

Designing Student and Workforce Outcomes Through Data and AI

From Insight to Action

Jamie Fogel

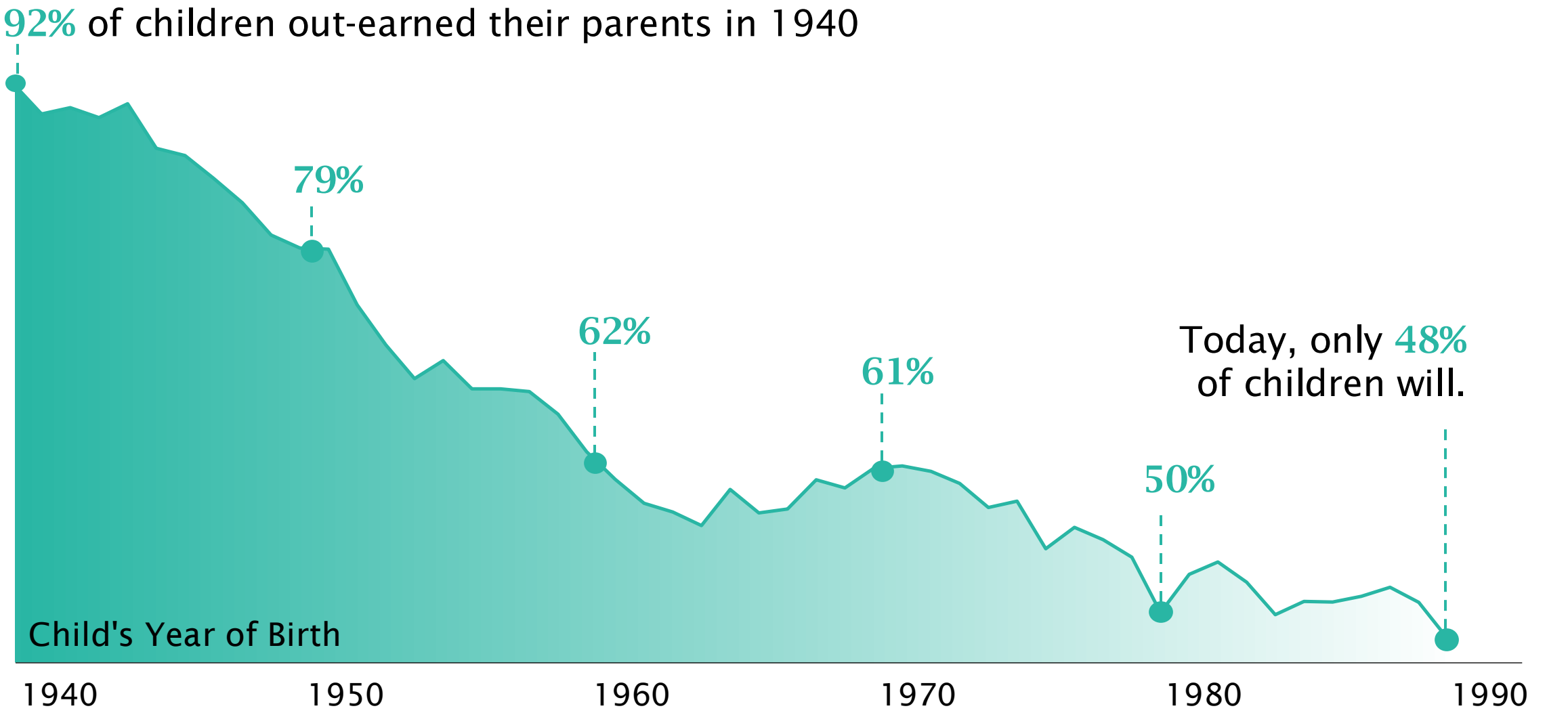
Harvard University

OPPORTUNITY INSIGHTS

CTData Conference | December 2026

The Fading American Dream: Upward Mobility by Birth Cohort

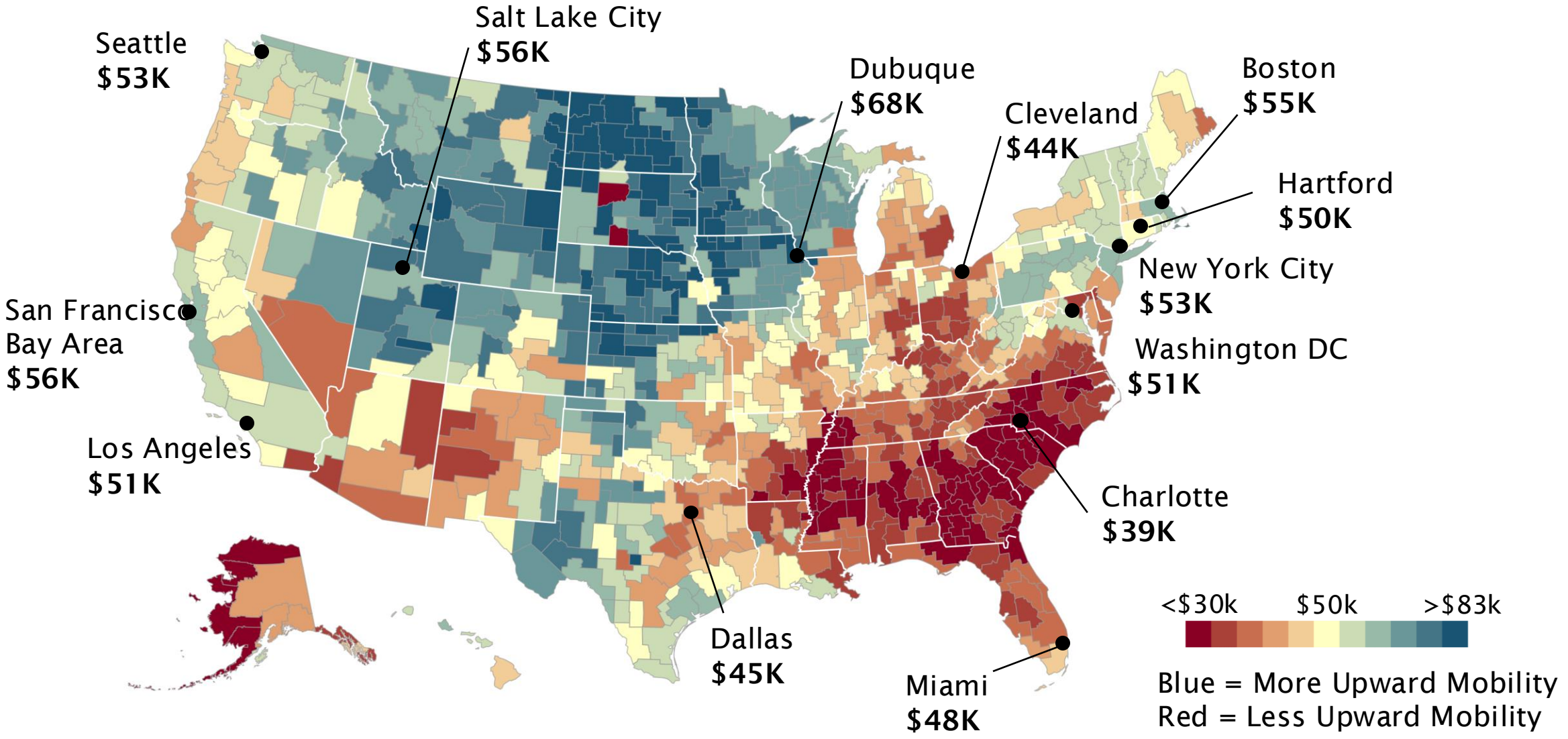
Percent of Children Earning More than Their Parents, by Year of Birth



Source: Chetty, Grusky, Hell, Hendren, Manduca, Narang (Science 2017)

The Geography of Upward Mobility in the United States

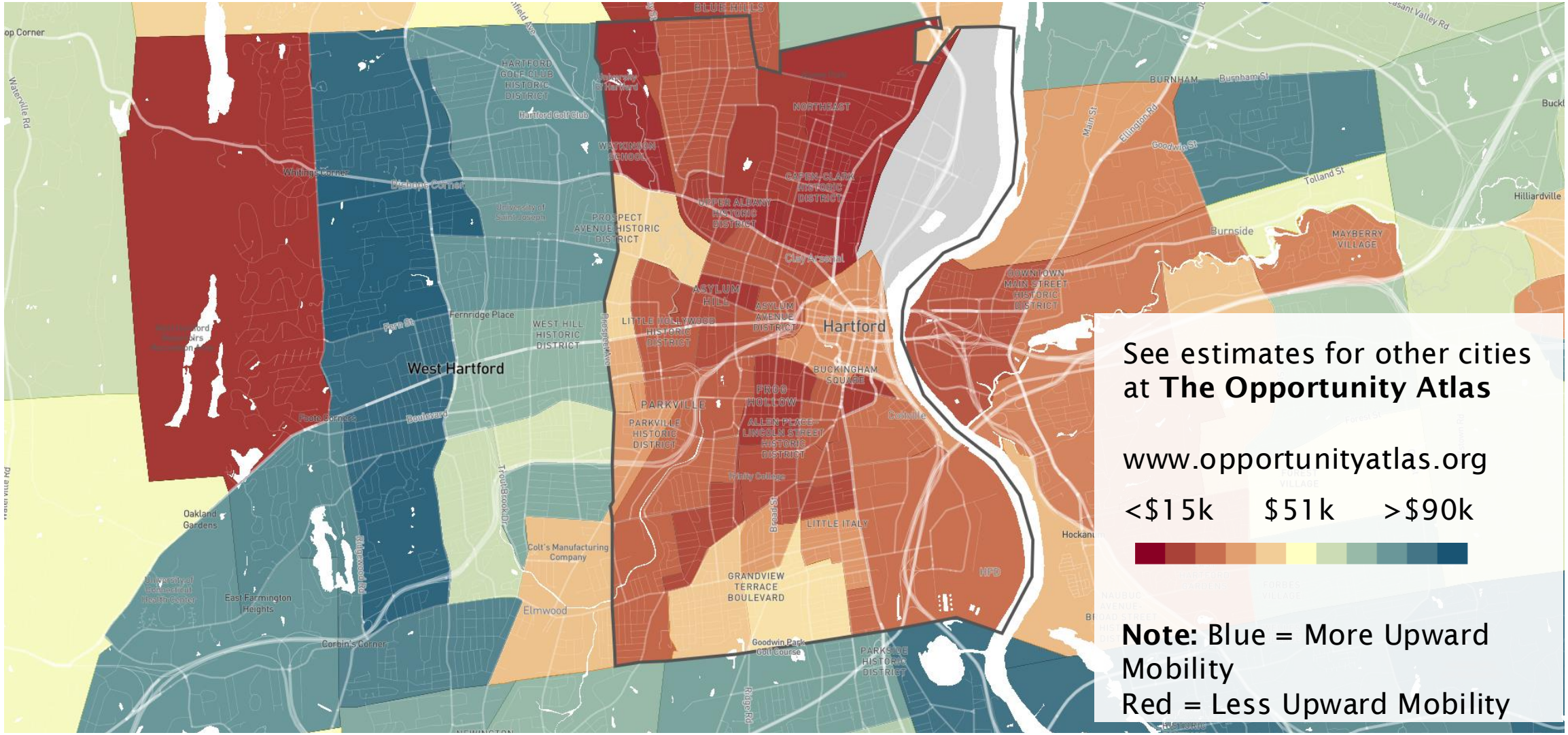
Average Household Income at Age 35 for Children whose Parents Earned $41K (25th percentile)



Source: Chetty, Hendren, Kline, Saez (QJE 2014)

The Geography of Upward Mobility in Hartford

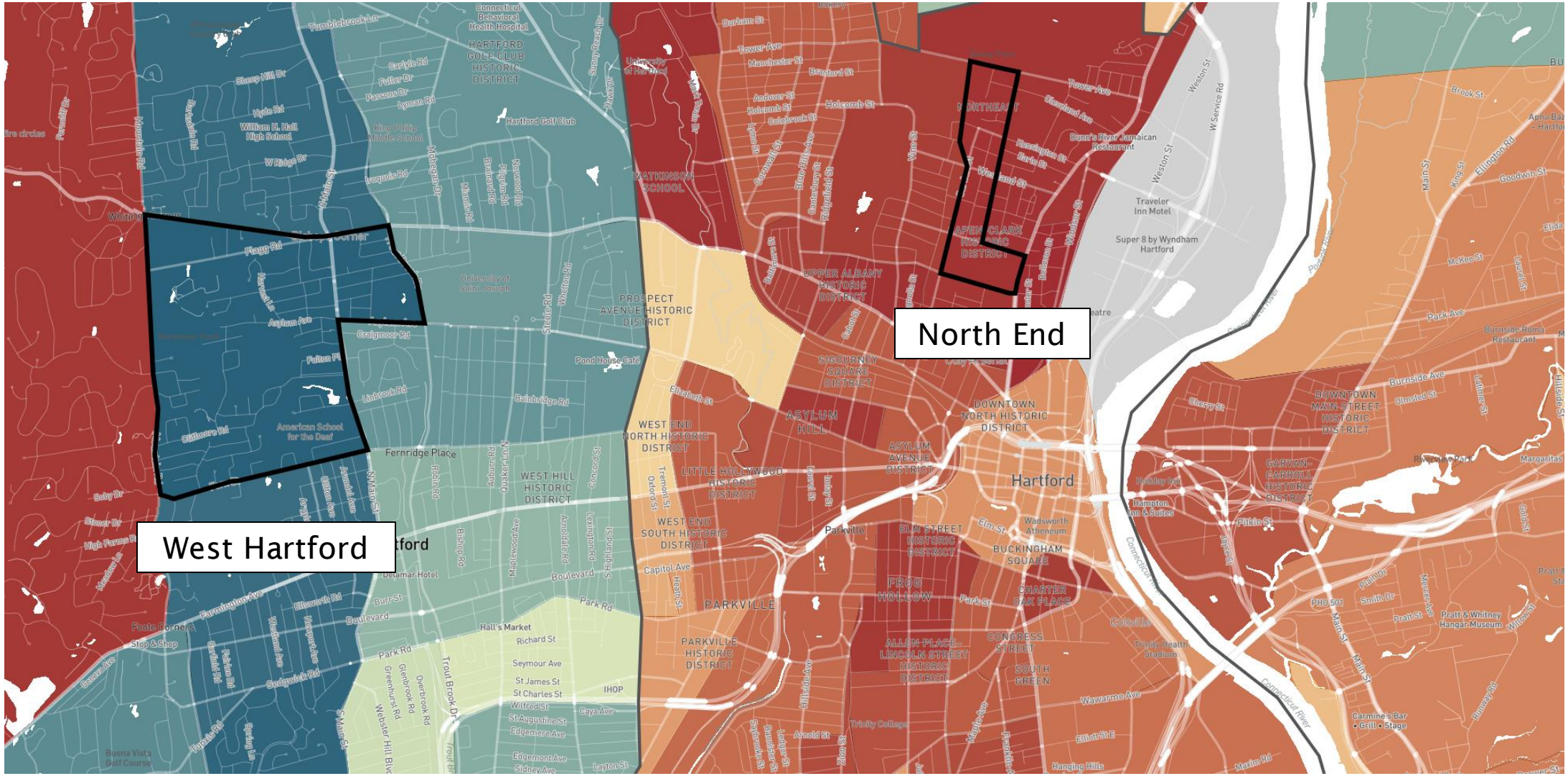
Average Income at Age 35 for Children whose Parents Earned $41K (25th percentile)



Source: Chetty, Friedman, Hendren, Jones, Porter (AER forthcoming 2025)

