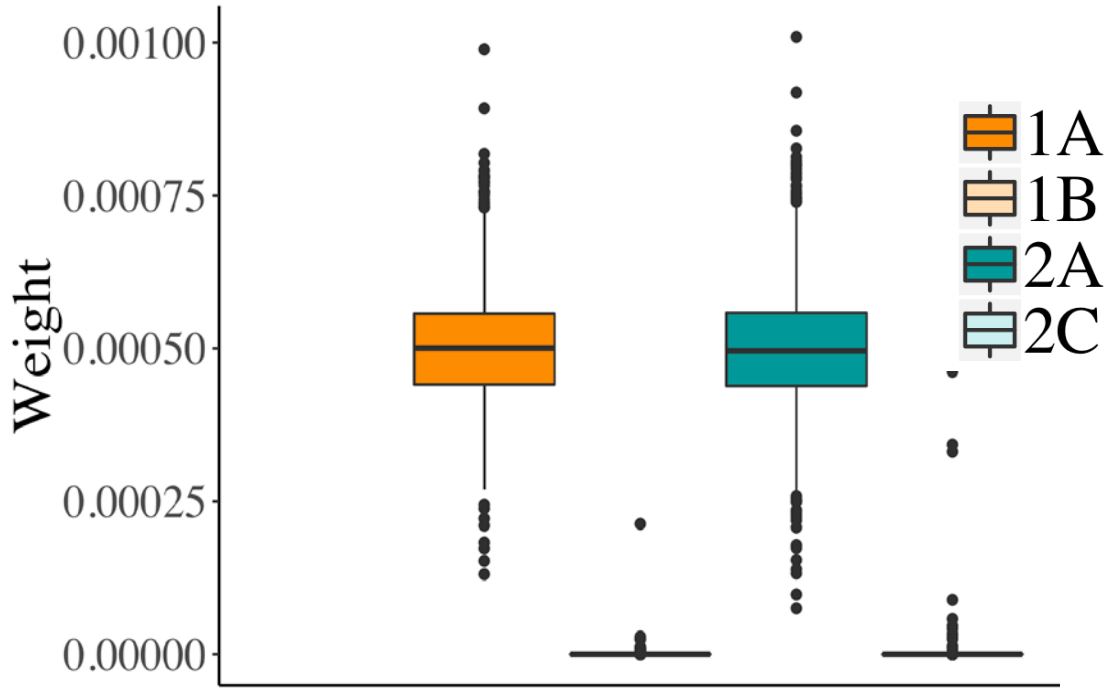


HYPOTHESIS 1: weights from Pop 1A/Pop 2A will be high; weights from Pop 1B/Pop 2C will be low

