Connecting the dots:
Statistics, causality, and policy

Frank Yoon, Ph.D.
Statistician

MATHEMATICA
Policy Research

Brief outline

- **Statistics and causal inference** (I'll talk)
  - Background and review
  - Definitions and concepts
  - Formalize *statistical intuition*
- **Case studies** (We'll talk)
  - Charter schools, job training, health insurance
  - Weigh the evidence critically
  - Strengthen *statistical intuition*
- **A hypothetical policy study** (You'll talk)
  - Exercise *statistical intuition*

What is statistics?

- Statistics is the discipline that concerns itself with the study of the nature of *data*.
- Elementary concepts
  - Distributions
  - Averages, variation, correlation
- Associations are described by *statistical models*.
  - Are student outcomes better in charter schools?
  - Do trained job seekers have more success than others?
  - How do health costs change after coverage expansion?

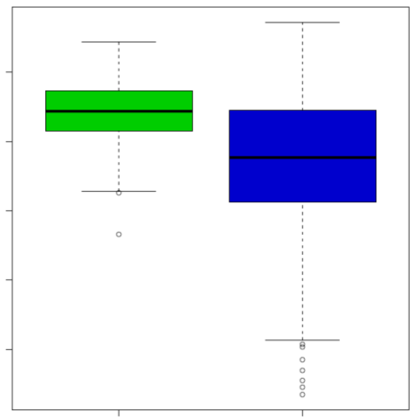
What is causality?

- Causation is the relationship between an event (*cause*) and a second event (*effect*), where the second event is understood as a consequence of the first.
- Inherently, the study causality concerns itself with logic.
- Causality is described by *causal statements*.
  - Do charter schools improve student outcomes?
  - Does job training benefit job seekers?
  - How does coverage expansion impact costs?

Statistics and causal inference

- "Correlation is not causation"

Statistical	**Causal**
Are student outcomes better in charter schools?	Do charter schools improve student outcomes?
Do trained job seekers have more success than others?	Does job training benefit job seekers?
How do health costs change after coverage expansion?	How does coverage expansion impact costs?



