

Louise Gatty
Alice Petillon
Charles Pyle
Antonio Raphaël
Anne Thébaud

Research and Data Plan

1. Motivate and introduce your research question.

(a) What is the research question?

Do prison approved vendors supplying incarcerated individuals practice higher markups on basic consumer goods compared to equivalent products sold in conventional retail markets (e.g. Walmart), and if so, to what extent?

(b) Provide context or background knowledge required to understand why this question is important.

A single contractor holds monopoly power over prison commissary services (food, hygiene products...) in the correctional facilities in California, meaning incarcerated individuals have no ability to choose among competing retailers. This raises concerns about excessive markups on those services.

Comparing prison commissary prices to those observed in competitive retail markets provides a way to assess whether monopoly power translates into systematically higher prices.

(c) Who cares about your research question? Why?

The issue raises ethical and distributional concerns. The question is therefore relevant to policy makers, prison administrators, consumer protection advocates, families of incarcerated individuals, economists studying market power and regulation, etc. More broadly, it informs public debate on the consequences of monopoly power for prices and consumer welfare.

(d) Why is this question not trivial?

We do not have strong doubts that markups are practiced in prison commissaries; however, the magnitude and structure of these markups are not trivial. In particular, it is unclear how large price differences are once comparable products are carefully matched across markets, and whether markups are uniform or vary systematically across product categories.

(f) Procure at least four academic or other (forums/newspaper/etc.) references around this topic: summarize them and use them in a way to showcase how they complement your work or raise questions which you will discuss.

We were not able to find references directly addressing product-level price markups in prison commissaries using systematic data. However, several studies provide important context by documenting the broader economic and social environment in which prison commissary markets operate. The [Sentencing Project](#) shows that mass incarceration deepens economic inequality and disproportionately affects low-income households. [Research from the Institute for Research on Poverty](#) highlights strong links between poverty, incarceration, and inequality, emphasizing how financial burdens associated with incarceration extend beyond prisoners themselves to their families.

These works do not measure commissary pricing directly, but they motivate our analysis by raising the question of whether monopoly pricing in prison commissaries constitutes an additional, under-documented channel through which incarceration contributes to economic inequality.

2. Describe your data collection strategy.

a.) What is the data source that you will use? What is the actual purpose of this data (i.e. platform for searching services xy)?

We will use two primary sources of data for this project. The first is the Walkenhorst Family Visits Catalogue for California Department of Corrections and Rehabilitation (CDCR) family visits. This is a PDF document which contains primarily food items, the brand of the item, the product name, the weight of the product in ounces, and how much it costs. This platform is for friends and family of CDCR inmates to purchase items that go into care packages for incarcerated individuals. Some catalogues are for quarterly deliveries to inmates, the catalogue we have chosen is specifically for items that are brought to an inmate during a family visit. Walkenhorst itself is a CDCR vendor/contractor that is the sole vendor for people to send food, clothes, and toiletries to incarcerated individuals.

The second data source will be scraped prices data from large US commercial vendor such as Target, Walmart, or Foodmax. We will scrap food prices and weights from a large, competitive market firm website using a California IP address. These websites are generally used for placing online orders.

b.) How do you get access? Will you scrape the data, or will you use an application-programming interface (API)?

We will access the Walkenhorst data by conducting PDF scraping to transform the PDF into a structured data set. We will access the market food prices via webscraping.

c.)What is one observation in your data (e.g. a price quote, a review, an appointment)? What are the different dimensions of the data (car model, doctor, city, time, versions,etc.)? Are the dimensions the same across all the data you collect?

One observation will be the price per ounce for a given product charged by the market and charged by Walkenhorts, the dimensions will be individual products at the given point in time that we scrape data from the website. The dimensions will be the same for all observations.

d.) Which is the information you will extract? What is the list of variables that you will have in the columns of a table you use for the analysis? Please provide a screenshot where this becomes more apparent.

Ideally an observation of our data will have the brand, the item, the Walkenhorst price, the “market” price, the weight in ounces, a transformed variable price per ounce, the type of food item (cracker, candy, cookie, cooking), whether or not the item is kosher, and potentially nutritional information.