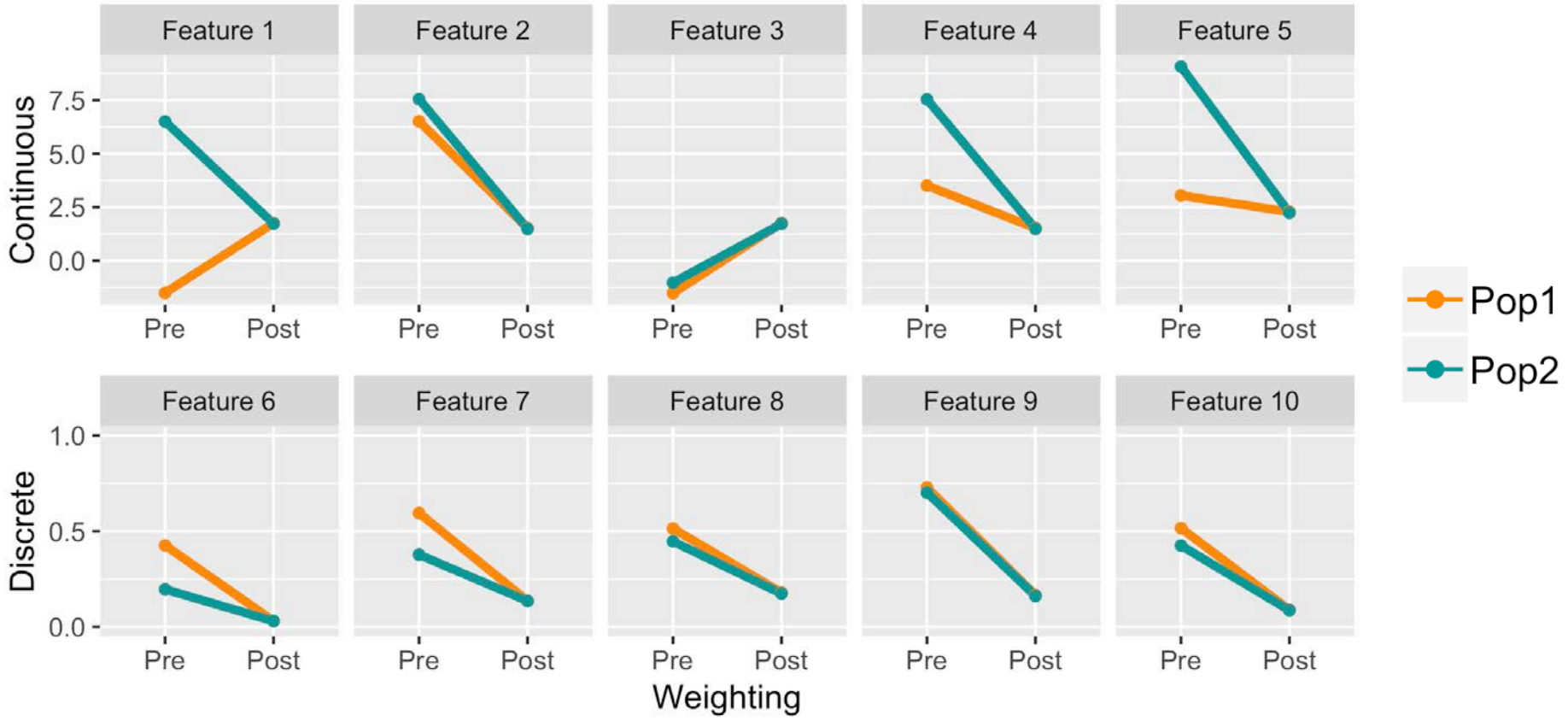
Average Weights over Training, by Subpopulations



Average Final Weights, by Subpopulations



HYPOTHESIS 2: weighting will make features more similar between populations



$ATE_{mixture} = 50$
$ATE_{overlap} = 70$

HYPOTHESIS 3: cGAN-weighted ATE will be less biased than comparators

Weighting Method
unweighted
cGAN
Inverse probability of treatment (IPTW)
Clipped IPTW
Binary regression propensity score
generalized boosted modeling of propensity scores McCaffrey 2004
covariate-balancing propensity scores Imai 2014
non-parametric covariate-balancing propensity scores Fong 2018
entropy balancing weights Hainmueller 2012
empirical balancing calibration weights Chan 2016
optimization-based weights Zubizarreta 2015

$ATE_{mixture} = 50$
$ATE_{overlap} = 70$

HYPOTHESIS 3: cGAN-weighted ATE will be less biased than comparators

Weighting Method	ATE
unweighted	50.03
cGAN	70.01
Inverse probability of treatment (IPTW)	92.00
Clipped IPTW	87.24
Binary regression propensity score	92.00
generalized boosted modeling of propensity scores McCaffrey 2004	84.51
covariate-balancing propensity scores Imai 2014	91.83
non-parametric covariate-balancing propensity scores Fong 2018	37.65
entropy balancing weights Hainmueller 2012	104.13
empirical balancing calibration weights Chan 2016	52.06
optimization-based weights Zubizarreta 2015	52.07

$ATE_{mixture} = 50$
$ATE_{overlap} = 70$

HYPOTHESIS 4: effective sample size of cGAN will be more reasonable than comparators

