People are Gross, and Saboteurs

A Statistically Insignificant Amount of Purell

Omar Al Taher, Carmen Easterwood, and Nikki Haas

1. Abstract

With current media reports about the flu epidemic, people are being encouraged to wash and sanitize their hands often. We explore hand sanitizer usage in various settings (retail, medical, gyms, kitchens) with and without a sign inviting the public to use the hand sanitizer. In our experiment, we use a randomized treatment/control schedule to study the effects of the sign, and weigh the hand sanitizer bottles daily. We also consider covariate effects from Google Trend data on flu and superbug searches. While we found a small naive effect in a basic OLS when the sign is introduced, the use of more appropriate time-series models fails to reject the null hypothesis that the sign has no effect on usage.

2. Introduction

The flu is a major health hazard that poses a significant risk to our health and our lives. In any given year as much as 20% of the US population is affected by the virus, with tens of thousands of hospitalizations and thousands of deaths[1].

Aside from the influenza vaccine, proper hand hygiene using soap and water as well as alcohol based rubs has been found to be effective in reducing the presence of some types of influenza virus on human hands[2]. Other studies also find a reduction in the transmission of the flu and other illnesses and absenteeism at work in response to hand hygiene programs incorporating alcohol based hand sanitizers[3,4,5].

This paper discusses an experiment to test the effectiveness of placing an indicator sign, pointing to the location of a hand sanitizer dispenser, on increasing the usage in multiple locations where people congregate. Considering the sobering statistics and the simplicity of the

---

[1] https://www.cdcfoundation.org/businesspulse/flu-prevention-infographic
[2] Grayson ML, Melvani S, Druce J, Barr IG, Ballard SA, Johnson PD, Mastorakos T, Birch C. Efficacy of soap and water and alcohol-based hand-rub preparations against live H1N1 influenza virus on the hands of human volunteers. Clin Infect Dis. 2009 Feb 1;48(3):285-91
[3] https://www.ncbi.nlm.nih.gov/pubmed/11029132 Hammond B, Ali Y, Fendler E, Dolan M, Donovan S. Effect of hand sanitizer use on elementary school absenteeism. Am J Infect Control. 2000 Oct;28(5):340-6
[4] Hübner NO, Hübner C, Wodny M, Kampf G, Kramer A. Effectiveness of alcohol-based hand disinfectants in a public administration: Impact on health and work performance related to acute respiratory symptoms and diarrhoea. BMC Infect Dis. 2010;10:250.
[5] Sandora TJ, Taveras EM, Shih MC, Resnick EA, Lee GM, Ross-Degnan D, Goldmann DA. A randomized, controlled trial of a multifaceted intervention including alcohol-based hand sanitizer and hand-hygiene education to reduce illness transmission in the home. Pediatrics. 2005 Sep;116(3):587-94.

proposed treatment, even a small treatment effect would have significant practical public health implications.

3. Experimental Design

The experiment is designed to reflect as much as possible a real life application of the treatment. Identical containers of Purell hand sanitizer were placed in 11 public spaces where subjects would be passing by as part of their daily life, and unlikely to be aware of the experiment being conducted. The locations were in Los Angeles, CA and Seattle, WA and covered various types as follows: a pediatrician's office (2 physical locations), a retail yarn store, two gyms, a residential building lobby, and a workplace (5 different break rooms on non-consecutive floors). In all of these locations, both the Purell containers and the treatment signs were inconspicuous due their generic look.

a. Treatments

In the treatment condition, an indicator sign was placed next to the bottle of Purell (see an example to the right). The sign was simple in wording "Sanitize Hands Here" and had an icon representing two hands in the act of sanitizing.

Due to the difficulty of measuring the number of subjects who are treated on any given day without raising privacy concerns (for example by using video recordings of the dispensers during use), this study depends on random assignment and multiple treatment periods to address the issue of balance in treatment and control periods.

b. Random Assignment

Each treatment and control period lasted between 1 and 6 days within a single calendar week. All locations started in control in the first week, but were then randomly assigned to be in either treatment or control in the second week.

- At the doctor's office, a coin was flipped to assign one entrance to treatment. The other entrance was kept in control.
- At the other nine locations, independent coins were flipped for each location and four were assigned to treatment for week 2 of the study.

After the second week, the assignment was flipped back and forth at each location for the following two week-long periods. The assignments went slightly awry at the five office locations, where the assistant who helped us collect the daily weight measurements forgot to flip the treatment and control assignments at the start of week 3. Those locations therefore had the same assignments for two weeks in a row, and were subsequently flipped for week 4. We experienced a similar treatment assignment error at the ECA gym and the yarn store, where the treatment was not switched on and off on the correct dates. (See Appendix A for the detailed assignment schedule.)

c. Outcome Measure

Our outcome measure is the amount of Purell used each day, measured in grams. The weight of the Purell bottles was measured daily at each location using a digital scale, and we subtracted the current day's weight from the previous day's weight to get the previous day's Purell usage. For example, in the table below, usage on Monday was 1850 - 1843 = 7 grams, and usage on Tuesday was 1843 - 1838 = 5 grams.

Date	**Weight of Purell Bottle (g)**
Monday at 9am	1850
Tuesday at 9am	1843
Wednesday at 9am	1838

Within each location, the bottle was always weighed at the same time of day to keep measurement periods consistent from day to day and week to week (for example, 8:30 am at the gym location before heading to work, and 9:30 am at the office locations after arriving at work).

