Final Report:

COVID-19 State-level Changes during 2020

Table of Contents

1. ANALYSIS OBJECTIVES

While there are an enormous range of questions that may arise within the context of COVID tracking information during 2020, the two questions we will assess are:

- Do states that aggressively relaxed distancing restrictions in mid-May have a different trajectory of positive cases than those states that delayed such measures?
- Did the case hospitalization rate (CHR) change from early in the pandemic to later, whether due to changes in underlying severity and/or the evolution of clinical management?

2. DEFINITIONS AND GENERAL CONSIDERATIONS FOR DATA ANALYSIS

- Two data sources:
    - COVID Tracking Project – Data on a wide range of COVID incidence metrics by state and day
    - State-level data on the population size and classification of distancing restrictions relaxation practices in mid-May 2020
- Constrain analysis data to
    - 50 states and DC
    - April and December experience (first and third "humps" of incidence during 2020)

3. DATA ANALYSES

3.1. Effect of relaxing distancing restrictions on incidence of positive cases

Calculate descriptive statistics and a box plot for the number of positive cases per capita by time (April vs. December) and re-opening approach (early vs. delayed). Compute a mixed effect regression model of positive cases per capita (100,000 cases / population) on re-opening approach treating state/DC as a random subject effect in the REPEATED statement using maximum likelihood estimation and an unstructured covariance structure.

3.2. Change in Case hospitalization rate (CHR) over time

Case hospitalization rate (CHR) is defined as [100 * (average daily hospitalizations in a month / total positive cases in a month)]. Calculate descriptive statistics for CHR during April2020, December2020 and the change in CHR within states (December minus April). Assess whether the change was statistically significantly different than 0 using a paired t-test.

4. RESULTS

4.1. Effect of relaxing distancing restrictions on incidence of positive cases

Distribution of Cases Per Capita by Reopening Approach Within Each Month

The MEANS Procedure

month=4

Analysis Variable : cases_per_capita								
Distance restrictions relaxed 1 is Early (never restricted or significantly relaxed restriction in mid-May2020) 0 is Delayed	N Obs	N	Mean	Median	Std Dev	Lower Quartile	Upper Quartile	Quartile Range
Delayed	20	20	416.4075440	191.9060349	430.7999337	131.2933472	549.3477377	418.0543905
Early	31	31	191.9482758	143.9100991	163.3161006	105.0734628	228.9871912	123.9137284

month=12

Analysis Variable : cases_per_capita								
Distance restrictions relaxed 1 is Early (never restricted or significantly relaxed restriction in mid-May2020) 0 is Delayed	N Obs	N	Mean	Median	Std Dev	Lower Quartile	Upper Quartile	Quartile Range
Delayed	20	20	1581.22	1685.51	655.9634049	989.3353732	2170.35	1181.01
Early	31	31	2012.95	1932.62	473.6618502	1711.80	2278.15	566.3557655

Figure 1. Descriptive statistics of the distribution of cases per capita by reopening approach within each month