Weighting Method	ATE	ESS Kish 1965
unweighted	50.03	8000
cGAN	70.01	3870
Inverse probability of treatment (IPTW)	92.00	6551
Clipped IPTW	87.24	6997
Binary regression propensity score	92.00	6551
generalized boosted modeling of propensity scores McCaffrey 2004	84.51	7207
covariate-balancing propensity scores Imai 2014	91.83	6686
non-parametric covariate-balancing propensity scores Fong 2018	37.65	11
entropy balancing weights Hainmueller 2012	104.13	65
empirical balancing calibration weights Chan 2016	52.06	65
optimization-based weights Zubizarreta 2015	52.07	114

1. simulation
2. application to clinical data

NewYork-Presbyterian

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Efficacy and Tolerability of Sitagliptin Compared with Glimepiride in Elderly Patients with Type 2 Diabetes Mellitus and Inadequate Glycemic Control: A Randomized, Double-Blind, Non-Inferiority Trial

Paul Hartley[1] · Yue Shentu[2] · Patricia Betz-Schiff[2] · Gregory T. Golm[2] · Christine McCrary Sisk[2] · Samuel S. Engel[2] · R. Ravi Shankar[2]

NIH U.S. National Library of Medicine
ClinicalTrials.gov

clinicaltrials.gov/ct2/show/NCT01189890

Hartley 2015

NewYork-Presbyterian

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Efficacy and Tolerability of Sitagliptin Compared with Glimepiride in Elderly Patients with Type 2 Diabetes Mellitus and Inadequate Glycemic Control: A Randomized, Double-Blind, Non-Inferiority Trial

Paul Hartley[1] · Yue Shentu[2] · Patricia Betz-Schiff[2] · Gregory T. Golm[2] · Christine McCrary Sisk[2] · Samuel S. Engel[2] · R. Ravi Shankar[2]

NIH U.S. National Library of Medicine
ClinicalTrials.gov

clinicaltrials.gov/ct2/show/NCT01189890

eligible patients

- diagnosis of Type II Diabetes Mellitus
- prescription to sitagliptin or glimepiride
- aged 65-80.

Hartley 2015

NewYork-Presbyterian

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Efficacy and Tolerability of Sitagliptin Compared with Glimepiride in Elderly Patients with Type 2 Diabetes Mellitus and Inadequate Glycemic Control: A Randomized, Double-Blind, Non-Inferiority Trial

Paul Hartley[1] · Yue Shentu[2] · Patricia Betz-Schiff[2] · Gregory T. Golm[2] · Christine McCrary Sisk[2] · Samuel S. Engel[2] · R. Ravi Shankar[2]

NIH U.S. National Library of Medicine
ClinicalTrials.gov

clinicaltrials.gov/ct2/show/NCT01189890

Hartley 2015

eligible patients

- diagnosis of Type II Diabetes Mellitus
- prescription to sitagliptin or glimepiride
- aged 65-80.

a **sub-sample** of sitagliptin users was taken to match the count of the glimepiride arm (N=608 vs N=144).

- not necessary

NewYork-Presbyterian

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Efficacy and Tolerability of Sitagliptin Compared with Glimepiride in Elderly Patients with Type 2 Diabetes Mellitus and Inadequate Glycemic Control: A Randomized, Double-Blind, Non-Inferiority Trial

Paul Hartley[1] · Yue Shentu[2] · Patricia Betz-Schiff[2] · Gregory T. Golm[2] · Christine McCrary Sisk[2] · Samuel S. Engel[2] · R. Ravi Shankar[2]

NIH U.S. National Library of Medicine
ClinicalTrials.gov

clinicaltrials.gov/ct2/show/NCT01189890

Hartley 2015

eligible patients

- diagnosis of Type II Diabetes Mellitus
- prescription to sitagliptin or glimepiride
- aged 65-80.

a **sub-sample** of sitagliptin users was taken to match the count of the glimepiride arm (N=608 vs N=144).