4. Data Pipeline

As we collected the daily weight measurements from our 11 locations, we entered the data into a Google spreadsheet with a sheet for each location. The sheet calculated daily Purell usage using the formula described above.

Once all the data was collected, it was combined into a CSV file along with some daily Google Trends data we felt was relevant (searches for "second wave flu" and searches for "drug resistant" -- see the Confounding Factors section for more information on why we collected this data). Finally we added a column specifying whether we thought an observation was suspicious (also discussed further in the Confounding Factors). After assembling all the data, we had the following information:

Date	3/17/18
Location	ECA Front Desk
Location Type	Gym
Treatment	0
Weight (g)	1518
Amount Used (g)	5
Collector	Nikki
Google “second wave flu” Trend	0
Google “drug resistant” Trend	15
Google “second wave flu” Trend for Previous Day	0
Google “drug resistant” Trend for Previous Day	65
Is Suspicious	0

5. Pilot Study

We began our experiment with a small pilot study at the two doctor’s office locations before rolling the experiment out to the other nine locations. We selected this location because we were initially discussing placing our bottles in locations where people might want to be clean, such as gyms, medical settings, and kitchens. Bottles of Purell were placed at each of the two locations, and each day we collected the weight measurements and confirmed that no one had stolen or tampered with them. We did not encounter any problems, so the next week we assigned one of the pilot locations to treatment and started our other nine locations with a week of the control condition. Unfortunately, as we will discuss later, some of our other locations did not go as smoothly as the doctor’s office.

In addition, a pilot analysis with mock data was run to see if an effect could be detected with only 30 data points. The mock data assumed a very high traffic area that for which an instructional sign had a clear effect. With this setup, we did see a statistically significant effect. Later, we discovered that in practice even very high traffic areas did not have nearly the usage we assumed in the mock data.

Results from the mock data model.

Out[163]:

OLS Regression Results

Dep. Variable:	wt_diff	**R-squared:**	0.714
Model:	OLS	**Adj. R-squared:**	0.706
Method:	Least Squares	**F-statistic:**	100.8
Date:	Sun, 04 Mar 2018	**Prob (F-statistic):**	0.00210
Time:	13:36:31	**Log-Likelihood:**	-183.20
No. Observations:	36	**AIC:**	370.4
Df Residuals:	34	**BIC:**	373.6
Df Model:	1		
Covariance Type:	cluster		

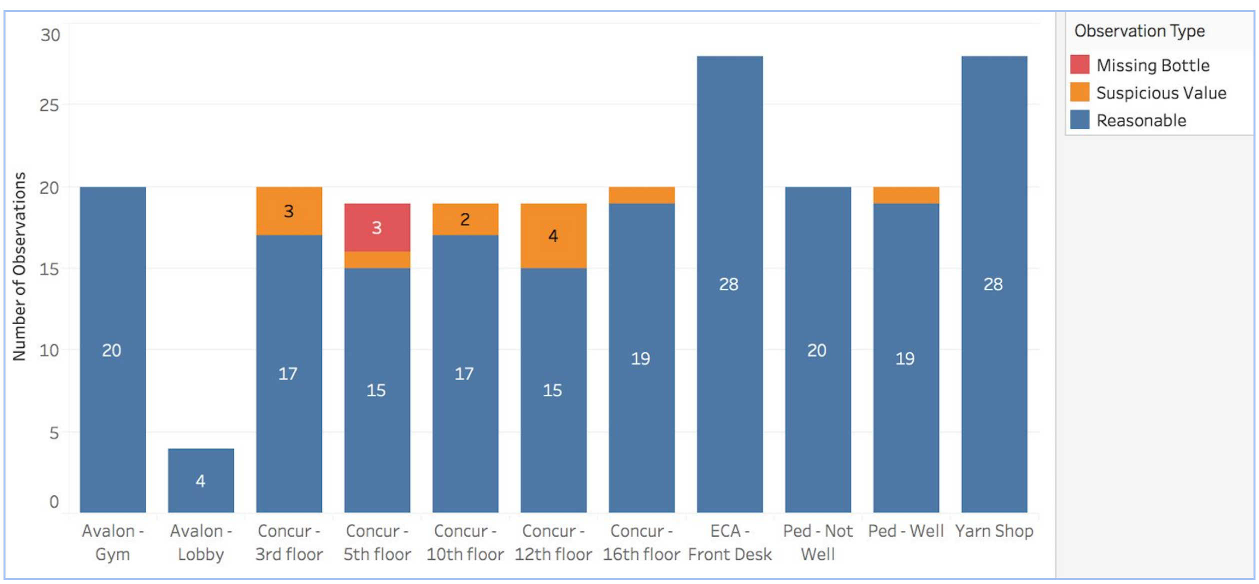
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	75.6413	12.387	6.107	0.009	36.222	115.061
blue	124.3587	12.387	10.040	0.002	84.939	163.778

Omnibus:	24.636	**Durbin-Watson:**	1.704
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	44.148
Skew:	1.739	**Prob(JB):**	2.59e-10
Kurtosis:	7.163	**Cond. No.**	2.56