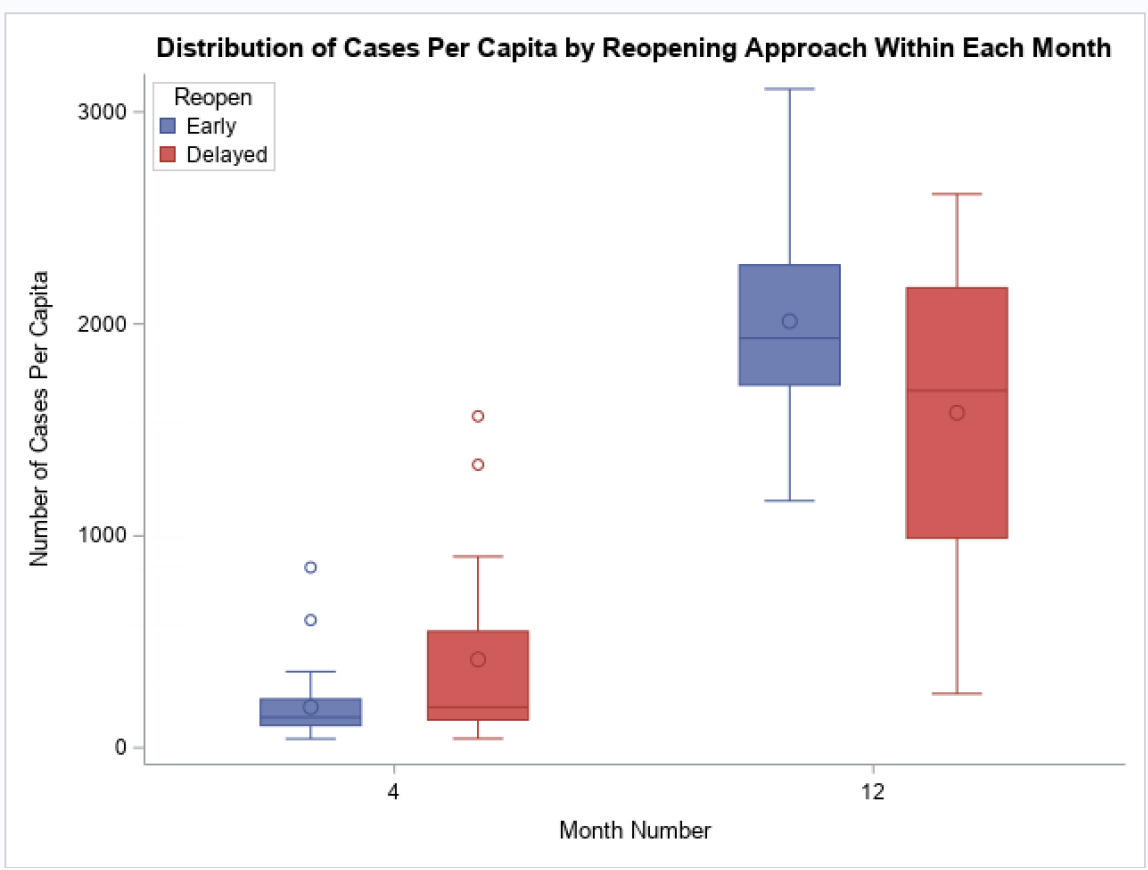


Figure 2. Distribution of Cases Per Capita by Reopening Approach Within Each Month

Insert clearly labelled results

4.2. Change in Case hospitalization rate (CHR) over time

The MEANS Procedure

month=4

Analysis Variable : case_hosp_rate						
N	Mean	Median	Std Dev	Lower Quartile	Upper Quartile	Quartile Range
51	0.2690196	0.2500000	0.1880240	0.1600000	0.3700000	0.2100000

month=12

Analysis Variable : case_hosp_rate						
N	Mean	Median	Std Dev	Lower Quartile	Upper Quartile	Quartile Range
51	0.0305882	0.0300000	0.0242001	0.0100000	0.0400000	0.0300000

The UNIVARIATE Procedure
Variable: chr_change

Moments			
N	51	Sum Weights	51
Mean	0.23843137	Sum Observations	12.16
Std Deviation	0.19013019	Variance	0.03614949
Skewness	0.49490144	Kurtosis	0.3688663
Uncorrected SS	4.7068	Corrected SS	1.80747451
Coeff Variation	79.7421039	Std Error Mean	0.02662355

Basic Statistical Measures			
Location		Variability	
Mean	0.23843	Std Deviation	0.19013
Median	0.22000	Variance	0.03615
Mode	-0.01000	Range	0.84000
		Interquartile Range	0.20000

Note: The mode displayed is the smallest of 2 modes with a count of 3.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	8.955656	Pr > |t|	<.0001
Sign	M	17.5	Pr >= |M|	<.0001
Signed Rank	S	626.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	0.73
99%	0.73
95%	0.61
90%	0.48
75% Q3	0.34
50% Median	0.22
25% Q1	0.14
10%	-0.01
5%	-0.05
1%	-0.11
0% Min	-0.11

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-0.11	12	0.50	5
-0.10	17	0.57	1
-0.05	41	0.61	25
-0.05	11	0.69	37
-0.03	10	0.73	38

