

Big Data Sources and the Telephone Point of Purchase Survey

Greg Barbieri

Economist

Consumer Prices and Price Indexes

FedCASIC

5/2/2016

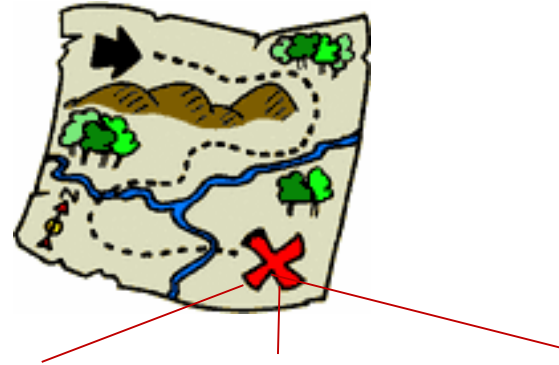


What is TPOPS

TPOPS generates the stores where prices are collected for the CPI

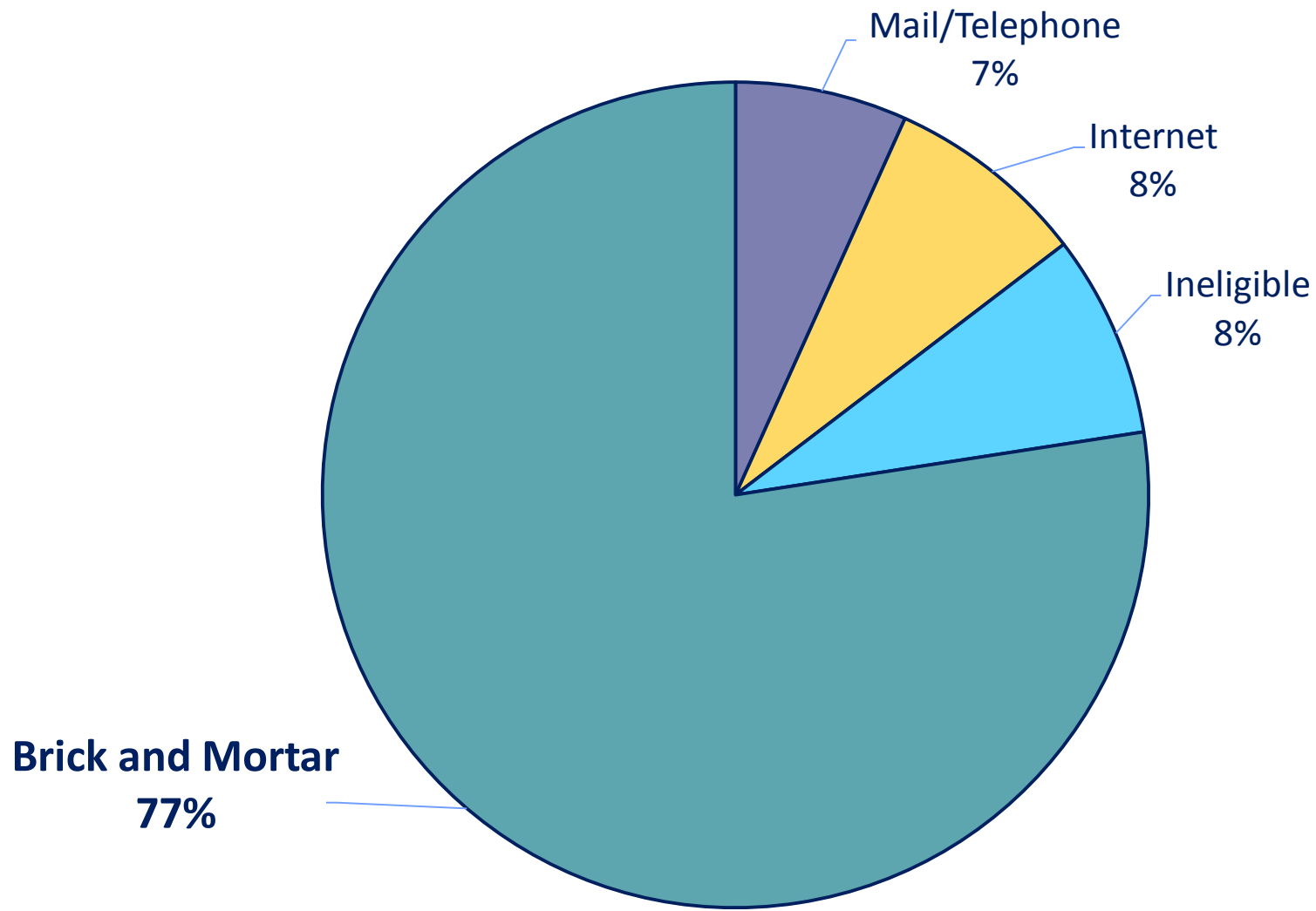
Store Location:

- ▶ Internet
- ▶ Telephone or Mail Order
- ▶ Brick and Mortar



Street Address, Cross Street, City, State

Retail Landscape



BLS Address Refinement

Stages of Refinement:

1. Shopping Center List
2. Outlet Co
3. Field Refir
4. Outlet Disaggregation

TPOPS
Report

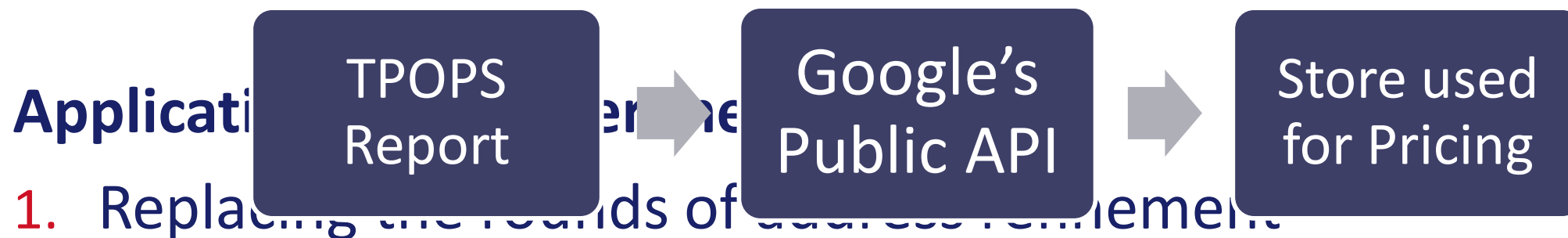
BLS Address
Refinement

Store used
for Pricing

Role of Big Data Sources

Tapping into Big Data sources (Google, Yelp, etc.)

- ▶ Name, address, geographic coordinates

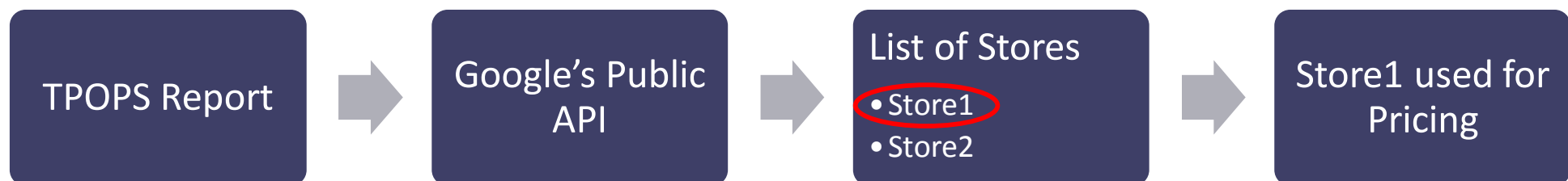


1. Replacing the records of address identification
2. Selecting stores by city and state

Role of Big Data Sources

Applications and Experiments:

2. Selecting stores by city and state



Experimental Data

- Mock TPOPS data were created for 10 categories in 10 cities
- “Dirty” store name and address
 - ▶ Removing details
 - ▶ Typos: extra letters, transpose, deletion

Cities	Categories
<ul style="list-style-type: none"> • Miami, FL • New York, NY • Washington, DC • Seattle, WA • Detroit, MI • San Francisco, CA 	<ul style="list-style-type: none"> • Grocery Stores • Clothing Stores • Electronic Stores • Restaurants • Legal Services • Accounting Services