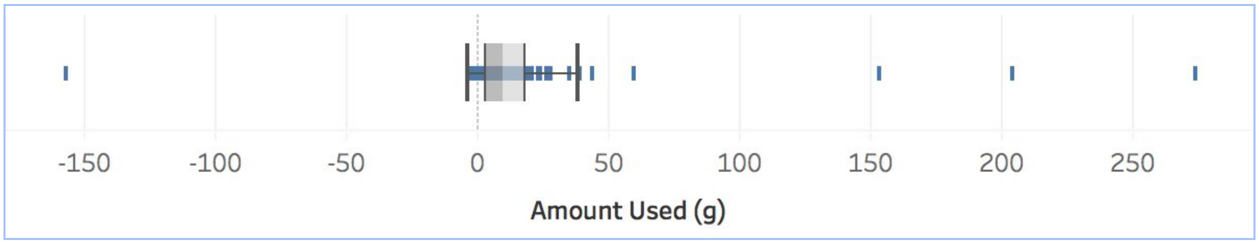
6. Data and Analysis of the Final Experiment

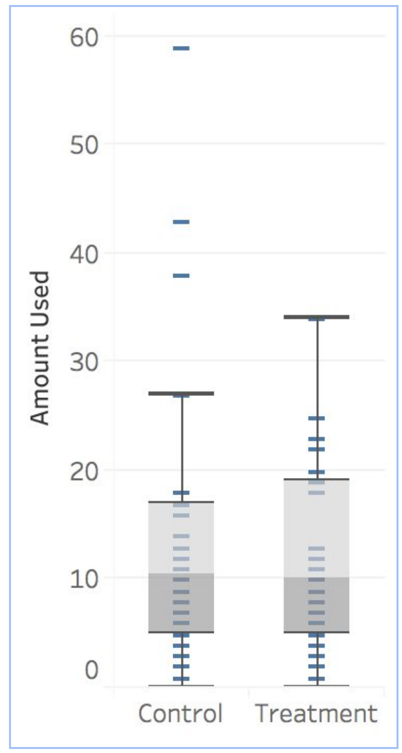
a. Data

In total, we have 217 observations of daily Purell usage from 11 locations. Unfortunately, 3 observations are missing and we don’t trust the observed value for 12 of them.



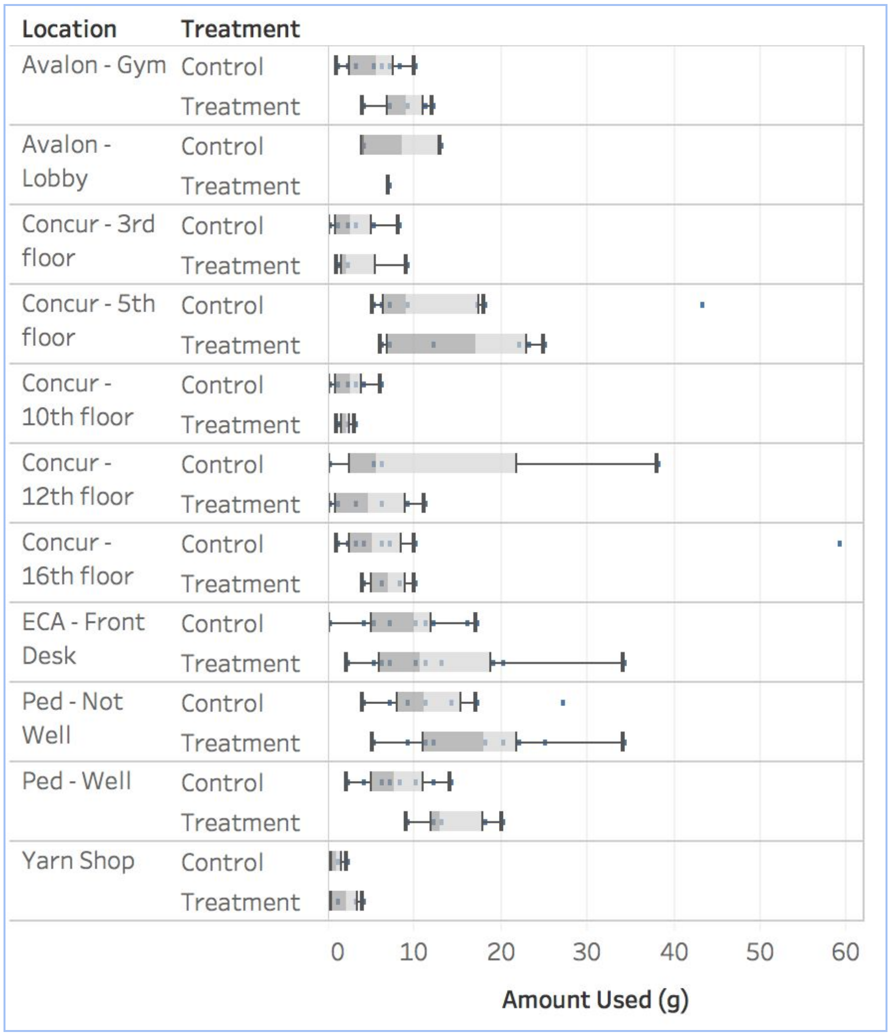
A box-and-whisker plot of the daily amount of Purell used shows how there are some extreme outliers, and even a couple of negative values.





Dropping all the negative values and the three largest positive values gives us a clearer picture of how daily Purell usage is distributed. Both treatment and control have means of about 10g per day, although the treatment distribution skews a little larger than the control distribution.

