Causality: Observational data

Instructor: Davide Proserpio

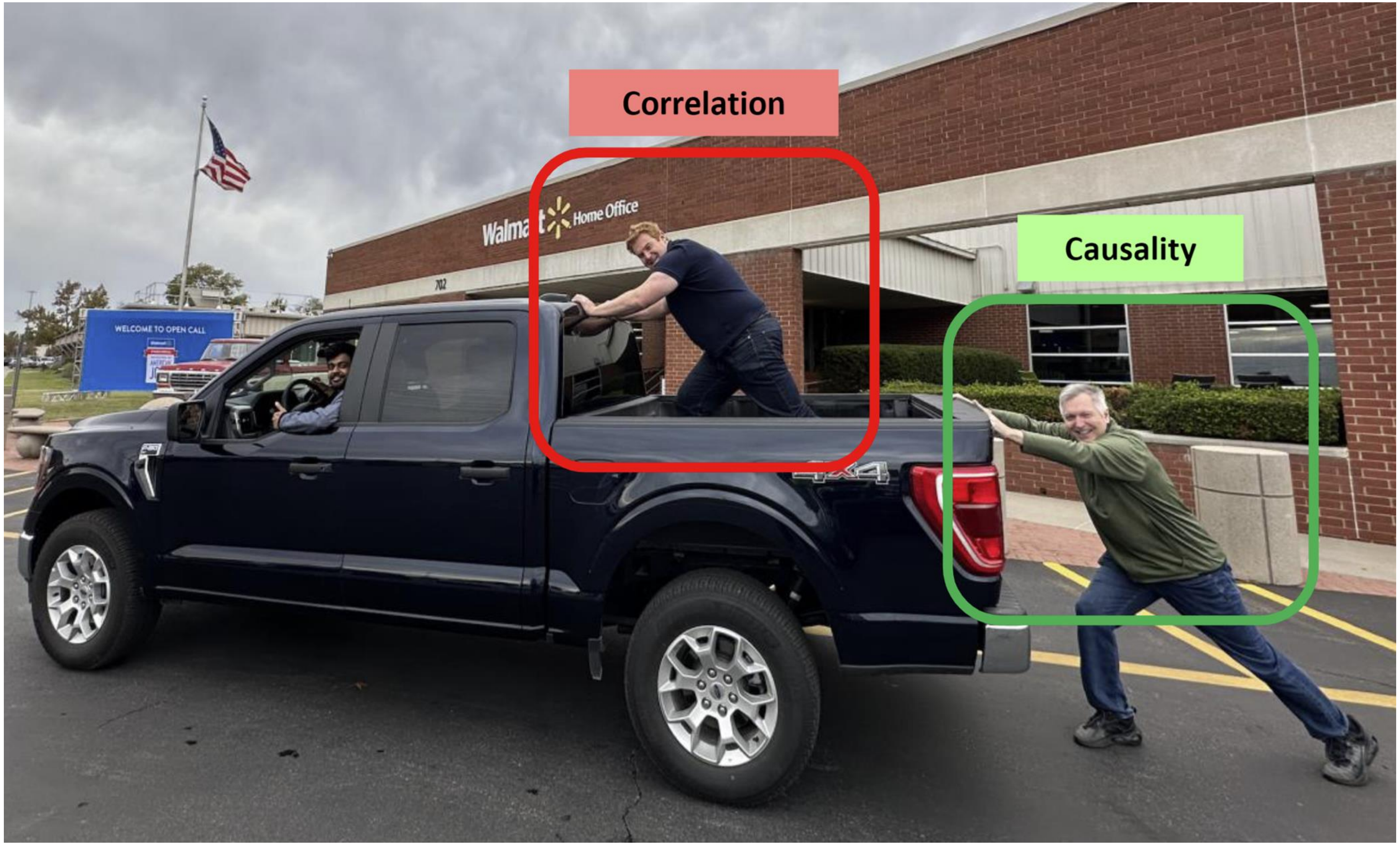
A few things

- Next Monday guest speaker (in person): Giovanni Marano, Analytics Senior Director at FanDuel
  - Leading the Marketing and Media Analytics team, responsible for measuring, testing and optimizing the Media investment strategy

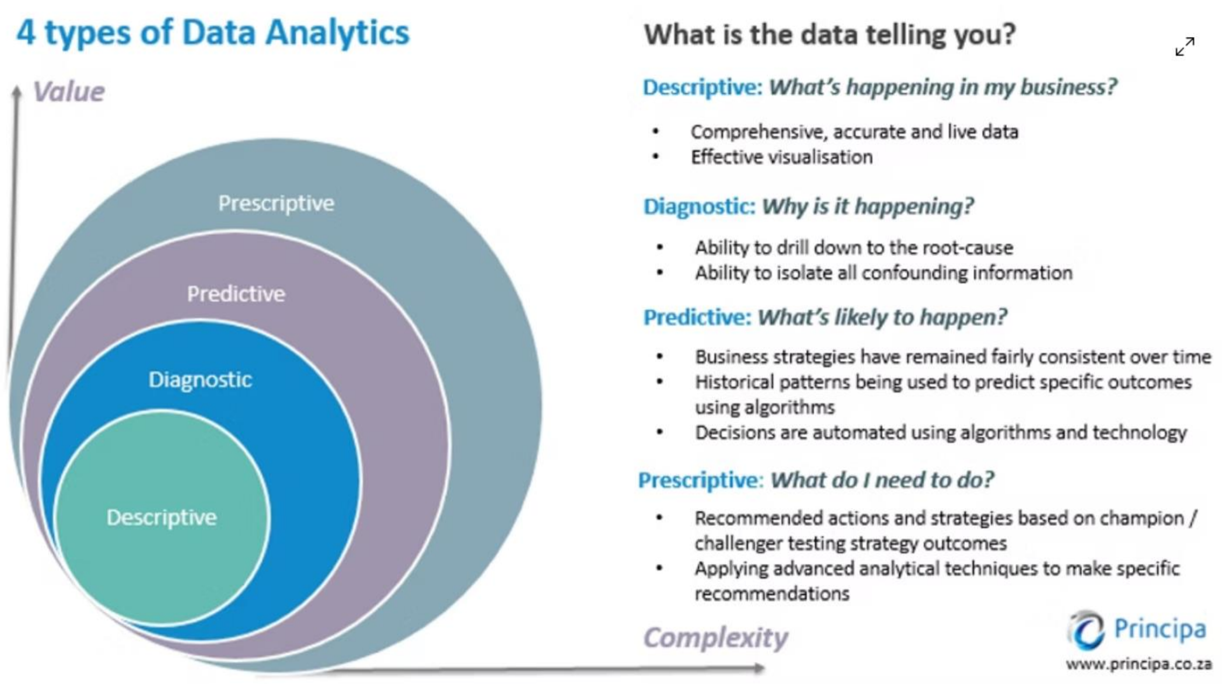
Why We Run Experiments

- Marketers constantly ask: *“Did my action* ***cause*** *a change?”*
- The problem: correlation ≠ causation.
- In general, what can we learn from a significant correlation?
  - “These two variables likely move together." Anything more requires assumptions.

Correlation ≠ causation



Why causality matters?



- Correlations are **descriptive** analytics (“facts”)
- Causality matters most for **diagnostic** and **prescriptive** analytics
- Causality can help build predictive models, but correlations suffice most of the time for **predictions**

The Fundamental Problem of Causal Inference

- For each customer (or product, retailer, etc.):
- We observe **what actually happened** under the chosen action.
- But we **never observe the counterfactual** (what would have happened otherwise).
- Causal inference is about approximating the **missing counterfactual.**

RCTs

- RCTs allow us to create the counterfactual world using
  - Random assignment of the treatment
  - Treated and controls are similar, other than one group received the treatment and the other one not

- As we saw last week, creating the right counterfactual is not always trivial!