- not necessary

37 features

- repeated measurements: the most recent result was selected.
- missing data: values were imputed

NewYork-Presbyterian

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Efficacy and Tolerability of Sitagliptin Compared with Glimepiride in Elderly Patients with Type 2 Diabetes Mellitus and Inadequate Glycemic Control: A Randomized, Double-Blind, Non-Inferiority Trial

Paul Hartley[1] · Yue Shentu[2] · Patricia Betz-Schiff[2] · Gregory T. Golm[2] · Christine McCrary Sisk[2] · Samuel S. Engel[2] · R. Ravi Shankar[2]

NIH U.S. National Library of Medicine
ClinicalTrials.gov

clinicaltrials.gov/ct2/show/NCT01189890

Hartley 2015

eligible patients

- diagnosis of Type II Diabetes Mellitus
- prescription to sitagliptin or glimepiride
- aged 65-80.

a **sub-sample** of sitagliptin users was taken to match the count of the glimepiride arm (N=608 vs N=144).

- not necessary

37 features

- repeated measurements: the most recent result was selected.
- missing data: values were imputed

hypothesis

1. cGAN will improve feature balance over comparator methods

Hypothesis 1: cGAN will improve feature balance over comparator methods

ASDM is a common metric of feature balance Austin 2011. A lower ASDM is indicative of feature balance

$$ASDM = \left| \frac{\bar{x}_{treatment} - \bar{x}_{control}}{\sqrt{\frac{s^2_{treatment} + s^2_{control}}{2}}} \right|$$