We initially tested a number of models that used a between-subjects design and found no significant effect, which is as we expected given the plot to the right. However, further analysis showed that the distribution of daily Purell use could vary significantly between locations, which ultimately led us to use a within-subjects design for our models. In the graphic below, the “Ped - Well” location has the most significant difference between treatment and control, although it does not reach the 5% level of significance. We can also see that the “Avalon - Lobby” data can’t be appropriately analyzed due to having only one observation in the treatment condition.



b. Models - Times Series and Power Analysis

Each Purell station was static in location over time, but changed from treatment to control during the experiment's duration. This experimental model requires us to consider each station's change over time as a time series analysis and as a within-treatment experiment design. The suspicious records mentioned in the previous section were dropped.

Over time, we can see that for each location the treatment week tends to be the highest. However, often a control week will be somewhat close to or higher than the treatment week.

Table: Average weekly usage per location

Location	Week	Treatment	Average Amount Used	Flu Trends
0	10	0	1.3	0.0
	11	0	0.3	0.0
	12	0	0.2	0.0
	13	0	0.2	61.0
	14	0	0.0	31.0
	14	**1**	**2.2**	3.0
1	10	0	5.5	0.0
	11	0	11.7	0.0
	12	**1**	**12.3**	0.0
	13	**1**	**11.7**	57.7
	14	0	10.8	3.8
	14	**1**	**10.0**	31.0
2	10	0	14.6	0.0
	11	**1**	**21.8**	0.0
	12	0	10.6	0.0
	13	**1**	**10.4**	73.2
3	10	0	10.6	0.0
	11	0	8.8	0.0
	12	**1**	**14.4**	0.0
	13	0	5.5	75.8
4	11	0	3.0	0.0
	12	**1**	**3.5**	0.0
	13	**1**	**1.0**	66.5
	14	0	1.5	3.8

Location	Week	Treatment	Average Amount Used	Flu Trends
5	11	0	17.5	0
	12	0	5.2	0
	13	0	5	73.2
	14	**1**	**6.4**	9.2
6	11	0	4.6	0
	12	0	5.8	0
	14	**1**	**8.6**	9.2
	15	0	5.6	3.2
7	11	0	12	0
	12	**1**	**4.5**	0
	13	**1**	**4.5**	80
	14	0	4	5
8	11	0	1.25	0
	12	0	3.6	0
	13	0	2	73.2
	14	**1**	**2**	5
9	11	0	12.5	0
	12	**1**	**18.5**	0
	13	**1**	**13**	73.2
	14	0	16	11.5
10	10	0	4	0
	11	0	13	0
	13	**1**	**7**	38
	14	0	4	0

Having such a small pool per location and seeing such high variability, we decided to run a power test on our data. A power test tells us given the data we have, do we have enough data to give us a meaningful result. If not, *how much* data would we need to get a meaningful result?

<u>Power Analysis Part I: Treating each location as a mini-experiment</u>

Design We treated each location as a its own mini-experiment. We rejected any data points that were suspicious such as measurement errors or dates with extremely high fluctuations or evidence of sabotage. For each location, we considered the treatment group size to be the number of days the location was in treatment, and the control group size was the number of days the location was in control. We then calculated the mean and standard deviation for each group per location. We ran a simulation of 10,000 experiments given those means and standard deviations and returned the power calculation back, defined as the proportion of simulations that resulted in the rejection of the null hypothesis. Next, we ran a simulation of 10,000 experiments to see how many days in control and treatment would each location have to be in to see a statistically significant difference between control and treat at least 80% of the time. We set the minimum at 30 days each. See Appendix C for coding details.

Results: The results show that for most of the locations, we did not have enough data points to make a determination. The statistical power of models for most of the locations was about 5%. While there were two locations that had enough of a difference to be statistically significant more than 80% of the time, one was a very low volume location and the only only had 25% of all days in control. To get enough data in general to get statistical significance, most locations needed the experiment to run from two months to two years. A few locations had such similar standard deviations and means that there was no number of days that allowed for convergence.

Power of Existing Data	Days required for Statistical Significance (80%)
Power for yarn shop location: 0.9568 treatment mean, sigma: 2.2, 1.64 control mean, sigma: 0.36, 0.73	With 30 days per group in yarn shop location, power = 1.000
Power for eca- front desk location: 0.0879 treatment mean, sigma: 11.87, 8.32 control mean, sigma: 10.46, 5.35	With 390 days per group in eca- front desk location, power = 0.816
Power for ped- not well location: 0.1558 treatment mean, sigma: 16.1, 9.39 control mean, sigma: 12.6, 6.54	With 85 days per group in ped- not well location, power = 0.800
Power for ped - well location: 0.8667 treatment mean, sigma: 14.4, 4.51 control mean, sigma: 8.5, 3.57	With 30 days per group in ped - well location, power = 1.000
Power for concur - 3rd floor location: 0.0502 treatment mean, sigma: 2.25, 2.76 control mean, sigma: 2.33, 2.65	Concur 3rd Floor location did not converge
Power for concur - 16th floor location: 0.0834 treatment mean, sigma: 6.4, 2.61 control mean, sigma: 8.64, 14.72	With 343 days per group in concur - 16th floor location, power = 0.800
Power for Avalon - Gym location: 0.6738 treatment mean, sigma: 8.6, 3.21 control mean, sigma: 5.33, 2.5	With 30 days per group in Avalon - Gym location, power = 0.994
Power for concur -12th floor location: 0.1408 treatment mean, sigma: 4.5, 3.85 control mean, sigma: 8.57, 13.24	With 92 days per group in concur -12th floor location, power = 0.801
Power for concur - 10th floor location: 0.0834 treatment mean, sigma: 2.0, 1.0 control mean, sigma: 2.36, 1.74	With 245 days per group in concur - 10th floor location, power = 0.800
Power for concur - 5th floor location: 0.0519 treatment mean, sigma: 14.57, 8.46 control mean, sigma: 14.25, 12.59	Concur 5th Floor location did not converge
Power for Avalon - Lobby location: 0.0 treatment mean, sigma: 7.0, nan Control mean, sigma: 7.0, 5.2	Avalon - Lobby Location did not converge