A → B

MATHEMATICA Policy Research

Statistics and causal inference

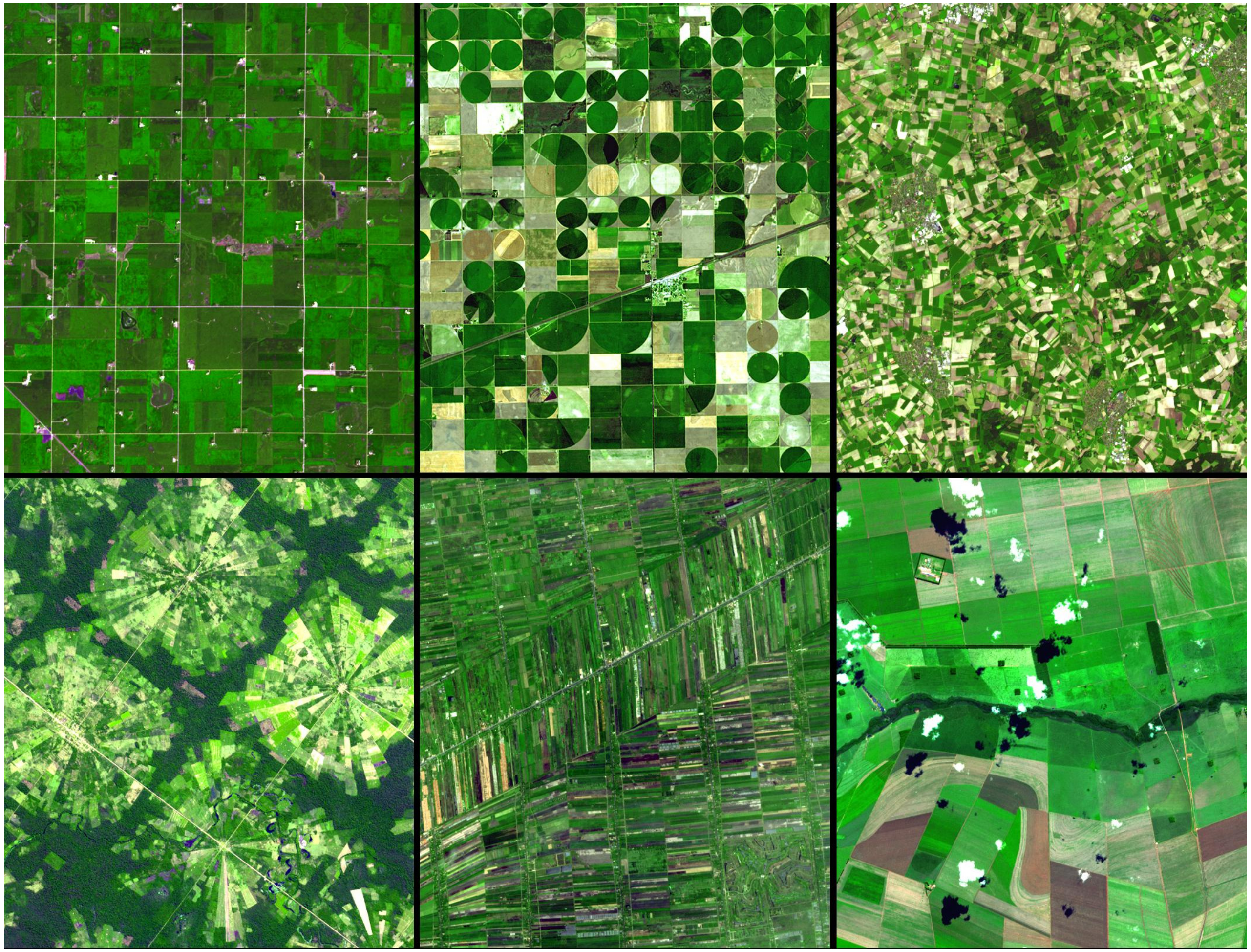


Photo credit: http://auracle.ca/news/

A simple study

A simple study

A simple study

A simple study

A simple study

A simple study

Causal framework

- **The Rubin Causal Model**
  - A study unit has two *potential outcomes.*
    - What would have happened under intervention
    - What would have happened under control
  - The causal effect is the difference between the two.
- **Fundamental Problem of Causal Inference**
  - We only get to see one outcome for a given unit.

Causal framework

- "No causation without manipulation"
- How are units assigned to intervention?
  - **Randomization**
    - Independence: the gold standard
    - Addresses all sources of bias
  - **Non-randomized assignment**
    - Decisions were made about intervention
    - Observed and unobserved biases