The Geography of Upward Mobility in Hartford

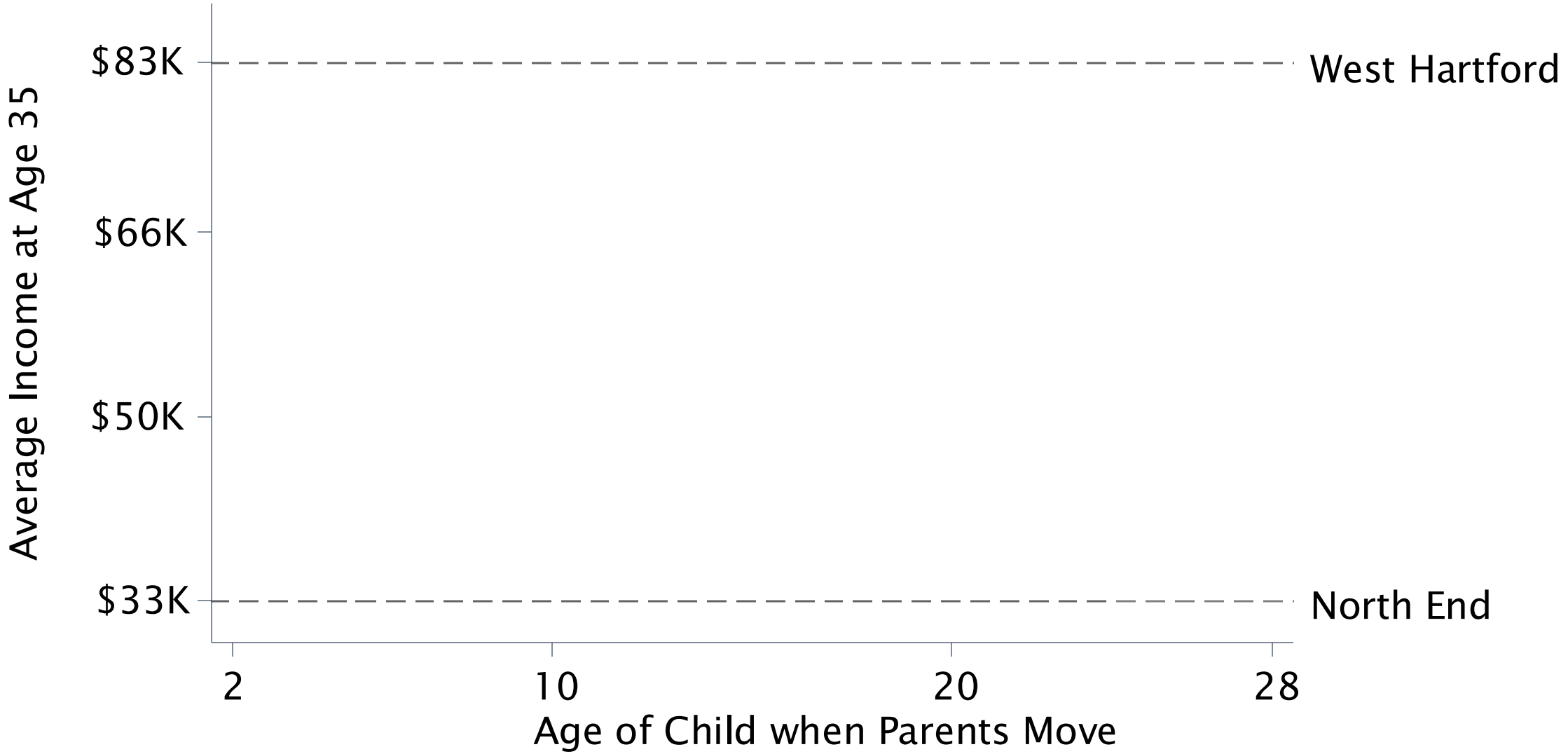
Average Income at Age 35 for Children Born in 1978 whose Parents Earned $41K (25th pctile)



Source: Chetty, Friedman, Hendren, Jones, Porter (NBER wp 2018)

Income Gain from Moving to a Higher Opportunity Neighborhood

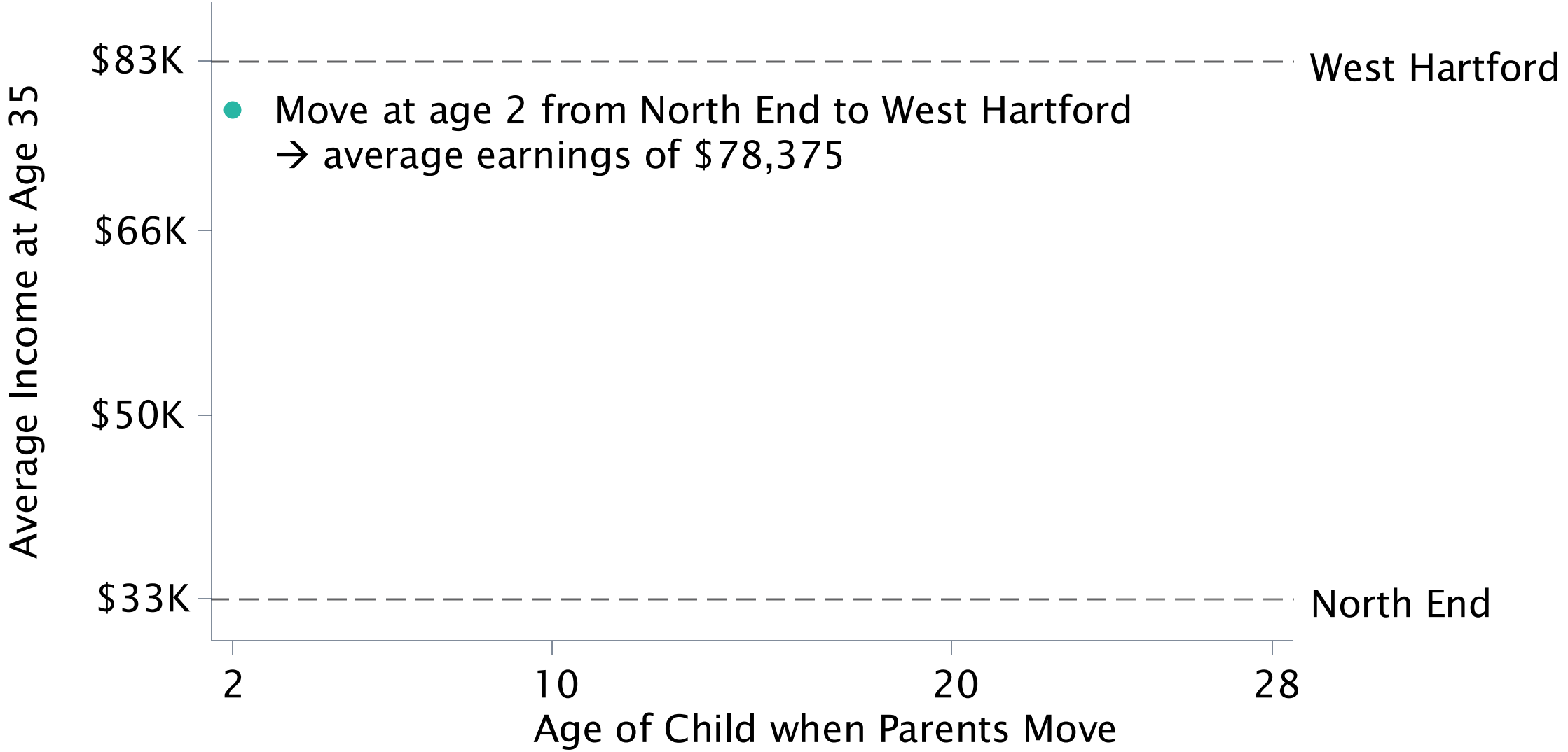
By Child's Age at Move



Source: Chetty and Hendren (QJE 2018)

Income Gain from Moving to a Higher Opportunity Neighborhood

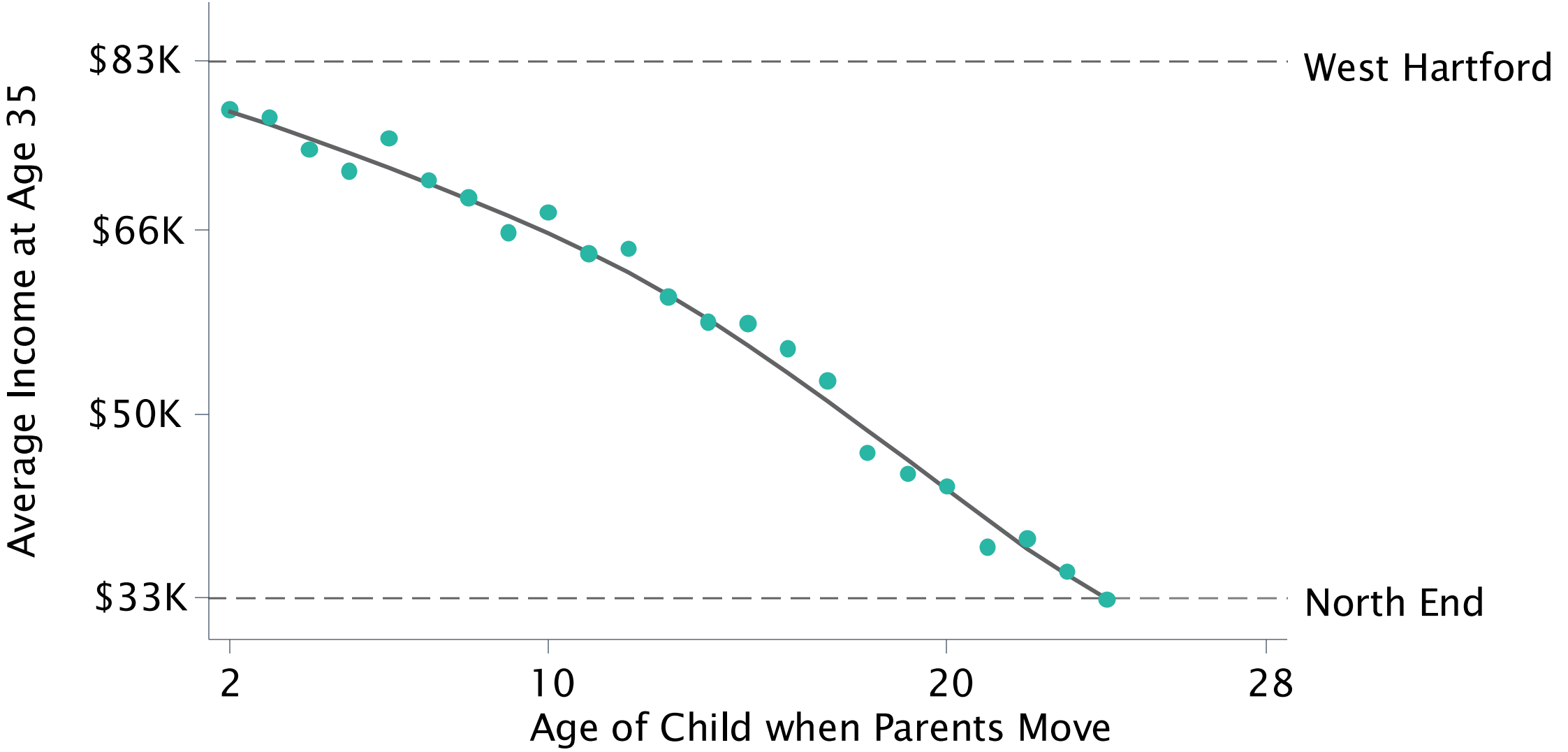
By Child's Age at Move



Source: Chetty and Hendren (QJE 2018)

Income Gain from Moving to a Higher Opportunity Neighborhood

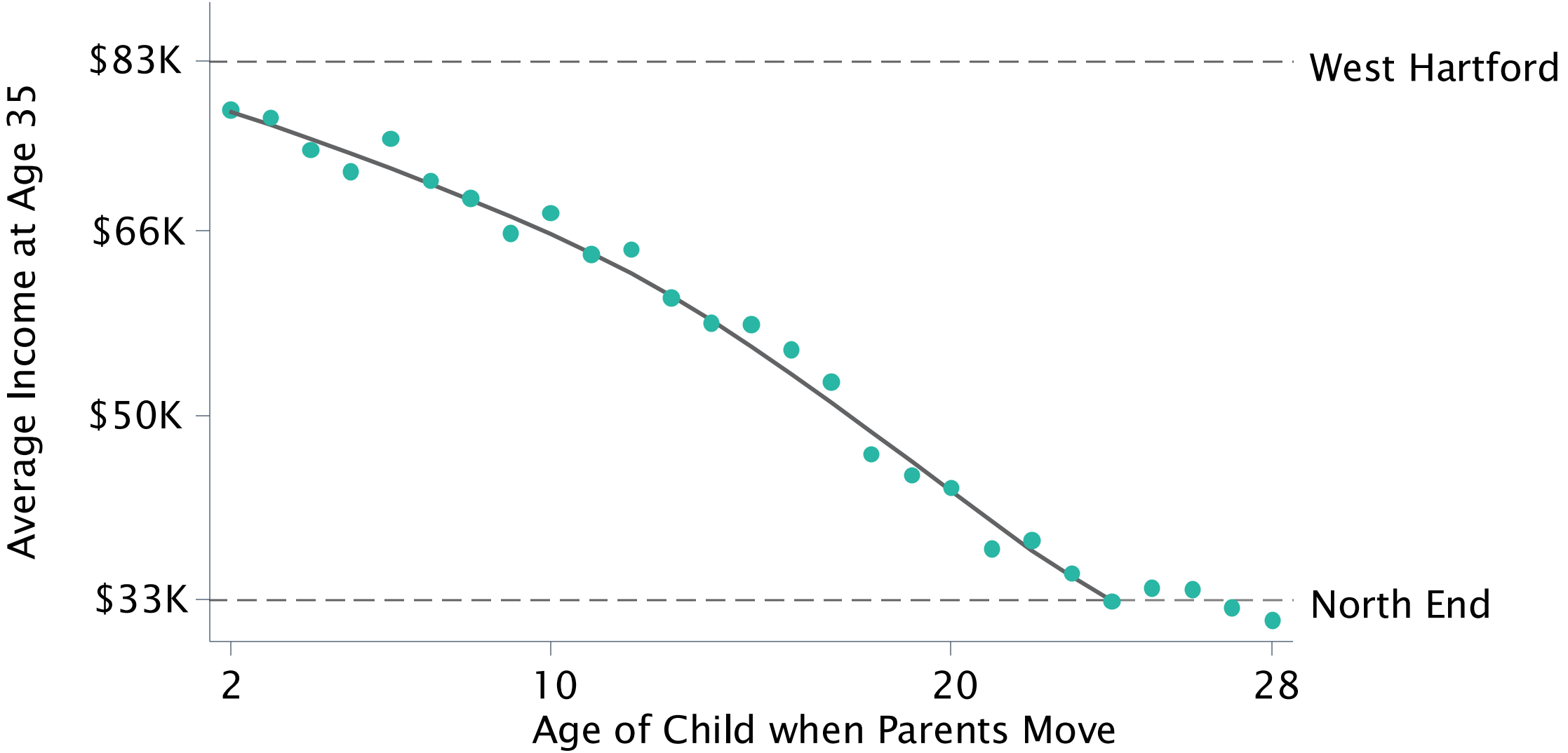
By Child's Age at Move



Source: Chetty and Hendren (QJE 2018)

Income Gain from Moving to a Higher Opportunity Neighborhood

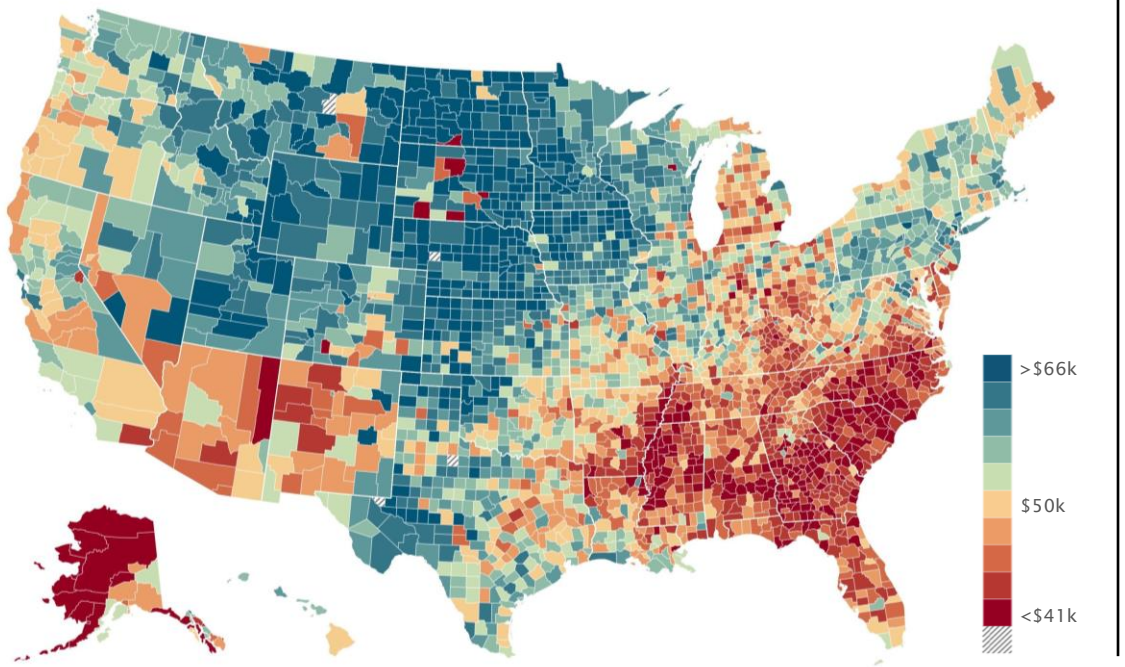
By Child's Age at Move



Source: Chetty and Hendren (QJE 2018)