Power Analysis Part II: Ignoring Location and looking at the Power of the set

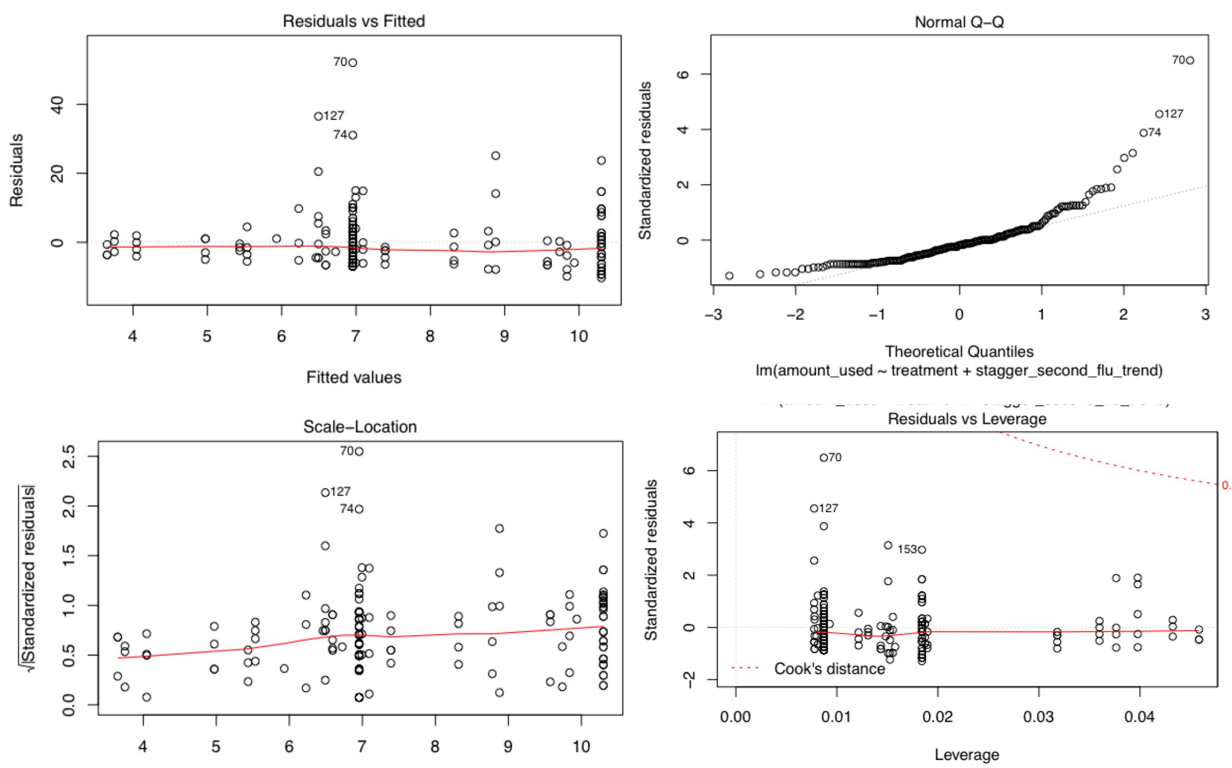
We have location-level detail on power, but perhaps looking at the whole set's power will tell us how the data's oddities would smooth out over all locations.

The Power Analysis for the entire dataset is 0.6784, which means that data similar to ours should be able to reject the null hypothesis that the sign had an effect on usage 67.8% of the time. Next, we looked to see how many days in control and treatment we would have needed to get a Power necessary to run our experiment. The algorithm shows we needed to have 137 Control and 137 Treat days to have an 80% of rejecting the null hypothesis. Our dataset has 72 valid treatment days and 129 valid control days. If we had a few more weeks, this experiment would have had an 80% chance of rejecting the null hypothesis.

c. Models - Linear Mixed Models

An OLS model based upon this data, removing suspicious terms and clustering on week and location returns a model with statistically significant treatment effect (see model summary on next page) and reasonable summary plots as shown below, but the explained variance is very low, at around 3.4%. More importantly, ordinary least squares is not the correct model to apply to the experiment as OLS assumes independence and this data is serially correlated due to the repeated observations on the same subjects (the locations).

Model graphs showing best OLS model



The nature of the experiment calls for panel data methods in order to allow for analysis and inference testing. The “plm” package in R is used to automatically process the time-dependent data and perform linear regression on the transformations. Two models are evaluated, a first difference model and a random effects model.

