

CISIL Methods Report: Question 2

Joseph T. Ornstein & Suhan Kacholia

June 30, 2022

Data Challenge Question

How does the level and duration of incentive (i.e., fare subsidy) affect riders' long-term use of and enrollment in low-income transit programs?

Research Design

We chose to study the period from October 2020 through March 2022 – following the reinstatement of fares in the King County Metro system – comparing transit riders who received a subsidized annual pass with those who received a \$10 stored value card. Because the assignment of subsidies is not randomized, to estimate their causal effect we must rely on the (somewhat less satisfying) Selection on Observables assumption (Rubin 1974) . We assume that potential outcomes are independent of treatment assignment conditional on the following confounding variables:

- Date when the card was issued
- Age of the cardholder
- Race of the cardholder
- Language spoken by the cardholder
- Median income in cardholder's Census tract
- Median age in cardholder's Census tract
- Percent white residents in cardholder's Census tract

By matching on these characteristics, we can recover a more credible average treatment effect than a simple comparison of means.

Supplementary Data

For each card ID, we compute two outcomes: the average weekly number of transit boardings, and the total dollar amount of individual sales (i.e. excluding organizational purchases) during the study period. For the independent variable, there were too few users in treatment conditions other than \$10 stored value and Subsidized Annual Pass, so we focus on those two levels.

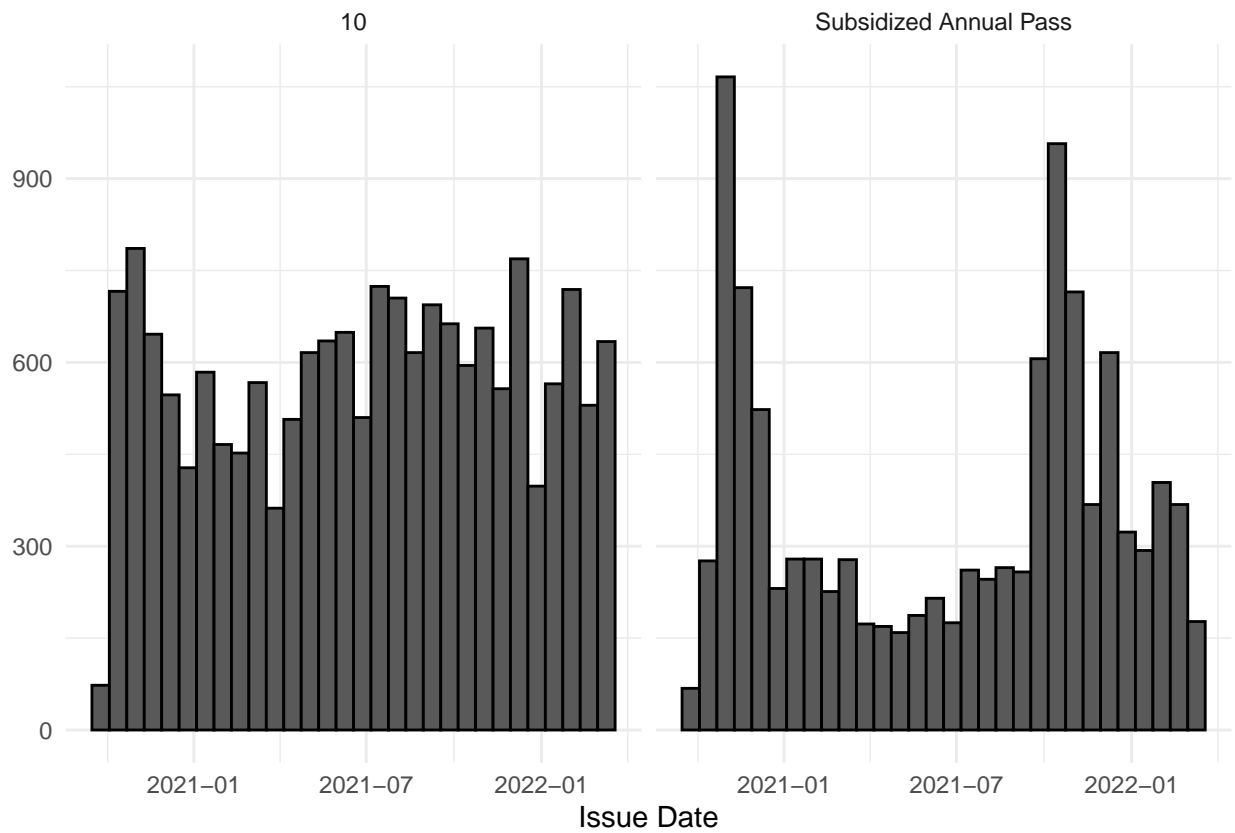
| Treatment Condition | Number of Cards |
|------------------------|-----------------|
| 10 | 17369 |
| 15 | 66 |
| 20 | 201 |
| 30 | 22 |
| 50 | 80 |
| 70 | 167 |
| Misc. Pass | 34 |
| Monthly Pass | 1 |
| Passport | 4 |
| Subsidized Annual Pass | 10883 |