RCTs vs. Observational data

Experiments (RCT)	**Observational Data**
We assign treatment randomly	Treatment is naturally occurring
Clean causal inference	Requires assumptions
Expensive, sometimes impossible	Cheap, abundant, covers many settings

Common Threat: Confounding

- A **confounder** is a variable that:
  - Affects treatment assignment
  - Affects the outcome
- Example:
  - We observe higher sales in stores that run promotions.
  - Does the promotion cause more sales?

Common Threat: Confounding

- A **confounder** is a variable that:
  - Affects treatment assignment
  - Affects the outcome
- Example:
  - We observe higher sales in stores that run promotions.
  - Does the promotion cause more sales?
    Maybe, or maybe stores promote when demand is high already.

The “Selection Problem”

- Customers receiving your marketing action are **not random**.
- Examples:
  - People who see an ad are more active shoppers.
  - Customers who redeem coupons are more price sensitive.
  - Sellers who adopt a new feature may be more sophisticated.
  - If we ignore selection: **we get biased estimates.**

Key Strategies with Observational Data

We try to replicate random assignment using a variety of techniques:

- Control variables / Regression
- Matching/Propensity Scores
- Difference-in-Differences (DiD)
- And many more

Approach 1: Regression with Controls

- We attempt to compare similar "units" by **adjusting for observed differences.**

$$Y_{it} = \beta Treatment_{it} + \gamma X_{it} + \epsilon_{it}$$

- Where:
  - X_{it} =confounders (price, category, seasonality,...)
  - **Key assumption:** We measured and controlled for **all confounders**.
  - **Risk:** Unobserved confounders → still biased.

- This is the simplest but weakest method.

Example: Coupon Campaign & Sales

- A retailer sends **coupons** to customers who are predicted to be **high spenders**.
- We want to estimate:
  - Treatment: Effect of receiving a coupon on spending
- So, we run a regression:

$$\text{Spend}_i = \beta Coupon_i + \gamma X_i + \epsilon_i$$

- Where X_i includes:
  - past purchases
  - demographics
  - category preferences
  - etc.

Example: Coupon Campaign & Sales

- A retailer sends **coupons** to customers who are predicted to be **high spenders**.
- We want to estimate:
  - Treatment: Effect *of receiving a coupon on spending*
- So, we run a regression:

$$\text{Spend}_i = \beta \cdot Coupon_i + \gamma X_i + \epsilon_i$$

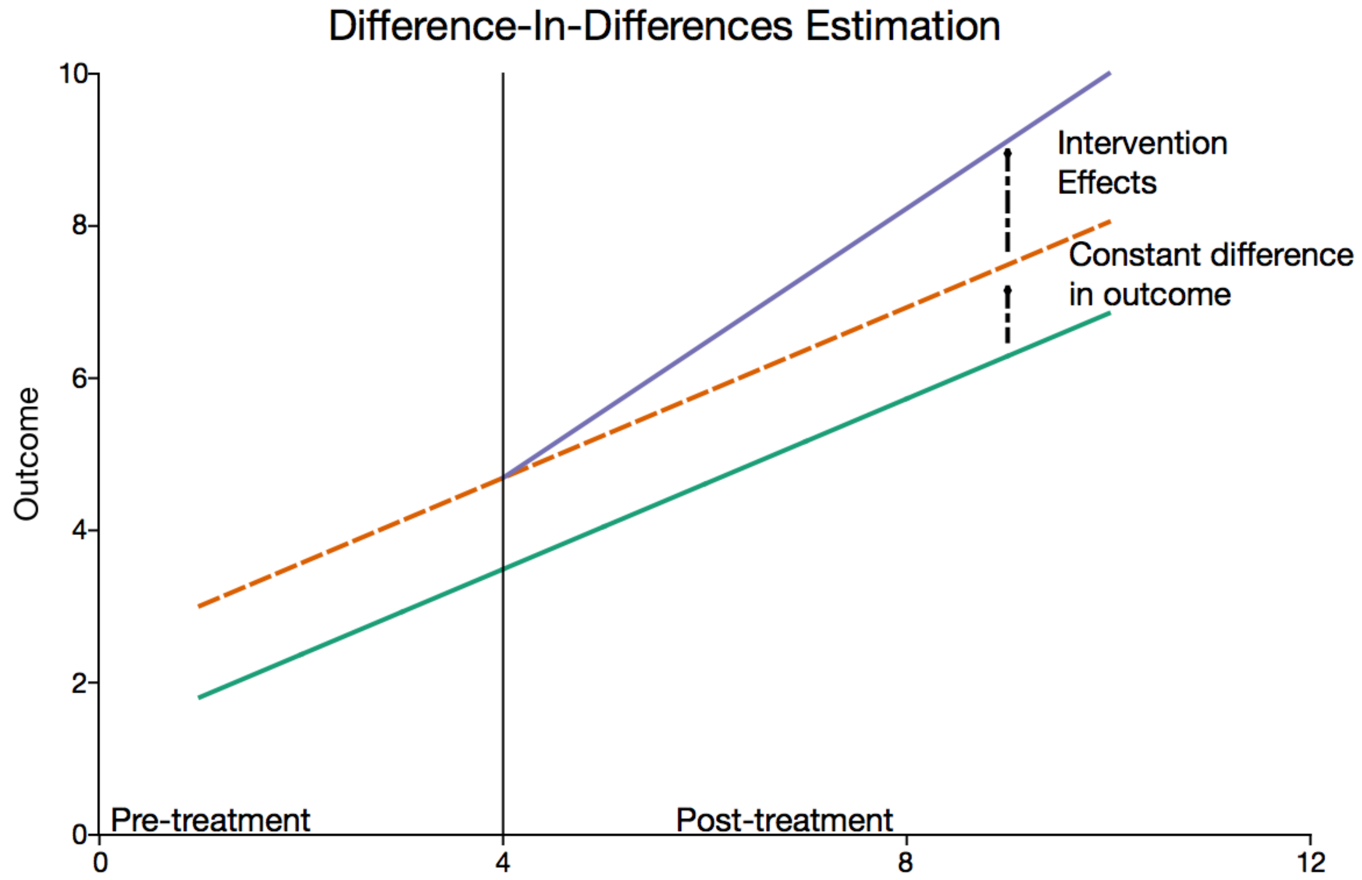
- Where X_i includes:
  - past purchases
  - demographics
  - category preferences
  - etc.
- **Problem:** The *reason* customers got the coupon is **their predicted future demand**, which we **cannot fully observe**.