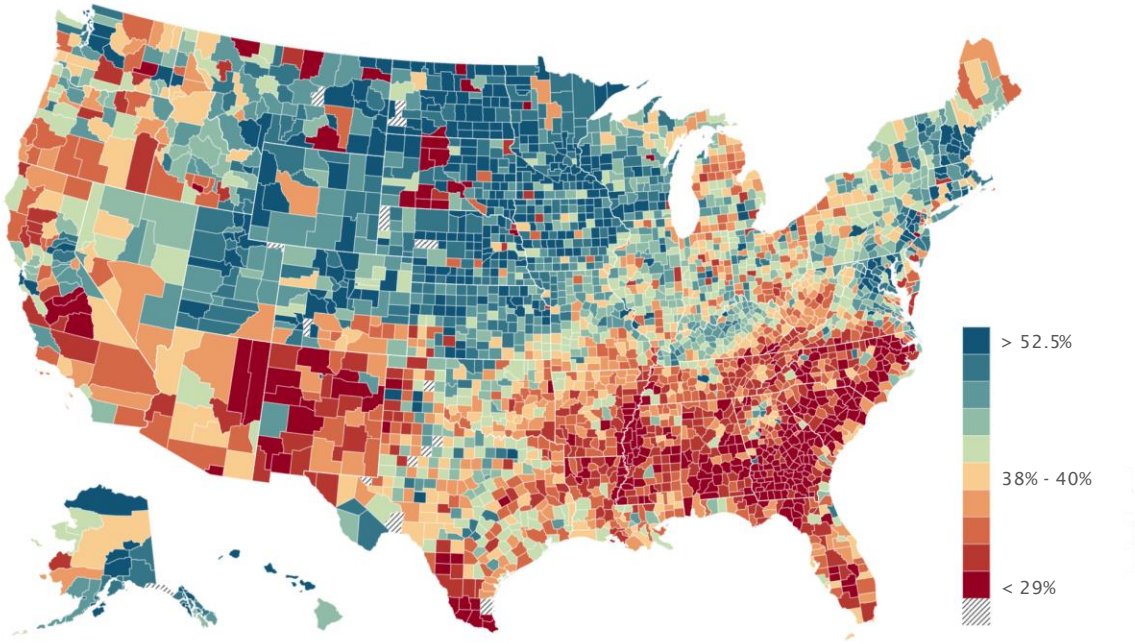
Economic Opportunity and Social Capital

Upward Mobility Based on Tax Records



Average Household Income at Age 35
for Children whose Parents Earned $41k

Economic Connectedness Based on Facebook Data



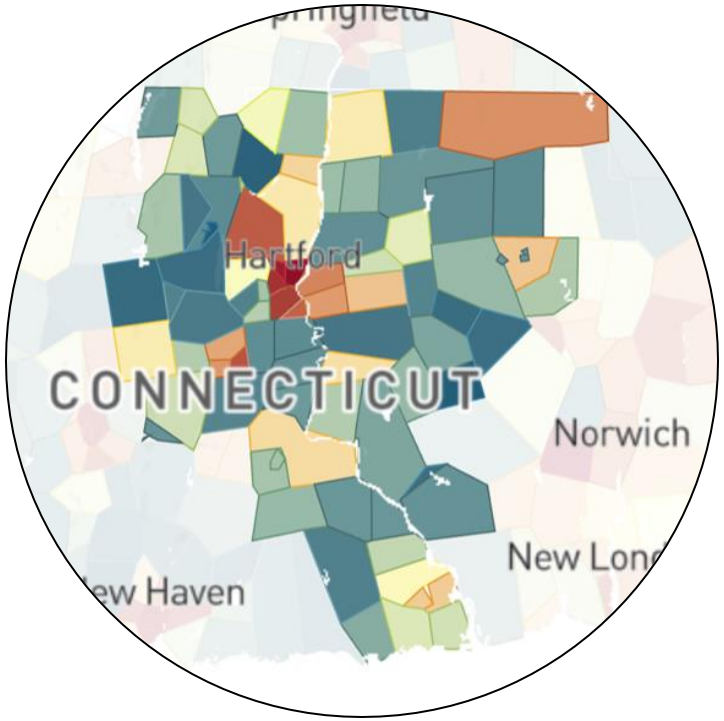
Share of Above-Median-Income Friends
among Below-Median-Income People

Source: Chetty, Friedman, Hendren, Jones, Porter (2020); Chetty, Jackson, Kuchler, Stroebel et al. (Nature 2022)

Economic Opportunity and Social Capital

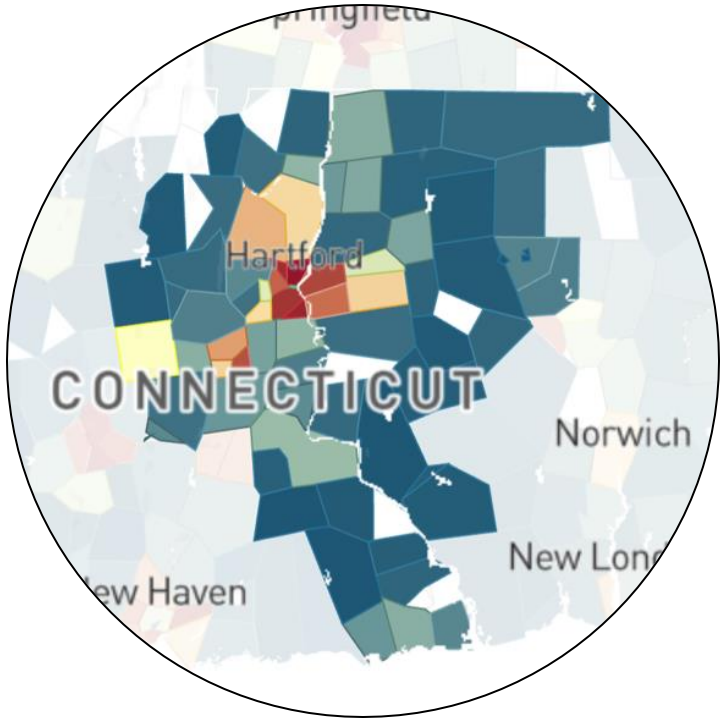
Hartford

Upward Mobility Based on Tax Records



Average Household Income at Age 35
for Children whose Parents Earned $41k

Economic Connectedness Based on Facebook Data



Share of Above-Median-Income Friends
among Below-Median-Income People

Source: Chetty, Friedman, Hendren, Jones, Porter (2020); Chetty, Jackson, Kuchler, Stroebel et al. (Nature 2022)

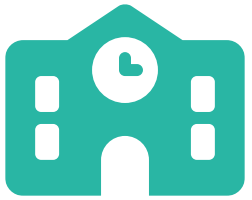
Characteristics of High-Mobility Neighborhoods



Lower Poverty Rates



More Stable Family Structure



Better School Quality



Greater Social Capital

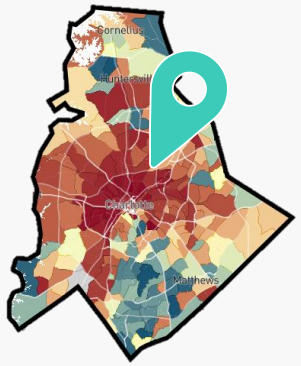
Source: Chetty, Friedman, Hendren, Jones, Porter (2020)

Three Policy Approaches to Increasing Upward Mobility

Reducing Segregation

Help Low-Income Families Move to High-Opportunity Areas

Place-Based Investments



Increase Upward Mobility in Low-Opportunity Areas

Higher Education & Workforce Training

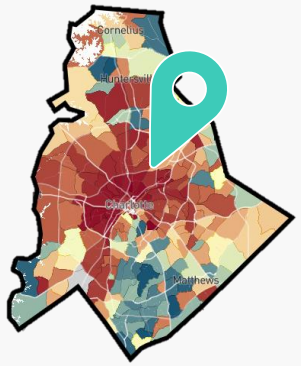
Amplify Impacts of Colleges and Training Programs on Mobility

Three Policy Approaches to Increasing Upward Mobility

Reducing Segregation

Help Low-Income Families Move to High-Opportunity Areas

Place-Based Investments



Increase Upward Mobility in Low-Opportunity Areas

Higher Education & Workforce Training

CLIMB

Amplify Impacts of Colleges and Training Programs on Mobility



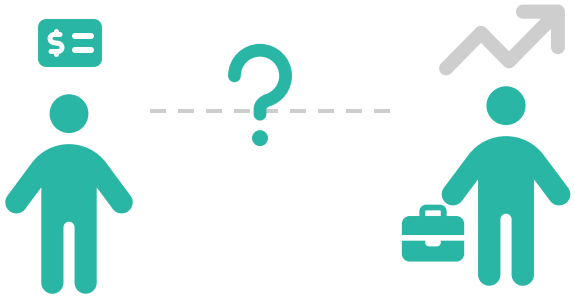
Source: Workforce Almanac



Source: Workforce Almanac

***Evaluated Using Randomized Trial**

Two Challenges in Scaling Programs to Promote Economic Mobility

Challenge	Lack of information on earnings impacts	Lack of capital to scale effective programs
		
Solution	Use digital twins in anonymized data to measure impact	Give private investors a share of earnings gains through tax revenues

How Can We Evaluate Program Impacts at Scale?

- Randomized trials provide gold standard evidence
- But it is infeasible to run RCTs for every program…
- One practical alternative: use observational (non-randomized) data to compare program participants to similar non-participants

History of Observational Estimators

- Many attempts to develop observational estimators over the past 40 years
  - General conclusion: observational estimates (matching, reweighting, regression) are unstable and often biased relative to estimates from RCTs
- Today, we have much larger administrative datasets and better technology
- Can we develop observational methods that work better in the age of big data using tools such as machine learning?



Part 1: Digital Twins



Data Sources

1. Tax + Census records covering the U.S. population (~327 million adults) from 1999-2024

    - Earnings measured from W-2 forms filed by employers

    - Other information from 1040 returns, information returns, and decennial Census: age, sex, Census tract, employers, industry, college attended

2. Enrollment data from workforce training and higher education programs linked to Tax + Census data using secure protocols by the Census Bureau

→ Linked longitudinal analysis dataset that is stripped of identifiers and anonymized

→ All results carefully reviewed by Census staff for confidentiality before disclosure



Case Study: Year Up United Job Training Program

- Begin with a case study of Year Up United job training program
- Revisit impacts of 2014 randomized trial involving 2,545 people who applied to Year Up United

Learn new skills to launch your career.

Year Up's job training program is tuition free, and offers access to today's best companies and a proven path to career success.

Zip Code

FIND TRAINING

Our approach

Three steps to a new job

1 Learn



The Learning & Development phase of our program provides you with skills based training from expert

2 Intern



Put the skills you learned into practice! We'll match you to a hands-on internship at a top

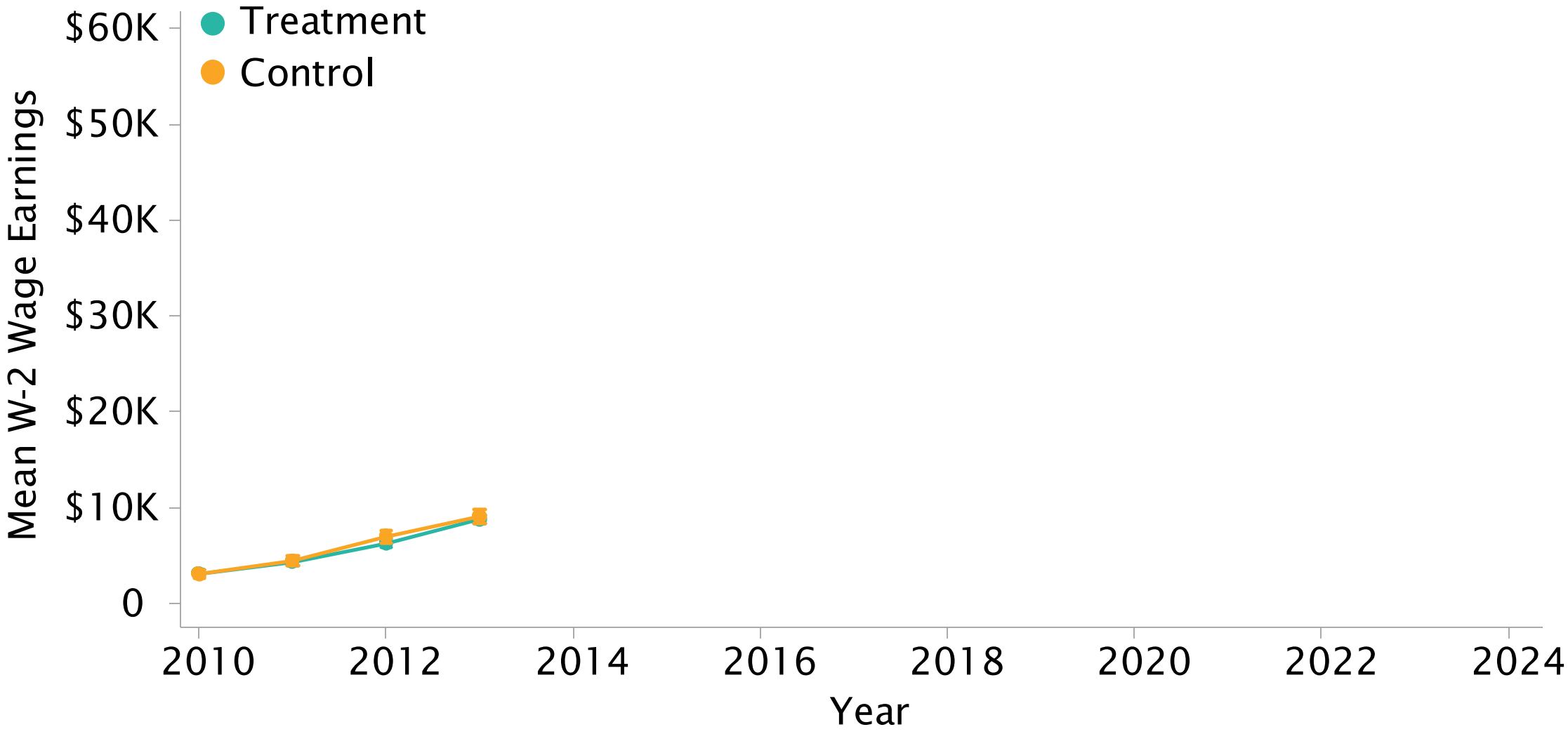
3 Land a job



80% of alumni are employed and/or enrolled in college within 4 months of graduation. Our support staff will