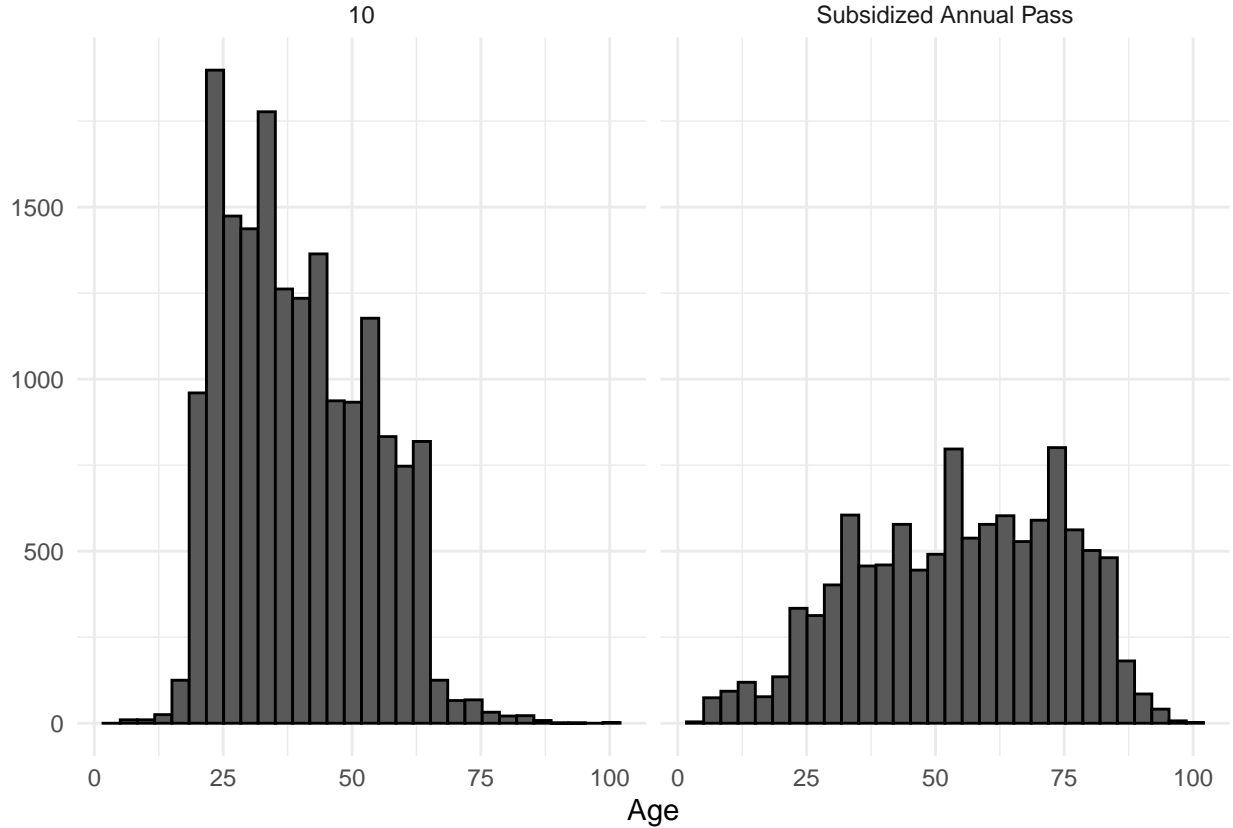
| Treatment Condition | Number of Cards |
|---------------------|-----------------|
| NA | 1882 |