Approach 2: Matching / Propensity Scores

- Goal: **Compare treated and untreated units that look similar before treatment.**
- Steps:
  - Estimate probability of receiving treatment (propensity score) based on observable characteristics
  - Match treated control units on these probabilities.
  - Compare outcomes.
  - **Good when:** We have many covariates that explain selection into the treatment.
- **Key assumption**: matching on observables accounts also for difference in unobservables

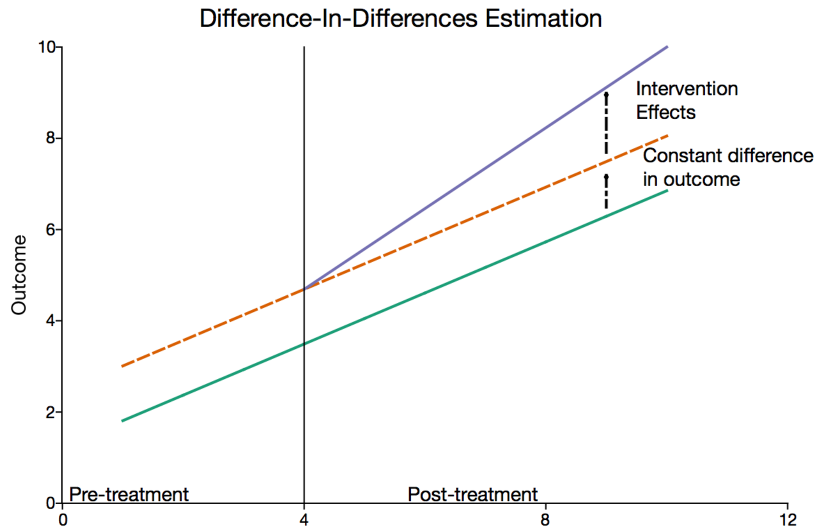
Approach 3: Difference-in-Differences

- Use **before/after** treatment variation + comparison to a control group



Approach 3: Difference-in-Differences

- Use **before/after** treatment variation + comparison to a control group



- **Key assumption:** Outcomes for treated and control groups would have evolved in the same way in the absence of the treatment **parallel trends** without treatment.

Approach 3: Difference-in-Differences

DiD suitable when:

- Treatment happens at a clear point in time (a policy change, product launch, feature rollout).
- Only some groups are affected, while others are not.
- We observe outcomes for both groups *before* and *after* treatment.

Approach 3: Difference-in-Differences

Potential questions that could be answered with DiD

- Shor-term rental policy impact on house prices
  - Some cities but not others ban Airbnb
- Impact of sustainability badges on sales
  - Amazon adds a new sustainability badge to some products but not others
- Impact free shipping affect sales
  - A retailer introduces free shipping for certain regions only
- Responses to reviews impact on future ratings
  - Some restaurants respond to reviews but not others

Approach 3: Difference-in-Differences

Potential questions that could be answered with DiD

- Shor-term rental policy impact on house prices
  - Some cities but not others ban Airbnb
- **Impact of sustainably badges on sales**
  - **Amazon adds a new sustainability badge to some products but not others**
- Impact free shipping affect sales
  - A retailer introduces free shipping for certain regions only
- Responses to reviews impact on future ratings
  - Some restaurants respond to reviews but not others

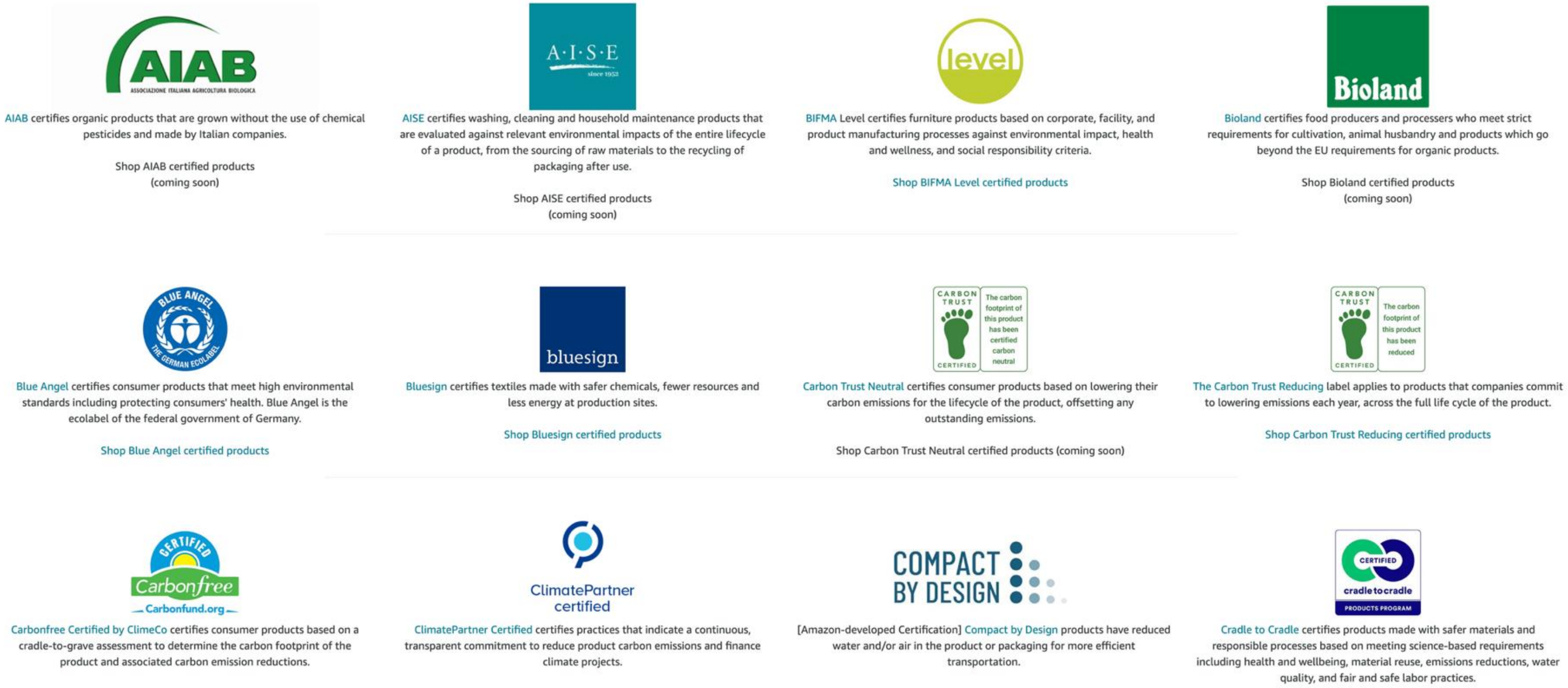
Example: Amazon’s Climate Pledge Friendly

- What is Climate Pledge Friendly (CPF)?
  - Amazon shopping program that helps customers discover products with sustainability features

Climate Pledge Friendly (CPF)

Amazon shopping program that helps customers discover products with sustainability features

- Qualification based on **60+** trusted sustainability certifications or Amazon's own certifications

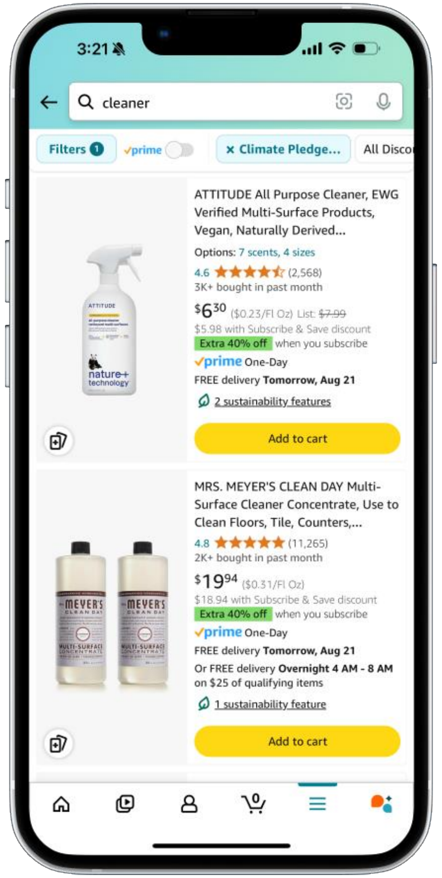


Climate Pledge Friendly (CPF)

