

TITLE*

SUBTITLE

Lorena Almaraz De La Garza, Amy Farrow, and Kumalasari Sondjaja

18 February 2021

Abstract

ABSTRACT

1 Introduction

2 Data

#TODO: citations for R and all packages used.

This experiment used two-stage stratified cluster sampling with survey data collection.

The population was all restaurants in Ontario.

A list of Ontario local health authorities (LHAs) that carry out food inspections was used to identify units that inspect restaurants. Using census data from 2016, these LHAs were sorted by population covered by the LHA, and the list was stratified into equally sized strata of small (less than 150,000 population), medium (150,000 to 400,000 population), and large (more than 400,000 population) LHAs. This list of stratified LHAs was the frame at the cluster level.

From each strata, two LHAs were randomly sampled to participate in the treatment, and two LHAs were randomly selected to participate in the control. This was the sample at the cluster level.

The clusters were used because pandemic shutdowns operated based on LHA, and the goal was to recreate the effect as closely as possible.

The stratification was used because there were very differently sized LHAs, and randomly selecting only 12 LHAs from a list of 33 left too high a likelihood of unequivalent treatment and control groups.

```
treatment_control_groups <- read_csv(here::here("outputs/treatment_control_groups.csv"))
```

```
##
## -- Column specification -----
## cols(
##   Group = col_character(),
##   Large = col_character(),
##   Medium = col_character(),
##   Small = col_character()
## )
```

```
treatment_control_groups
```

*Code and data are available at: github.com/amycfarrow/ontariorestaurantclosuresexperiment.

```
## # A tibble: 4 x 4
##   Group      Large      Medium      Small
##   <chr>      <chr>      <chr>      <chr>
## 1 Treatment Hamilton    Haliburton, Kawartha, Pine Ridge ~ Algoma
## 2 Treatment Simcoe Muskoka Windsor-Essex County Timiskaming
## 3 Control   Durham Region Southwestern Ontario Brant County
## 4 Control   Region of Water~ Sudbury and Districts Northwestern He~
```

Once the treatment and control LHAs were selected, each corresponding Food Inspection unit was contacting, and a list of all registered restaurants in each LHA was obtained. Each restaurant was listed by name and address. Once collected into one dataframe, this was the frame at the unit level.

```
all_units_data <- read_csv(here::here("outputs/data/all_units_data.csv"))
```

```
##
## -- Column specification -----
## cols(
##   name = col_character(),
##   address = col_character(),
##   unit = col_character(),
##   group = col_character()
## )
all_units_data
```

```
## # A tibble: 13,017 x 4
##   name          address          unit group
##   <chr>        <chr>        <chr> <chr>
## 1 17 Restaurant - Restaurant 2-Woodward Ave. Blind River ON~ algo~ treatm~
## 2 A&W - Restaurant 496 Causley Street Blind Rive~ algo~ treatm~
## 3 A&W - Restaurant 121 Great Northern Rd Sault S~ algo~ treatm~
## 4 A&W - Restaurant 659 Great Northern Rd Sault S~ algo~ treatm~
## 5 A&W - Restaurant 201 Highway 17 White River ON~ algo~ treatm~
## 6 Absolutely Delicious - Restaura~ 2200 Queen Street East Sault ~ algo~ treatm~
## 7 Agree Outpost Camp Food - Resta~ PO Box 624 Wawa ON POS 1K0 algo~ treatm~
## 8 Airdale Hunting and Fishing Lod~ 1 Whitefish Lake Road Maness algo~ treatm~
## 9 AlgomaTrad - Restaurant 1249 F&G Line Road Richards L~ algo~ treatm~
## 10 Annettes Diner - Restaurant B-4683-Highway 17 Spragge ON ~ algo~ treatm~
## # ... with 13,007 more rows
```

A simple random sample of 15% of the treatment list and 15% of the control list was randomly selected to be surveyed. This was the sample at the unit level.