Study sample

MATHEMATICA Policy Research

Study sample

MATHEMATICA Policy Research

Randomized experiment

Randomized experiment

Observational study

Observational study

Observational study

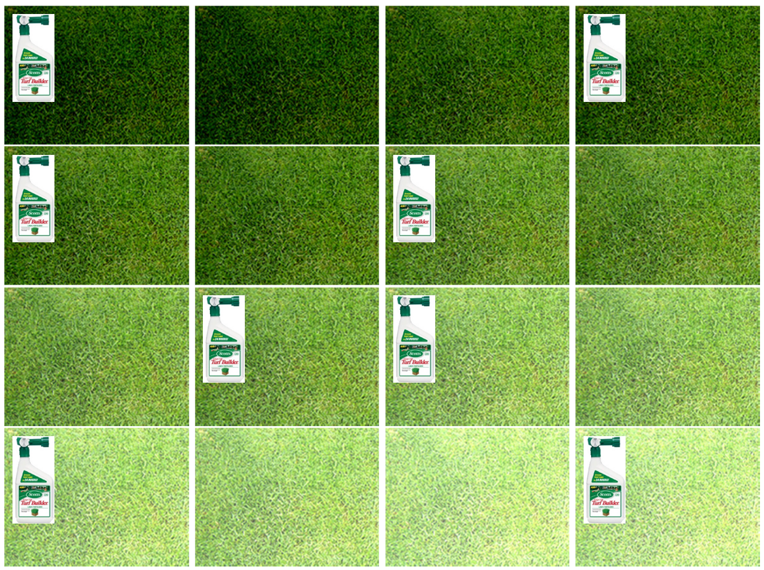
Observational study

Internal validity

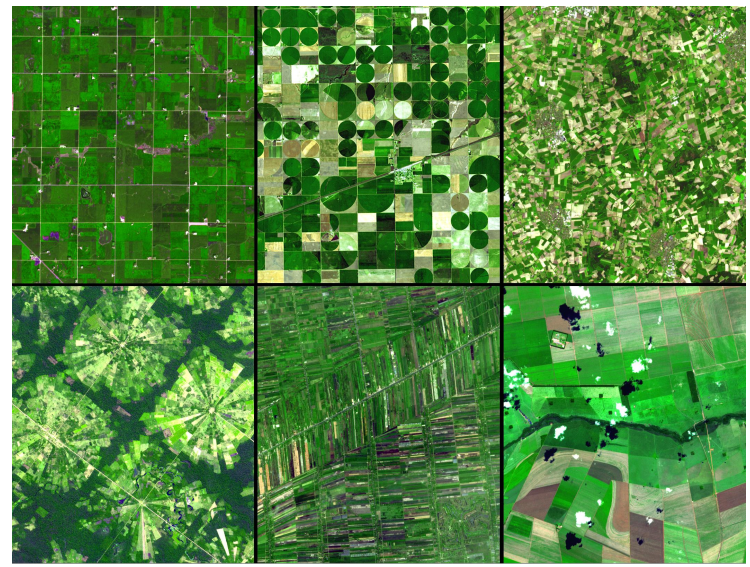
- Can we attribute the observed difference between the two groups to the intervention?
- Randomization guarantees internal validity.
- In an observational study adjustments are needed.
  - Assume no hidden biases
    - *Strongly ignorable treatment assignment*
    - **How big is this assumption?**
  - Remove observed biases
    - Regression
    - Matching

External validity

- What would happen outside study conditions?



versus



Bias matters

- Internal validity is needed to claim that the intervention works in the study.
- External validity is needed to claim that the intervention would work elsewhere.
- A good evidence base has
  - Lots of studies with internal validity…
  - Representing lots of real-world conditions.

What works for whom?

- The **average effect** is often reported.
  - But does it represent everyone?

MATHEMATICA Policy Research

What works for whom?

- **Heterogeneity**
  - In a study the impacts of intervention vary along the study units, whose characteristics matter.

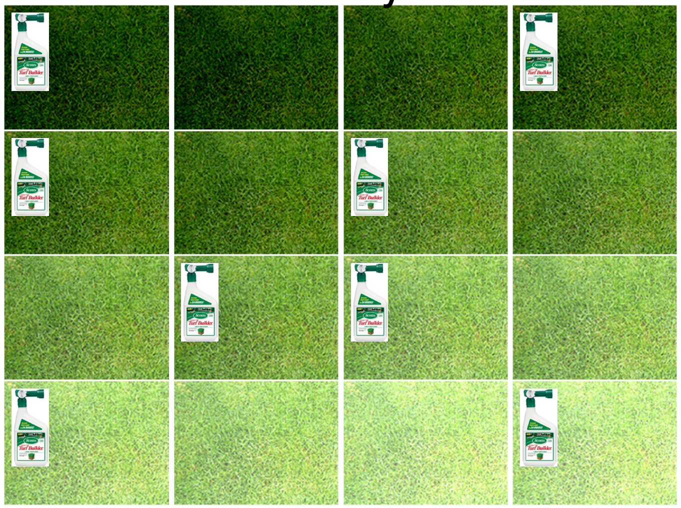
MATHEMATICA Policy Research

Weighing the evidence

- Internal validity is a must.
- External validity makes the result more useful.
- Usually, the average doesn't describe it all.