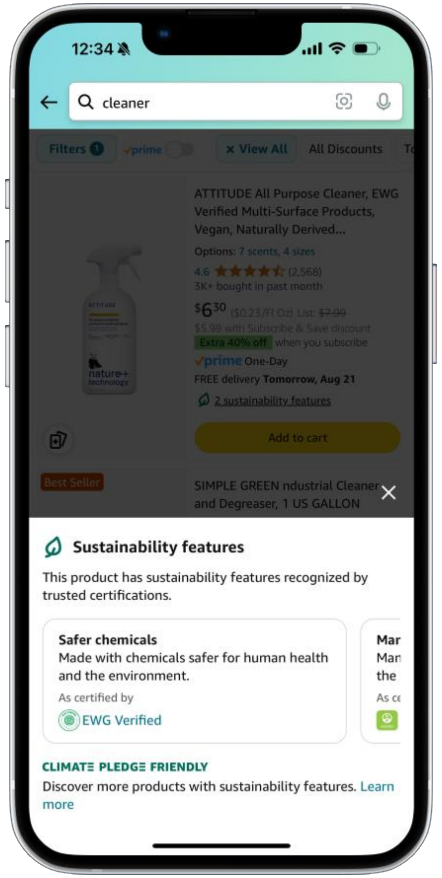
Amazon shopping program that helps customers discover products with sustainability features

- Qualification based on **60+** trusted sustainability certifications or Amazon's own certifications
- Available in **14** countries
- **2.2M** products so far
- **271,000** selling partners participating
- Customers can discover CPF products using a search filter, product recommendations, or a dedicated storefront

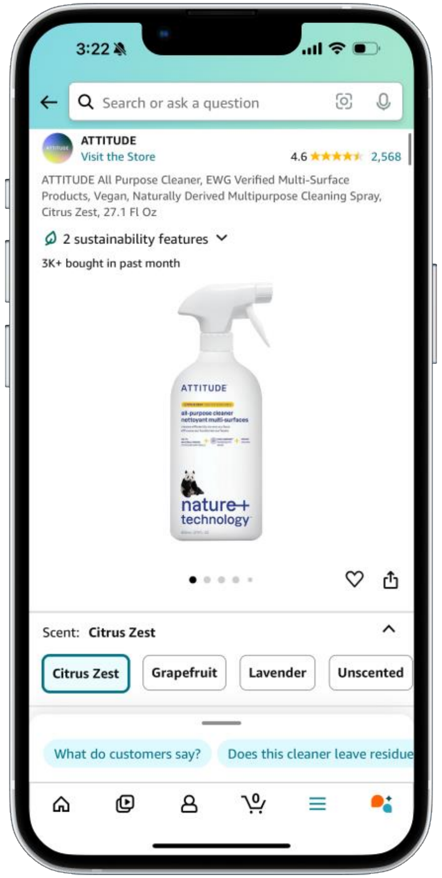
Climate Pledge Friendly (CPF)



Search



Click

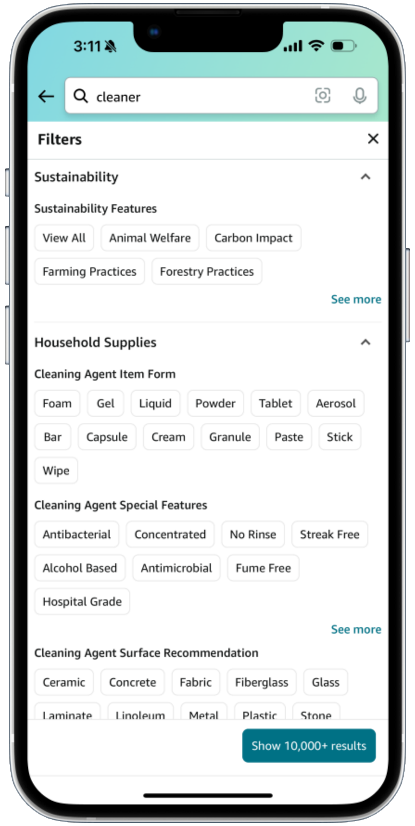


Product page

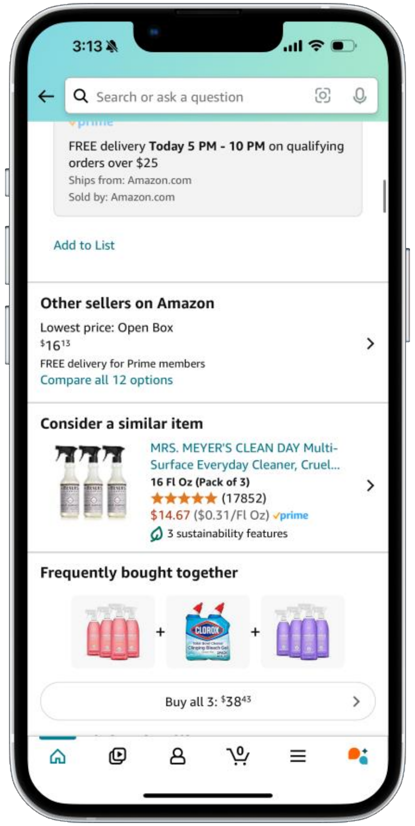
Climate Pledge Friendly (CPF)

Customers can discover CPF products using:

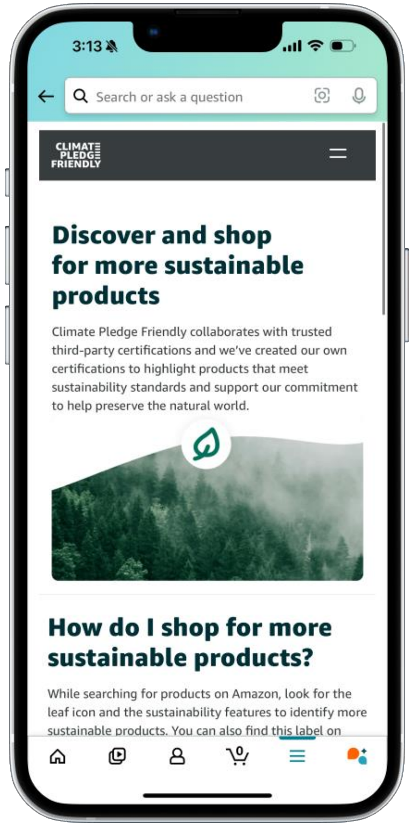
Filter



Reccs.



Dedicated store



The value-action gap

Gap between stated preferences and actual behavior

- While consumers claim to prioritize sustainability there is ongoing debate about the extent to which these stated preferences translate into real-world purchase decisions

The value-action gap

Gap between stated preferences and actual behavior

- While consumers claim to prioritize sustainability there is ongoing debate about the extent to which these stated preferences translate into real-world purchase decisions

Whether programs like CPF convert customer demand to purchases depend on several factors

- Comprehension and trust (Delmas and Grant, 2014)
- Cultural and political beliefs (Aneja et al., 2023; Kim and Liu, 2023),
- Willingness to pay

Amazon’s question

Does the CPF program **causally** affect consumer purchase behavior?