This selection was used to create a panel, so the same restaurants would be surveyed for the first survey and the second survey. Attempting to sample only 15% of the restaurants allowed time and money to be spent on follow-up and multiple methods of data collecting, reducing the non-response bias.

```
table_for_surveys <- read_csv(here::here("outputs/table_for_surveys.csv"))
```

```
##
## -- Column specification -----
## cols(
##   name = col_character(),
##   address = col_character(),
##   unit = col_character(),
##   group = col_character(),
##   verify = col_character(),
##   ID = col_double()
```

```
## )
```

```
table_for_surveys
```

```
## # A tibble: 2,006 x 6
```

```
##   name                address          unit group verify   ID
##   <chr>              <chr>          <chr> <chr> <chr> <dbl>
## 1 A&W - Restaurant    121 Great Northern Rd Sa~ algo~ treat~ 4Y9    605
## 2 A&W - Restaurant    659 Great Northern Rd Sa~ algo~ treat~ 5Y1    835
## 3 A&W - Restaurant    201 Highway 17 White Riv~ algo~ treat~ 3G0    840
## 4 Antico - Restaurant  6 Village Ct Sault Ste. ~ algo~ treat~ 6K1    749
## 5 Aurora's Westside - Rest~ 300 Second Line W Sault ~ algo~ treat~ 2J5   1313
## 6 Blueberry Hill Motel and~ D-2528 D Highway 17 N Go~ algo~ treat~ 1E0   1664
## 7 Casey's Bar & Grill (Wat~ 360 Great Northern Road ~ algo~ treat~ 4Z7   1795
## 8 Country Kitchen - Restau~ A-9538 Highway 129 Thess~ algo~ treat~ 1L0    253
## 9 Ernie's Coffee Shop - Re~ 13 Queen Street East Sau~ algo~ treat~ 1Y4    477
## 10 Fat Bastard Burrito - Re~ 150 Churchill BLVD Sault~ algo~ treat~ 3Z9    787
## # ... with 1,996 more rows
```

The randomly sampled restaurants were all assigned ID numbers in a random order.

The table for surveys was used to generate 2,006 mailers to be sent to each restaurant on the list. Each mailer was a small envelope containing a sheet that explained the survey, provided a link to the survey, and provided a QR code that went to the same place as the link. There was also a copy of the survey contained in a mailer envelope and a contact number, allowing restaurant owners to complete the survey by phone or mail if they did not feel able to complete it online. This procedure was repeated for Survey 1 and Survey 2. If there was no response within two weeks, the restaurant was contacted by phone, and again if there was no response five days later.

Each link and paper survey was personalized to the restaurant's ID number. The link lead to the survey, which asked for confirmation of the last three letters of the restaurant's postal code. This was to ensure that no mailing mistakes were made, and to prevent duplicate data collection. Once verified, the survey-taker was given the main survey.

This survey cost «««<»»»»»»»»»».

The data from the survey was tied to the ID number and last three digits of the postal code, but not the restaurant name or address. This was to ensure data privacy. The identifying information was kept in a separate dataframe (Table for Surveys) than the survey answers (Survey 1 Data, Survey 2 Data).

Survey 1 was conducted June 3rd to 27th 2021, and it asked about May 2021.

Survey 1 Data was used to confirm that the two-stage stratified cluster sampling had created treatment and control groups that were roughly equivalent.

Survey 1 collected the following data:

- Demographic information
- Disability status
- Indigenous status
- Visible minority (non-Indigenous) status
- Gender identity
- Type of service provided (dine-in, take-out, or both)
- Revenue in May 2021
- Employees
- Number of full-time employees (30 hours/week or more)
- Number of part-time employees (less than 30 hours/week)

Distributions for revenue and tables for key demographics were compared.