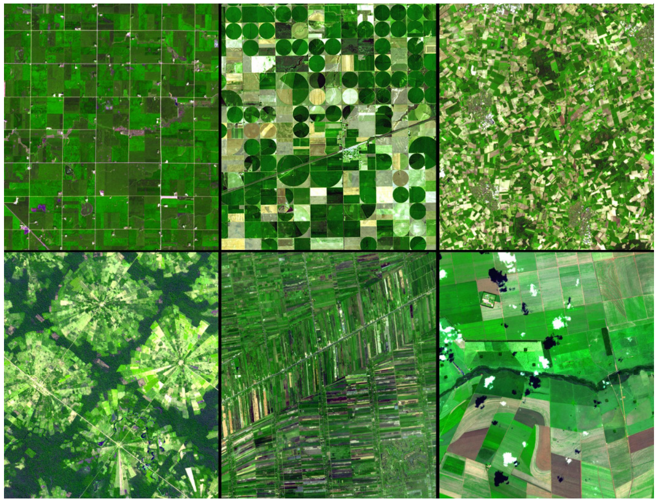
Weighing the evidence

Randomized study



Observational study

The real world



What is policy?

Case studies

- The effect of charter-school management organizations on student achievement
- The effect of voucher models on employment and earnings
- The effect of mental health parity on costs and service use

Charter-school management organizations

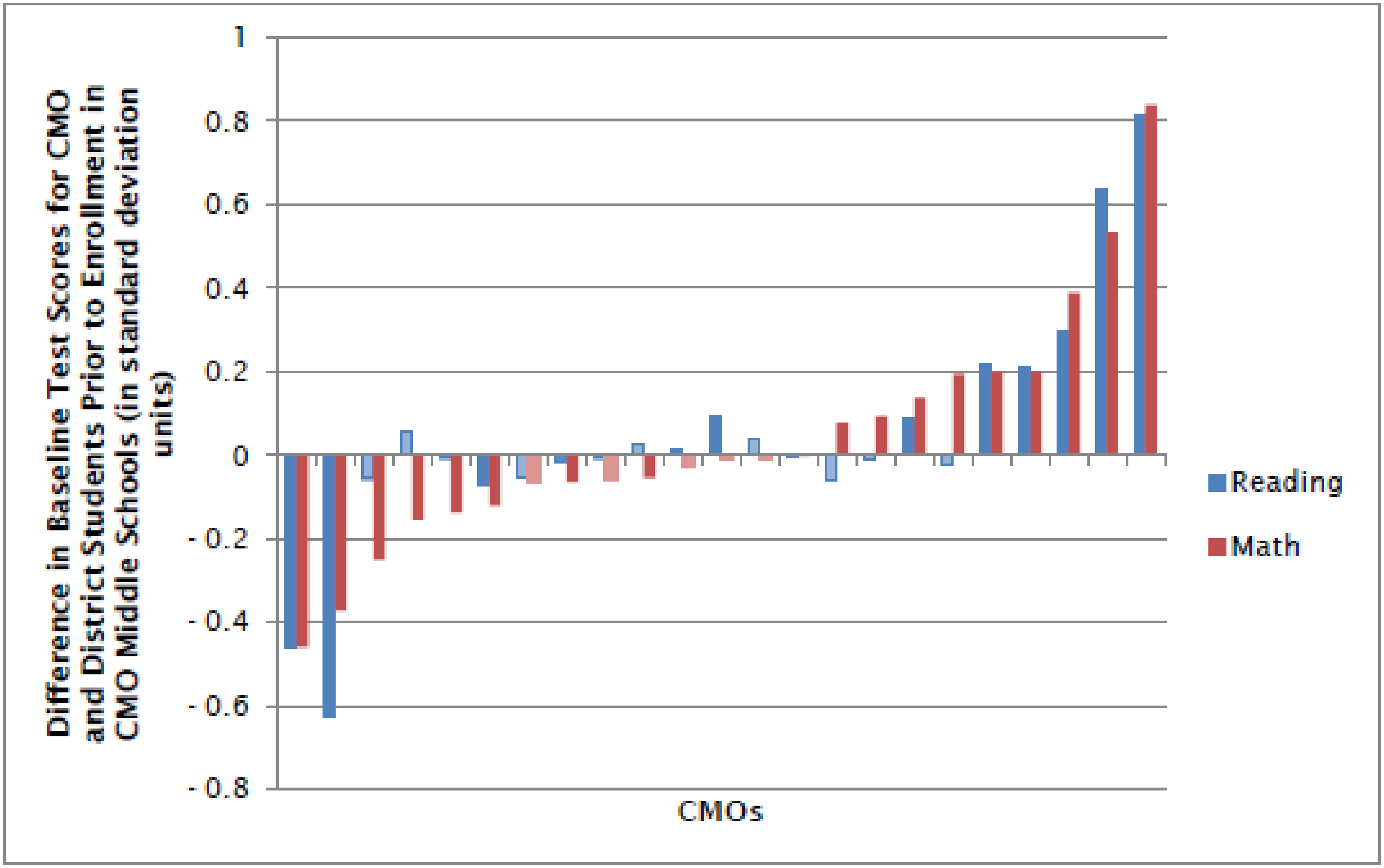
- Charter-school management organizations (CMOs) establish and operate multiple charter schools.
- The National Study of CMO Effectiveness
  - Middle-school outcomes
    - Mathematics and reading test scores after 1-3 years
    - Social studies and science test scores after 3 years
  - Data sources
    - Student-level administrative records
    - 2010 Survey of CMO and District Principals