Estimating the causal impact of CPF

Estimating the causal impact of CPF

Main issues:

- (Unobservable) differences between CPF and not CPF products may affect purchase behavior
- Products joining CP may implement additional actions (e.g., marketing strategies) that increase sales

Estimating the causal impact of CPF

Main issues:

- (Unobservable) differences between CPF and not CPF products may affect purchase behavior
- Products joining CP may implement additional actions (e.g., marketing strategies) that increase sales

We try to address these issues in several ways

- Exploit CPF adoption → Difference-in-Differences
- Focus on a relatively short time window (12 weeks) before and after adoption to limit the possibility of confounders affecting the results
- Account for several factors affecting product performance

Data

Random sample of products from two populations

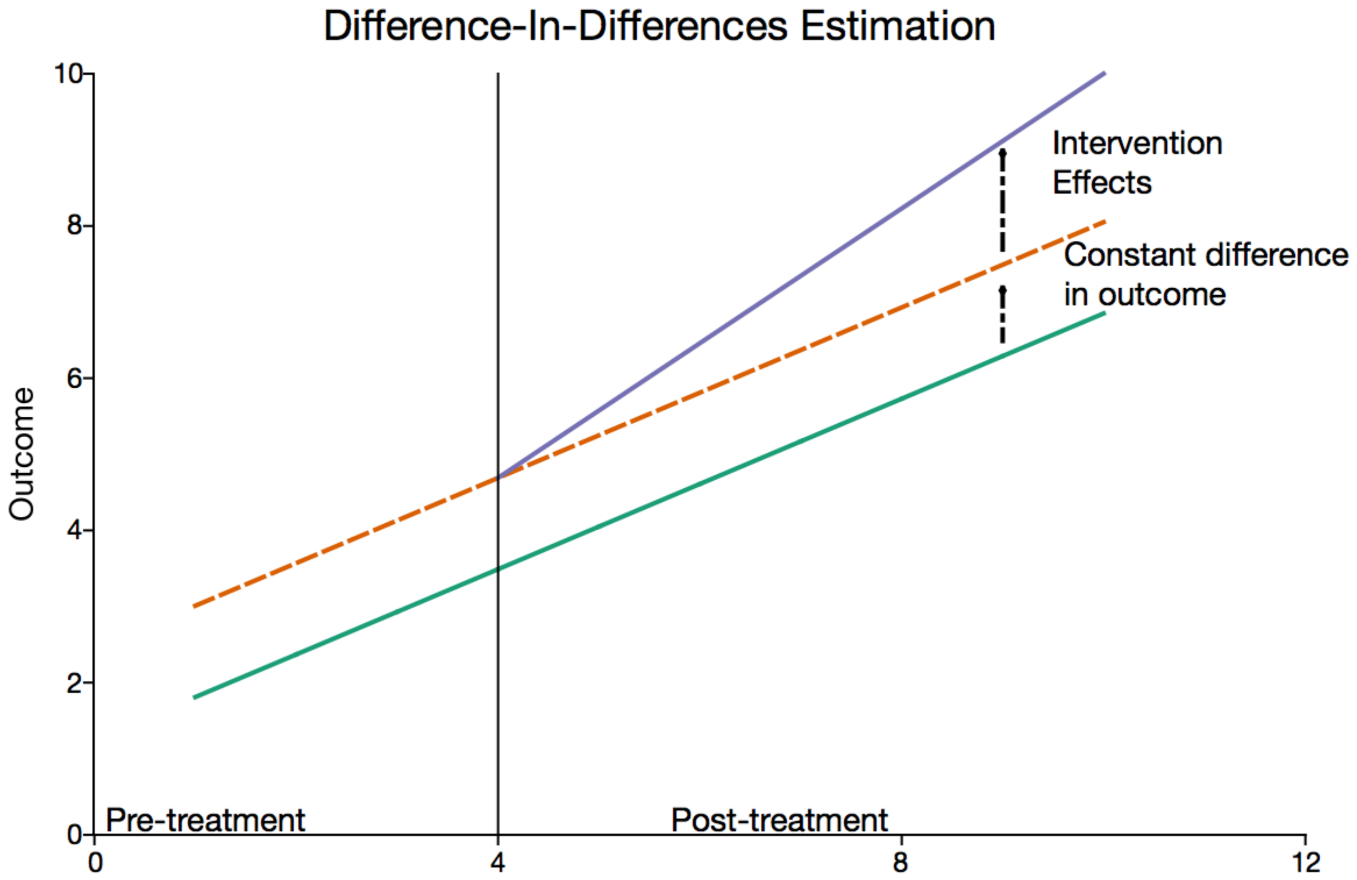
- Products eventually joining the program
- Products that never joined the program

Dataset: **~45k** products (**~36k** eventually CPF) across consumables, hardlines, and softlines

- Weekly data from July 2021 to January 2024
- About 4M product-week observations
- Two outcomes: Weekly Sales and Net Shipped Units (units sold)
- A lot of controls: price, discount, ad spend, ratings and reviews