Merged Dataset

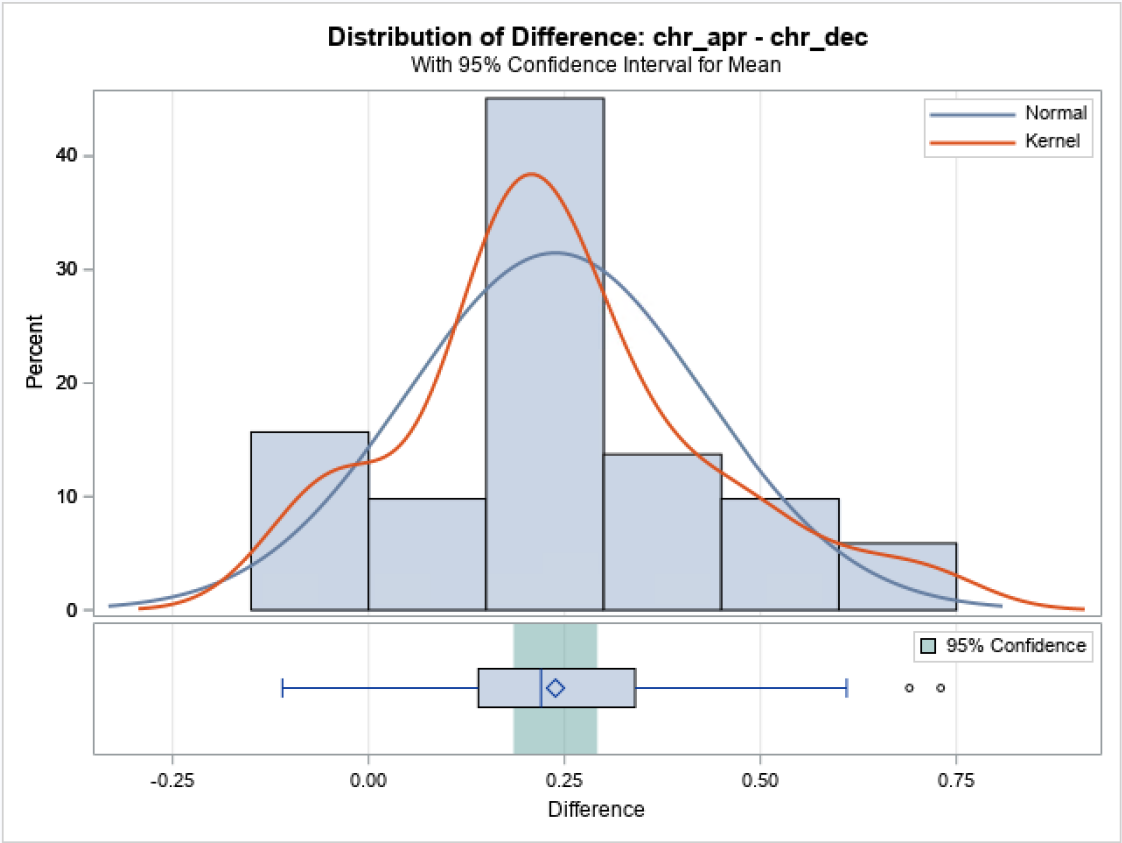
The TTEST Procedure

Difference: chr_apr - chr_dec

N	Mean	Std Dev	Std Err	Minimum	Maximum
51	0.2384	0.1901	0.0266	-0.1100	0.7300

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.2384	0.1850	0.2919	0.1901	0.1591	0.2363

DF	t Value	Pr > |t|
50	8.96	<.0001



Insert clearly labelled results

5. CONCLUSIONS

- Do states that aggressively relaxed distancing restrictions in mid-May have a different trajectory of positive cases than those states that delayed such measures?

For states with an early reopening, the mean number of cases per capita witnessed a steeper increase from April to December as compared to the states with a delayed reopening (figures 1 and 2). Additionally, the median number of cases per capita, lower quartile, and upper quartile increased more for early than late "reopen states." In April, the lower quartile for both early and delayed reopen states were rather close (105 and 131, respectively), yet in December, the difference between the two lower quartiles was over 715 cases per capita, with early reopening far greater. Along with the large increase in the mean and median number of cases per capita, this may suggest that reopen policies influence the number of cases per capita. It should be noted that the range of 1181 cases per capita for delayed opening is quite large and may hint at other factors involved with influence on cases per capita state-by-state.

- Did the case hospitalization rate (CHR) change from early in the pandemic to later, whether due to changes in underlying severity and/or the evolution of clinical management?

The mean case hospitalization rate, which described the ratio of hospitalizations to total positive cases, decreased by a range of 0.2384 (~24%) from April to December. Despite increased cases per capita in December, the hospitalization rate decreased. It is possible that the number of hospitalizations remained the same while case numbers increased and points to clinical adaptations to managing virus patients; but further analysis and detail of the data is required.

The distribution of the change in case hospitalization rate from April to December in all states was slightly positively skewed based on the results from the ttest procedure. Thus, more states witnessed a range of CHR between April and December below the median and mean CHR change.

Brief paragraph for each of the two domains of analysis summarizing our observations

6. APPENDIX 1 – ANALYSIS DATASET SPECIFICATIONS

Insert the specifications upon which your analysis datasets were based; assure consistency with other specifications (sources, dataset names, variable names) and results provided. This may be an image of the Excel file or importing the contents into a table in your word processing document just as long as it is clearly presented.

