When (and how much) can we trust quasi-experimental evaluations of policies?

David Frisvold (University of Iowa)

Policy Evaluation Network Workshop

September 7, 2021

Focus of the workshop: Tradeoffs and synergies of model-based (structural) and design-based (quasi-experimental) methods

- Focus of this presentation: Issues related to quasi-experimental methods, with an emphasis on SSB taxes

- Agenda:
  1. The question of interest
  2. Variation of estimates
  3. Outcomes, data options, and data availability
  4. Defining the counterfactual
  5. Methods in quasi-experimental evaluations
  6. Additional challenges
  7. Concluding thoughts

- Quasi-experimental methods
  - What was the impact of a tax on an outcome?
  - Example: What was the impact of the 1.5 cents per ounce tax in Philadelphia on beverage purchases and consumption for Philadelphia residents? (Cawley et al., 2019; JHE)
  - Informs policymakers about what happened and can provide an understanding about consumer behavior
  - Context specific
  - Evidence from a variety of contexts needed to inform policymakers considering new taxes

- Model-based estimates
  - What would be the impact of a tax on an outcome?
  - Informs policymakers about what would happen and can provide an understanding about consumer behavior
  - Assumptions are more explicit
  - Value of the information to policymakers depends on the applicability, credibility, and completeness of the model

Issues related to quasi-experimental methods, with an emphasis on SSB taxes

- Agenda:
  1. The question of interest
  2. **Variation of estimates**
  3. Outcomes, data options, and data availability
  4. Defining the counterfactual
  5. Methods in quasi-experimental evaluations
  6. Additional challenges
  7. Concluding thoughts

- Estimates of similar taxes using the same methods can vary throughout a country
- Estimates based on similar methods for the same outcome of the impact of the same tax can vary a lot
- Why?

