

The Promise and Peril of Large Administrative Datasets





As illustrated in an evaluation of opioid use disorder treatment

Jason R Williams

Addictions, Drug & Alcohol Institute, UW School of Medicine

Center for Statistics and the Social Sciences seminar, 22 May 2024

Team effort: Caleb Banta-Green, Jeanne Sears, Anthony Floyd, & Mandy Owens.
Sponsor: The Paul G Allan Philanthropies

We acknowledge that we are on the ancestral lands of the Coast Salish peoples, who have lived in the Salish Sea basin and the North Cascades watershed from time immemorial. We honor with gratitude the lands, waters, and tribes of this region, including the Duwamish, Suquamish, Tulalip, and Muckleshoot, among others.    

- 1 Focus of the Presentation
 - What This Presentation is Missing
 - What This Presentation is About
- 2 Large Administrative Datasets in Research
 - The Promise
- 3 One Reality
 - Context
 - Problems and solutions
- 4 Summary

- 1 Focus of the Presentation
 - What This Presentation is Missing
 - What This Presentation is About
- 2 Large Administrative Datasets in Research
 - The Promise
- 3 One Reality
 - Context
 - Problems and solutions
- 4 Summary

What This Presentation is Missing

Fancy statistics

What This Presentation is Missing

<Pause for exodus>

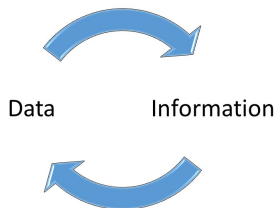
What This Presentation is Missing

"The all-cause mortality rate difference between the intervention and comparison group was tested in a logistic regression model that accounted for propensity score weighting and included history variables used in estimating the propensity score as covariates, followed by marginal effects estimation to calculate an average risk ratio."

What This Presentation is About

What This Presentation is About

Data-information cycle



What This Presentation is About

Data–information cycle

- One person's information is another person's data.
- When your boss says, "data", it may be on a different level than you are thinking about.

What This Presentation is About

Data-information cycle



- 1 Focus of the Presentation
 - What This Presentation is Missing
 - What This Presentation is About
- 2 Large Administrative Datasets in Research
 - The Promise
- 3 One Reality
 - Context
 - Problems and solutions
- 4 Summary

Large administrative datasets in research and evaluation

Known advantages

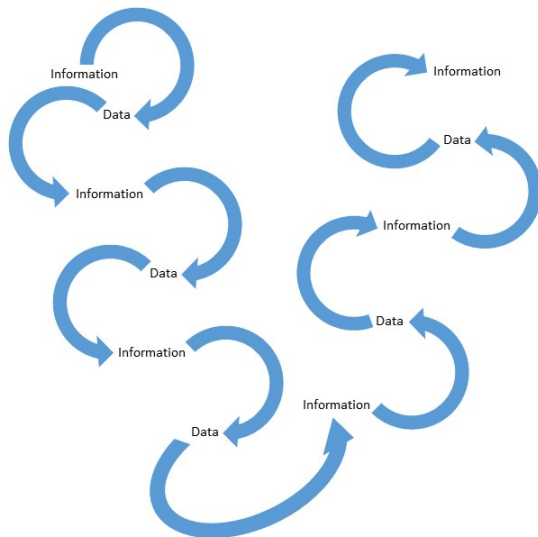
- Established and motivated data entry
- Measures of process, output, and outcome (?) that are relevant to practice should be relevant to evaluation
- Low cost
- Large N confers legitimacy, power

Large administrative datasets in research and evaluation

Known disadvantages

- Missing data
- Poor data documentation
- Measures for practice might not be valid indicators for research

Large administrative datasets in research and evaluation



- 1 Focus of the Presentation
 - What This Presentation is Missing
 - What This Presentation is About
- 2 Large Administrative Datasets in Research
 - The Promise
- 3 One Reality**
 - Context
 - Problems and solutions
- 4 Summary

STUDY PROTOCOL

Open Access



The Community-Based Medication-First program for opioid use disorder: a hybrid implementation study protocol of a rapid access to buprenorphine program in Washington State

Caleb J. Banta-Green^{1,2,3*} , Mandy D. Owens¹, Jason R. Williams¹, Jeanne M. Sears^{2,3,4}, Anthony S. Floyd¹, Wendy Williams-Gilbert⁵ and Susan Kingston¹

Abstract

Background: Opioid use disorder (OUD) is a serious health condition that is effectively treated with buprenorphine. However, only a minority of people with OUD are able to access buprenorphine. Many access points for buprenorphine have high barriers for initiation and retention. Health care and drug treatment systems have not been able to provide services to all—let alone the majority—who need it, and many with OUD report extreme challenges starting and staying on buprenorphine in those care settings. We describe the design and protocol for a study of a rapid access buprenorphine program model in six Washington State communities at existing sites serving people who are unhoused and/or using syringe services programs. This study aimed to test the effectiveness of a Community-Based Medication-First Program model.