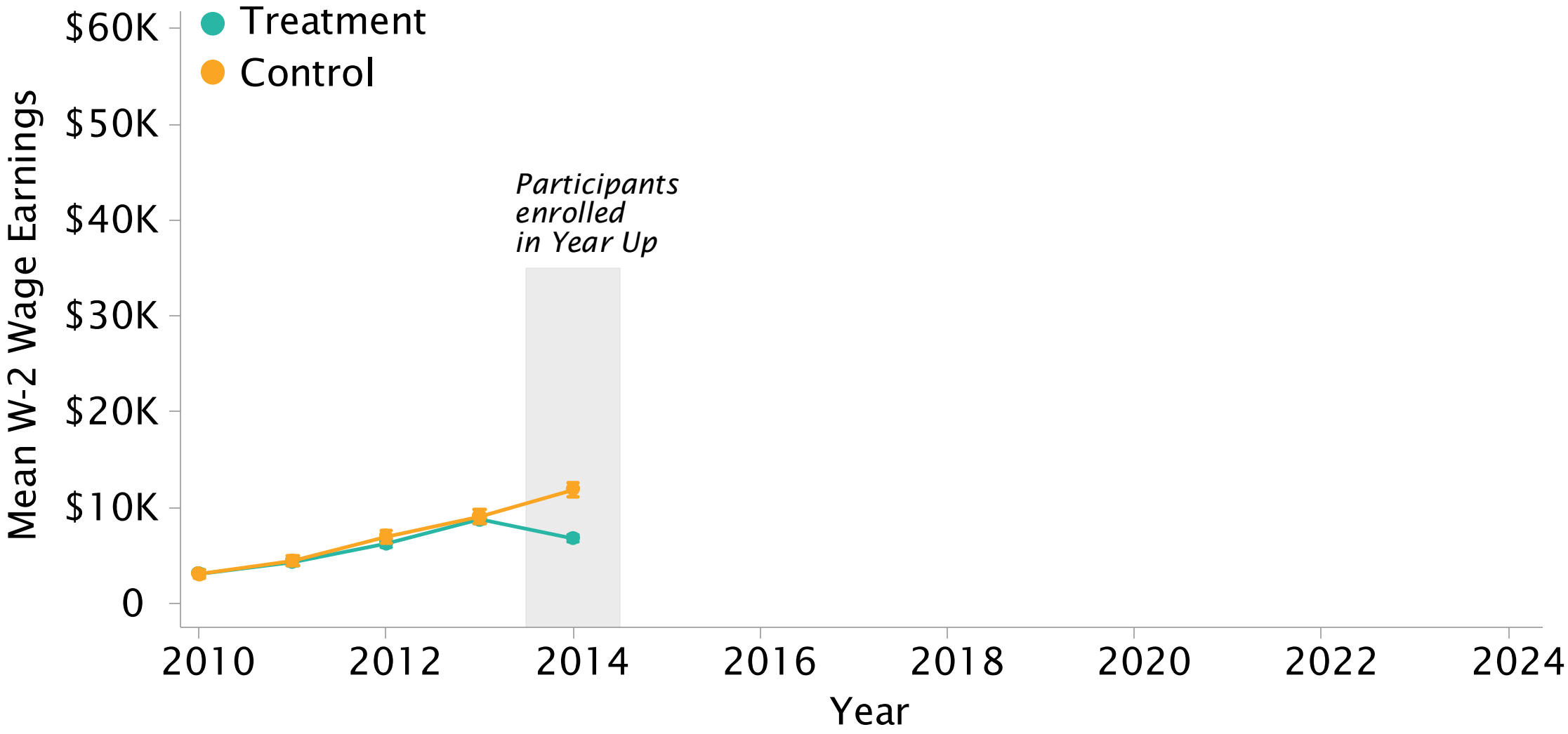
Impact of Year Up United on Earnings

Evidence from a PACE Randomized Trial in 2014



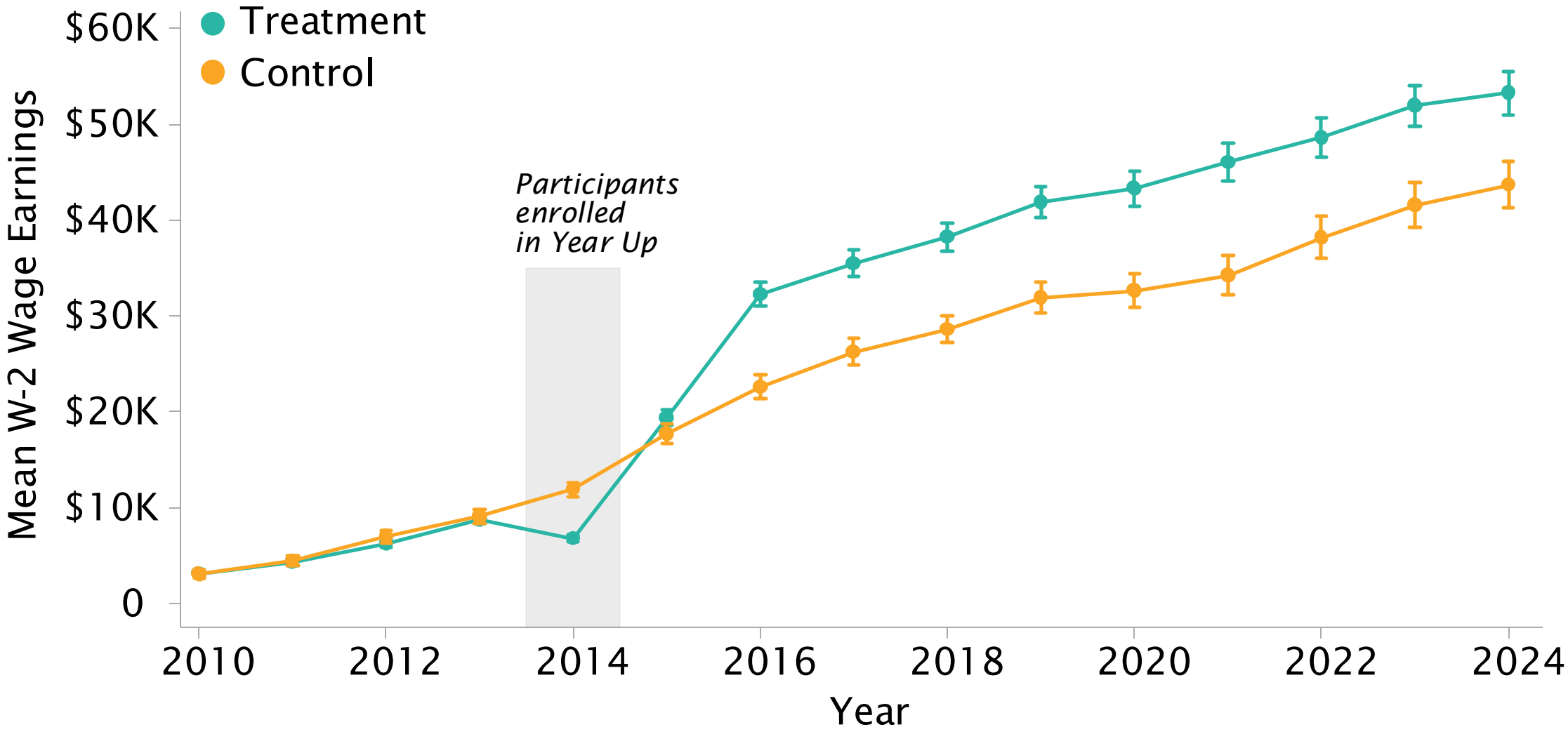
Impact of Year Up United on Earnings

Evidence from a PACE Randomized Trial in 2014



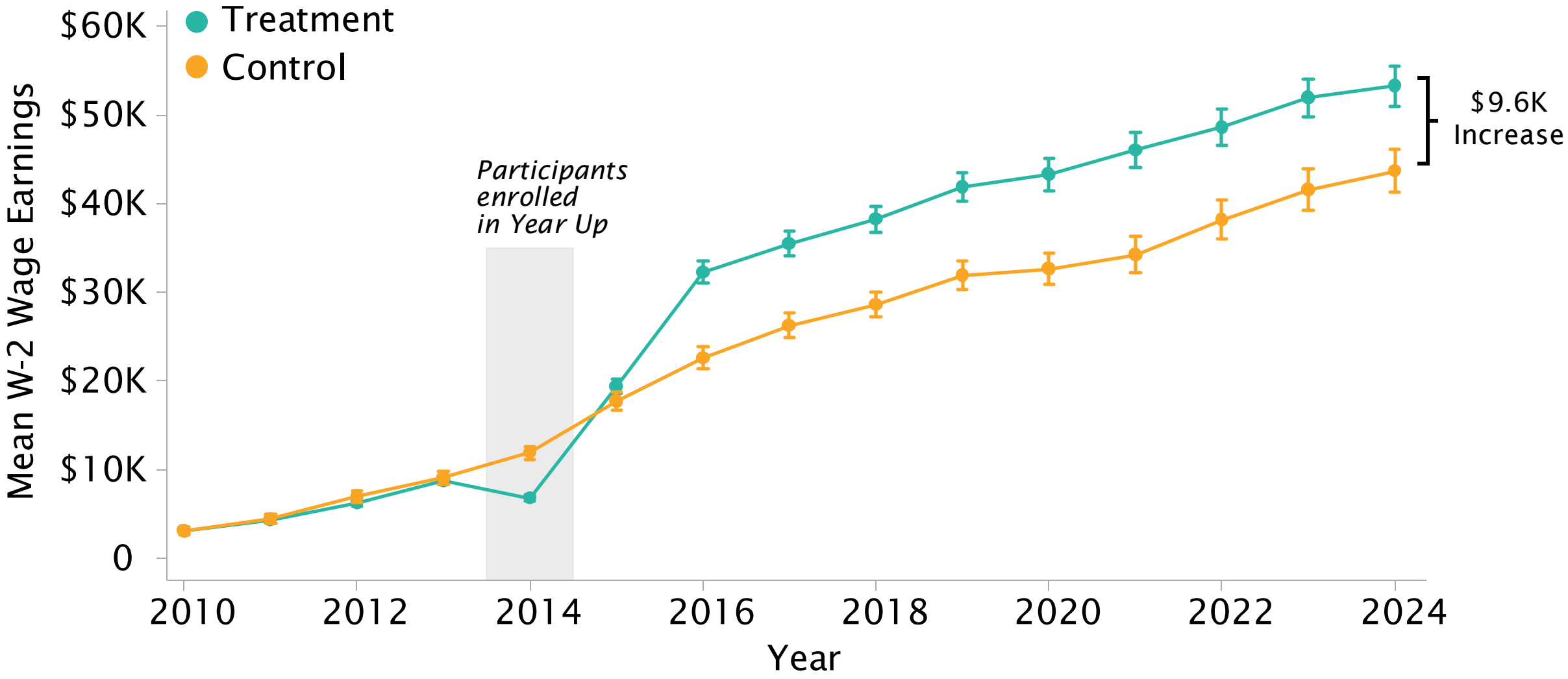
Impact of Year Up United on Earnings

Evidence from a PACE Randomized Trial in 2014



Impact of Year Up United on Earnings

Evidence from a PACE Randomized Trial in 2014



Observational Estimators

- Ideally, would run an experiment like we did for Year Up United for every one of the 5000+ programs in every year

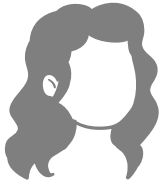
- As a feasible alternative, try to replicate experimental results using observational data that we already have, without relying on randomization

Constructing Digital Twins Using Machine Learning

Year Up Participants

A 

B 

C 

D 

E 

Constructing Digital Twins Using Machine Learning

Year Up United Participants

A



Demographics	Age 23 Male Single 1 child
Neighborhood	North End
Employment	
2018	$20K (McDonalds)
2019	$17K (Walmart)
2020	$18K (McDonalds)
2021	$22K (Walmart)
2022	$21K (Walmart)

270 Million American Adults

Anonymized Tax + Census Records



Twins are non-participants who:

Live in **North End**
Age 23
Have 1 child
Worked at McDonalds and Walmart
Earned **$20K, $17K, $18K, $22K, $21K**

Constructing Digital Twins Using Machine Learning

Year Up United Participants

A



Demographics	Age 23 Male Single 1 child
Neighborhood	North End
Employment	
2018	$20K (McDonalds)
2019	$17K (Walmart)
2020	$18K (McDonalds)
2021	$22K (Walmart)
2022	$21K (Walmart)

270 Million American Adults

Anonymized Tax + Census Records



Near Twins: non-participants who

Live in **North End or Frog Hollow**
Are between ages **22-26**
Have 1 child
Worked at **McDonalds or Target**
Earned between **$16K-$22K** in 2018-22

Constructing Digital Twins Using Machine Learning

Year Up United Participants

A



Demographics	Age 23 Male Single 1 child
Neighborhood	North End
Employment	
2018	$20K (McDonalds)
2019	$17K (Walmart)
2020	$18K (McDonalds)
2021	$22K (Walmart)
2022	$21K (Walmart)

270 Million American Adults

Anonymized Tax + Census Records



Near Twins: non-participants who

Live in **North End or Frog Hollow**
Are between ages **22-26**
Have 1 child
Worked at **McDonalds or Target**
Earned between **$16K-$22K** in 2018-22

Constructing Digital Twins Using Machine Learning

Year Up Participants

A



Demographics	Age 23 Male Single 1 child
Neighborhood	North End
Employment	
2018	$20K (McDonalds)
2019	$17K (Walmart)
2020	$18K (McDonalds)
2021	$22K (Walmart)
2022	$21K (Walmart)

270 Million American Adults

Anonymized Tax + Census Records



Synthetic Digital Twin



Near Twins: non-participants who

Live in **North End or Frog Hollow**
Are between ages **22-26**
Have 1 child
Worked at **McDonalds or Target**
Earned between **$16K-$22K** in 2018-22

Predict **counterfactual future earnings** using machine learning model (Gradient Boosted Tree)

Experimental Validation

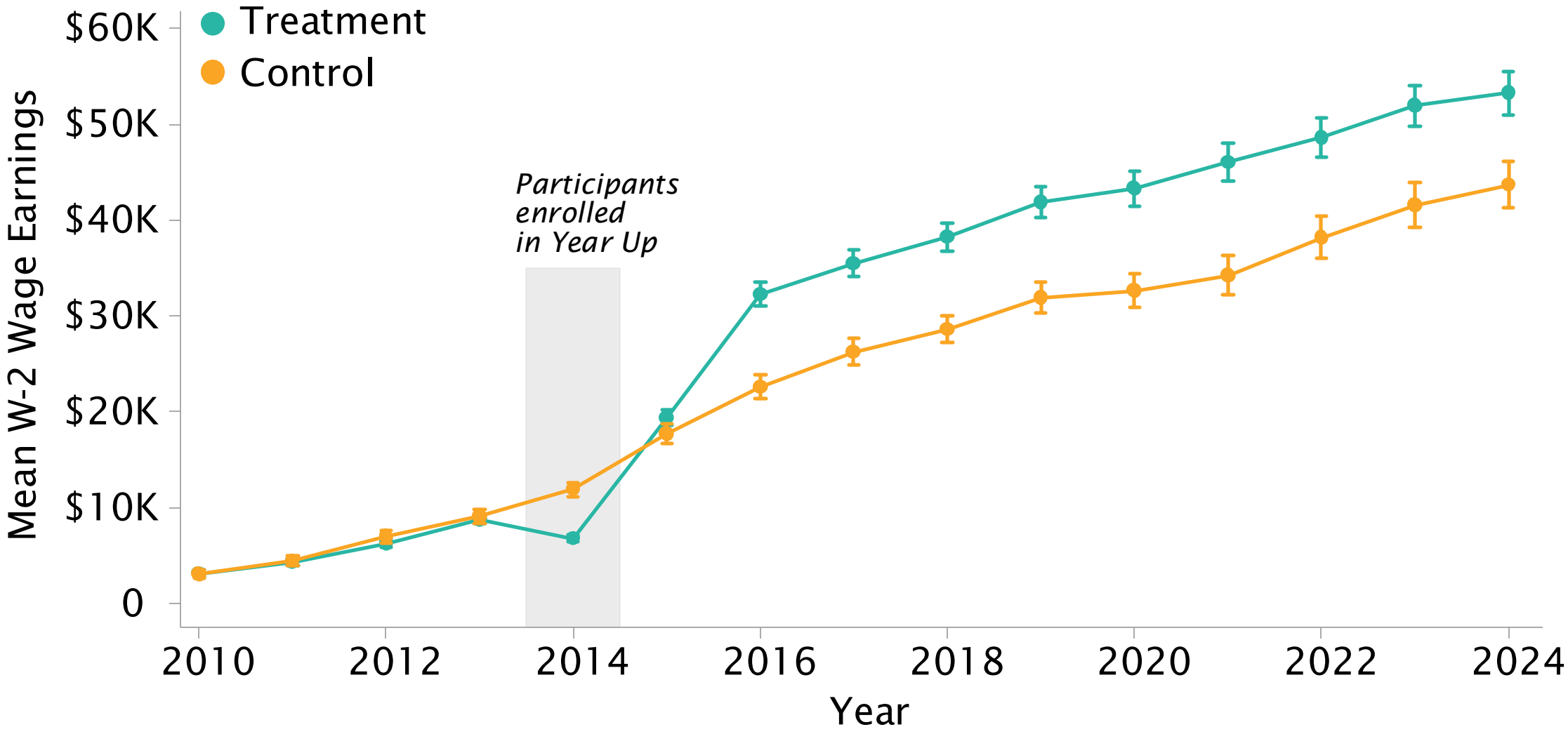


Evaluating Machine Learning Estimates: Comparisons to RCT Estimates

- No guarantee that machine learning model captures all differences between program participants and non-participants
  - Data are rich, but we still cannot measure some things that may matter, like motivation or grit or family circumstances
- Key test: validate estimates against ground truth from randomized experiments