We also recode the language spoken variable to four levels: English, Spanish, Chinese, and Other. We made no other changes to the data provided by the Causal Inference for Social Impact Lab’s Data Challenge – all the necessary code to replicate our analysis is available on a public GitHub repository.

Methods

There are several significant covariate imbalances between the treatment and control groups. Riders who received a subsidized annual pass were significantly more likely to be senior citizens, less likely to speak Spanish, and their cards were more likely to be issued around the winter holiday season.





| Treatment Condition | % White | % Spanish-speaking |
|------------------------|----------|--------------------|
| 10 | 37.37118 | 7.167943 |
| Subsidized Annual Pass | 28.54911 | 1.157769 |

To adjust for these imbalances, we create a matched control group using entropy balancing (Hainmueller 2012). This procedure selects a set of non-zero weights such that the first several statistical moments of the weighted control group’s covariates (listed above) are exactly equal those in the treatment group.

Here are the means of a selection of observed variables before and after matching.

| Variable | Treated | Control (Before Matching) | Control (After Matching) |
|--------------------|---------|---------------------------|--------------------------|
| % White | 26.18 | 35.84 | 26.18 |
| % Spanish-Speaking | 1.00 | 8.32 | 1.00 |
| Mean Age | 57.43 | 39.72 | 57.43 |

Findings

After constructing the matched control group, we estimated the Average Treatment Effect on the Treated Group (ATT) using Weighted Least Squares. We find that riders who received a subsidized annual pass boarded transit roughly 0.54 more times per week during the study period than riders who only received a \$10 stored value card. These riders also spent \$123 less on passes and card loadings during the study period than the riders who received a \$10 preloaded card (as would be expected, since their pass allowed them to ride fare-free for a year).

| | Weekly Boardings | Total Sales |
|-------------|------------------|---------------------|
| (Intercept) | 7.432 (0.077) | 144.485 (2.156) |
| treated | 0.538 (0.109) | −123.060 (3.049) |
| Num.Obs. | 15 752 | 7106 |

Conclusion

Our study has a number of important limitations. Because data does not come from a randomized control trial, it is difficult to make conclusive claims about the causality of different subsidy types. The type of riders who received annual subsidized passes differed in important respects to those who did not. While we try to control for this the best we can, the results from the randomized control trial (conducted separately) may better establish the relationship between type of subsidy and ridership. We also only study two types of subsidy, as the other levels lack sufficient data after the reinstatement of fares in October 2020. Lastly, the study duration was in the midst of a global pandemic. This limits how applicable the data might be to other, more typical time periods.

Despite these limitations, our results provide some evidence that more generous subsidies encourage ridership over an extended period of time (a year and a half). This ridership increase occurs while simultaneously reducing the amount the subsidized pass holders spend on transit services. Short of eliminating fares (the subject of Research Question 1), such annual pass subsidies are a promising route towards increasing ridership and improving equity.

References

- Hainmueller, Jens. 2012. “Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies.” *Political Analysis* 20 (1): 25–46.
- Rubin, Donald B. 1974. “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.” *Journal of Educational Psychology* 66 (5): 688–701. <https://doi.org/10.1037/h0037350>.