After Survey 1 data was collected, on June 28th, shutdowns were announced for the six treatment LHAs. The shutdowns ran from July 1st to July 14th, inclusive. This length of time was considered to be the minimum effective length for a shutdown to stop circulation of the virus. During the shutdown, all restaurants in the treatment LHAs were officially banned from offering dine-in and patio services. Take-out and delivery were permitted. This type of partial shutdown was selected because it has been the most common type of shutdown since the pandemic began.

Survey 2 was conducted August 3rd to 27th 2021, and it asked about July 2021.

Survey 2 collected the following data:

- Demographic information
- Disability status
- Indigenous status
- Visible minority (non-Indigenous) status
- Gender identity
- Type of service provided (dine-in, take-out, or both)
- Closures (none, temporary, or permanent)
- Revenue in May 2021
- Employees
- Number of full-time employees (30 hours/week or more)
- Number of part-time employees (less than 30 hours/week)

```
survey_1 <- read_csv(here("outputs/data/survey_1.csv"))
```

```
##
## -- Column specification -----
## cols(
##   ID = col_double(),
##   group = col_character(),
##   verify = col_character(),
##   service_type = col_character(),
##   disability = col_character(),
##   woman = col_character(),
##   indigenous = col_character(),
##   visible = col_character(),
##   revenue = col_double(),
##   ft = col_double(),
##   pt = col_double()
## )
```

```
survey_2 <- read_csv(here("outputs/data/survey_2.csv"))
```

```
##
## -- Column specification -----
## cols(
##   ID = col_double(),
##   group = col_character(),
##   verify = col_character(),
##   service_type = col_character(),
##   disability = col_character(),
##   woman = col_character(),
##   indigenous = col_character(),
##   visible = col_character(),
##   shutdown = col_character(),
##   revenue = col_double(),
##   ft = col_double(),
```

```
## pt = col_double()
## )
```

2.1 Survey 1 results

```
treat1 <- subset(survey_1, group == "treatment")
control1 <- subset(survey_1, group == "control")
t.test(treat1$revenue, control1$revenue)
```

```
##
## Welch Two Sample t-test
##
## data: treat1$revenue and control1$revenue
## t = -0.022943, df = 1218.9, p-value = 0.9817
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -4015.895 3923.055
## sample estimates:
## mean of x mean of y
## 69743.43 69789.85
```

```
survey_1 %>%
  group_by(group) %>%
  summarise(mean_revenue = mean(revenue, na.rm = TRUE),
            median_revenue = median(revenue, na.rm = TRUE),
            n = n())
```