Price estimates vary across cities, studies, and data type

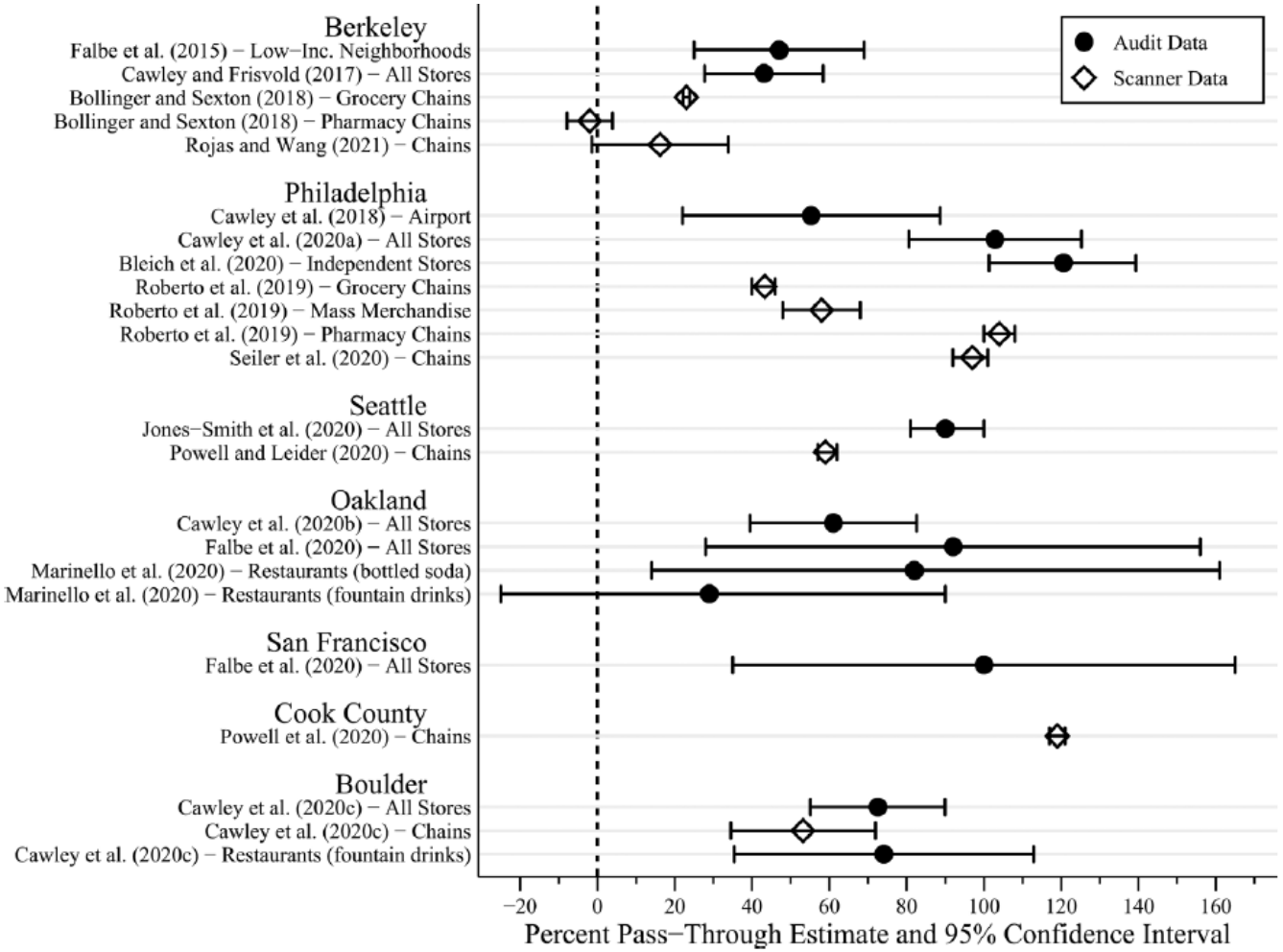


Figure 1. Estimated pass-through of SSB taxes, by city, study, and data type

Note: This figure shows the pass-through estimates with the 95% confidence interval for papers in the literature on SSB taxes, including the estimates from this paper. The pass-through estimates are calculated as the point estimate of the change in price, in cents per ounce, divided by the amount of the local tax. The figure shows the primary estimate for all stores or restaurants in the sample for all SSBs, if available; otherwise, the row headings describe the store types or beverage types corresponding to the estimate.

Cawley, Frisvold, Jones, Lensing *AJAE* (2021)

- Estimates of similar taxes using the same methods can vary throughout a country

- Estimates based on similar methods for the same outcome of the impact of the same tax can vary a lot
  - Type of data
    - Audit data (researcher-collected), scanner data, etc.
  - Sample
    - All stores, select store types, select neighborhood demographics
    - Retail stores, restaurants
  - Comparison groups
    - Adjacent counties
    - Distinct cities in the same region
    - Similar stores from anywhere in the country matched on pre-tax trends
      - Note: country-wide taxes often don't have a comparison group

Issues related to quasi-experimental methods, with an emphasis on SSB taxes

- Agenda:
  1. The question of interest
  2. Variation of estimates
  3. **Outcomes, data options, and data availability**
  4. Defining the counterfactual
  5. Methods in quasi-experimental evaluations
  6. Additional challenges
  7. Concluding thoughts

- Outcomes
  - Prices
  - Purchases/Sales
  - Consumption
  - Health outcomes
  - Other outcomes

- Outcome of interest informs the type of data that need to be used
  - Ex: sales – use scanner data; consumption – use survey data
  - Scanner data are commonly available for large chain retailers only
  - Time lags in data availability
  - Data availability also influences the selection of outcomes

- Health outcomes (obesity) are of key interest
  - Few quasi-experimental studies on the impact of SSB taxes on obesity

- Data types (purchases and sales)
  - Observation/survey of customers exiting stores
    - Advantages: Can gather data from all stores and all store types
    - Disadvantages: Costly, does not capture non-retail purchases
  - Consumer panel data (Ex: Nielsen/IRI, Kantar World Panel)
    - Advantages: Observe purchases from every location
    - Disadvantages: Are all purchases recorded? Small sample size
  - Scanner data (Nielsen/IRI, Retailer-specific data)
    - Advantages: Observe all items sold, weekly data
    - Disadvantages: Selected store types, does not capture non-retail purchases, quicker access is costly
- Purchase location provides evidence on cross-border purchases (tax evasion)