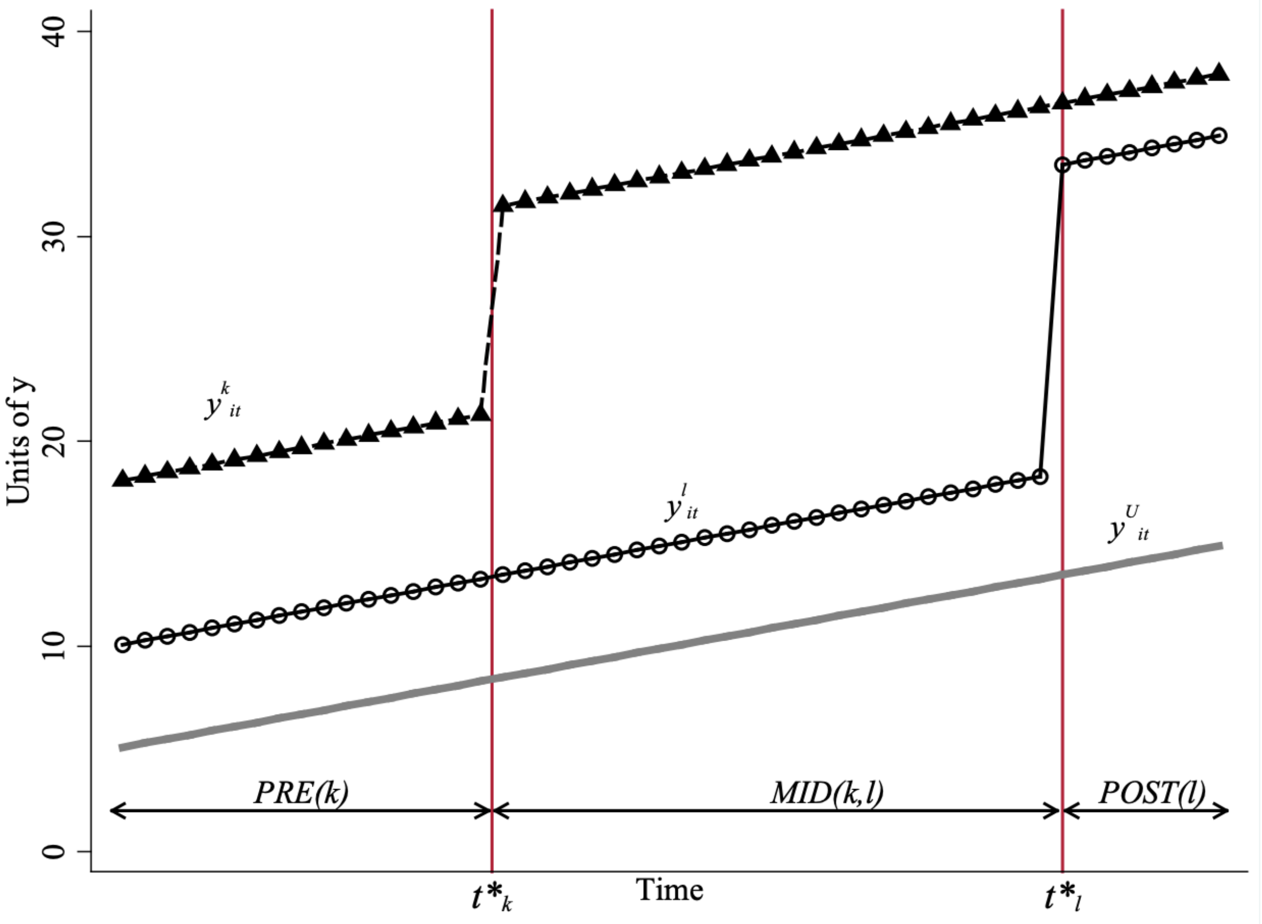
Approach 3: Difference-in-Differences

- Use **before/after** treatment variation + comparison to a control group



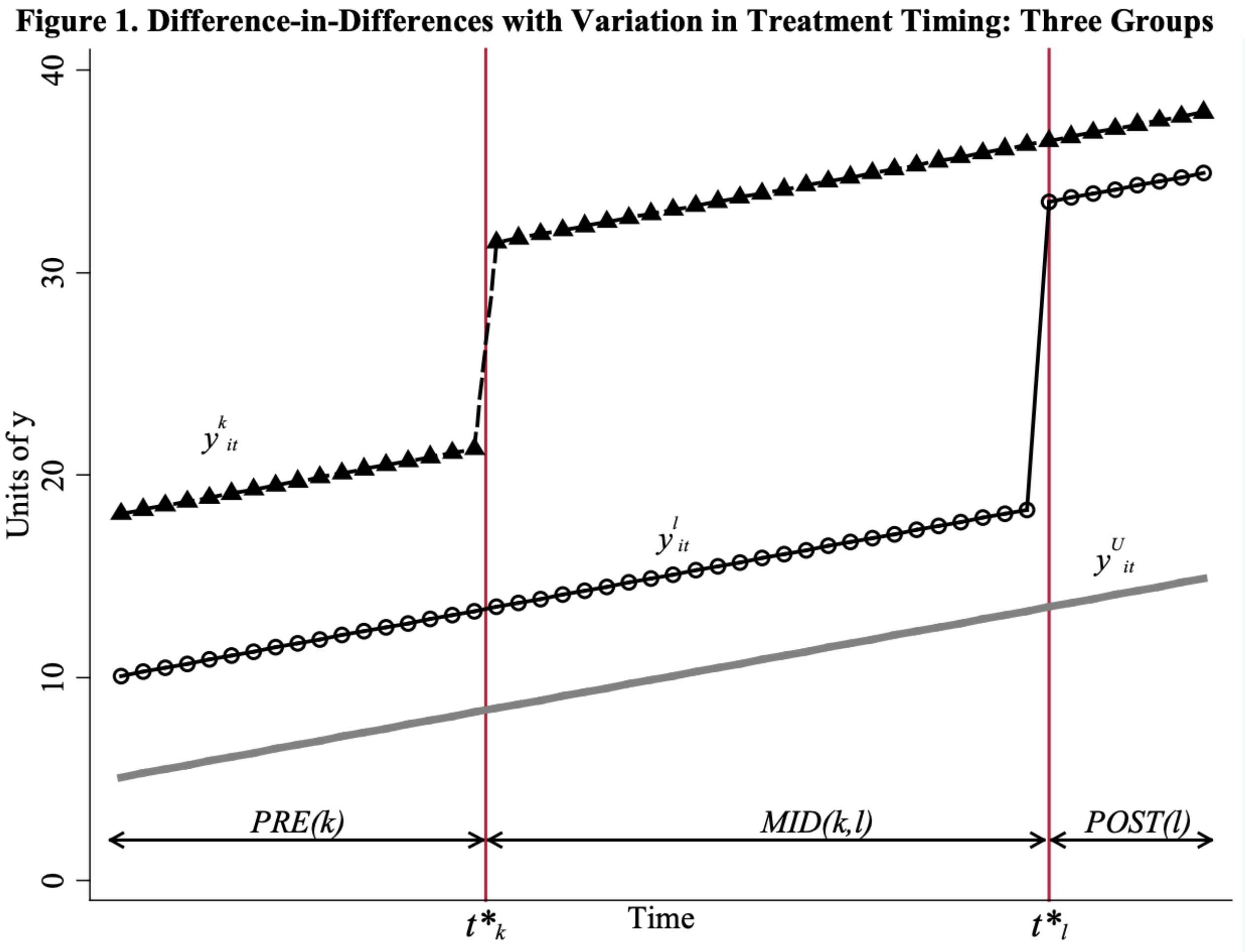
Identification strategy: Staggered DD

Figure 1. Difference-in-Differences with Variation in Treatment Timing: Three Groups



* Figure from Goodman-Bacon (2021)