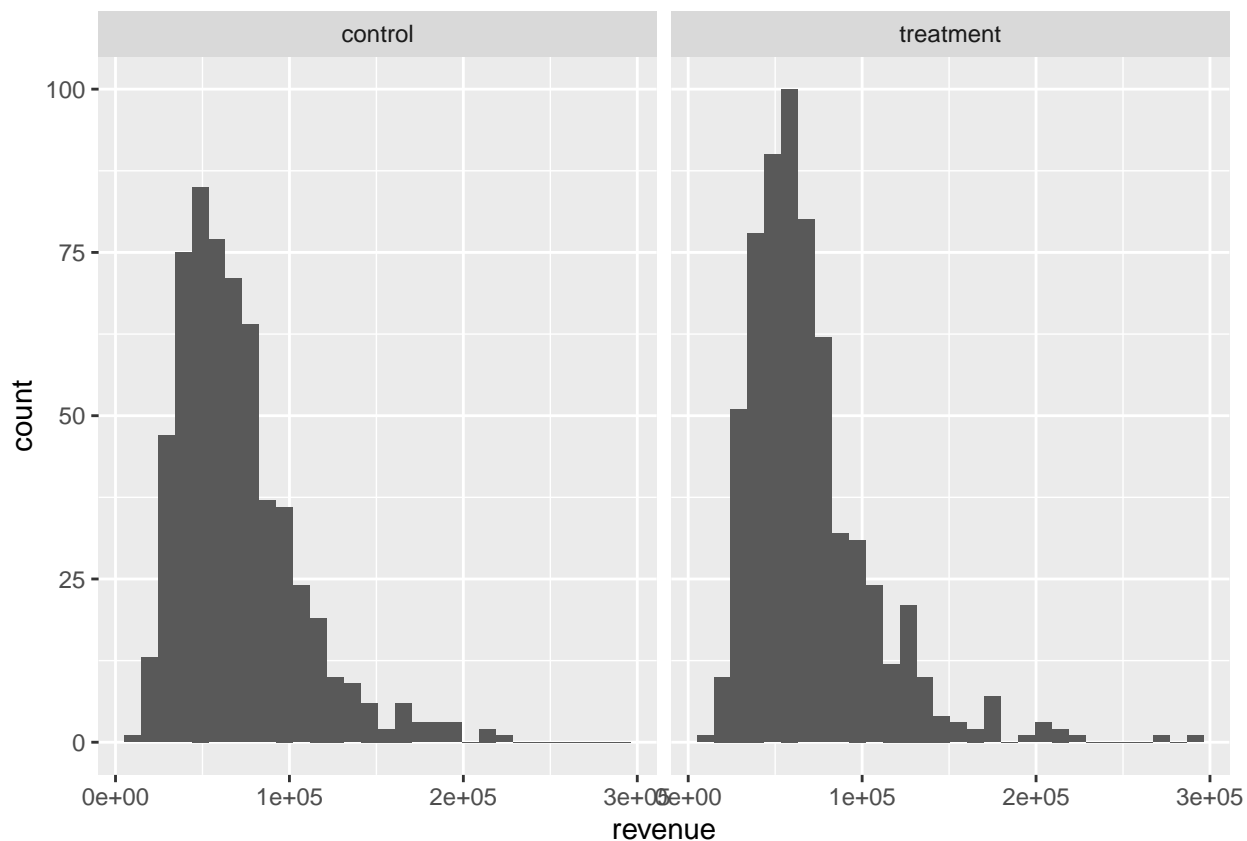
```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
## # A tibble: 2 x 4
##   group mean_revenue median_revenue n
##   <chr>      <dbl>      <dbl> <int>
## 1 control    69790.      63092.   594
## 2 treatment  69743.      61588.   628
```

```
survey_1 %>%
  ggplot(aes(x = revenue)) +
  geom_histogram() +
  facet_wrap(~group)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



Survey 1 data as baseline

Counterfactuals established.

2.2 Survey 2 results

Survey 2 data to contrast treatment and control.

```
survey_2 %>%
  group_by(group) %>%
  summarise(mean_revenue = mean(revenue, na.rm = TRUE),
            median_revenue = median(revenue, na.rm = TRUE),
            n = n())

## 'summarise()' ungrouping output (override with '.groups' argument)

## # A tibble: 2 x 4
##   group    mean_revenue median_revenue     n
##   <chr>         <dbl>         <dbl> <int>
## 1 control      67646.         59965.   617
## 2 treatment   46626.         39996.   618

treat2 <- subset(survey_2, group == "treatment")
control2 <- subset(survey_2, group == "control")
t.test(treat2$revenue, control2$revenue)