- Purchases ≠ consumption
  - Different individuals within a household
  - Purchases are often measured from retailers only
  - Beverages consumed from a variety of sources (restaurants, school, etc.)
- More sources of data on purchases than consumption
- Primary (only?) data source for consumption is a survey
  - Lagged release & concerns of coverage of public surveys
    - US surveys: NHANES, BRFSS, etc.
  - Measurement issues
  - How to construct a representative sample

- Measurement of consumption is difficult
  - Survey of frequency of consuming types of beverages

    "During the past month, how often did you drink soda?"

  - Concerns
    - Units consumed vs. ounces consumed
    - Recall error among adults and children
    - Is there non-classical measurement error?

- Constructed measures
  - Daily consumption
  - Any consumption
  - Frequency of consumption
  - Summary measures (grams of added sugars consumed, etc.)

Issues related to quasi-experimental methods, with an emphasis on SSB taxes

- Agenda:
  1. The question of interest
  2. Variation of estimates
  3. Outcomes, data options, and data availability
  4. **Defining the counterfactual**
  5. Methods in quasi-experimental evaluations
  6. Additional challenges
  7. Concluding thoughts

- What would the outcome be for individuals who are exposed to the tax, if they had not been exposed to tax?

- Possibilities to construct a counterfactual scenario
  1. Randomly assign individuals/stores to be taxed
  2. Trends of untaxed beverages
  3. Trends of the same individuals/stores prior to the tax
  4. Trends of similar individuals/stores near the area with the tax
  5. Trends of similar individuals/stores farther area from the taxed area

1. Randomly assign individuals/stores to be taxed

   - Those not taxed provide the counterfactual outcomes
   - Probably not feasible for national/sub-national policies

2. Trends of untaxed beverages

   - Assumes that the trends of untaxed beverages after the tax would be similar to the trends of taxed beverages if there was not a tax
   - Ignores substitution
     - If people drink fewer SSBs, do they drink less liquid or drink something else?

3. Trends of the same individuals/stores prior to the tax
   - Most common scenario for national taxes
   - Assumes that the outcomes of individuals/stores would continue on the same trends if the tax was not implemented

   - Advantages
     - Counterfactual is based off of the same individuals/stores
   - Disadvantages
     - Other policies or changes to demand or supply may occur around the same time
       - High frequency data helps to overcome this limitation

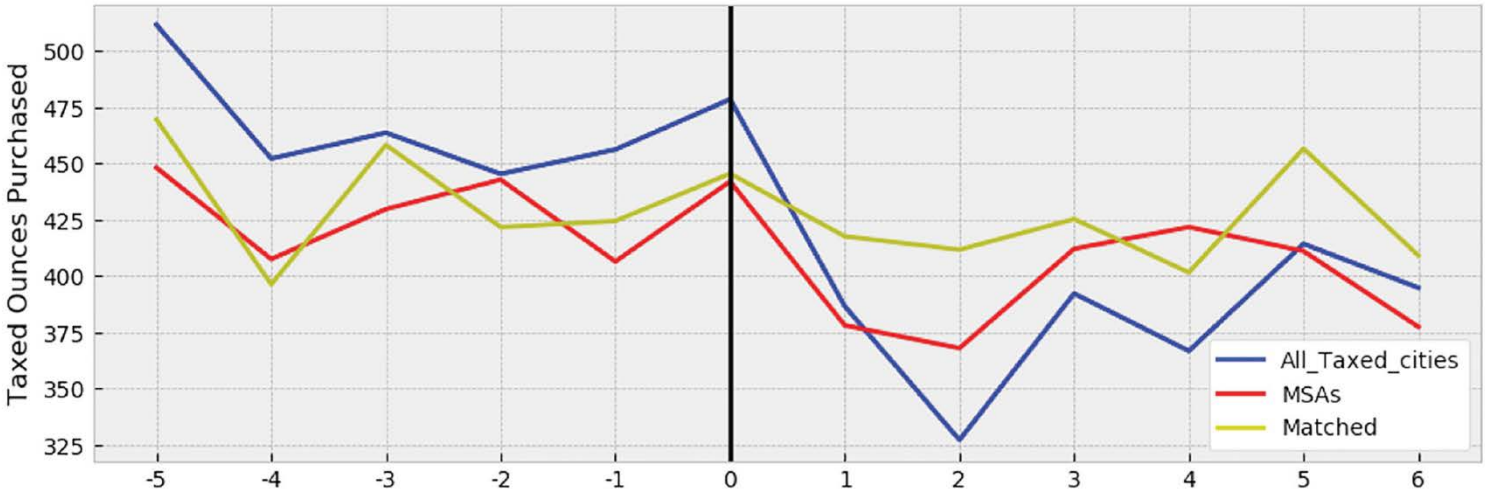
4. Trends of similar individuals/stores near the area with the tax
   - Common scenario for sub-national taxes
   - Assumes that the trends of taxed and nearby individuals/stores would be similar after the tax, if the tax was not implemented