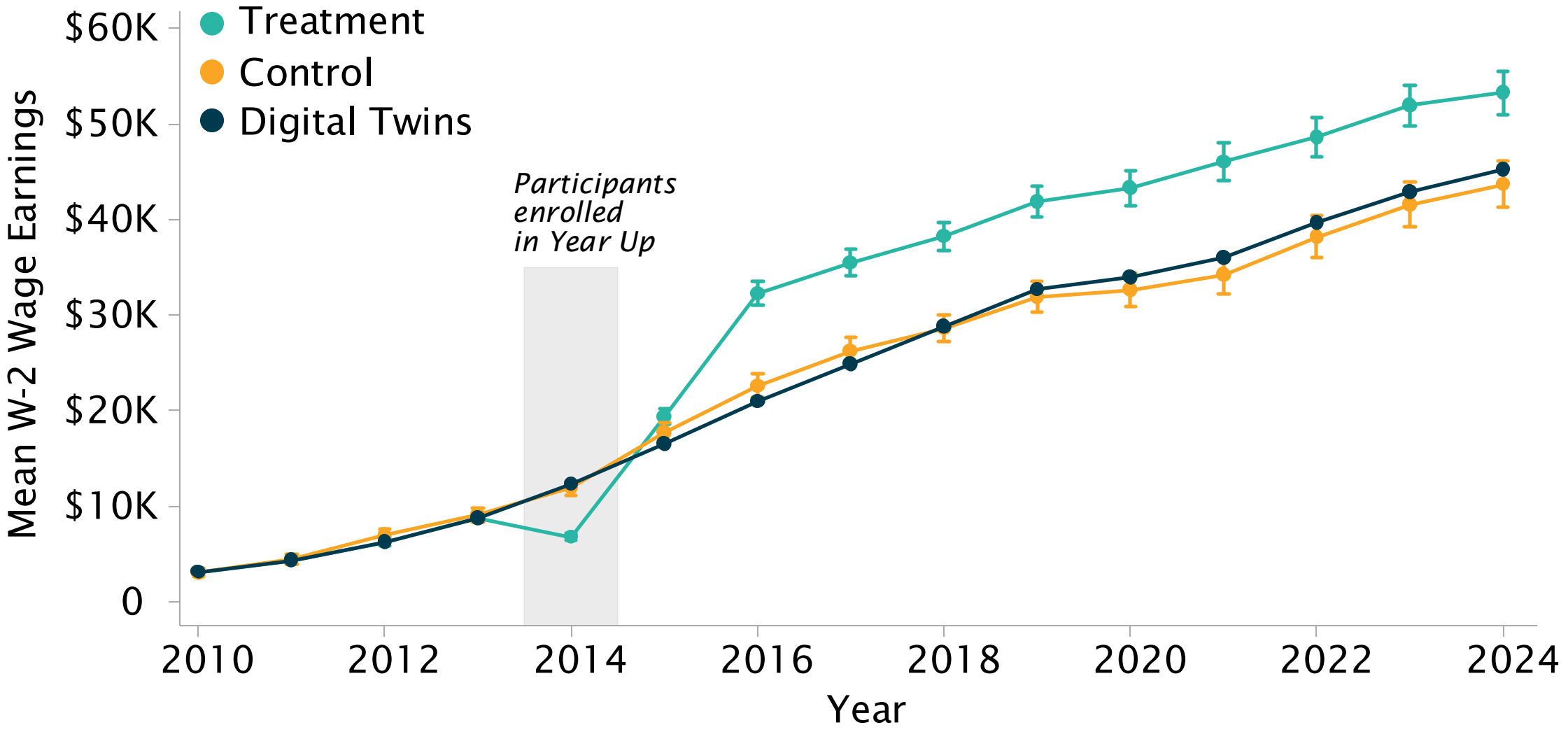
Impact of Year Up United on Earnings

Evidence from PACE Randomized Trial in 2014



Digital Twins Counterfactual Almost Identical to RCT Control

Evidence from PACE Randomized Trial in 2014

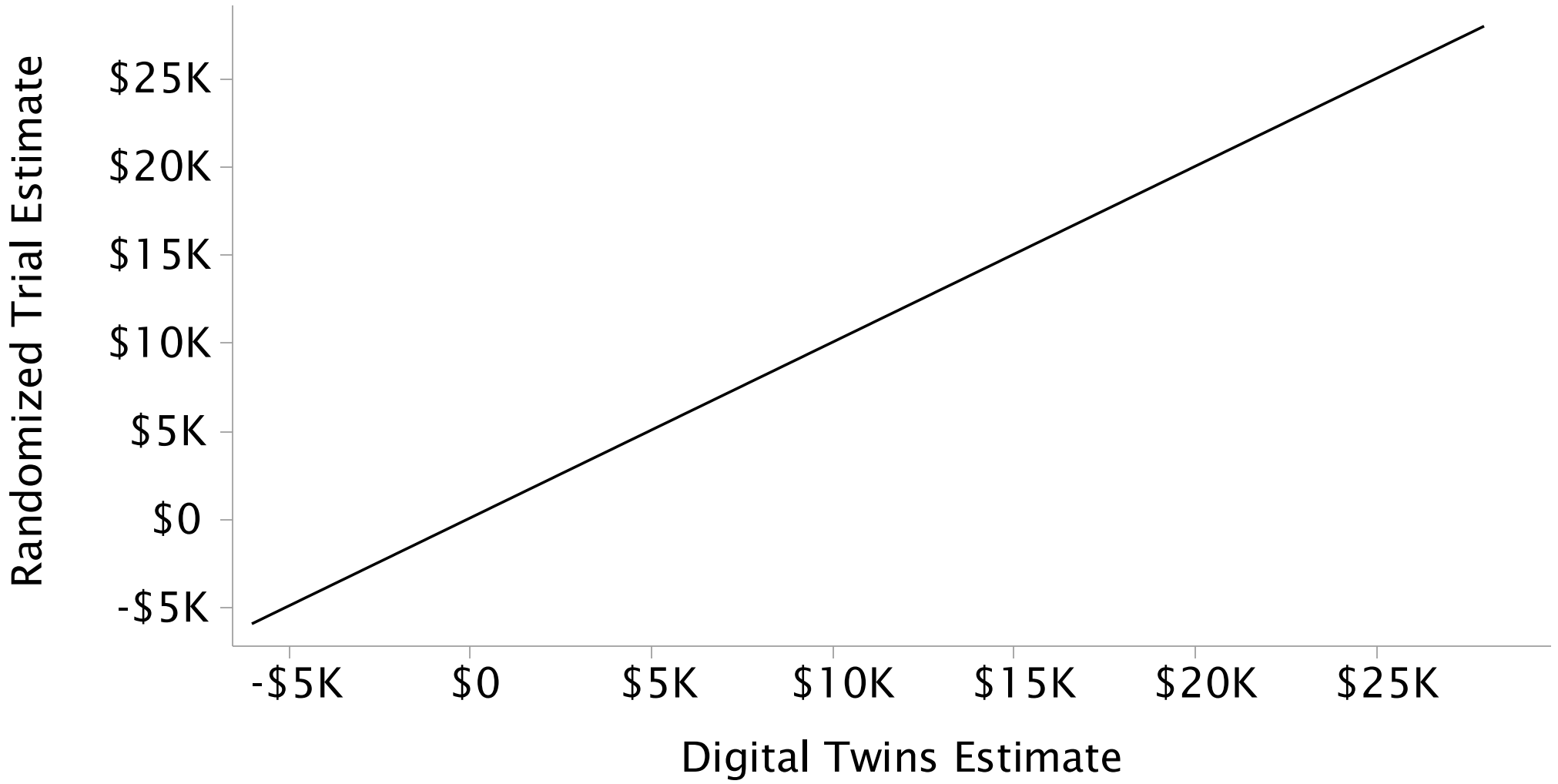


37 Randomized Trials of Job Training Programs Across America



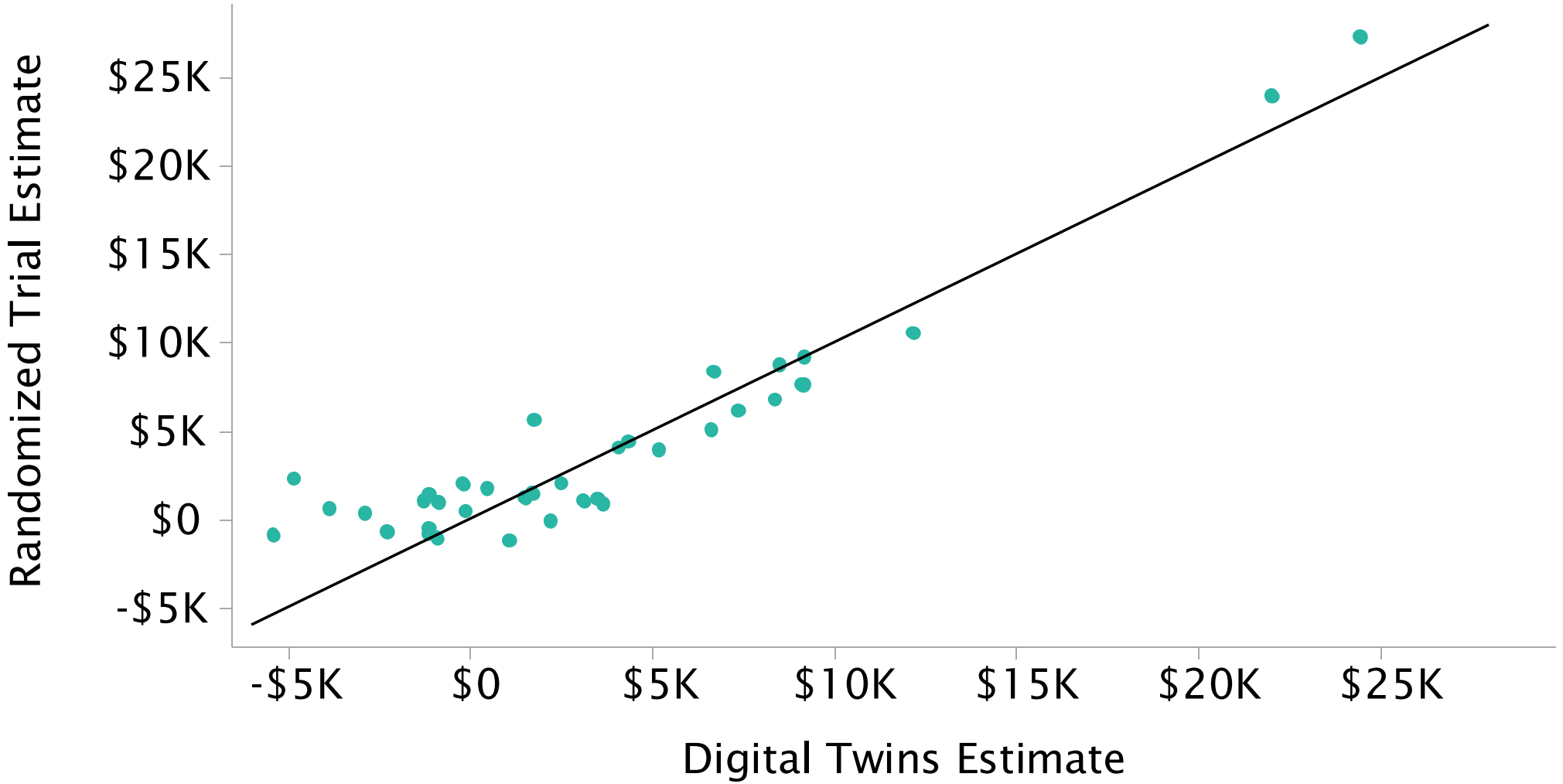
Effects of Workforce Training Programs on Earnings

Observational vs. Randomized Trial Estimates for Programs Implemented in 37 Sites



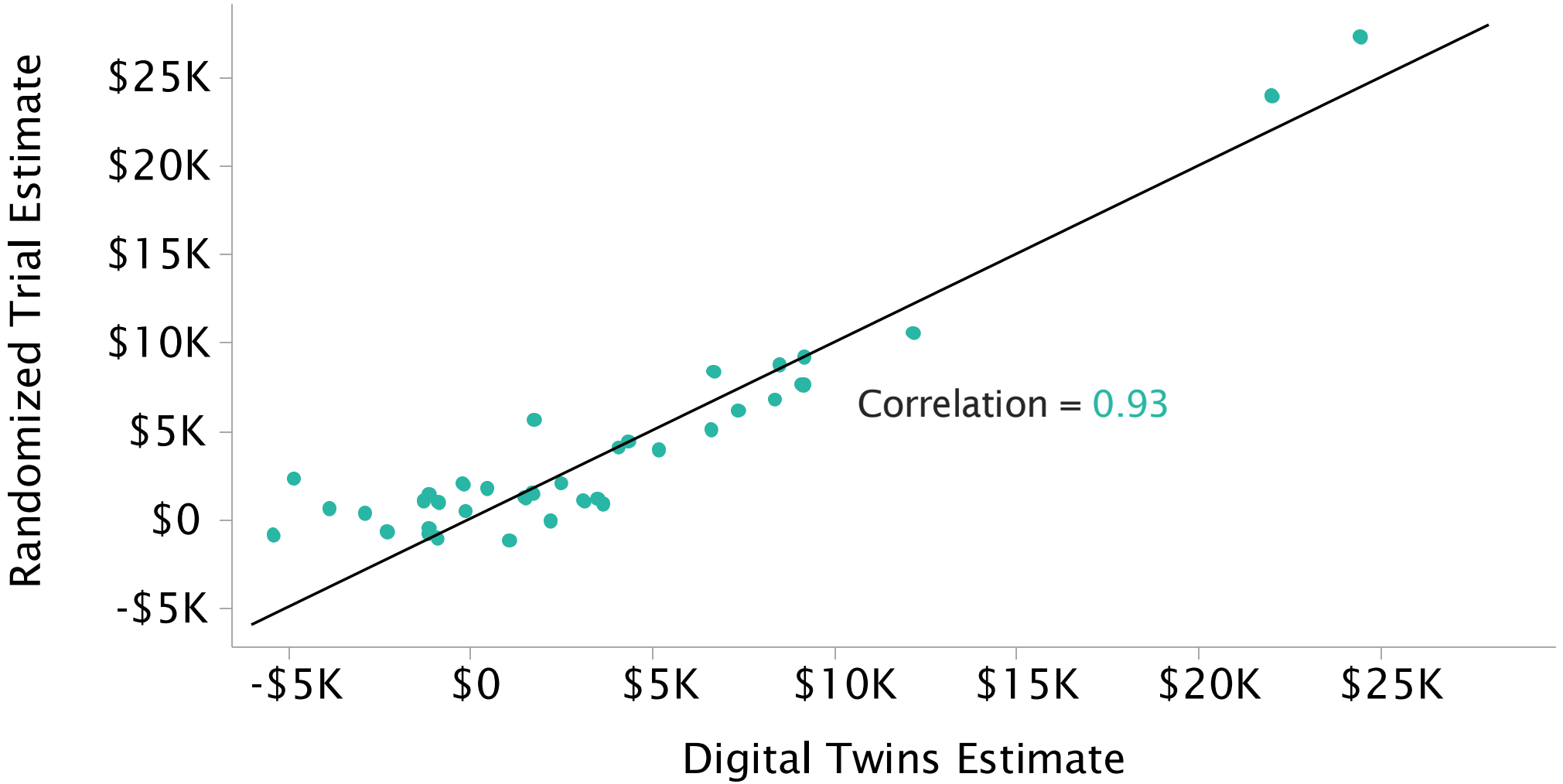
Effects of Workforce Training Programs on Earnings

Observational vs. Randomized Trial Estimates for Programs Implemented in 37 Sites