2025 FAMILY VISITATION CATALOG			WALKENHORST'S		
Item #	Description	Price	Item #	Description	Price
3111-153	Charleston Chew Chocolate 2.1 oz.	\$1.25	86383-173	Ice Breakers Cinnamon Mints 1.5 oz.	\$2.95
3111-151	Charleston Chew Strawberry 2.1 oz.	\$1.25	86383-172	Ice Breakers DUO Watermelon Mints 1.3 oz.	\$2.95
3111-152	Charleston Chew Vanilla 2.1 oz.	\$1.25	86383-191	Ice Breakers Sparkling Pineapple Mango Seltzer Mints 1.5 oz.	\$2.95
92426-110	Charms Fluffy Stuff Cotton Candy - Original 3.5 oz.	\$2.75	22669-282	Jell-O Pudding Cups 1.55 oz.	\$1.45
22669-237	Chocolate Lovers Mini Variety Pack 8.9 oz	\$6.95	22669-281	Jell-O Pudding Cups 3.5 oz.	\$2.75
86383-085	Chocolatey Payday King Size 3.1 oz.	\$2.45	22669-285	Jell-O Sour Candy Squares Berry Blue 4.5 oz.	\$2.45
62267-250	Chunky 1.4 oz.	\$1.25	86383-300	Jolly Rancher Assorted Fruit Flavors 7 oz.	\$3.25
92426-018	Cookie Dough Bites Chocolate Chip 3.1 oz.	\$2.25	86383-083	Jolly Rancher Gummies Original 7 oz.	\$2.95
62267-233	Crunch 1.55 oz.	\$1.25	86383-084	Jolly Rancher Gummies Sours 6.5 oz.	\$2.95
58182-199	Crunch Buncha Crunch 8 oz.	\$6.45	86383-175	Jolly Rancher Gummies Very Berry 6.5 oz.	\$2.95
62267-251	Crunch Fun Size 10 oz.	\$6.75	86383-176	Jolly Rancher Hard Candy Tropical 6.5 oz.	\$2.95
58421-017	Darrell Lea Rocklea Road Block 6.4 oz	\$4.95	46266-276	Jovy Chili Covered Mango & Watermelon Candy 6 oz.	\$2.50
58182-002	Dove Dark Chocolate Bar 1.3 oz.	\$1.25	3110-711	Joyva Kosher Halvah Marble 1.75 oz.	\$1.25
58182-003	Dove Milk Chocolate Bar 1.44 oz.	\$1.25	92426-149	Junior Mints 2.6 oz.	\$2.45

[Shop all RITZ](#)

RITZ Fresh Stacks Zesty Herb Crackers Travel Snacks - 11.8oz

★★★★★ 222 ∨

\$4.29 (\$0.36/ounce)

See 1 deal for this item



Label info

Serving Size: 16 g	
Serving Per Container: about 21	
Amount Per serving:	
Calories: 80	
	% Daily Value*
Total Fat 4g	5%
Saturated Fat 1g	5%
Trans Fat 0g	
Cholesterol 0mg	0%
Sodium 100mg	4%
Total Carbohydrate 10g	4%
Dietary Fiber 0g	0%
Sugars 1g	
Added Sugars 1g	2%
Protein 1g	
Vitamin D 0mcg	0%
Calcium 20mg	0%
Iron 0.6mg	4%
Potassium 20mg	0%
* Percentage of Daily Values are based on a 2,000 calorie diet.	

Ingredients:

Unbleached Enriched Flour (Wheat Flour, Niacin, Reduced Iron, Thiamine Mononitrate [Vitamin B1], Riboflavin [Vitamin B2], Folic Acid), Soybean And/Or Canola Oil, Palm Oil, Sugar, High Fructose Corn Syrup, Oat Fiber, Leavening (Calcium Phosphate, Baking Soda), Salt, Dried Onion, Dried Garlic, Dried Parsley, Dried Thyme, Dried Marjoram, Dried Oregano, Soy Lecithin, Natural Flavor, Citric Acid.' '

Allergens & Warnings:

CONTAINS A BIOENGINEERED FOOD INGREDIENT, CONTAINS: WHEAT, SOY.

e.) What is your strategy to create the sample and in order to increase the number of data points? How and by what do you filter the data (e.g. some category, a set of cities, etc.)? How many data points do you expect to collect in the end?

The maximum number of data points for the analysis will be around ~3,600 (an estimated number of items in the Walkenhorst catalogue). However we will scrape far more observations from companies such as Target and Walmart. Rather than try and target the exact same items from the catalogue, we will scrape all items by category, and then match products by name.

f.) Will your data coverage be complete? If your coverage is not complete, what is the rate of coverage, and is it "representative" or the "x percent most relevant"?