   - Advantages
     - Nearby individuals/stores may experience the same local shocks to demand or supply
     - Ex: Changes in economic conditions, exposure to same public debate about the tax/same local advertising market, same state policies, weather
   - Disadvantages
     - Spillover effects: may shop in the taxed area and be exposed to the tax
       - Bias estimate downwards

5. Trends of similar individuals/stores farther area from the taxed area
   - Common scenario for sub-national taxes
   - Assumes that the trends of taxed and farther away individuals/stores would be similar after the tax, if the tax was not implemented

   - Advantages
     - No spillover effects
   - Disadvantages
     - Individuals/stores do not experience the same local shocks to demand or supply

Influence of comparison group choice: nearby and distant individuals



Notes: This figure shows the trends in the monthly average volume of taxed beverages purchased (top panel) and untaxed beverages purchased (bottom panel) for the six months before and six months after the implementation of each city beverage tax for households in the cities and the two comparison groups: (1) households in the MSAs but outside of the treatment cities and (2) matched households in cities nationwide. The taxes were implemented on the first day of month 1. The vertical line at month 0 distinguishes between the pre- and post-tax periods.

Cawley, Frisvold, and Jones, *Health Economics* (2020)

Issues related to quasi-experimental methods, with an emphasis on SSB taxes

- Agenda:
  1. The question of interest
  2. Variation of estimates
  3. Outcomes, data options, and data availability
  4. Defining the counterfactual
  5. **Methods in quasi-experimental evaluations**
  6. Additional challenges
  7. Concluding thoughts

- How to estimate the effect of the tax on the selected outcomes?

1. Difference-in-differences (most common)
2. Event study (closely related to #1)
3. Value-added (related to #1)
4. Matching estimators
5. Synthetic control method (related to #2 and #4)
6. Regression discontinuity in time
7. Interrupted time series (closely related to #6)

Focus on #1 and #2 for today

$$Y_{it} = \alpha_0 + \alpha_1 Post_t + \alpha_2 Treated_i + \alpha_3 Treated_i * Post_t + \alpha_X X_{it} + \varepsilon_{it}$$

- Post = observation is after the tax (binary)
- Treated = observation is in the taxed area (binary)
  - The comparison group used to define the counterfactual has a value of 0
- α_3 is the DD estimate
  - The change over time of the outcome in the taxed area, relative to the change over the same time period in the untaxed area

- The impact of the Philadelphia tax on prices
  - Outcome: price/ounce of taxed beverages
  - Tax: 1.5 cents/ounce on SSBs and diet drinks, Jan 2017
  - Taxed group: Stores in Philadelphia
  - Comparison group: Stores outside of Philadelphia within the metro area and in the same state
    - Matched to stores in Philadelphia on store type and local neighborhood characteristics
  - Audit data (self-collected)
    - Nov & Dec 2016 (pre-tax)
    - Nov & Dec 2017 (post-period)