Charter-school management organizations

- Intervention: students enrolled in CMO schools
  - Selected by lottery
- Control: students enrolled in the same public school district
  - Matched on: pre-entry test scores, race, gender, free/reduced-price lunch status

Charter-school management organizations

Figure II.13. Incoming Reading and Math Scores for CMO Students and Their District Peers



MATHEMATICA Policy Research

Charter-school management organizations

	Math		Reading	
CMO	CMO	Comp.	CMO	Comp.
N	-0.33	-0.27	-0.50	-0.45
C	0.00	-0.01	0.01	-0.01
H	0.44	0.36	0.52	0.46
E	0.43	0.46	0.34	0.34
D	-0.43	-0.37	-0.25	-0.22
Q	-0.24	-0.22	-0.17	-0.17
T	-0.03	0.01	-0.14	-0.09
L	0.11	0.23*	0.26	0.38*
P	-0.02	0.01	0.05	0.05
M	-0.02	0.00	0.02	0.03

MATHEMATICA
Policy Research

Charter-school management organizations

Figure IV.1. Distribution of Test- Score Effect Sizes After Two Years in Math

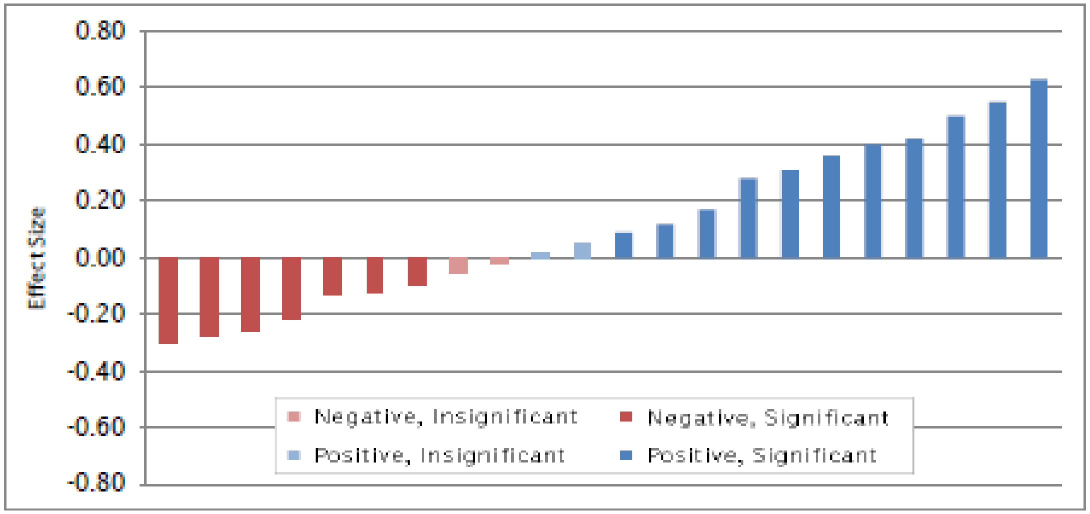
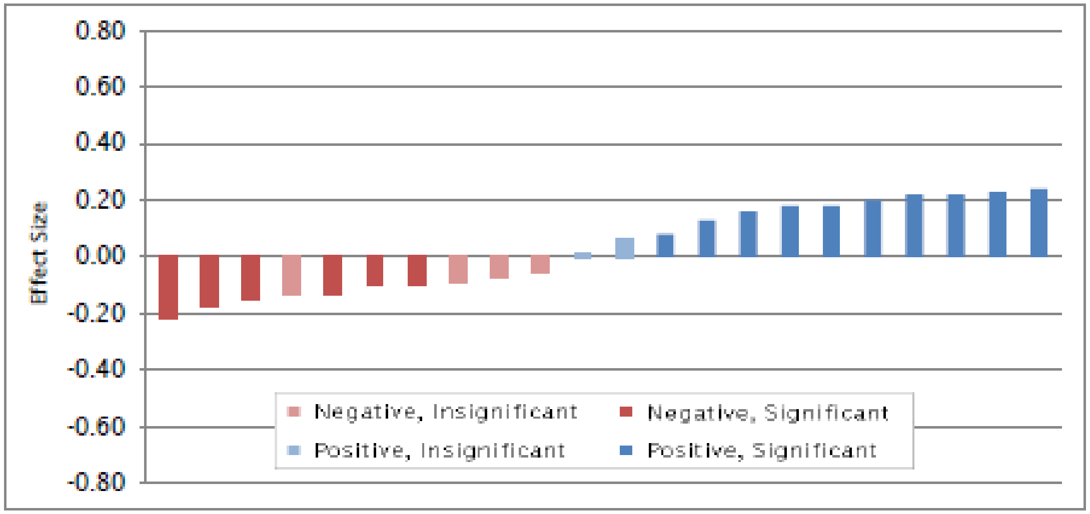


Figure IV.2. Distribution of Test- Score Effect Sizes After Two Years in Reading



MATHEMATICA Policy Research

The Workforce Investment Act

- In 2000 the Workforce Investment Act (WIA) replaced the Job Partnership Training Act
  - WIA aimed to transform the employment and training system by streamlining service delivery and providing universal access to services.
  - It promoted customer choice, attempted to more fully engage businesses, and changed service provided to youth.
- Training programs financed through Individual Training Accounts (ITAs).

The Workforce Investment Act

- DOL launched the ITA experiment in 1999 to provide state and local workforce agencies with systematic assessment of alternative approaches to ITAs and for estimating impacts.