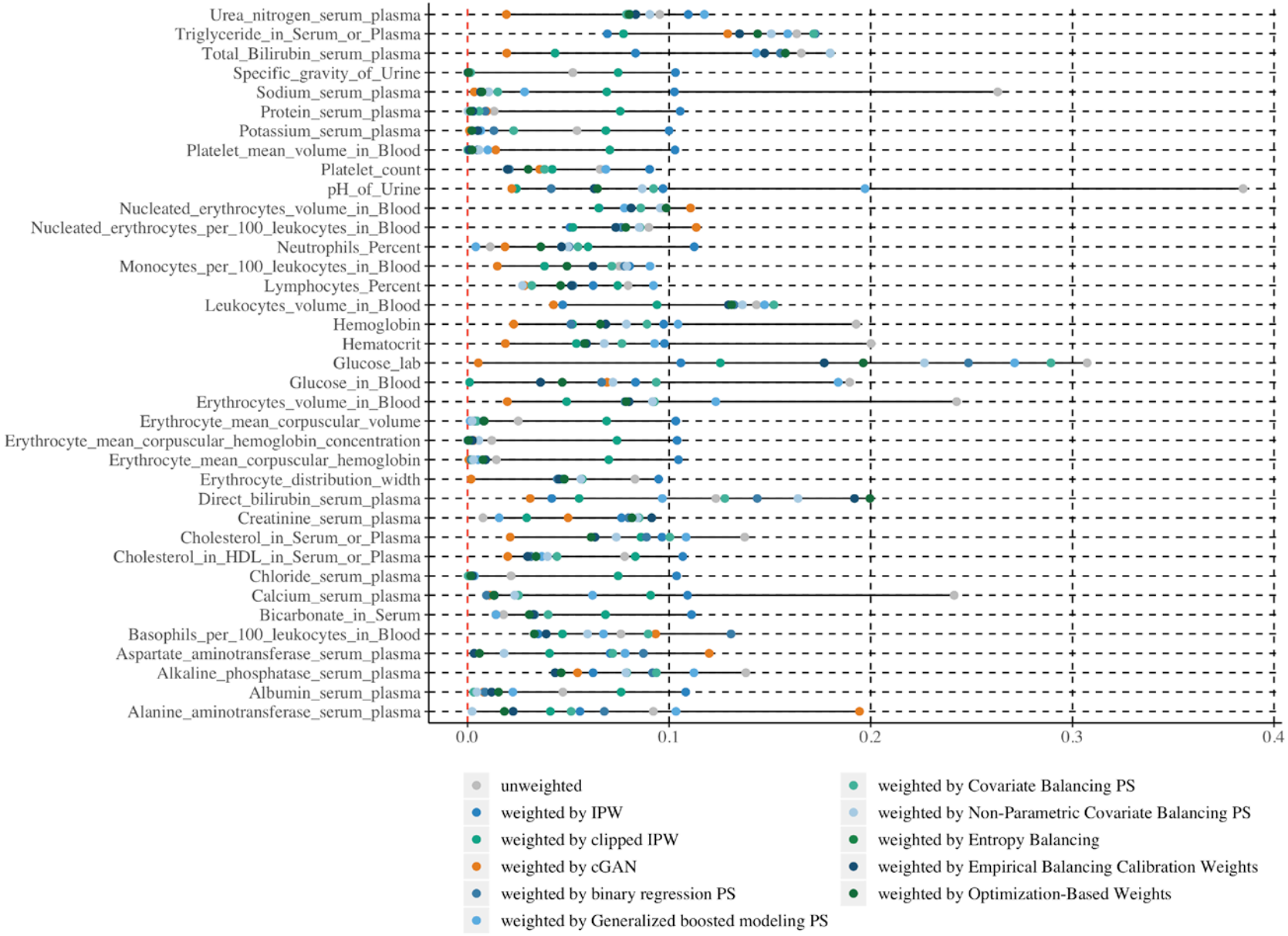
same comparators as simulation

Hypothesis 1: cGAN will improve feature balance over comparator methods

ASDM is a common metric of feature balance Austin 2011. A lower ASDM is indicative of feature balance

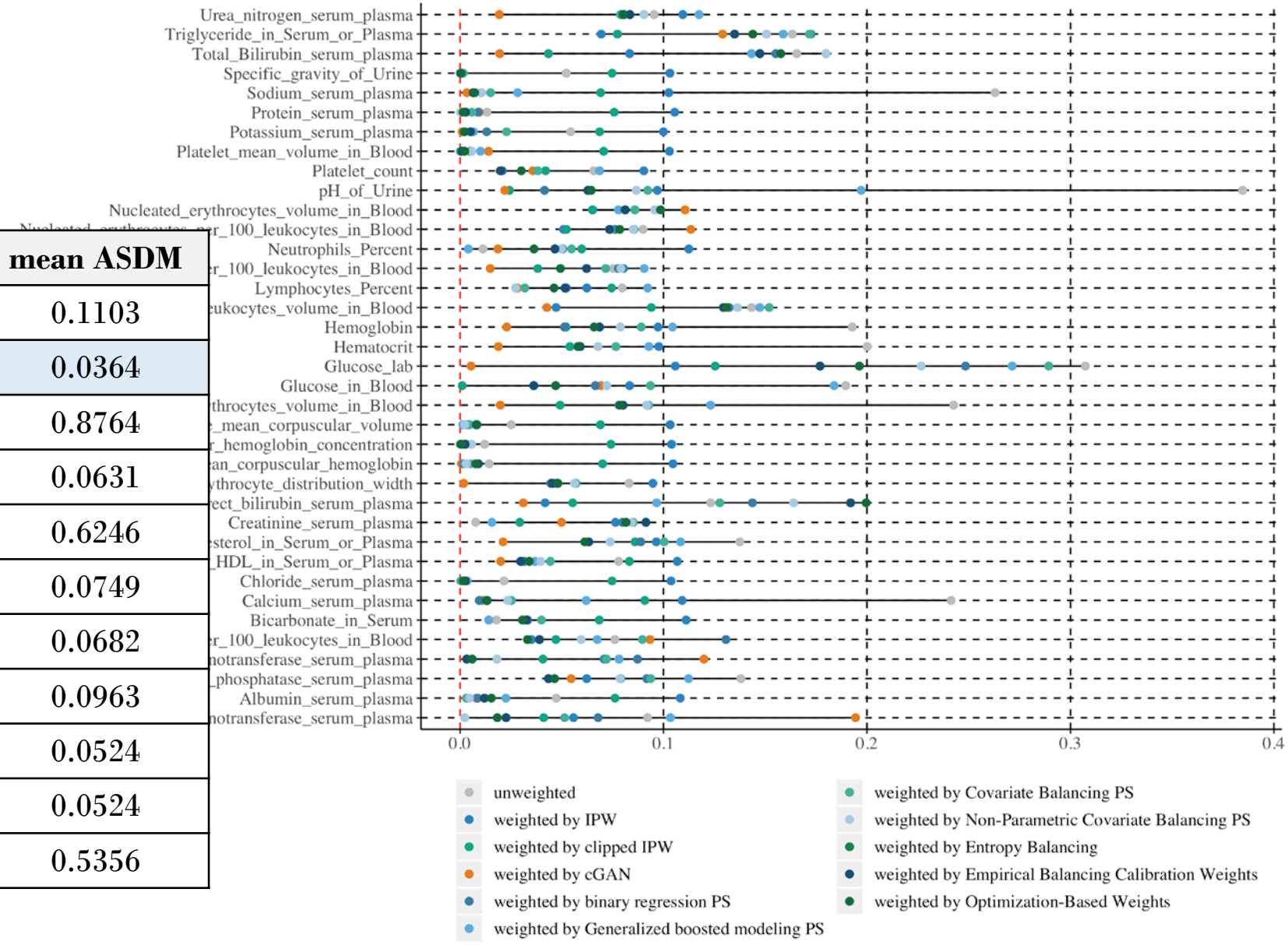
$$ASDM = \left| \frac{\bar{x}_{treatment} - \bar{x}_{control}}{\sqrt{\frac{s^2_{treatment} + s^2_{control}}{2}}} \right|$$