Store locations

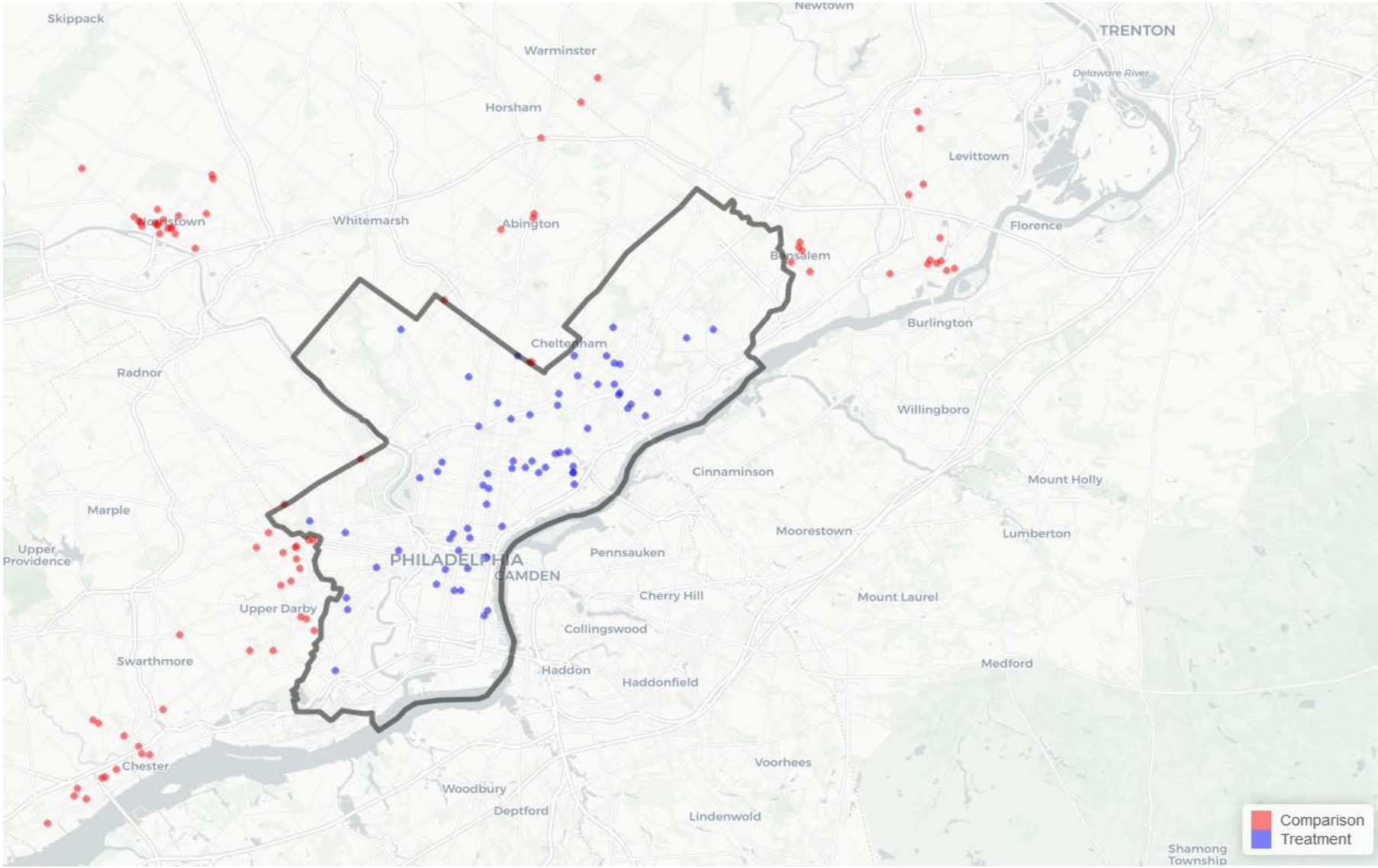


Table 1. Average price per ounce of beverages before and after the implementation of the SSB tax in Philadelphia.