- 8,000 training-eligible WIA customers in 8 sites
- Randomly assigned to 1 of 3 ITA approaches varying on intensity of counseling, authority of counselor over customer's choice and over ITA amount

The Workforce Investment Act

- **Approach 1: Structured Customer Choice.** This most directive approach required customers to receive intensive counseling, and counselors had considerable discretion to customize the amount of the ITA investment. On one hand, counselors were expected to constrain customers by steering them to training with a high expected return, and they could reject customers' choices that did not fit this criterion. On the other hand, counselors also had much greater discretion to set higher ITA amounts (up to a maximum of $8,000 in most sites) if they felt expensive training was a sound investment for certain customers.
- **Approach 2: Guided Customer Choice.** This approach, similar to what most workforce agencies adopted in the transition to WIA, involved mandatory counseling. However, counseling was less intensive than under the preceding approach. Counselors could not reject customers' choices if the chosen provider was on the state's approved list. The amount of the ITA award was fixed at $3,000 to $5,000, depending on the site.
- **Approach 3: Maximum Customer Choice.** This approach, the least structured of the three, did not require customers to participate in counseling after being found eligible for WIA-funded training, but they could request and receive it. Customers received a fixed ITA award of $3,000 to $5,000, depending on the site (as in the preceding approach). Counselors could not reject customers' choices if the provider was on the state's approved list.