The first-difference model discards the assumption that there is no correlation between the individual subject and the independent variables and explicitly allows for unobserved effects to correlate with them. This approach does require that the error term be uncorrelated with the independent variables in all time periods, which is a reasonable assumption in this case. This model shows no statistically significant treatment effect.

Finally, a random effects model is fitted using the “plm” package. A stronger assumption is needed here that unobserved effects are uncorrelated with the independent variable. This model also shows no statistically significant treatment effect.

One thing that was interesting is that both panel data models show that stagger_second_flu_trend is statistically significant which might hint at the fact that CDC announcements and other media coverage can change people’s hand sanitizing behavior. Unfortunately, the effect is practically insignificant at a fraction of a gram.

	Dependent variable:		
	amount_used		
	OLS	panel linear	
	Clustered	First-Difference	Random Effects
	(1)	(2)	(3)
treatment	3.344**	11.709	14.726
	(1.454)	(21.126)	(20.333)
stagger_second_flu_trend	-0.033	0.140*	0.120*
		(0.071)	(0.071)
Constant	6.956***		27.893**
	(1.462)		(13.242)
Observations	200	35	46
R2	0.044	0.118	0.089
Adjusted R2	0.034	0.091	0.046
Residual Std. Error	8.048 (df = 197)		
F Statistic	4.490** (df = 2; 197)	4.390** (df = 1; 33)	2.091 (df = 2; 43)
Note:			*p<0.1; **p<0.05; ***p<0.01

d. Non-compliance

We do not have non-compliance concerns with this experiment due to the nature of the model design. This was confirmed with our advising professor.

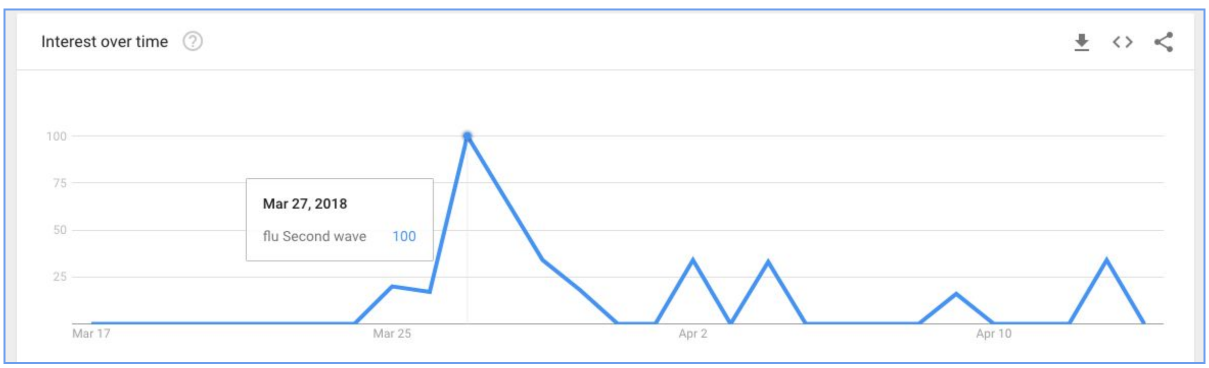
e. Attrition

During the experiment run, we had one bottle stolen and one sign stolen. We dealt with these data points by excluding them from our model.

7. Confounding Factors and Covariates

a. CDC announcements - Flu & Superbugs

Right before the experiment run, the Centers for Disease Control (CDC) released a press release about a resurgence of the flu this winter[6]. During the experiment, several popular news outlets released articles based upon this press release. The most prominent of these articles was released by TIME on March 27[7]. The article states that the more severe flu strain H3N2 was dominant, but now influenza B is becoming more prominent and triggering a second wave of the flu. Within seven days of this article the purell bottles saw large spikes in use in several locations. In addition to influenza epidemics, there was a great deal of coverage from the FDA, CDC, and news outlets about so-called "superbugs," bacteria that are not affected by traditional antibiotic medications.



Google Trends output for "Flu Second Wave" in March and April 2018 in the United States[8]

b. Google Trends

To study the effects media had on our experiment, we needed to find one comprehensive measurement of effects. Google Trends[9] has long been a resource used by statisticians to understand what current events are concerning to the public over defined time periods. We pulled data regarding influenza and superbugs to use as a proxy for how people get their news and follow up for details. It's difficult to measure when and if local news outlets in our selected

---

[6] https://www.cdc.gov/media/releases/2018/a0209-widespread-flu-activity.html
[7] http://time.com/5216990/influenza-b-flu-season-2018/
[8] https://trends.google.com/trends/explore?date=today%201-m&geo=US&q=flu%20Second%20wave
[9] https://trends.google.com/trends/

markets talked about the health concerns, but we can understand what people have on their minds through Google Trends. We used the United States aggregates for the 90 days leading up to and including our study on “Flu Second Wave” and “Drug Resistant.” We found that people generally had heightened concerns for superbugs than the flu but we found there was a very large spike for Flu Second Wave occurring simultaneously with the TIME article.

8. Limitation

a. Measurement error

During the experiment run, it became apparent that our scales were not finely tuned to measure grams with much accuracy. Some days a scale would show that a bottle of Purell had 0 or -1 change in weight.This is likely due a limitation on the repeatability and precision of the scales used and probably indicates a 0g use day.

b. Sabotage

During the experiment run, one of the bottles (Concur 12th Floor) gained 158 grams or 87 pumps’ worth of Purell. This same bottle was also known by an employee on the 12th floor to be part of an experiment. We hypothesize that this bottle had an adulterant added to it so we removed that datapoint from consideration.