	Stores in Philadelphia			Comparison stores		
	2016	2017	Difference	2016	2017	Difference
Taxed beverages	7.474 (0.203) [1,253]	9.332 (0.226) [1,052]	1.859 (0.303) [2,305]	7.865 (0.200) [1,562]	8.257 (0.207) [1,366]	0.392 (0.288) [2,928]

DD Estimate: 1.859 – 0.392 = 1.467 cents/ounce

Tax rate is 1.5 cents/ounce (98% pass through)

Cawley, Frisvold, Hill and Jones, *JPAM* (2020)

After controlling for store and product fixed effects:

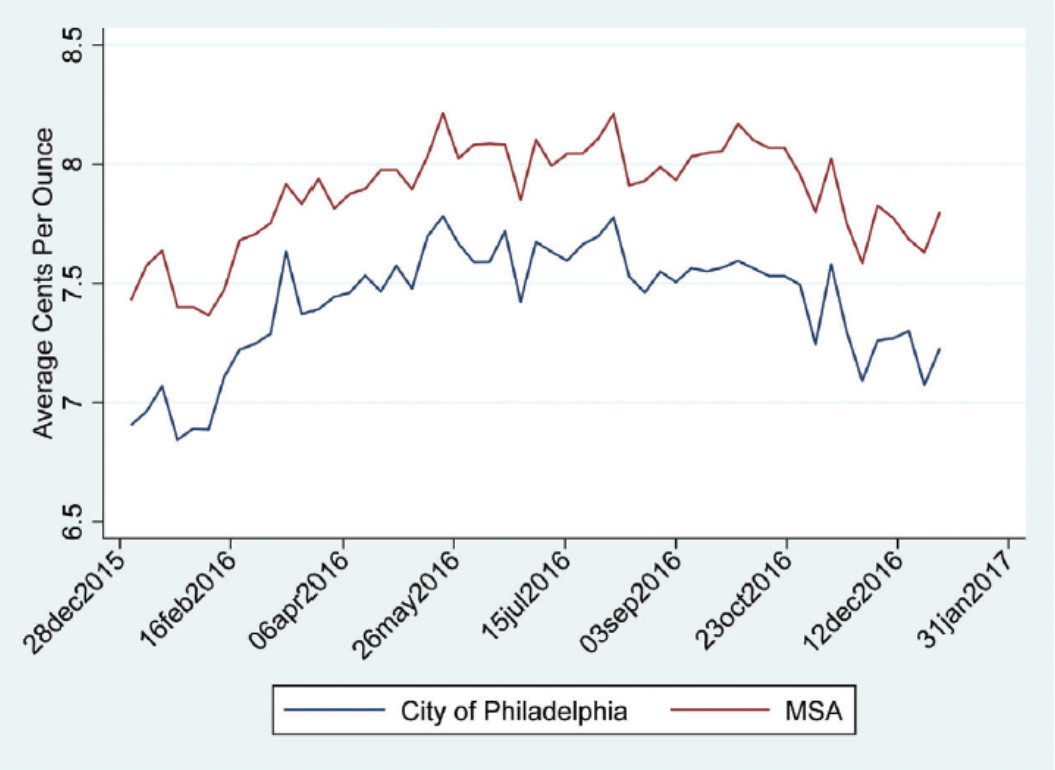
Table 2. The impact of the SSB tax on prices.

	Full sample
All taxed beverages	1.544 (0.171) [5,233]

Cawley, Frisvold, Hill and Jones, *JPAM* (2020)

- *One* assumption in this example
  - $\Delta Y_{Philadelphia} = \Delta Y_{Comparison}$ if the tax had not occurred
  - Untestable, but is there evidence consistent with this assumption?
  - Are pre-tax characteristics similar?
  - Are pre-tax levels of the outcomes similar?
  - Are the trends in the outcomes prior to the tax similar?
    - Called the parallel trends assumption
    - Cannot examine with the audit data
    - Can examine with data from another set of stores
      - Post-period data was not available for this sample

DD: Parallel trends assumption



Notes: This figure shows the average weekly price per ounce of regular and diet soda in retail stores in Philadelphia and in the Philadelphia MSA in the 12 months leading up to the tax (January 2016 to December 2016). The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.
Source: Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (U.S.), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.