Identification strategy: Staggered DD

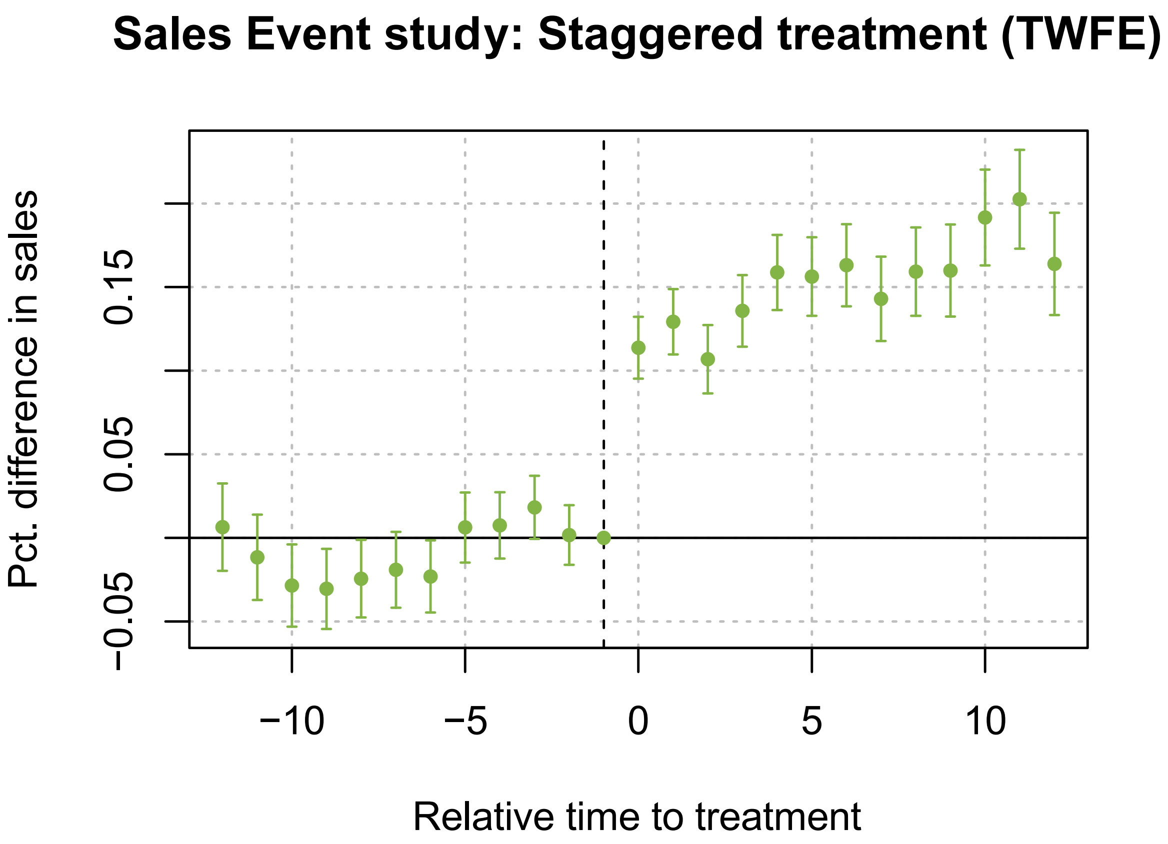


Both nevertreated and not-yet-treated act as controls

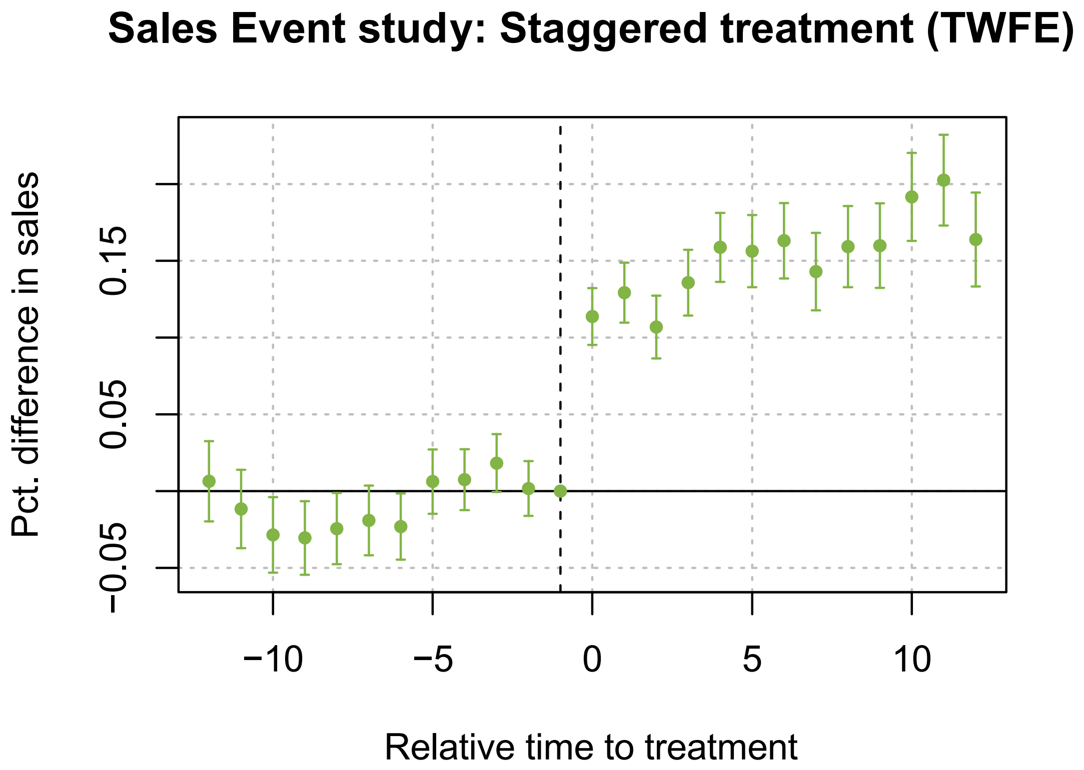
* Figure from Goodman-Bacon (2021)

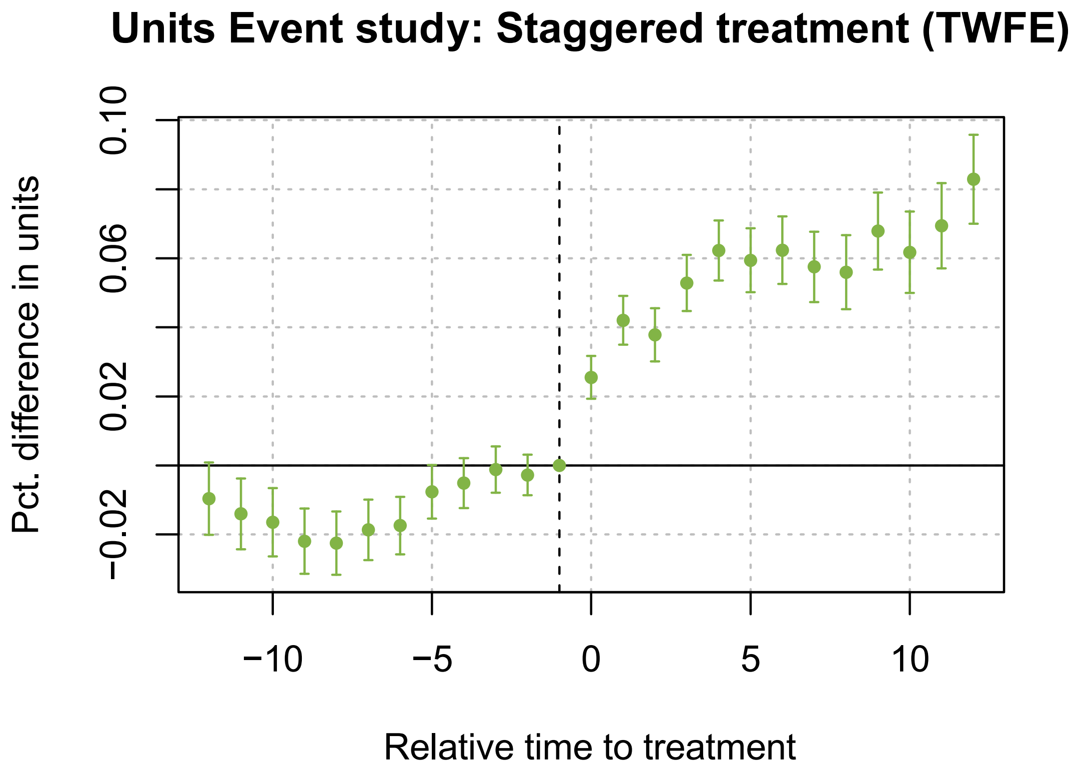
Results: Differences in outcomes between treated and control groups

Results: Differences in outcomes between treated and control groups



Results: Differences in outcomes between treated and control groups





Main results: Estimates w/o controls

	log Sales	log Units
	(2)	(3)
After	0.146***	0.059***
	(0.009)	(0.004)
Observations	1,365,441	1,365,441
R^2	0.663	0.777

Note: *p<0.1; **p<0.05; ***p<0.01

Estimates obtained using TWFE. All models include product and year-week fixed effects. Standard errors reported in parentheses are clustered at the product level.

Main results: Estimates with controls

Controls: Price, promotion, ad spend, ratings and reviews

Main results: Estimates with controls

Controls: Price, promotion, ad spend, ratings and reviews

	log Sales	log Units
	(2)	(3)
After	0.118***	0.043***
	(0.008)	(0.003)
Observations	1,242,984	1,242,984
R^2	0.704	0.813

Note: *p<0.1; **p<0.05; ***p<0.01

Estimates obtained using TWFE. All models include product and year-week fixed effects. Standard errors reported in parentheses are clustered at the product level.

Main results: Estimates with controls

Controls: Price, promotion, ad spend, ratings and reviews

	log Sales	log Units
	(2)	(3)
After	0.118***	0.043***
	(0.008)	(0.003)
Observations	1,242,984	1,242,984
R^2	0.704	0.813

Note: *p<0.1; **p<0.05; ***p<0.01
Estimates obtained using TWFE. All models include product and year-week fixed effects. Standard errors reported in parentheses are clustered at the product level.

Differences ranging from **~20%** for
sales to **27%** for shipped units

Can we do better?