<u>Original Information</u>		<u>Mock TPOPS Report</u>	aters Stores
Alan C. Young & Associates, P.C.	Alan C Young Associates	Ala C Youkng Associatevs	
7310 Woodward Ave #740, Detroit, MI 48202, United States	Woodward Ave, Detroit, MI	Woodward Ae, Detroit, MI	

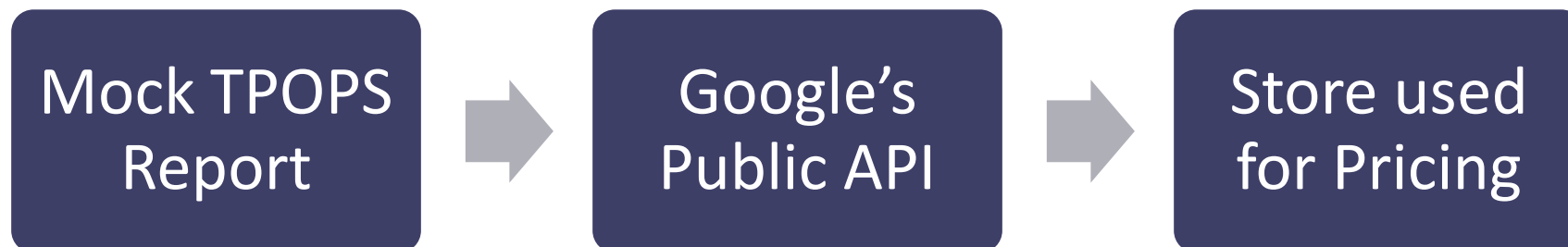


Expt 1: Replacing BLS Refinement

Search: Store Name, Street Address, City and State.

BD Source: Google

Data: Mock TPOPS data.