Methods: We are conducting a hybrid effectiveness-implementation study of a rapid access buprenorphine model



The Meds First evaluation

"The Medication-First model of care is based upon. . .

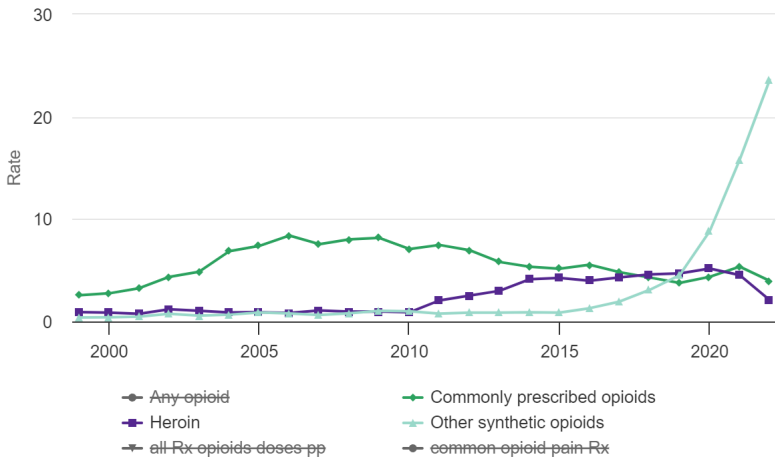
- rapid, typically same-day, access to medications;
- convenient, non-appointment-based care;
- no exclusions for poly-substance use;
- no counseling mandates (but services readily available); and
- ongoing, easy-to-access care."

The Meds First evaluation

- Community-based, 6 sites throughout state
- Nurse care managers
- Care navigators
- Walk-in. . . and walk out with a prescription for buprenorphine
- Began August 2019

The Meds First evaluation

Statewide opioid death rates



Analysis by UW ADAI. For data sources, see text or adai.uw.edu/WAdata

"It'll be easy," he said.

"It'll be easy," he said. "We've got admin data."

"The primary aim is to test the effectiveness of the intervention on morbidity and mortality outcomes. These outcomes are documented with Washington State data systems including arrest records, Medicaid claims for emergency and inpatient health care utilization, statewide hospitalization data, and the state death registry."

"It'll be easy," he said. "We've got admin data."

"A comparison group will be identified from state records of people with indicators of OUD. Construction of the comparison group will align characteristics such as geography, demographics, historical rates of arrests, MOUD history, and health care utilization, using restriction and propensity score techniques."

"It'll be easy," he said. "We've got admin data."

- Buprenorphine prescriptions (Prescription Monitoring Program)
- Medical encounters (Provider One)
- Other prescription records from Provider One ("Rx file")
- Arrests and charges (Client Services Database)
- Hospitalizations (CSDB)
- Buprenorphine, naltrexone, and methadone (CSDB)
- Age, county of residence, and death flag (CSDB)
- OUD flags (CSDB)

Treatment group

- Initiated treatment, and then recruited for study
- N=825 ages 18–70
- State linked them to state records (Link King) → 814

Potential comparison group

- State appeared to provide anyone 18–70 as of August 2019 who had an OUD flag in the past 7 years
- N=106475

Potential comparison group

- State appeared to provide anyone 18–70 as of August 2019 who had an OUD flag in the past 7 years
- N=106475

Sheer size of data

- Comparison Rx file: 4.1GB SAS file
- Medical encounters: 9 pieces, 7.8GB total

- Layers of pre-processing
 - Drop observations and variables you do not need
 - Drop those deceased before MF began and those who ever had a script associated with a MF provider (possibly MF participants not enrolled in evaluation)
 - Break large datasets down → more manageable chunks
 - E.g., pull out all medical encounters that involved an OUD or drug poisoning diagnosis: many GB → 669MB ready for later use
- Trial code with samples of data

Bupe and the PMP

- PMP: All filled prescriptions in state, completed by pharmacies
- Rx file: Prescriptions paid by Medicaid in Provider One