Dataset Name:	work.cases			
Data Source(s):	work.state_covid, work.covid_2020			
Unique Key(s):	month, state			
Purpose:	This data set will be used to analyze the effect of state reopening on pertinent Covid data, in this case, cases per			
Variable Name	**Variable Label**	**Type/Format**	**Specifications**	**Comments/QC Finding(s) & Resolution**
month	April or December	Nominal, Characters	cases.month derived from state_covid.date	
state	State or District of Columbia	Nominal, Characters	state_info.state variable	constrained to 50 states and
reopen	Distance restrictions relaxed: Early (1) and Delayed (0)	Categorical	state_info.reopen variable	formatted so that 1 = Early and 0 = Delayed
case_hosp_rate	The ratio of average daily hospitalizations to total positive cases in a month by state	Continuous	10000 * cases.pos_total/state_info.pop	cases.pos_total is the sum of positive cases and calculated from covid_2020.positive
cases_per_capita	The number of positive cases per capita	Continuous	ROUND(100 * cases.avghosp/cases.pos_total, 0.01)	cases.avghosp obtained from average of cases.hosp_total over state_covid.num_days. Cases.hosp_total is the sum of covid_2020.hospitalized
Dataset Name:	work.change			
Data Source(s):	work.cases			
Unique Key:	state			
Purpose:	This data set will be used to analyze the differene in case hospitalization rate by state			
Variable Name	**Variable Label**	**Type/Format**	**Specifications**	**Comments/QC Finding(s) & Resolution**
state	State or District of Columbia	Nominal, Characters	state_info.state variable	constrained to 50 states and
chr_apr	Case hospitalization rate by state in April	Continuous	change.chr_apr is taken from the April cases.case_hosp_rate	
chr_dec	Case hospitalization rate by state in December	Continuous	change.chr_apr is taken from the December cases.case_hosp_rate	
chr_change	Change in case hospitalization rate between April and December, by state	Continuous	change.chr_apr - change.chr_dec	

7. APPENDIX 2 – DATA ANALYSIS SPECIFICATIONS

Insert the specifications upon which your <u>analyses</u> were based; assure consistency with other specifications (sources, dataset names, variable names) and results provided.

<u>DATABASE OVERVIEW</u>

- Data Source(s):

- state_info.csv: work.state_info contains 51 observations on the state-level data on the population size and classification of distancing.
- covidtrackingapr2020_dec2020.csv: work.covid_2020 is a dataset of Covid cases from 4/1/2020-12/31/2020 for every state or district of Columbia.

Other Sources/Datasets Used but not present in analysis:

- work.state_covid: a merged data set between work.state_info and work.covid by state.

- Analysis Dataset(s)
  - work.cases
    - Seriality: state month
    - Contents:
      - This dataset rolls up the data from work.merged_data, then calculates cases per capita, average number of hospitalized, and case hospitalization rate based on the variables: total hospitalized, number of days, total positive cases, and total population.
      - Contains variables state, month, reopen, cases_per_capita, and case_hosp_rate.
      - cases_per_capita: the number of positive cases per capita.
        - Formula: 100000*cases.total positive cases / state_info.population
      - case_hosp_rate: case hospitalization rate calculated by month.
        - Formula: 100*(cases.average daily hospitalized in a month/cases.total positive cases in a month)
  - work.change
    - Seriality: state
    - Contents
      - Taking the original CHR from work.cases_monthstate, this analysis dataset transposes the data to view CHR by state as well as calculates the difference between case hospitalization rate from April to December within each state.
      - Contains variables chr_apr, chr_dec, and chr_change.
      - chr_change: change in case hospitalization rate within states between April and December.

- Formula: (changes.CHR April) – (changes.CHR December)

DATA ANALYSES

Examination of cases per capita by reopening approach within each month

- Inputs
  - work.cases: calculates cases per capita and case hospitalization rate.
- Outputs and associated SAS PROC details
  - *Descriptive statistics of the number of positive cases per capita by reopening approach within each month.*
    - Variables: cases.cases_per_capita, cases.reopen, cases.month
    - Via a means procedure in SAS, the output displays number of observations, mean, median, standard deviation, quartile 1, quartile 3, and the interquartile range for variable cases per capita for reopening strategy based on each month.
  - *Box plot of descriptive statistics on the number of positive cases per capita by reopening approach within each month*
    - Variables: cases.cases_per_capita, cases.reopen, cases.month
    - *Via* an sgplot procedure in SAS, the boxplot was calculated from case per capita in category month and group reopen.

Analysis of case hospitalization rate during the different months and within each state

- Inputs:
  - work.change: use of change in case hospitalization rate by state.
- Outputs and associated SAS proc details:
  - *Descriptive statistics of case hospitalization rate during April, December, and between states*
    - Usage of data set work.changes.
    - change.chr_apr, change.chr_dec, change.chr_change
    - Through a means procedure in SAS, the number of observations, mean, median, standard deviation, 1st quartile, 3rd quartile, and interquartile range is calculated for variable case_hosp_rate.
  - *Paired t-test output from procedure ttest in SAS for change in case hospitalization rate to assess whether CHR was statistically significant from 0.*

- Usage of data set work.change
- Variables: change.chr_change
- *Input the change in CHR within a paired ttest procedure to calculate and plot associated distribution of data and data summaries.*

8. REFERENCES

- US Census 2010
  - https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html
- The COVID Tracking Project
  - https://covidtracking.com/data/