See section 6e - Attrition for details on the missing bottles and signs.

9. Findings/Conclusion

The between subject naive design showed statistical significance, but only accounted for 3.4% of the variability in the set. The within-subject designs (random effects model) did not show a statistical significance. Several other variants of the models including Google Trends and staggered Google Trends as covariates did not improve matters. The power analysis shows that for most locations, we had less than a 20% chance of finding statistical significance with the data we had. We are left with the conclusion that we fail to reject the null hypothesis that a sign encouraging passers-by to use hand sanitizer has no effect.

10. Appendices

a. Intended Measurement and Treatment Schedule

As discussed in Section 3b, the treatment schedule was not implemented exactly as planned, but in general this schedule was followed.

	Pilot Study							Regular Study - Week 1						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Location	March 4	March 5	March 6	March 7	March 8	March 9	March 10	March 11	March 12	March 13	March 14	March 15	March 16	March 17
Omar's Pediatrician - Well		Control	Control	Control	Control	Control			Control	Control	Control	Control	Control	
Omar's Pediatrician - Not Well		Control	Control	Control	Control	Control			Treat	Treat	Treat	Treat	Treat	
Yarn Store									Control	Control	Control	Control	Control	Control
Nikki's Gym									Control	Control	Control	Control	Control	Control
Carmen's Gym									Control	Control	Control	Control	Control	
Carmen's Apartment Lobby								Control						
Concur Floor 3									Control	Control	Control	Control	Control	
Concur Floor 5									Control	Control	Control	Control	Control	
Concur Floor 10									Control	Control	Control	Control	Control	
Concur Floor 12									Control	Control	Control	Control	Control	
Concur Floor 16									Control	Control	Control	Control	Control	
Number of Observations	0	2	2	2	2	2	0	1	10	10	10	10	10	2

	Regular Study - Week 2							Regular Study - Week 3 (also Spring Break!)						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Location	March 18	March 19	March 20	March 21	March 22	March 23	March 24	March 25	March 26	March 27	March 28	March 29	March 30	March 31
Omar's Pediatrician - Well		Treat	Treat	Treat	Treat	Treat			Control	Control	Control	Control	Control	
Omar's Pediatrician - Not We		Control	Control	Control	Control	Control			Treat	Treat	Treat	Treat	Treat	
Yarn Store		Control	Control	Control	Control	Control	Control		Treat	Treat	Treat	Treat	Treat	Treat
Nikki's Gym		Treat	Treat	Treat	Treat	Treat	Treat		Control	Control	Control	Control	Control	Control
Carmen's Gym		Control	Control	Control	Control	Control								
Carmen's Apartment Lobby	Control													
Concur Floor 3		Treat	Treat	Treat	Treat	Treat			Control	Control	Control	Control	Control	
Concur Floor 5		Treat	Treat	Treat	Treat	Treat			Control	Control	Control	Control	Control	
Concur Floor 10		Control	Control	Control	Control	Control			Treat	Treat	Treat	Treat	Treat	
Concur Floor 12		Treat	Treat	Treat	Treat	Treat			Control	Control	Control	Control	Control	
Concur Floor 16		Control	Control	Control	Control	Control			Treat	Treat	Treat	Treat	Treat	
Number of Observations	1	10	10	10	10	10	2	0	9	9	9	9	9	2

	Regular Study - Week 4							Regular Study - Week 5						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Location	April 1	April 2	April 3	April 4	April 5	April 6	April 7	April 8	April 9	April 10	April 11	April 12	April 13	April 14
Omar's Pediatrician - Well														
Omar's Pediatrician - Not We														
Yarn Store														
Nikki's Gym														
Carmen's Gym		Treat	Treat	Treat	Treat	Treat			Control	Control	Control	Control	Control	
Carmen's Apartment Lobby	Treat							Control						
Concur Floor 3		Treat	Treat	Treat	Treat	Treat								
Concur Floor 5		Treat	Treat	Treat	Treat	Treat								
Concur Floor 10		Control	Control	Control	Control	Control								
Concur Floor 12		Treat	Treat	Treat	Treat	Treat								
Concur Floor 16		Control	Control	Control	Control	Control								
Number of Observations	1	6	6	6	6	6	0	1	1	1	1	1	1	0

b. Raw Data

Figure 1: All data. Suspicious data noted in red.