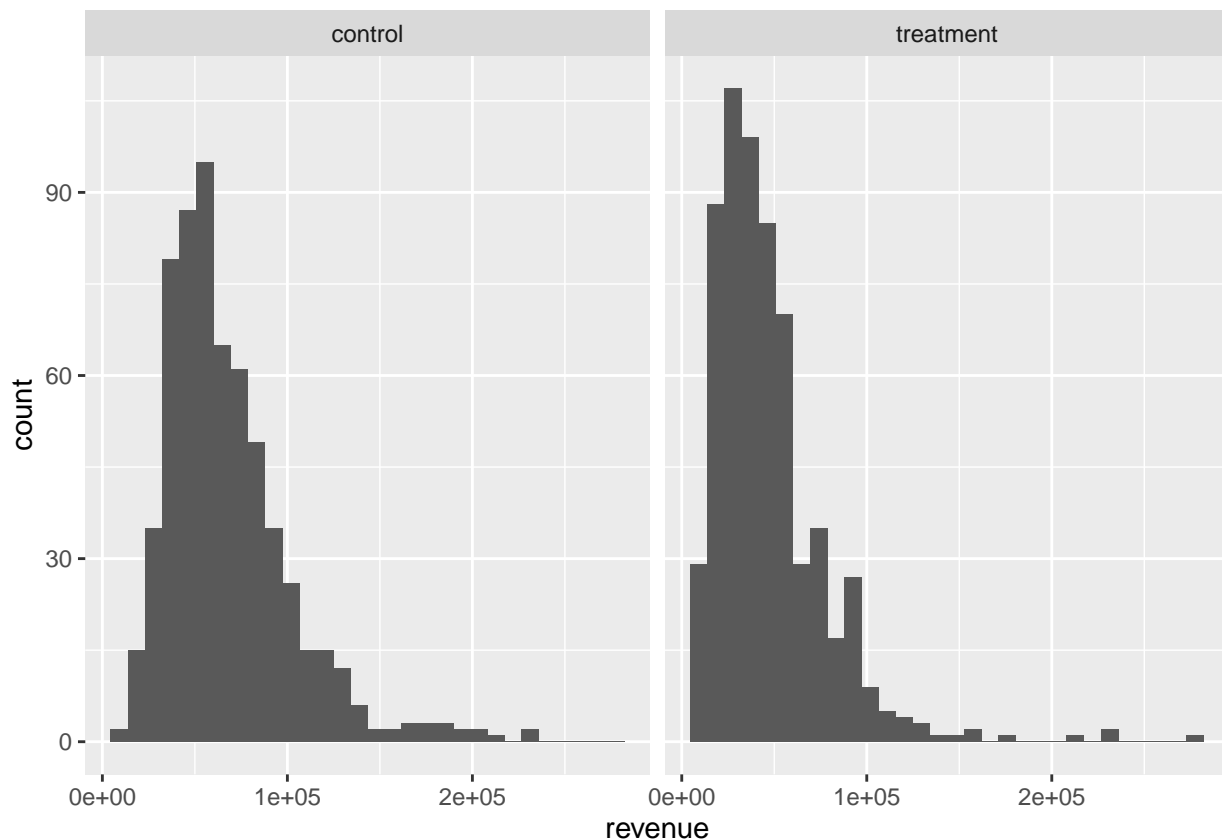
##
## Welch Two Sample t-test
##
```

```
## data: treat2$revenue and control2$revenue
## t = -11.436, df = 1216.8, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -24625.44 -17413.38
## sample estimates:
## mean of x mean of y
## 46626.49 67645.90
```

```
survey_2 %>%
  ggplot(aes(x = revenue)) +
  geom_histogram() +
  facet_wrap(~group)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).
```



3 Discussion

3.1 Overview

3.2 Findings

3.2.1 FINDING ONE

Closures had a negative impact on revenues and employment counts

Suggestion: funding for businesses and temporary unemployment

3.2.2 FINDING TWO

Closures had a stronger negative impact on dine-in only restaurants

Suggestion: technical support for restaurants to increase takeout business

3.2.3 FINDING THREE

Closures had a stronger negative impact on indigenous or visible minority-owned businesses.

Suggestion: grants for minority business owners

3.3 Limitations

External validity, internal validity, cluster sampling. Unavoidable because shutdowns were by health unit for COVID, and we were trying to replicate them. It was ethically unsound to run the experiment on the entire province, so sampling was necessary.

Legally registered businesses only.

Survey non-response.

Self-reported information.

3.4 Future Directions

For statistical analysis, this would require propensity score weighting.

Appendix

4 References