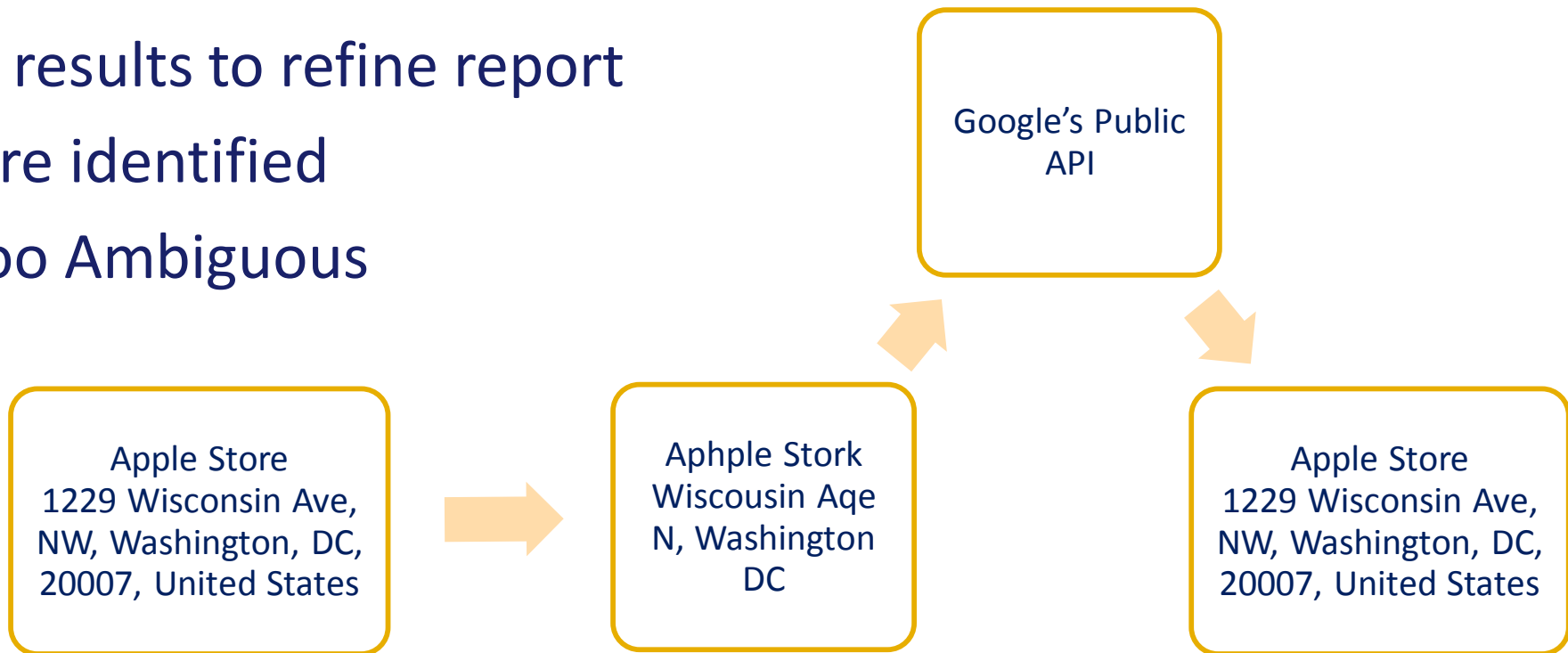


Expt 1: Measuring Results

Number of results:

- ▶ Zero: No results to refine report
- ▶ One: Store identified
- ▶ Many: Too Ambiguous

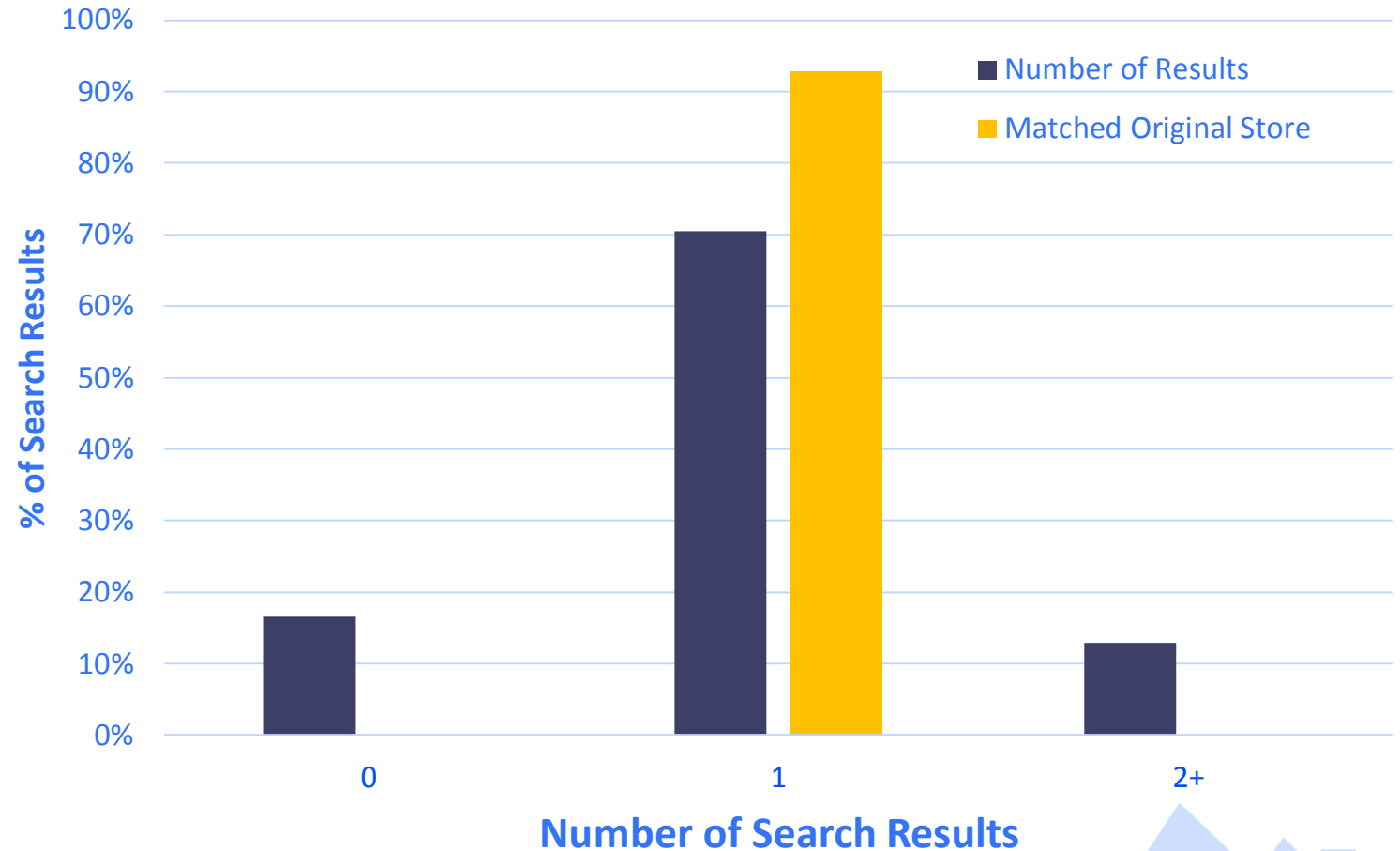
Matching:



Expt 1: Results

- Stores Searched: 1,917
- Overall Performance:
 - ▶ Zero: 16%
 - ▶ One: 71%
 - ▶ Many: 13%

Distribution of Search Results



Expt 2: Selecting Stores by City and State

Search: Store name, City, and State

Measuring Results:

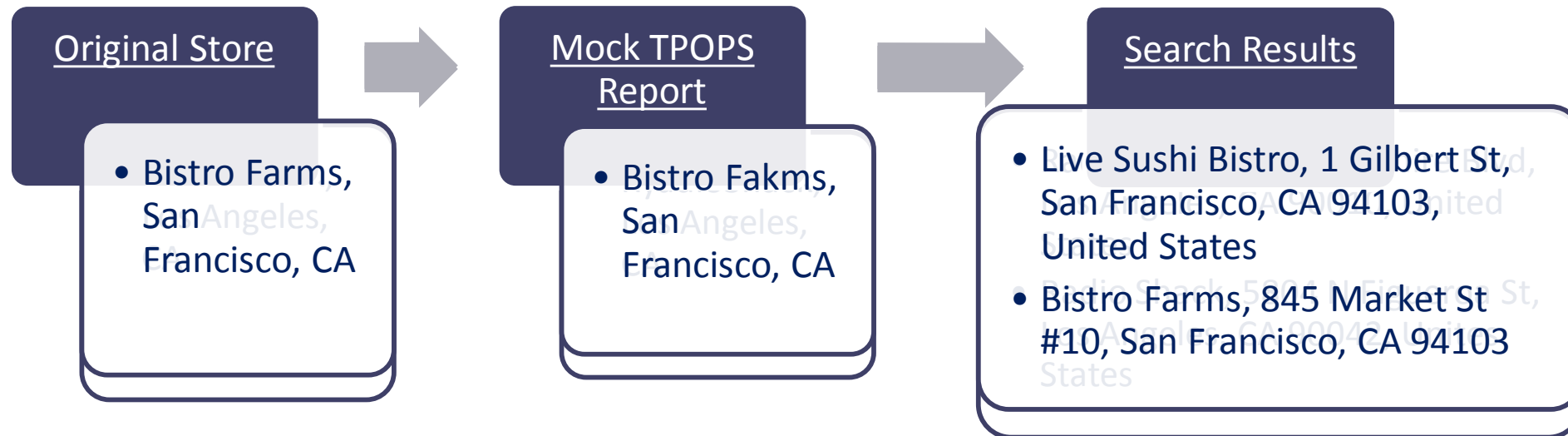
- ▶ Goal is to determine feasibility of selecting a store from many
- ▶ Focus will be on scenarios with multiple search results.