Bupe and the PMP

CID	dateWritten	dateFilled	quantity	daysSupply	dispenserD NPI	NPI lookup
1564	10/20/2020	10/20/2020	2		1 FE5239517	
1564	10/20/2020	10/21/2020	12		6 FE2656669	
1564		10/27/2020	14		7	1114036423 THRIFTY PAYLESS INC
1564	11/3/2020	11/3/2020	14		7 FE5239517	
1564	11/10/2020	11/10/2020	14		7 FE5239517	
1564		11/17/2020	28	14		1114036423 THRIFTY PAYLESS INC
1564		11/30/2020	14	7		1114036423 THRIFTY PAYLESS INC
1564		12/7/2020	14	7		1114036423 THRIFTY PAYLESS INC
1564		12/14/2020	42	21		1114036423 THRIFTY PAYLESS INC
1564		1/4/2021	14	7		1114036423 THRIFTY PAYLESS INC
1564		1/10/2021	15	7		1114036423 THRIFTY PAYLESS INC
1564		1/18/2021	16	8		1114036423 THRIFTY PAYLESS INC
1564		1/25/2021	28	14		1114036423 THRIFTY PAYLESS INC
1564	2/8/2021	2/8/2021	28		14 FE5239517	
1564	2/22/2021	2/22/2021	28		14 FE5239517	
1564	3/8/2021	3/8/2021	28		14 FE5239517	
1564	3/22/2021	3/22/2021	4		2 FE5239517	
1564	3/24/2021	3/24/2021	28		14 FE5239517	
1564	3/31/2021	4/7/2021	14		7 FE5239517	
1564	4/15/2021	4/15/2021	21		7 FE5239517	
1564	4/22/2021	4/22/2021	28		14 FE5239517	
1564	5/6/2021	5/6/2021	28		14 FE5239517	
1564		5/20/2021	10	5		1750490066 THRIFTY PAYLESS INC
1564		5/25/2021	20	14		1750490066 THRIFTY PAYLESS INC

Bupe and the PMP

How big a problem?

- Any bupe in 90 days post induction? Don't see almost 13% having bupe.
 - PMP only: 74.9%
 - Rx file only: 77.9%
 - Combined: 87.7%
- Days missed (90 for those with no bupe)? 5-6 days difference on average.
 - PMP only: 59.3 days
 - Rx file only: 57.6 days
 - Combined: 53.1 days

Main problems in matching

How to represent the counterfactual?

- 1 Treatment group has a start date. No start date for "normal services" comparison.
- 2 Treatment group self-selected based on identified need in that moment. Need to document need and find similar people in the potential comparison group. See #1.

Main problems in matching

How to represent the counterfactual?

Trx group member									
Month 263					Month 275				
T	T	T	T	T	MF	B	B		

- T = touchpoint
 - Poisoning treated by bystander naloxone, EMS, or at ED
 - Any medical encounter
 - Treatment: office, OTP
 - Jail-based evaluation
- Most should filter up to CSDB → OUD flag

Main problems in matching

How to represent the counterfactual?

Trx group member							
Month 263				Month 275			
T	T	T	T	T	MF	B	B

Comparison group member							
Month ???				Month ???			
T	T	T	T	T	?	?	?

Main problems in matching

How to represent the counterfactual?

- 1 Treatment group has a start date. No start date for "normal services" comparison.
- 2 Treatment group self-selected based on identified need in that moment. Need to document need and find similar people in the potential comparison group. See #1.

Main problems in matching

State data not showing need

- OUD flags (CSDB)
- Buprenorphine prescriptions (Prescription Monitoring Program)
- Medical encounters (Provider One) with an opioid poisoning diagnosis
- Medical encounters with an OUD diagnosis
- Other prescription records from Provider One
- Buprenorphine, naltrexone, and methadone (CSDB)

Main problems in matching

State data not showing need

- OUD flags (CSDB)
- Buprenorphine prescriptions (Prescription Monitoring Program)
- Medical encounters (Provider One) with an opioid poisoning diagnosis
- Medical encounters with an OUD diagnosis
- Other prescription records from Provider One
- Buprenorphine, naltrexone, and methadone (CSDB)

Main problems in matching

History as of when?

- "Just compare their histories."
- Match people with a documented start date to people with an OUD flag in the past 7 years.