Figure A1. Average Weekly Price per Ounce of Regular and Diet Soda in the 12 Months Prior to the Philadelphia Beverage Tax (January 2016 to December 2016).

Cawley, Frisvold, Hill and Jones, *JPAM* (2020)

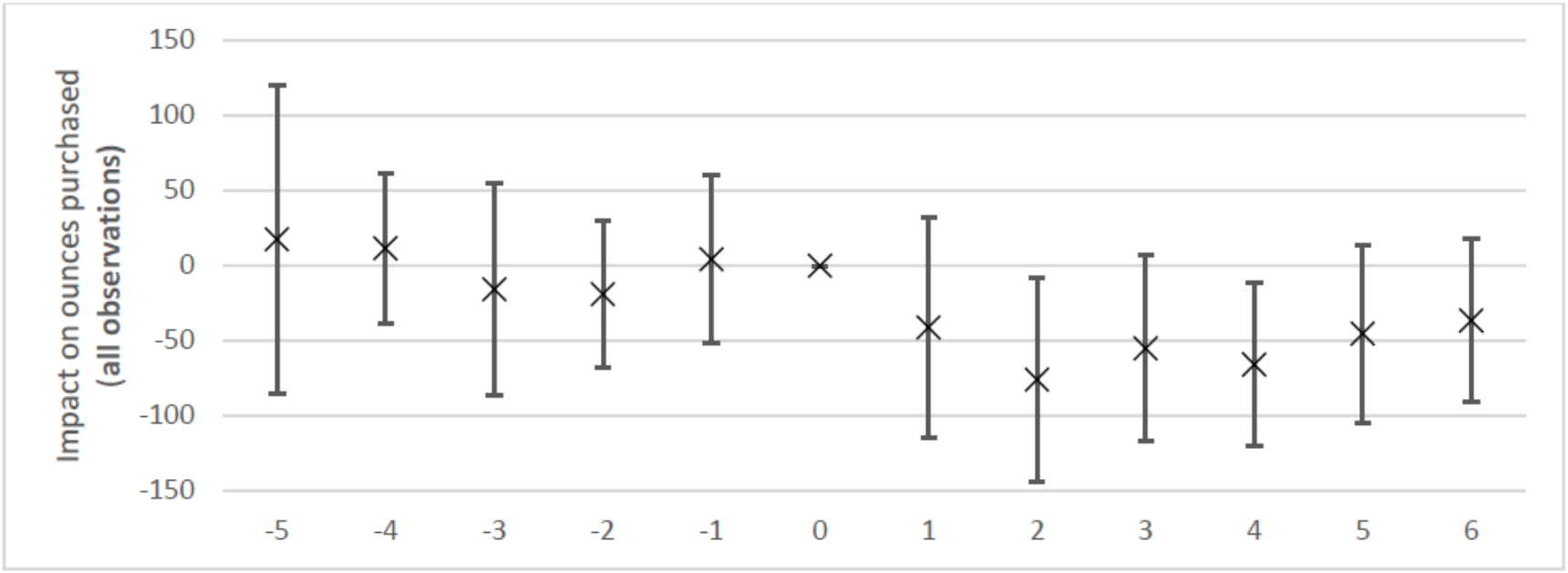
- Estimate how the impact changes over time
  - Interactions with the tax variable and week or month

- Staggered implementation
  - When the taxes are implemented in different months
  - Use relative time (months until or since the tax was implemented) instead of calendar time

$$Y_{hct} = \alpha_0 + \sum_r \alpha_r 1(t - T_c = r) + \delta_h + \gamma_t + \varepsilon_{hct}$$

r = relative time (months since tax implemented in time T)

Figure 2: Impact of SSB Taxes on Purchases of Taxed Beverages (Ounces/Month)



Cawley, Frisvold, and Jones, *Health Economics* (2020)

Issues related to quasi-experimental methods, with an emphasis on SSB taxes

- Agenda:
  1. The question of interest
  2. Variation of estimates
  3. Outcomes, data options, and data availability
  4. Defining the counterfactual
  5. Methods in quasi-experimental evaluations
  6. **Additional challenges**
  7. Concluding thoughts