Expt 2: Measuring Results

Matching Criteria Changes:

- ▶ Matching using store name, city and state



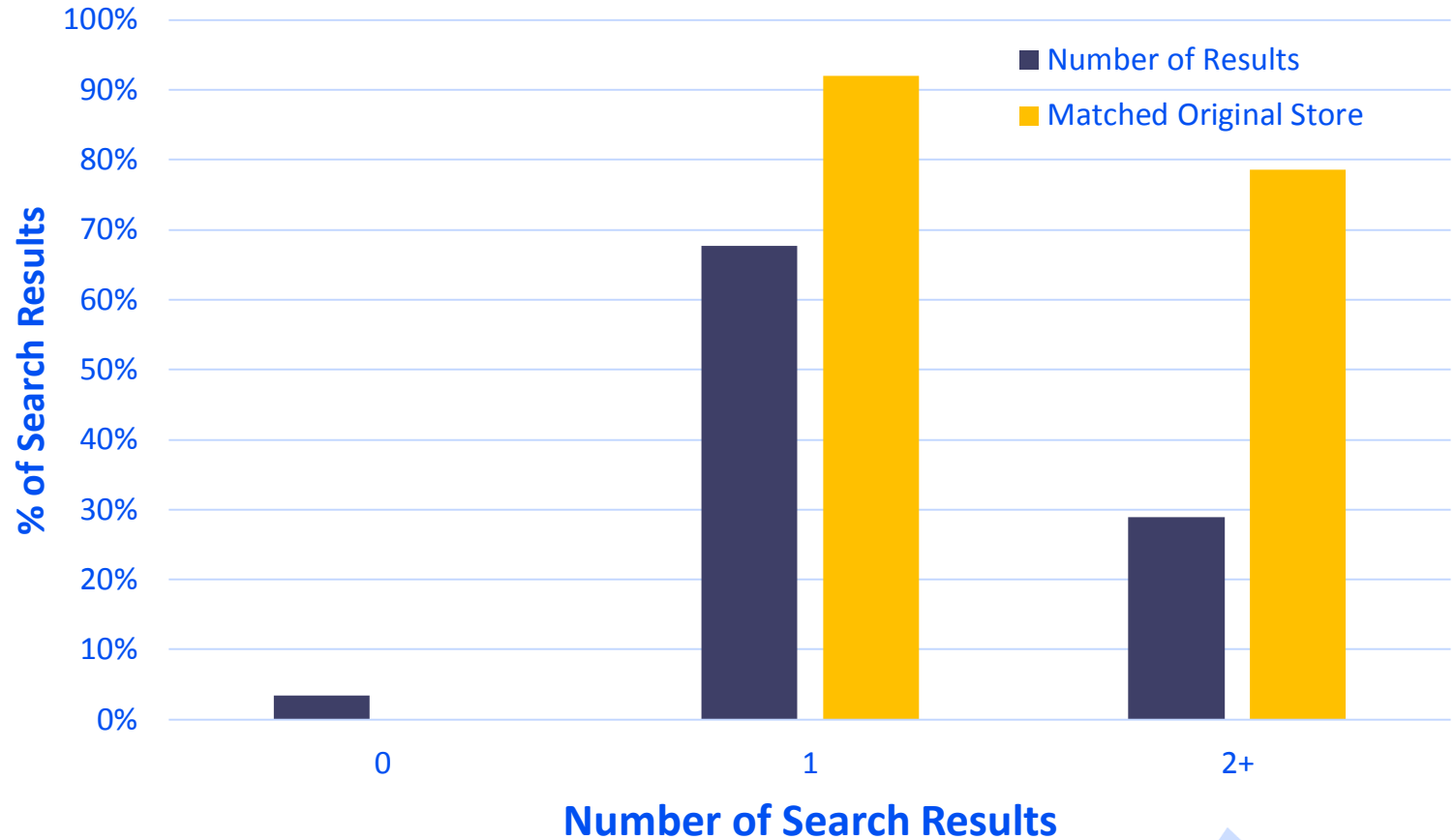
Expt 2: Results

■ Stores Searched: 1,907

■ Overall Performance:

- ▶ Zero: 3%
- ▶ One: 68%
- ▶ Many: 29%

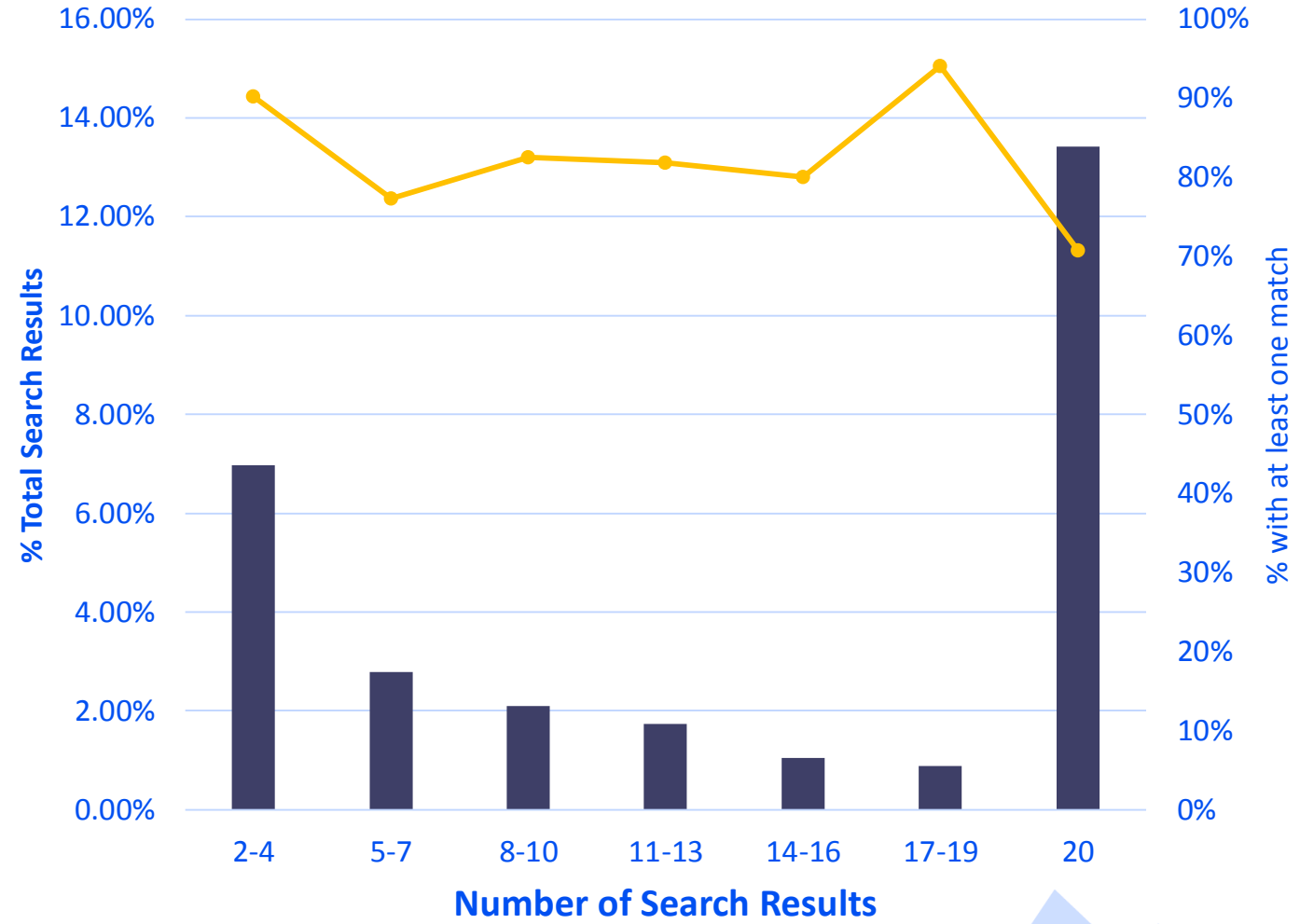
Distribution of Search Results



Expt 2: One-to-Many

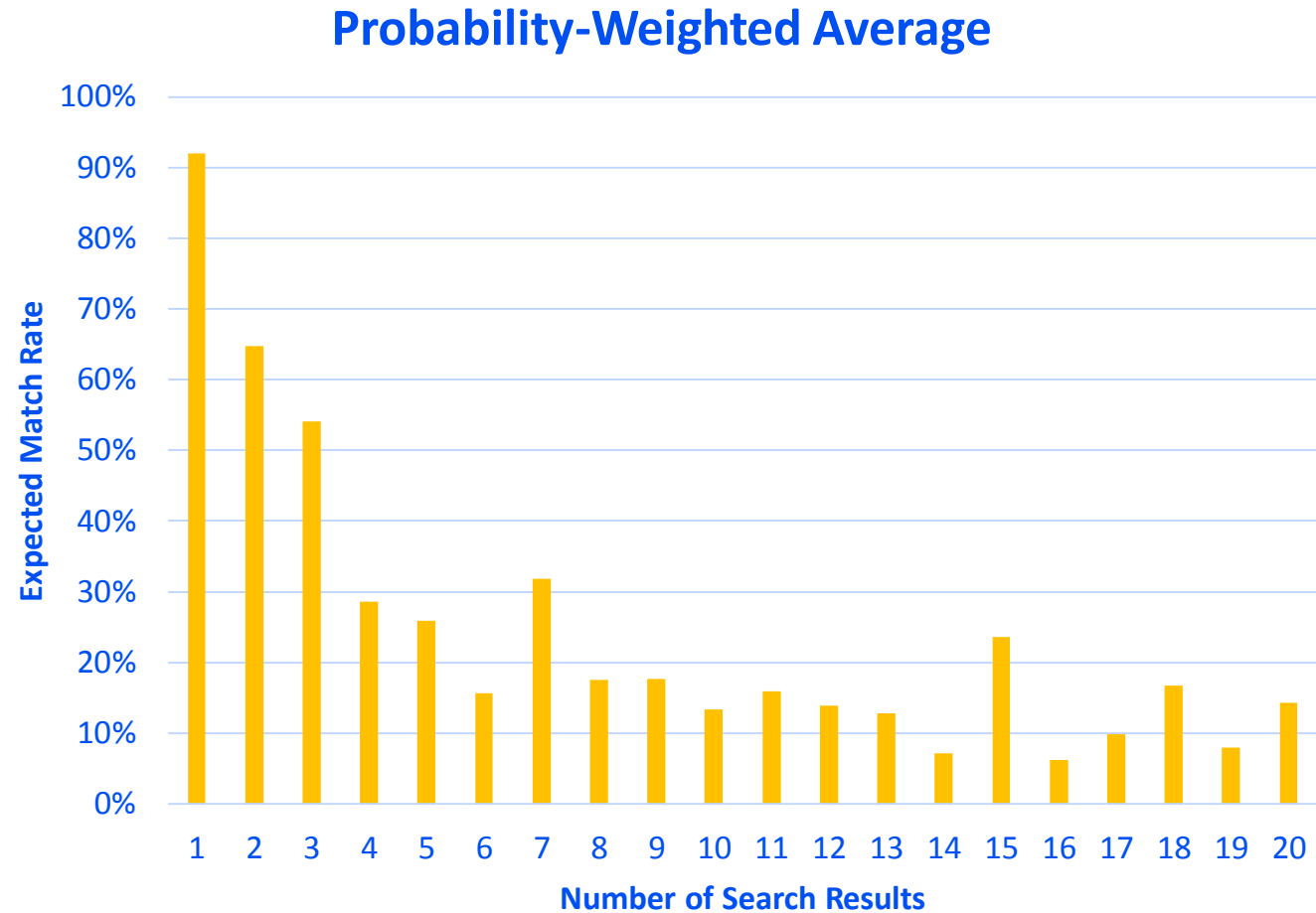
- Constant match rate
- Sample quality in question
 - ▶ ↑ Results means ↓ prob of selection.

Distribution of Search Results



Expt 2: Expected Match Rate

- Expected match rate: sum of match rate * proportion of incidence
- If EPS, then expected match rate is also probability of selection



Hurdles

- Legal – Google Terms of Service
 - ▶ Saving BLS data
 - ▶ Saving Google data
- Methodology
 - ▶ Selecting from many stores with different probabilities
- Representativeness
 - ▶ Data quality and incentives

Contact Information

Greg Barbieri

Economist

Consumer Prices and Price Indexes

202-691-5342

Barbieri.Greg@bls.gov