Solution: Two-stage matching

- Stage 1:
 - Anchor potential matches from the broad pool to treatment observations in time.
 - Select matches based on granular historical characteristics.
- Stage 2: PSM with finer characteristics

Solution: Two-stage matching

- Stage 1:
 - 1 For each enrolled treatment group member (N=670), we assign all comparison pool members a "start date" matching that of the treatment observation.
 - 2 Select pool based on county of residence during the "start" month, using a logical list of matching counties for each county of residence of the treatment observation.
 - 3 Drop those dead as of month prior to start month (remember, death is just 0/1 in month).
 - 4 Create a broad-net monthly indicator of any MOUD need, calculate total months with MOUD need and months since last observe MOUD need.
 - 5 Match on those two measures: needs a window/some fuzziness
 - 6 Keep up to 15 matches: If more than 15, randomly select.
 - 7 Go back to #1 and repeat for next enrollee.

Solution: Two-stage matching

- Stage 1 implications:
 - Comparison group becomes individuals with an assigned start date paired to a treatment group member.
 - Now that we have a start date, we can construct outcomes.
- Stage 1 results:
 - 14.96 matches per treatment group member
 - 10021 comparison observations (8350 unique individuals)

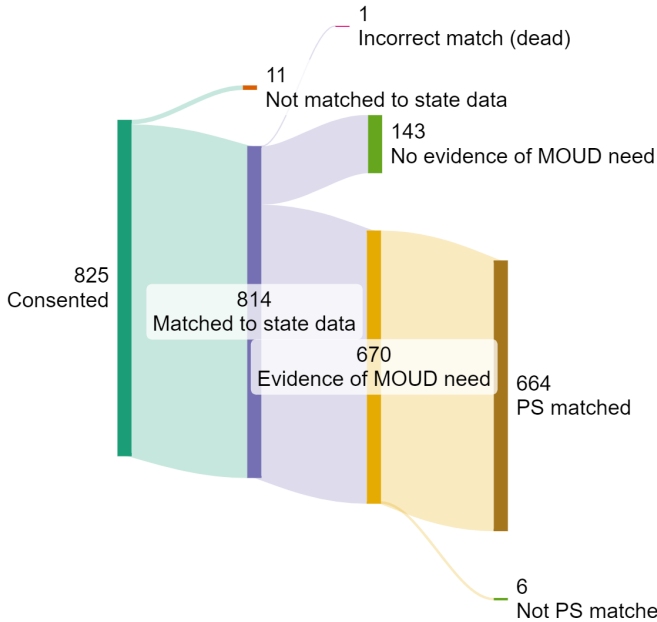
Solution: Two-stage matching

- Stage 1: Anchor potential matches to treatment observations in time & match based on granular historical characteristics.
- Stage 2: PSM with finer characteristics in a common 24-month pre-treatment period
 - 1 months of Medicaid eligibility in last 24
 - 2 number of opioid poisonings
 - 3 months with a poisoning
 - 4 number of OUD/poisoning-related ED visits
 - 5 number of OUD/poisoning-related inpatient visits
 - 6 months with any buprenorphine
 - 7 months with any naltrexone
 - 8 months with any methadone
 - 9 months with any MOUD
 - 10 months with any inpatient hospitalizations, all cause
 - 11 months with any all-cause ED visits
 - 12 months with any arrest recorded
 - 13 female versus male, per RDA
 - 14 age in start month, per RDA

Solution: Two-stage matching

- Stage 1 results:
 - 14.96 matches per treatment group member
 - 10021 comparison observations (8350 unique individuals)
- Stage 2 results:
 - Our selected solution: optimal full matching with a linear function
 - Standardized mean differences for each predictor after matching were less than 0.06.
 - Retained the highest number of treatment ($n = 664$) and comparison group ($n = 9893$) members.

Lost treatment group members



Result results!

Outcome, 12 months	Contrast	p-value
Mortality (RR, all-cause)	0.323	0.039
Months with buprenorphine	2.199	0.000
Months with any MOUD	1.790	0.000
Months with a poisoning	0.039	0.068
Months with any inpatient hospitalization	0.342	0.003
Months with any ED visit	-0.072	0.415

Notes: All models include the covariates used for PSM and interact treatment with each of those covariates. Standard errors are estimated using the delta method and are robust and clustered at the PS (stage 2) matching level and the (relatively sparse) stage 1 matching level.

- 1 Focus of the Presentation
 - What This Presentation is Missing
 - What This Presentation is About
- 2 Large Administrative Datasets in Research
 - The Promise
- 3 One Reality
 - Context
 - Problems and solutions
- 4 Summary

- Large admin datasets may provide multiple sources of information on key indicators.
- With **LARGE** datasets, pre-processing and testing of code on small samples will be necessary.
- As always: KYD!
 - Interrogate who and what is included in each data source and who and what is excluded



MISTAKES

IT COULD BE THAT THE PURPOSE OF YOUR LIFE IS
ONLY TO SERVE AS A WARNING TO OTHERS.