Cross-market DD

One of the main concerns with main DD is that results may be driven by unobservable differences between CPF and non-CPF products

This strategy's goal is to reduce this type of concern

Cross-market DD

We compare outcomes for the **same products** across two market (countries)

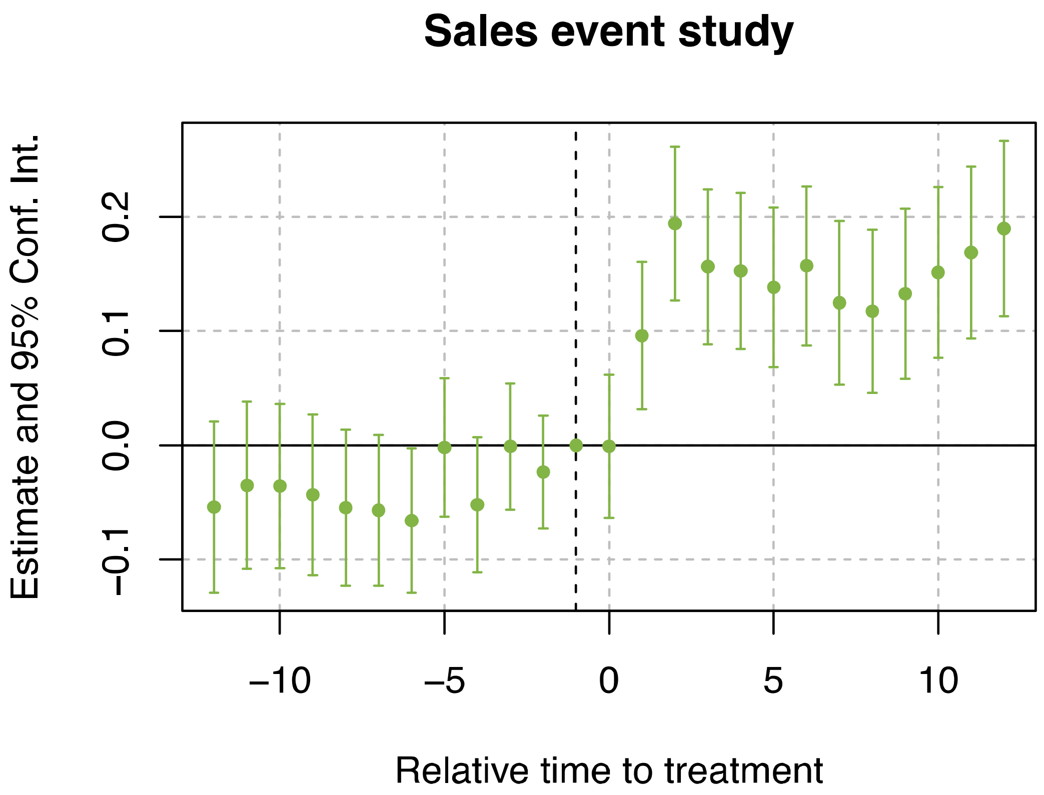
- UK and Germany

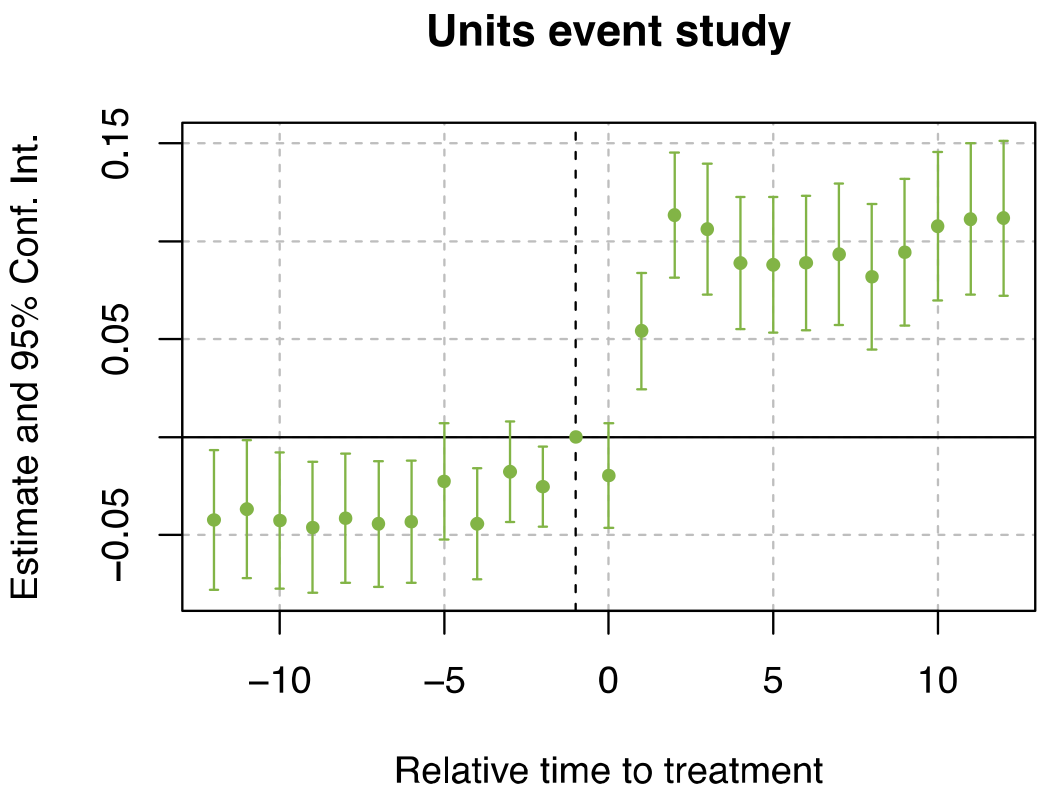
Each product is CPF in once country but not the other

- Products need to certify in each country
- Some certifications are country-specific due to different standards
  - E.g., Carbon certification

- Note that estimated effects are LATE (Local ATE)
  - Effect for products that comply with the selection criteria

Cross-market DD





Cross-market DD: Results

	log Sales	log Units	log Sales	log Units
	(2)	(3)	(5)	(6)
After × Treated	0.170***	0.119***	0.086***	0.077***
	(0.023)	(0.012)	(0.016)	(0.008)
After	0.001	−0.022**	0.012	−0.017**
	(0.019)	(0.010)	(0.015)	(0.007)
Treated	0.882***	0.486***	0.134***	0.045***
	(0.037)	(0.020)	(0.021)	(0.011)
Controls	No	No	Yes	Yes
Observations	167,040	167,040	166,869	166,869
R^2	0.585	0.649	0.765	0.825

Note: *p<0.1; **p<0.05; ***p<0.01

Estimates obtained using the cross-country DD. All models include product-pair and year-week fixed effects. Standard errors reported in parentheses are clustered at the product level.

Where CPF works best

Where CPF works best

CPF products are part of three categories

1. **Consumables** are non-durable products that are regularly consumed and replenished such as beauty, health and personal care, cleaning supplies, and grocery.
2. **Hardlines** are durable goods typically made of rigid materials like plastic, metal, and wood such as small appliances, hardware, automotive parts, sporting goods, and toys.
3. **Softlines** are products related to fashion including apparel and accessories. Examples include clothing, footwear, handbags, and luggage

Where CPF works best

Consumables experience larger effects (similar to Borin et al., 2011)

Why?

- Consumables are non-durable products such as food and personal care items
- These types of products emphasize consumer safety and require stringent quality controls → ensure products are safe
- CPF may be seen by consumers as an additional check that validates the quality of these products

Conclusions

CPF program does affect consumers purchase decisions

Including controls, we observe a 12.5% increase in sales and 4.5% increase in net shipped units in the 12 weeks after joining CPF

Making it easier for consumers to discover more sustainable products

- Allows them to make purchases that are more aligned with their values
- Benefits brands who join sustainability programs

Key Takeaways

- Observational data can be used for causal inference, but it’s hard.
- Main challenge: **selection bias** / confounding due to non-random assignment
- Methods differ mainly in **how they reconstruct the counterfactual**.
- Always **diagnose assumptions** and show evidence they likely hold