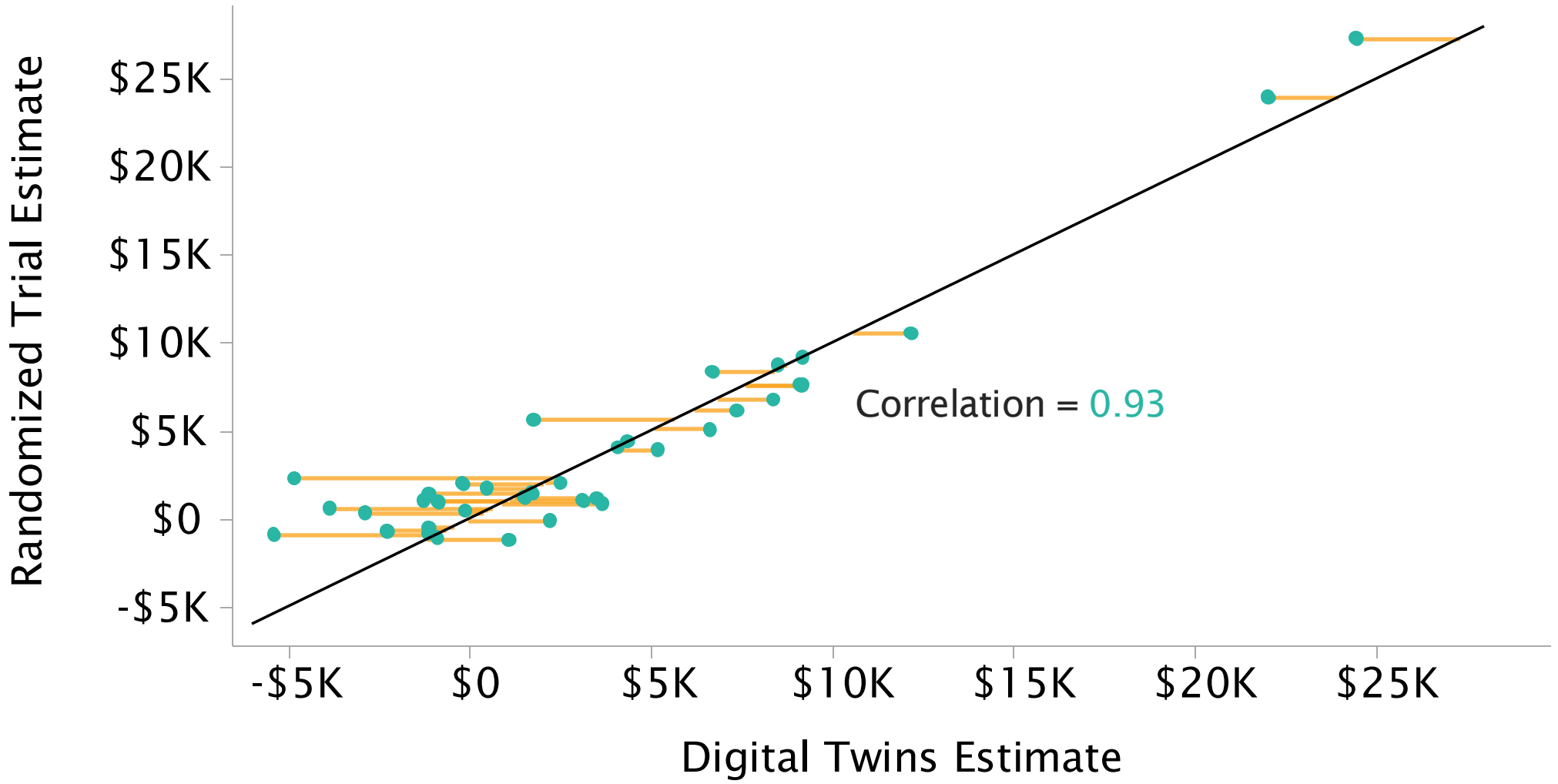
Effects of Workforce Training Programs on Earnings

Observational vs. Randomized Trial Estimates for Programs Implemented in 37 Sites



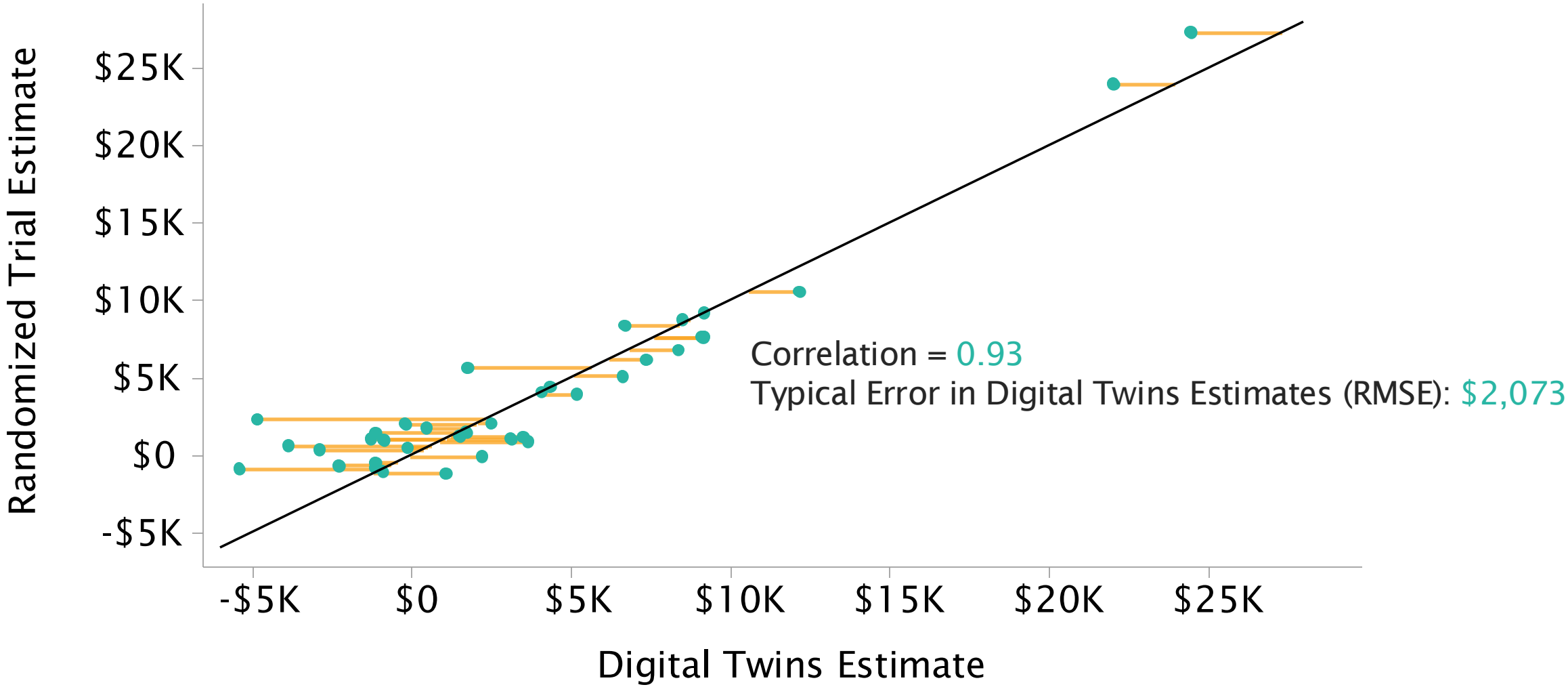
Effects of Workforce Training Programs on Earnings

Observational vs. Randomized Trial Estimates for Programs Implemented in 37 Sites



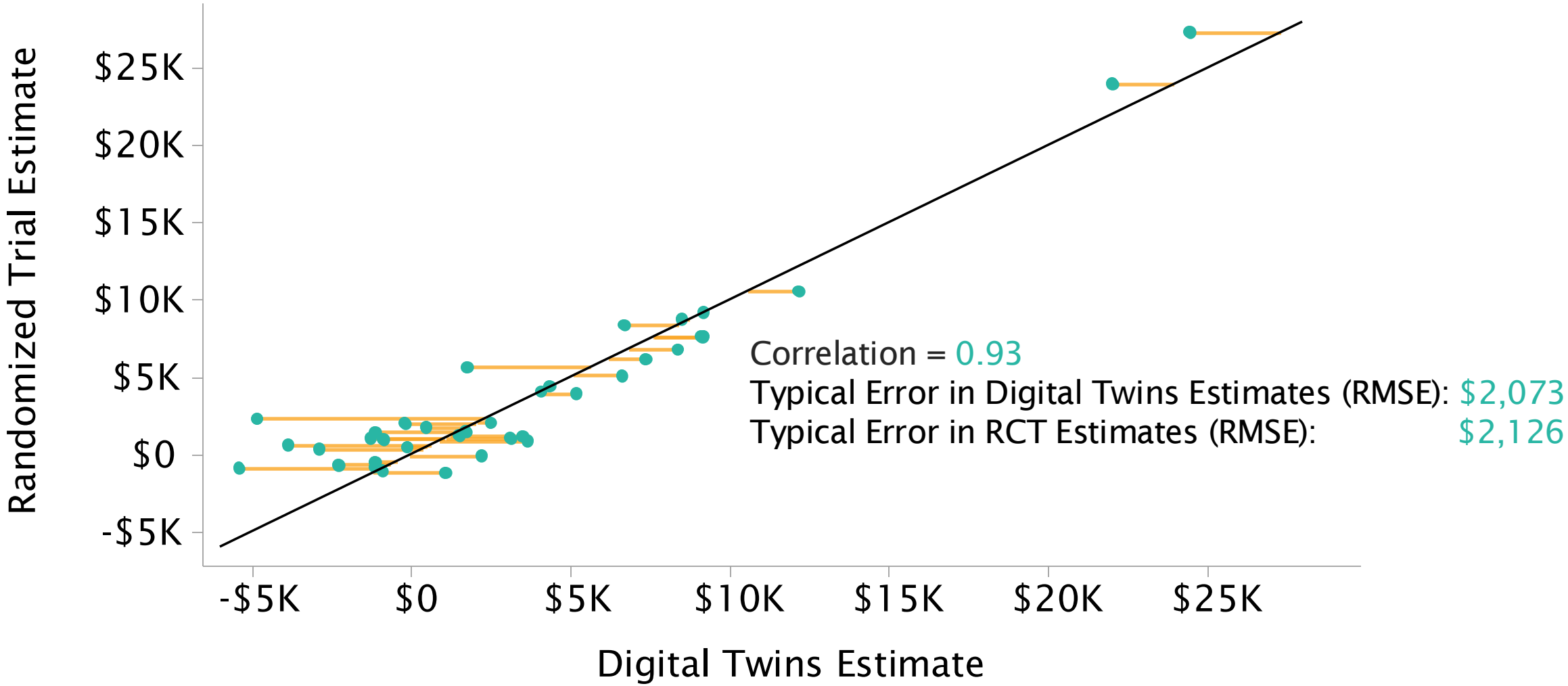
Effects of Workforce Training Programs on Earnings

Observational vs. Randomized Trial Estimates for Programs Implemented in 37 Sites



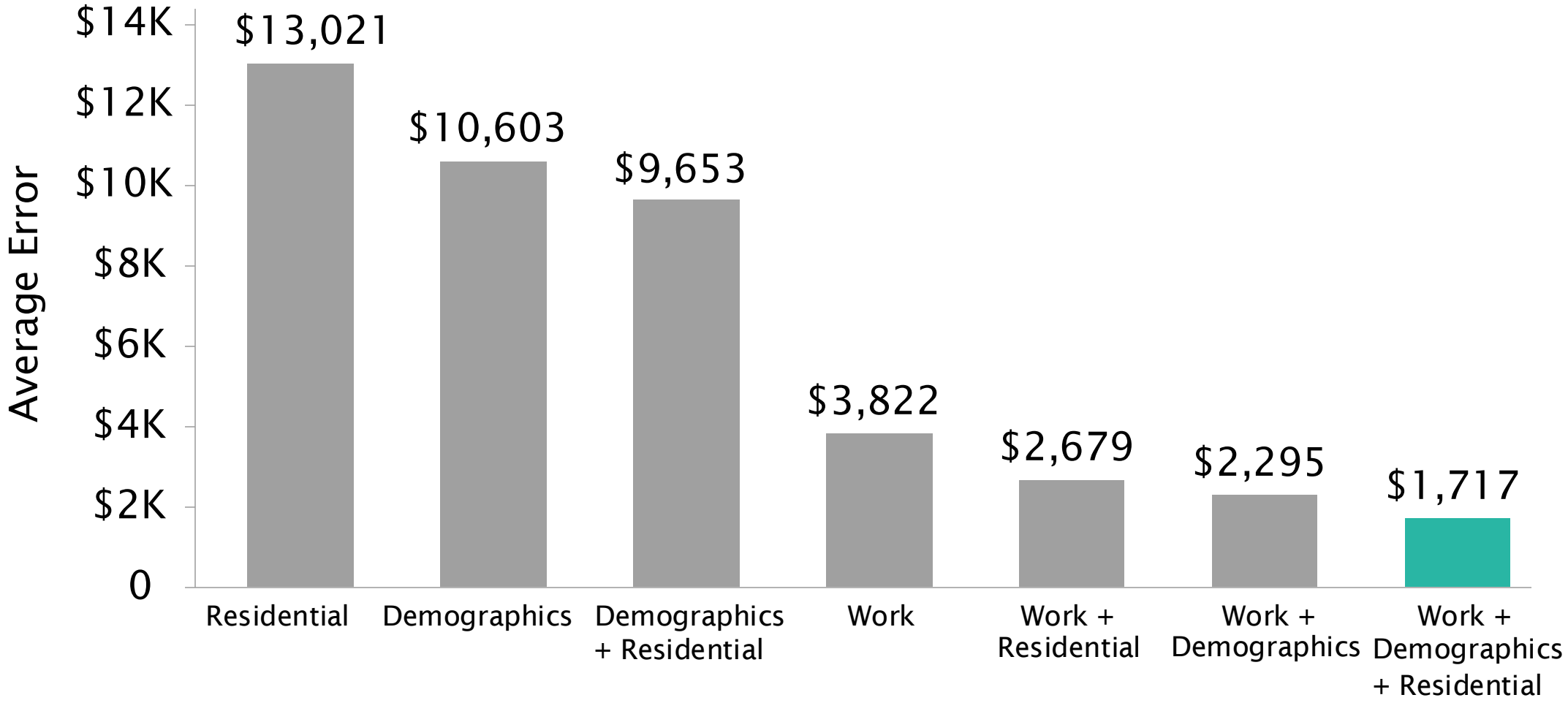
Effects of Workforce Training Programs on Earnings

Observational vs. Randomized Trial Estimates for Programs Implemented in 37 Sites



Noise-Adjusted RMSE by Covariates Used to Predict Earnings

Role of Work History, Demographics, and Residential History



Part 2: Using Digital Twins to Evaluate Programs at Scale

Applying Digital Twins Estimator to Evaluate Other Programs

- Now apply digital twins estimator to programs where we do not have experimental estimates

- Key assumption: new programs/participants are sufficiently similar to programs in validation set

Two Applications of Digital Twins Method

1. Other sites and cohorts for programs for which we have one experimental estimate already

2. Other programs for which we have no experimental estimate

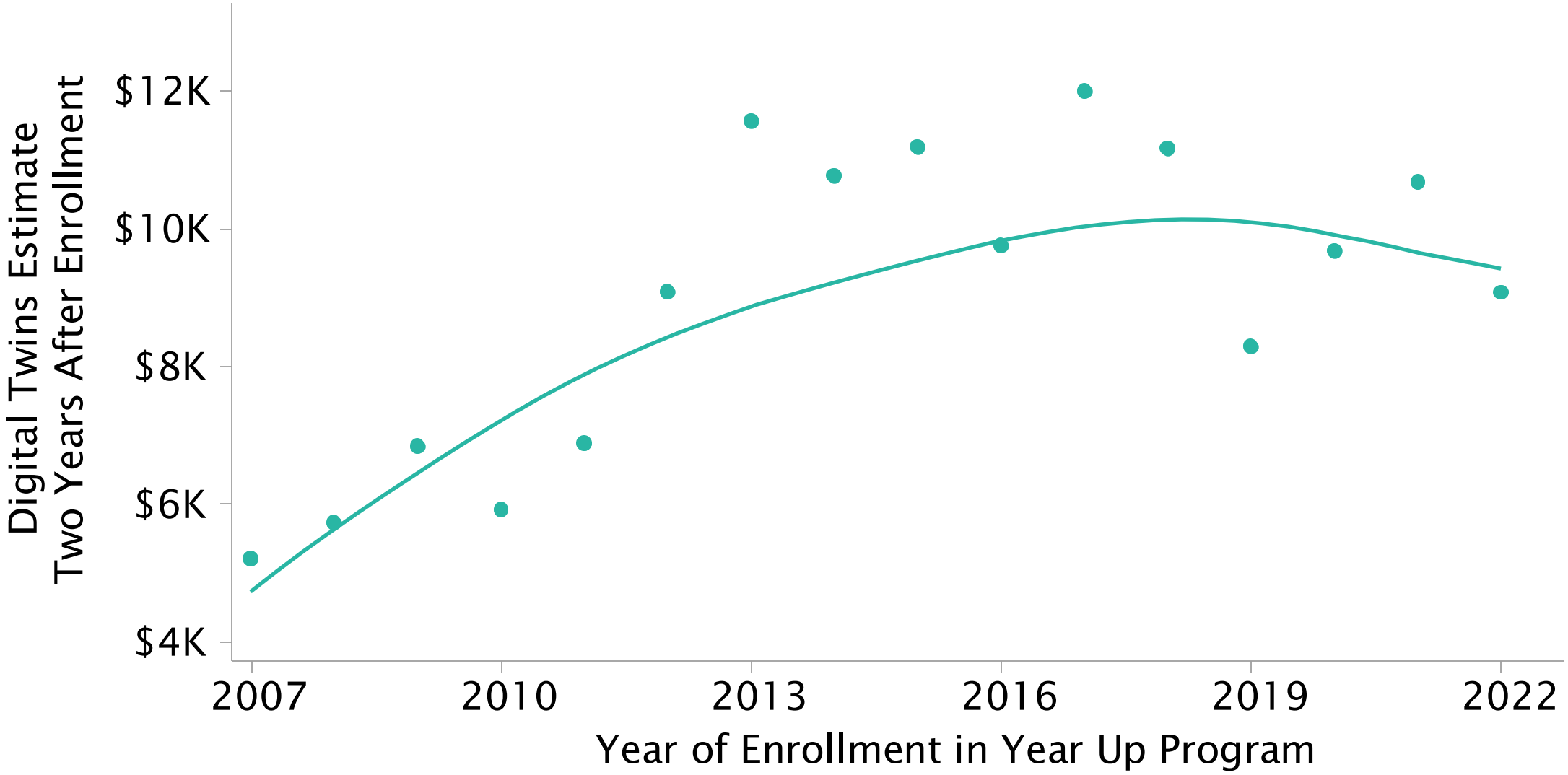
Two Applications of Digital Twins Method