same comparators as simulation



Hypothesis 1: cGAN will improve feature balance over comparator methods

Weighting Method	mean ASDM
unweighted	0.1103
cGAN	0.0364
IPTW	0.8764
clipped IPTW	0.0631
binary regression	0.6246
generalized boosted modeling	0.0749
covariate-balancing	0.0682
non-parametric covariate-balancing	0.0963
entropy balancing weights	0.0524
empirical balancing calibration	0.0524
optimization-based weights	0.5356



ASDM

Conclusions

the experiments suggest that Counterfactual χ-GAN is an effective method of learning feature balancing weights to support counterfactual inference!

the experiments suggest that Counterfactual χ-GAN is an effective method of learning feature balancing weights to support counterfactual inference!

the Counterfactual χ-GAN could provide an alternative means to causal inference from observational data.

Conclusions

the experiments suggest that Counterfactual χ-GAN is an effective method of learning feature balancing weights to support counterfactual inference!

the Counterfactual χ-GAN could provide an alternative means to causal inference from observational data.

furthermore, if we assume that all potentially confounding variables are observed and included as features, average treatment effect estimates from Counterfactual χ-GAN weighted models may be less biased.

Limitations
GANs are unstable
Parameter tuning is hard
What is the best way to assess convergence?
discrete data - gradients are unbiased, but high variance

Conclusions

Limitations	Future Directions
GANs are unstable	application to clinical data. compare to RCT. need multisite collaborators
Parameter tuning is hard	
What is the best way to assess convergence?	assessing variance of outcome. this requires a more complex simulation
discrete data - gradients are unbiased, but high variance	

References

Thank you

Natnicha Vanitchanant

Rajesh Ranganath

Adler J. Perotte

aja2149@cumc.columbia.edu

http://people.dbmi.columbia.edu/AJA/

https://www.linkedin.com/in/ameliajaveritt

@AJAveritt

back up

learning for discrete data

We leverage a **score function estimator**. This score function-based estimator exchanges a gradient of an expectation for an expectation of a gradient which we can make an unbiased Monte Carlo estimate and incorporate into a modified stochastic backpropogation procedure.

$$\nabla_{\pi_a} \mathbb{E}_{\hat{q}_t(x;\pi_a)}[f(x)] = \mathbb{E}_{\hat{q}_t(x;\pi_a)}[f(x) \nabla_{\pi_a} \log \hat{q}_t(x;\pi_a)]$$

Glasserman 2003; Fu 2006; Schulman 2015