- Already discussed:
  - Measurement error in the outcomes
  - Bias, related to comparison group

- Additional challenges:
  - Is the sample representative?
  - Standard errors
    - Are observations independent within a city or over time?
    - Clustering standard errors at the geographic level of the tax
      - Not feasible with one or two geographic units
      - Wild-cluster bootstrap with few clusters
  - Understanding the reasons behind changes in behavior
    - Guidance from theory
    - Complementarity of qualitative interviews with quantitative analysis

Clustering standard errors

	All observations		All observations
Taxed beverages		*Philadelphia*	
Tax rate		Tax rate	
Point estimate	−53.00	Point estimate	−84.07
95% confidence interval		95% confidence interval	
Clustered at household level	[−86.04, −19.97]	Clustered at household level	[−130.08, −38.05]
Wild-cluster bootstrap, area clusters	[−93.35, 41.74]	Wild-cluster bootstrap, area clusters	[−101.9, −65.99]
Wild-cluster bootstrap, city clusters	[−84.76, 17.67]		
Pretax mean	432.92	Pretax mean	455.13
Observations	17,364	Observations	9,948
Households	1,447	Households	829

- Cluster at the household level
  - Account for correlations over time in the purchases of each household
- Area clusters = 12 clusters
  - Four treatment cities, 8 control groups (MSA and matched HH from 7 different cities)
- City clusters = 4 clusters
  - Combines the treatment city and its 2 control groups into a cluster

Cawley, Frisvold, and Jones, *Health Economics* (2020)

Issues related to quasi-experimental methods, with an emphasis on SSB taxes

- Agenda:
  1. The question of interest
  2. Variation of estimates
  3. Outcomes, data options, and data availability
  4. Defining the counterfactual
  5. Methods in quasi-experimental evaluations
  6. Additional challenges
  7. **Concluding thoughts**

- When (and how much) can we trust quasi-experimental evaluations of policies?
  - How much? More trustworthy if …
    - Transparent about assumptions, measurement and demonstrates robustness
    - Data/code are available
    - Pre-specification plans are an interesting idea
  - When?
    - Trustworthy for the time/location being studied
    - A literature provides guidance about more general impacts

- Complementarities and value of model-based and quasi-experimental studies
- Paper of interest
  - Pick a tax and dataset (example: Philadelphia and Nielsen Retail Scanner Data)
  - Model-based estimates
    - Using only pre-tax data
    - Vary assumptions
      - QUAIDS, etc.
      - Different instruments
  - Quasi-experimental estimates
    - Use pre- and post-tax data
    - Vary assumptions
      - Different specifications/comparison groups
  - How similar are the range of estimates?
    - Does the existing literature for this same tax fall within this range?

References

Cawley, John, David Frisvold, Anna Hill, and David Jones. 2020. "The Impact of the Philadelphia Beverage Tax on Prices and Product Availability." *Journal of Policy Analysis & Management,* 39(3): 605-628.

Cawley, John, David Frisvold, Anna Hill, and David Jones. 2019. "The Impact of the Philadelphia Beverage Tax on Purchases and Consumption by Adults and Children." *Journal of Health Economics*, 67: 10225.

Cawley, John, David Frisvold, David Jones, and Chelsea Lensing. Forthcoming, 2021. "The Pass-Through of a Tax on Sugar-Sweetened Beverages in Boulder, Colorado." *American Journal of Agricultural Economics*.

Cawley, John, David Frisvold, and David Jones. 2020. "The Impact of Sugar-Sweetened Beverage Taxes on Purchases: Evidence from Four City-Level Taxes in the U.S." *Health Economics*, 29(10): 1289-1306.

Cawley, John, Anne Marie Thow, Katherine Wen, and David Frisvold. 2019. "The Economics of Taxes on Sugar-Sweetened Beverages: A Review of the Effects on Prices, Sales, Cross-Border Shopping, and Consumption." *Annual Review of Nutrition*, 39:317-338.

For additional papers, data, and code:

https://www.biz.uiowa.edu/frisvold/soft-drink-taxes/