1. Other sites and cohorts for programs for which we have one experimental estimate already

2. Other programs for which we have no experimental estimate

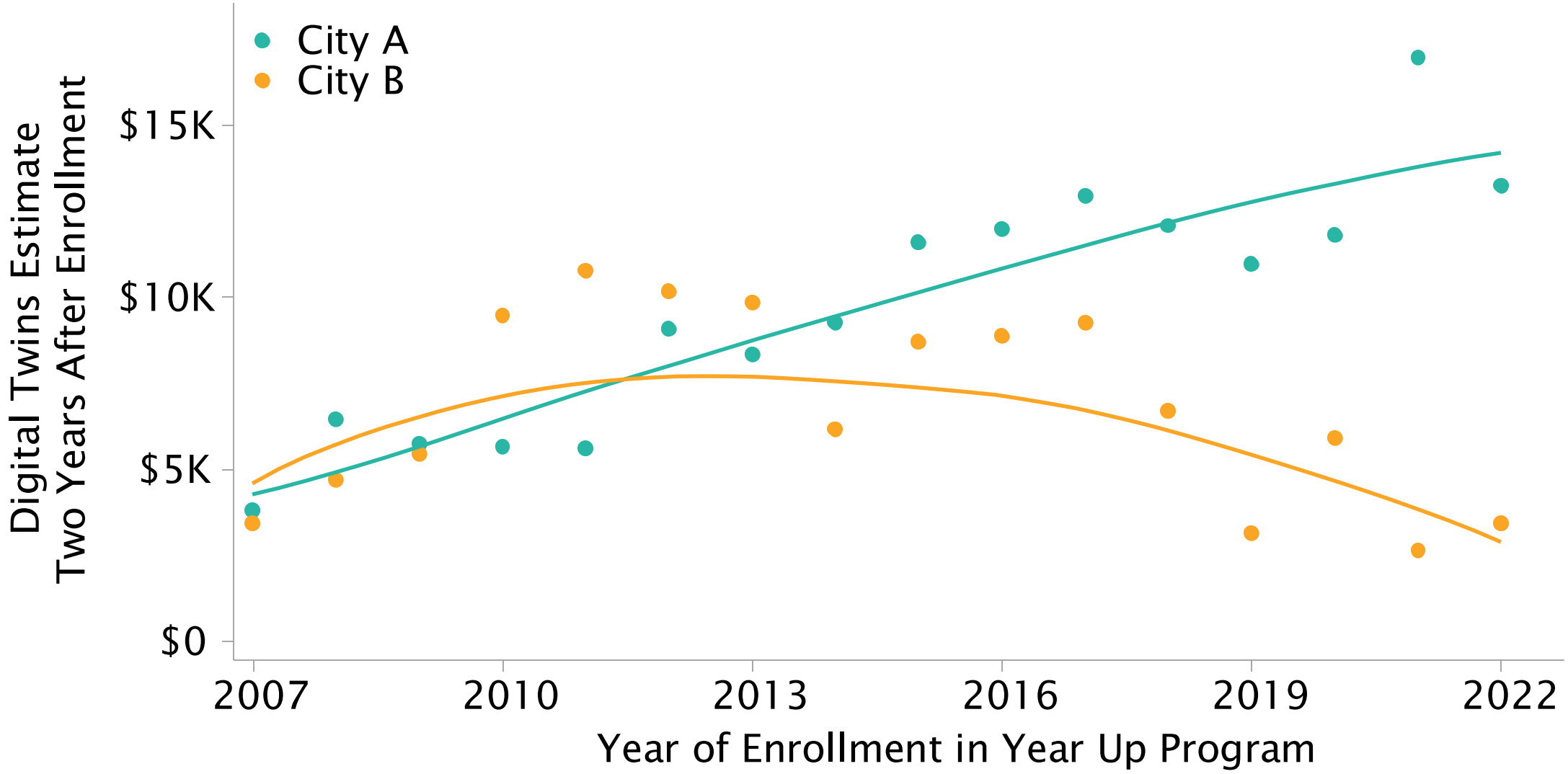
Using Digital Twins for Real-Time Outcome Measurement

Impacts of Year Up United Program on Earnings, By Cohort of Program Enrollment



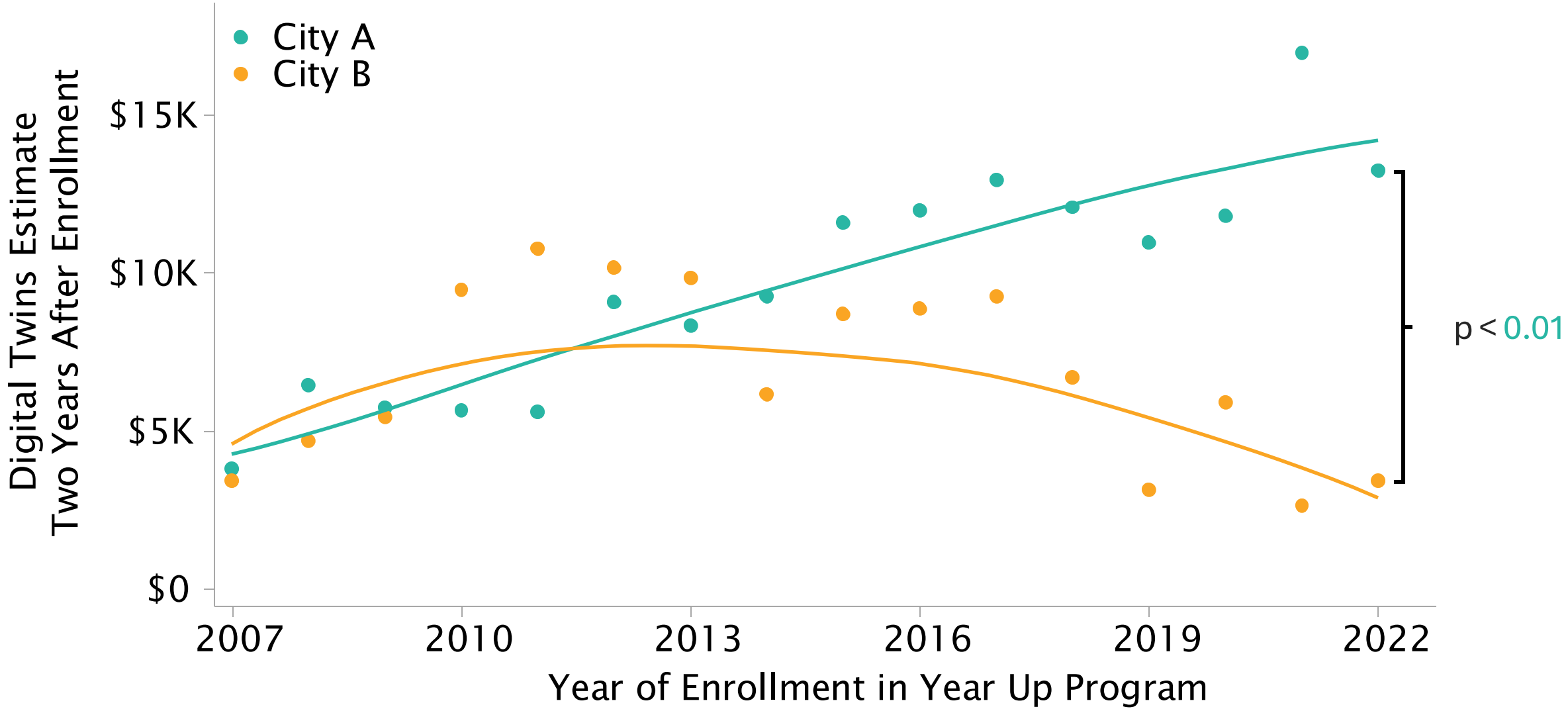
Using Digital Twins for Real-Time Outcome Measurement

Impacts of Year Up United Program on Earnings, By Site and Cohort of Program Enrollment



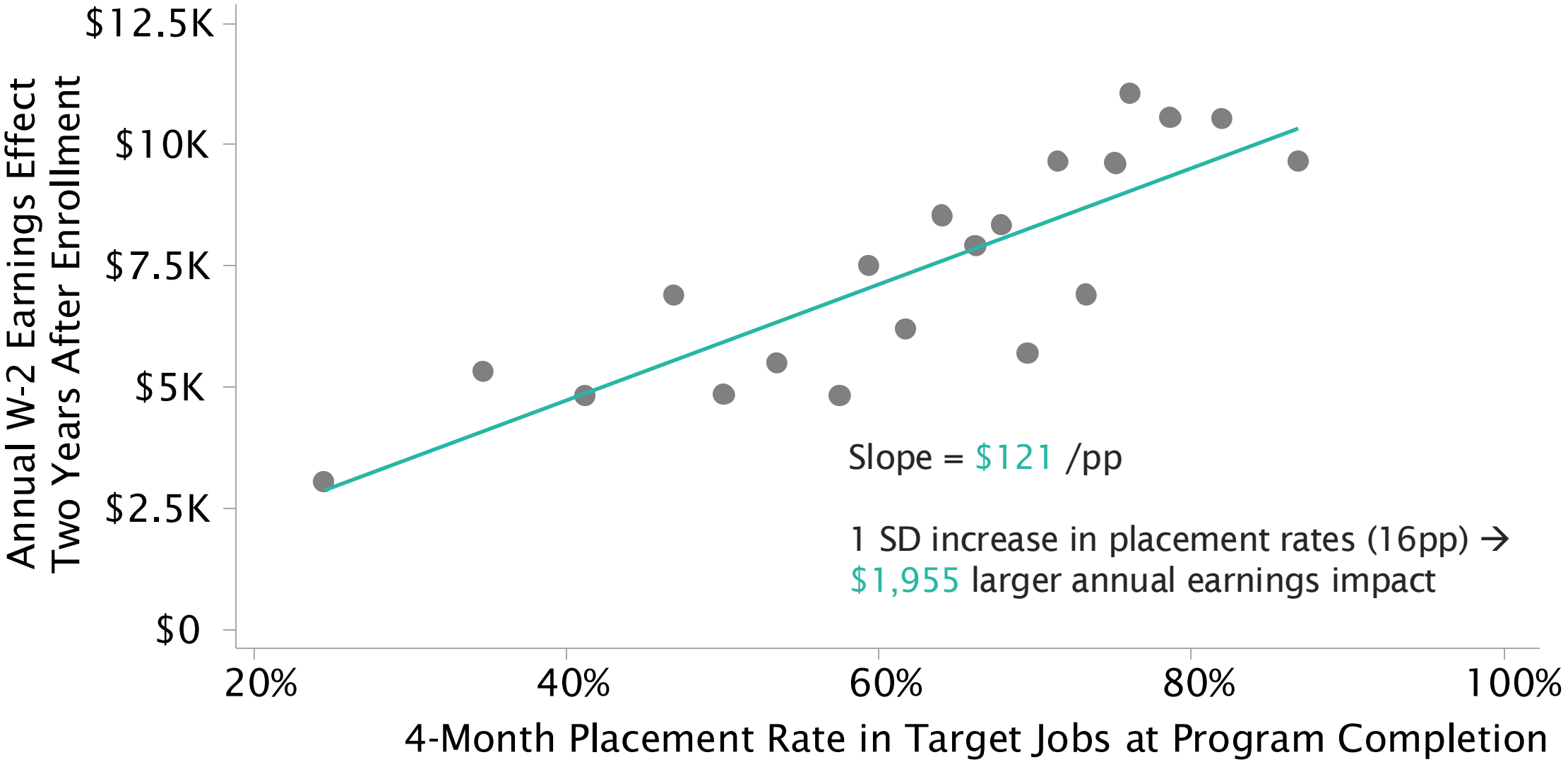
Using Digital Twins for Real-Time Outcome Measurement

Impacts of Year Up United Program on Earnings, By Site and Cohort of Program Enrollment



Impacts of Year Up United on Earnings vs. Job Placement Rates

By Site and Year of Enrollment: 271 Site-By-Enrollment-Year Cells



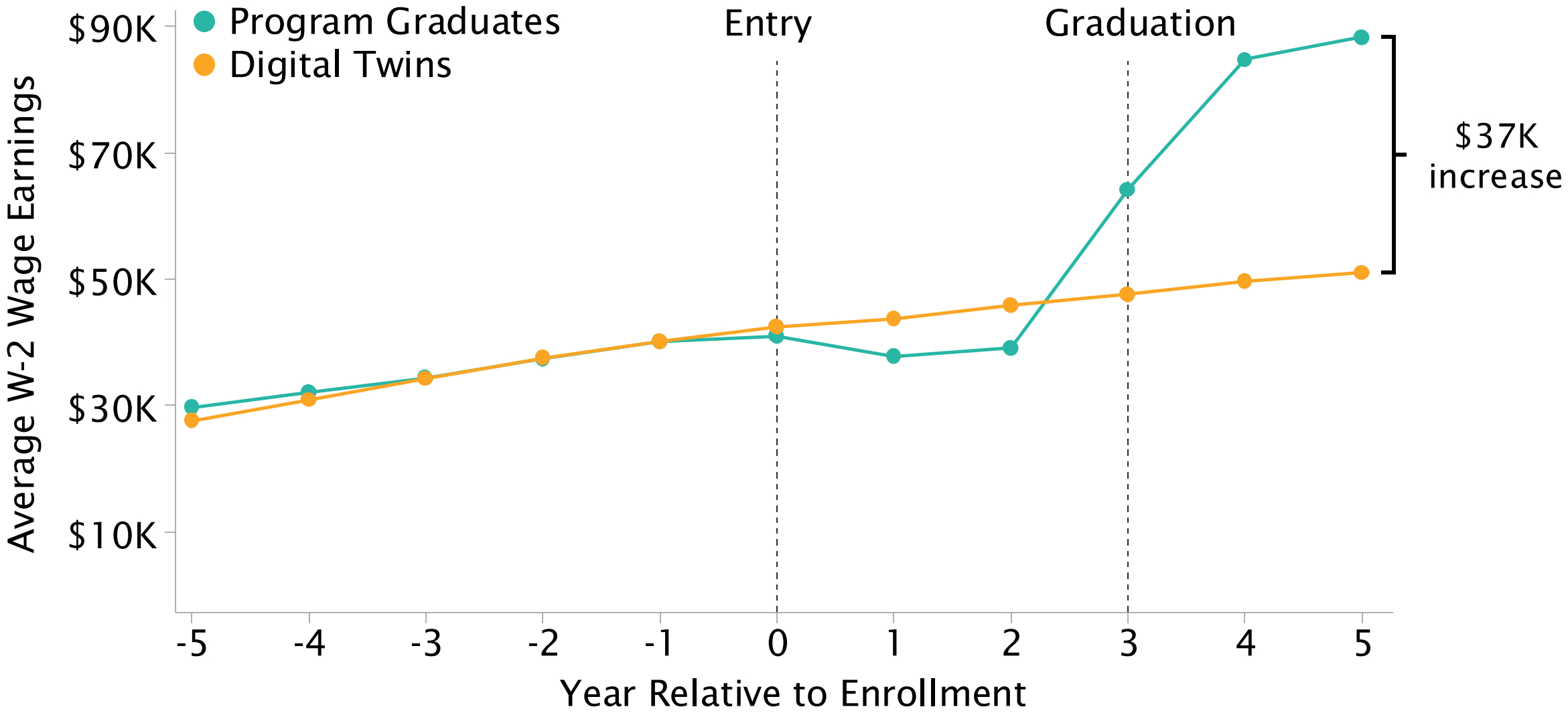
Two Applications of Digital Twins Method