The Workforce Investment Act

- Maximum choice
  - More likely to attend ITA orientation and use ITA
  - 6-7% higher participation rates
  - Entered training two weeks sooner on average
  - Less likely to participate in counseling after orientation
- Structured choice
  - Not fully implemented as designed – counselors were reluctant to comply!
  - $1,308 higher average earnings than maximum choice

Mental health parity

- Role of insurance: protection from catastrophic financial loss due to severe illness
- Traditionally, mental health and substance use disorder (MH/SUD) benefits more limited than general medical benefits
  - Limits disproportionately affect persons with severe illness
- In era of managed MH/SUD care
  - Are benefit limits needed to manage costs?
  - Will patients receive needed care?

Mental health parity

- In 2001 MH/SUD parity implemented in Federal Employees Health Benefit Program plans (FEHBP)
  - Improved insurance protection without increasing costs
  - Not associated with changes in access to care or MDD
- Mental Health Parity and Addiction Equity Act
  - Became Federal law in 2008
  - Regulations implemented in 2010

Mental health parity

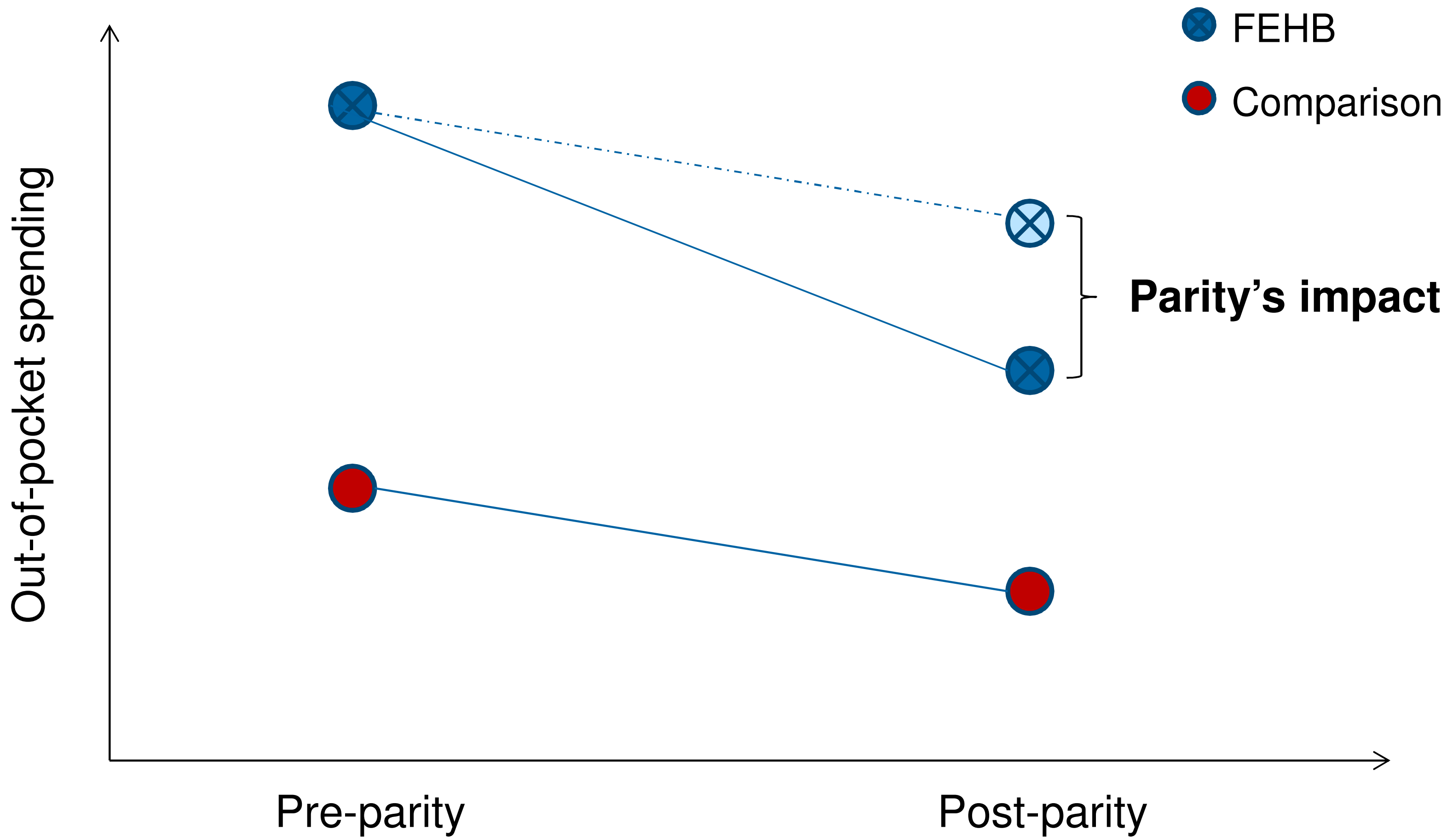
	FEHBP (N=19,094)		Comparison (N=10,521)	
Characteristic	%		%	
Female	64.9		68.5	
Employee	61.9		63.6	
Geographical Region				
Northeast	11.4		16.7	
South	64.3		15.9	
Midwest	6.8		59.5	
West	17.6		7.9	
Mean age [s.d.]	46.2 [8.3]		43.6 [11.0]	
Psychiatric Diagnosis	N	%	N	%
Bipolar Disorder	2,557	13.4	1,177	11.2
Major Depression	10,412	54.5	5,245	50.0
Adjustment Disorder	6,125	32.1	4,099	39.0