Our data coverage should be fairly complete, we will not target the full range of products offered by Walkenhorst, as the full PDF catalogue is too difficult to scrape as it lacks structure, however we should have fairly comprehensive coverage of all of the different food items offered by Walkenhorts. It should be representative as there is only one CDCR vendor that offers this service, and the data we will scrape will represent the total "population." The prices data scraped from a company like Walmart should also be representative of market prices.

3. Describe your first ideas for a possible outline for the analysis.

(a) Do you expect any steps to pre-process the data that might be time-consuming?

Yes.

Extracting structured data from the Walkenhorst PDF and from a retailer's website will require substantial cleaning. We may face inconsistent text formatting, line breaks, and merged columns. We will need to manually verify that product names, brands, prices, and weights are correctly parsed and aligned, and potentially standardise units (eg: ounces vs multi-pack weights).

Also, matching both datasets will be a challenging step as the names, packaging sizes and brands of the products may differ slightly between Walkenhorst and Californian retailers. We will therefore need to do keyword-based matching (or fuzzy matching) and verify manually, which may be time consuming.

We may also need to do price normalising since package sizes differ across sellers. We will need to compute a common metric (price per ounce). We may also need to categorise products (eg: biscuits, sweets, premade meals) to later be able to analyse the price differences by category.

(b) What are the outcomes that you want to study, and by which variation in the data do you want to explain it?

The primary outcome of interest is the price per ounce of a product, and more specifically the price differential between Walkenhorst and the competitive market for comparable goods.

Variation comes from the fact that we have :

- Two types of venders: one that can do monopolistic pricing and one that faces competitive market pricing. We therefore expect to have variation in the prices.
- Different product characteristics: different categories of food, different nutritional information, different brands. These also influence the prices.

Conceptually, we are asking whether holding product characteristics constant, prices are systematically higher in the monopolistic prison vendor market, and whether this difference varies across product types.

We could also look at the markups of Walkenhorst, assuming that $P = MC$ in the competitive market equilibrium (following equation), and seeing how this changes through the different categories/brands of goods.

$$\text{Markup (Walkenhorst, \%)} = \frac{P_{\text{Walkenhorst}} - MC_{\text{good}}}{MC_{\text{good}}} \times 100 = \frac{P_{\text{Walkenhorst}} - P_{\text{competitive retailer}}}{P_{\text{competitive retailer}}} \times 100$$

(c) What Kind Of Tables Would You Produce? What kind of visualization could be insightful?

We will produce boxplots to compare prices per ounce for Walkenhorst and market vendors across different product categories, which will allow us to visualize price differences and markups. We may also include bar charts showing average markup by category for a clear summary of relative price differences. Additionally, we will produce a table reporting regression results to formally test the effect of the vendor on price. An example of a regression to test this could be :

$$\text{PricePerOunce}_i = \alpha + \beta_1 \text{Walkenhorst}_i + \sum_k \gamma_k \text{Category}_{ik} + \sum_j \delta_j \text{Brand}_{ij} + \sum_m \theta_m \text{Nutrition}_{im} + \phi \text{Size}_i + \epsilon_i$$

We will test the significance of β_1 to determine whether Walkenhorst prices are systematically higher than market prices.

(d) Finally, describe a possible result of your study, or a type of result, and why you think it is interesting. This could be a public policy recommendation or managerial implication, for example.

A plausible result is that Walkenhorst prices are significantly higher per ounce than comparable products sold in competitive retail markets, even after controlling for product category, brand, nutritional information, size of packs... We may also find that markups are particularly large for certain categories (e.g. snacks or specialty items) or for goods with fewer close substitutes (the number of substitutes could be proxied as the number of items in the category).

This result would be interesting because it provides empirical evidence of how monopolistic market structures in carceral settings can lead to higher consumer prices, disproportionately affecting incarcerated individuals and their families. From a public policy perspective, the findings could support arguments for:

- Increased price regulation of prison vendors,
- Allowing multiple vendors to operate within correctional systems,
- Or greater transparency and oversight in state procurement contracts.

More broadly, the study would contribute to discussions about the economic burden placed on incarcerated populations and their families, highlighting how lack of competition can translate into materially worse outcomes even for basic consumer goods.