1. Other sites and cohorts for programs for which we have one experimental estimate already

2. Other programs for which we have no experimental estimate

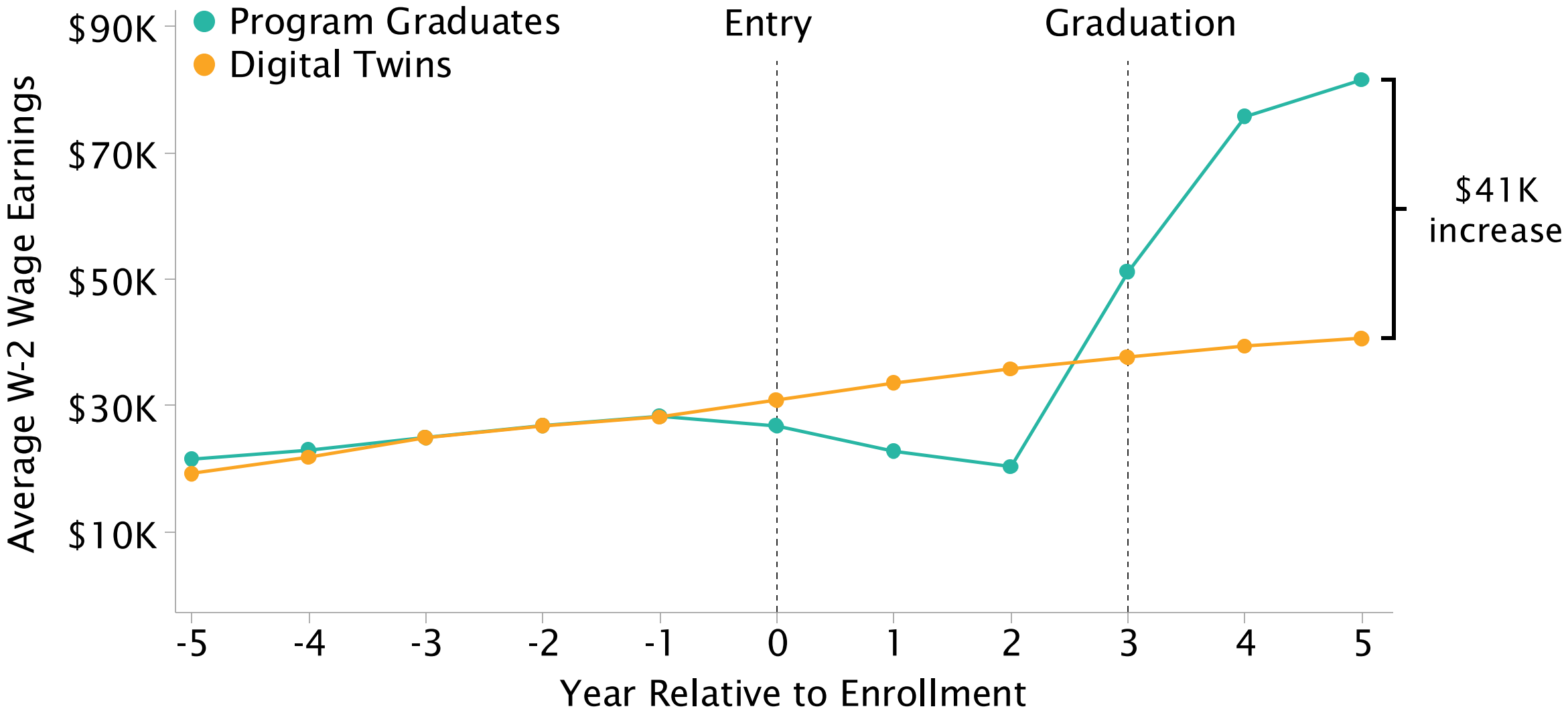
Impact of WGU Pre-Licensure Nursing Program

WGU Enrollees who Graduate 3 Years After Entry vs. Digital Twins



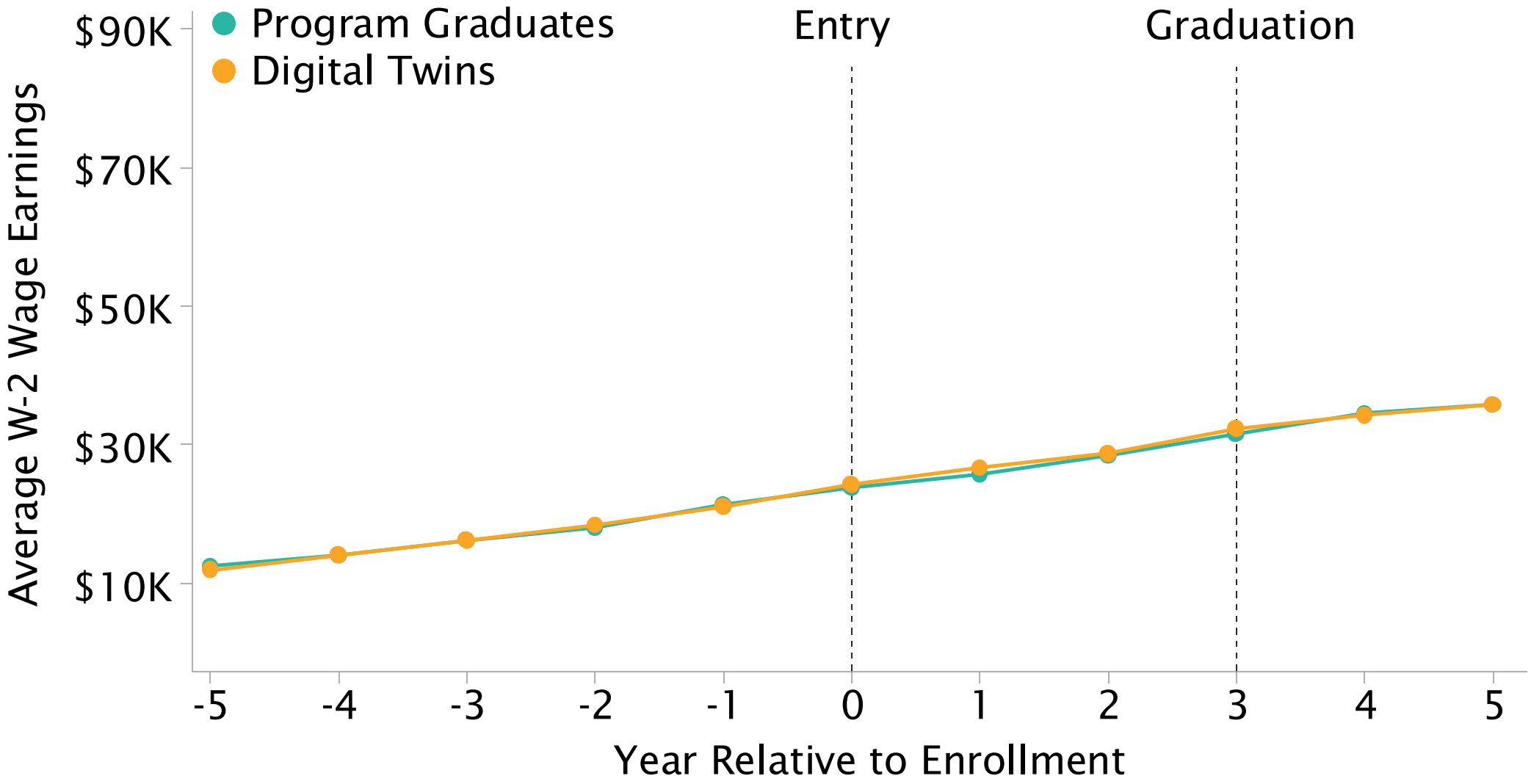
Impact of Dallas College Nursing Associate Degree

Dallas College Enrollees who Graduate 3 Years After Entry vs. Digital Twins



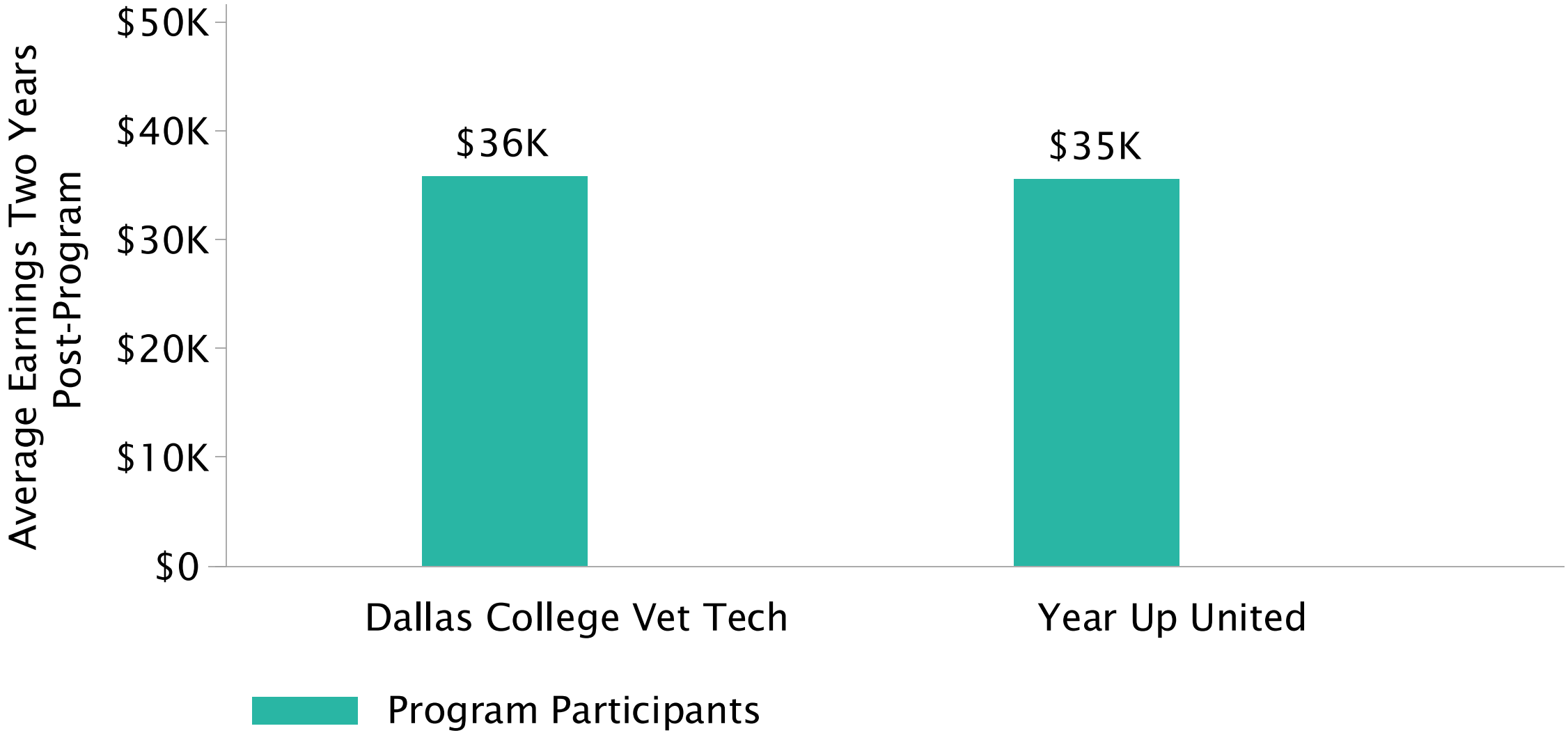
Impact of Dallas College Veterinary Technician Program

Dallas College Enrollees who Graduate 3 Years After Entry vs. Digital Twins



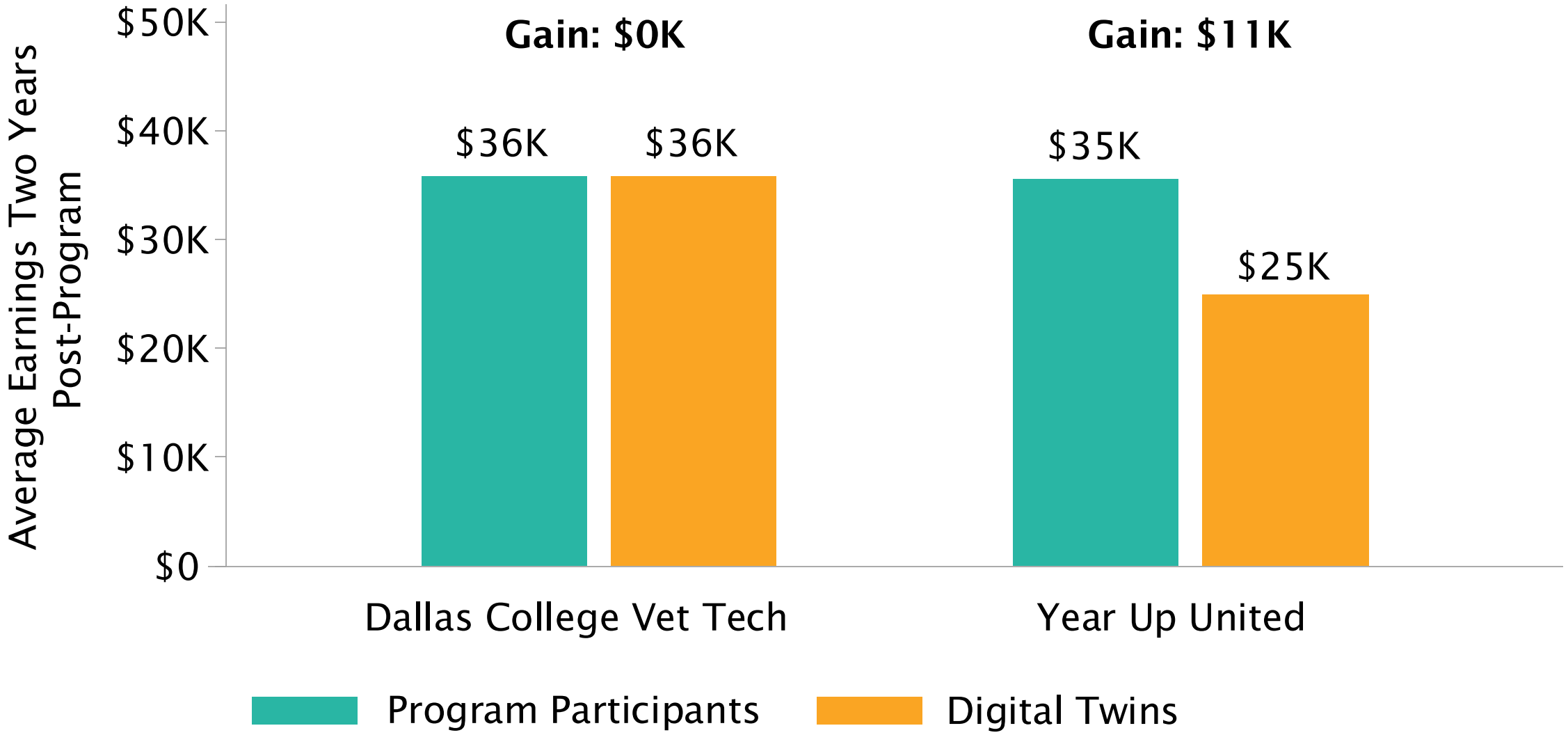
Importance of Measuring Earnings Gains Relative to Digital Twins

Dallas College Vet Tech Program vs. Year Up United



Importance of Measuring Earnings Gains Relative to Digital Twins

Dallas College Vet Tech Program vs. Year Up United



Putting this Into Practice and Conclusion



Benefits of Having Estimates of Program Effects at Scale

1. Use larger set of treatment effect estimates to learn about mechanisms

2. Build tools to scale effective programs, measuring impact in real time (Opportunity Capital)

    - Some programs are so effective that they can pay for themselves through increases in income tax revenue

    - Opportunity Capital: allow investors to fund program expansions and realize a return on investment only if programs are effective

Piloting Opportunity Capital Investments

Federal Demonstrations



DOL+ED FIPSE **$24M** pilot applications: Year Up, Per Scholas, MA DHE

Treasury SIPPRA ~**$20M** MA+TX pilot in development

State Demonstrations



Utah state bill passed in 2026 for **$5M** pilot

Exploring interest in other states

Philanthropic Investment



Philanthropic investment in outcome-based payments

Blue Meridian+Surgo **$15M** match to federal funds

Pilots implemented in partnership with Social Finance

Social Return on Private Investment



$10M

upfront investment
by private investors





$200M

gain in worker earnings
over 20 years

Conclusion

1. New methodology for determining what works in job training and related areas
    - Promote and support continuous improvement
    - Direct funds to effective programs
2. Build on Connecticut's data infrastructure
    - Digital twins method works well in 1% sample of U.S.; likely will work in specific states
    - Explore linking program enrollment records to state earnings and demographic data

Ways to learn more

1. Explore our data: https://www.opportunityatlas.org

2. Interested in collaborating/sharing your data? More info at this QR code

Thank you for attending:

Designing Student and Workforce Outcomes Through Data and AI

From Insight to Action

Please Fill out our Session Survey



2026 CT data Conference

#CTDataConf | ctdata.org