Mental health parity

Baseline spending ($) per person	FEHBP Mean	FEHBP (s.d.)	Comparison Mean	Comparison (s.d.)
Total spending				
Bipolar disorder	$3,347	(5,503)	3,536	(5,306)
Major depression	1,907	(2,943)	2,354	(3,055)
Adjustment disorder	807	(894)	931	(951)
Out-of-pocket spending				
Bipolar disorder	$841	(1,263)	387	(557)
Major depression	568	(657)	302	(348)
Adjustment disorder	339	(334)	153	(199)

Mental health parity

- Difference-in-difference



Mental health parity

Out-of-pocket spending ($)	Diff-in-diff (99% CI)
Bipolar Disorder	-149 (-217, -85)
Major Depression	-100 (-123, -77)
Adjustment Disorder	-69 (-84, -54)

Hypothetical case study

Evidence and proof

If you insist on strict proof (or strict disproof) in the empirical sciences, you will never benefit from experience, and never learn from it how wrong you are.

Sir Karl Popper
The Logic of Scientific Discovery

Many thanks!

- Senate Office of Education and Training
  - Megan Daly and Charles Revak
- The American Statistical Association
  - Steve Pierson
- Mathematica Policy Research
  - Carmen Ferro and Jennifer de Vallance

Contact me (just down the street)

Frank Yoon, Statistician

Mathematica Policy Research

1100 1st Street, NE

Washington, DC 20002

fyoon@mathematica-mpr.com

(202) 554-7518

www.mathematica-mpr.com

References

- Busch AB, Yoon FB, Barry CL et al. (2012), "The effects of mental health parity on spending and utilization for bipolar, major depression, and adjustment disorders," *American Journal of Psychiatry*, accepted for publication.
- Decker PT, Berk JA (2011), "Ten years of the Workforce Investment Act (WIA): interpreting the research on WIA and related programs," *Journal of Policy Analysis and Management* 30, pp. 906-926.
- Furgeson J, Gill B, Haimson J et al. (2012), "Charter-school management organizations: diverse strategies and diverse student impacts," Princeton, NJ: Mathematica Policy Research and Seattle: University of Washington Bothell, The Center on Reinventing Public Education.
- Holland PW (1986), "Statistics and causal inference," *Journal of the American Statistical Association* 81, pp. 945-960.
- Shadish WR, Clark MH, Steiner PM (2008), "Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments," *Journal of the American Statistical Association* 103, pp. 1334-1343.