	Avalon - Gym	Avalon - Lobby	Concur - 3rd floor	Concur - 5th floor	Concur - 10th floor	Concur - 12th floor	Concur - 16th floor	ECA - Front Desk	Ped - Not Well	Ped - Well	Yarn Shop
March 5, 2018									17	10	
March 6, 2018									11	7	
March 7, 2018									9	14	
March 8, 2018									27	8	0
March 9, 2018									9	14	2
March 10, 2018								0			2
March 11, 2018		4						11			
March 12, 2018	5		0				1	10	12	6	2
March 13, 2018	5		8	18	1	5	7	7	20	6	0
March 14, 2018	7		5	17	1	5	-2	17	34	10	0
March 15, 2018	5		1	9	0	0	59	16	18	12	0
March 16, 2018	1		1	6	3	38	3	10	25	10	0
March 17, 2018								5			0
March 18, 2018		13						17			
March 19, 2018	3		1	25	4	3	2	6	11	18	0
March 20, 2018	6		2		6	6	10	7	7	12	0
March 21, 2018	2		9		3	9	7	20	4	9	0
March 22, 2018	8		-4		1	-1	1	13	14	13	1
March 23, 2018	10		2	12	4	0	6	19	17	20	0
March 24, 2018								10			0
March 25, 2018								11			
March 26, 2018			1	23	2	273	4	6	9	2	0
March 27, 2018			1	22	0	11	3	34	5	10	0
March 28, 2018			1	6	2	3	6	20	5	203	0
March 29, 2018			1	7	2	3	6	5	11	4	0
March 30, 2018			-1	7	4	1	6	2	22	6	0
March 31, 2018								5			1
April 1, 2018		7						10			
April 2, 2018	7		-1	5	-1	-158	8	10			0
April 3, 2018	12		3	7	1	0	4	4			1
April 4, 2018	11		0	43	2	6	6	11			4
April 5, 2018	9		2	152	3	6	10	12			0
April 6, 2018	4		1	9	-1	-3	4	16			3
April 7, 2018								9			3
April 8, 2018		4									
April 9, 2018	7										
April 10, 2018	7										
April 11, 2018	7										
April 12, 2018	2										
April 13, 2018	5										

Figure 2: Suspicious data removed. Treatment highlighted dark blue and control in light blue.

	Avalon - Gym	Avalon - Lobby	Concur - 3rd floor	Concur - 5th floor	Concur - 10th floor	Concur - 12th floor	Concur - 16th floor	ECA - Front Desk	Ped - Not Well	Ped - Well	Yarn Shop
March 5, 2018									17	10	
March 6, 2018									11	7	
March 7, 2018									9	14	
March 8, 2018									27	8	0
March 9, 2018									9	14	2
March 10, 2018								0			2
March 11, 2018		4						11			
March 12, 2018	5		0				1	10	12	6	2
March 13, 2018	5		8	18	1	5	7	7	20	6	0
March 14, 2018	7		5	17	1	5		17	34	10	0
March 15, 2018	5		1	9	0	0	59	16	18	12	0
March 16, 2018	1		1	6	3	38	3	10	25	10	0
March 17, 2018								5			0
March 18, 2018		13						17			
March 19, 2018	3		1	25	4	3	2	6	11	18	0
March 20, 2018	6		2		6	6	10	7	7	12	0
March 21, 2018	2		9		3	9	7	20	4	9	0
March 22, 2018	8				1		1	13	14	13	1
March 23, 2018	10		2	12	4	0	6	19	17	20	0
March 24, 2018								10			0
March 25, 2018								11			
March 26, 2018			1	23	2		4	6	9	2	0
March 27, 2018			1	22	0	11	3	34	5	10	0
March 28, 2018			1	6	2	3	6	20	5		0
March 29, 2018			1	7	2	3	6	5	11	4	0
March 30, 2018				7	4	1	6	2	22	6	0
March 31, 2018								5			1
April 1, 2018		7						10			
April 2, 2018	7			5			8	10			0
April 3, 2018	12		3	7	1	0	4	4			1
April 4, 2018	11		0	43	2	6	6	11			4
April 5, 2018	9		2		3	6	10	12			0
April 6, 2018	4		1	9			4	16			3
April 7, 2018											3
April 8, 2018		4									
April 9, 2018	7										
April 10, 2018	7										
April 11, 2018	7										
April 12, 2018	2										
April 13, 2018	5										

c. Code[10]

Code for ETL:

```
from pathlib import Path
import pandas as pd
import numpy as np
import random

import statsmodels.formula.api as sm
import math
from sklearn import linear_model

from matplotlib import pyplot as plt
import seaborn as sns
sns.set(color_codes=True)
%matplotlib inline


import datetime
import sys

### for debugging purposes
###sys.version
### should say:
###'3.6.3 |Anaconda custom (64-bit)| (default, Oct  6 2017, 12:04:38) \n
###[GCC 4.2.1 Compatible Clang 4.0.1 (tags/RELEASE_401/final)]'


### begin ETL

d = pd.read_csv('../data/raw_data.csv')
d_working = d.dropna(how='any')


### create numeric mapping for locations
unique_locs = d.Location.value_counts().index.tolist()
loc_map = [x for x in range(len(unique_locs))]
loc_set = dict(zip(unique_locs, loc_map))
#loc_set

d_working['numeric_location'] = d_working['Location'].map(lambda x: loc_set[x] if x in
loc_set.keys() else -1)

### get dummies for locations
dum = pd.get_dummies(d_working['numeric_location'])
dum.columns = ['loc_%s' % (x) for x in dum.columns.tolist()]
d_working = pd.concat([d_working, dum], axis = 1)

### create week aggregates

def week_num(date_string):
```

[10] All code can be found on GitHub https://github.com/mathaholic/sanitization_habits_statistical_analysis

```
    date = datetime.datetime.strptime(date_string,'%m/%d/%y')
    return datetime.date.isocalendar(date)[1]

d_working['week_number'] = d_working['Date'].apply(week_num)

### create a set geared to reject all suspicious datapoints
d_tsa = d_working[d_working['is_suspicious'] ==0]
```

Power Code

```
### find the power of our experiment per location
import scipy.stats

d_tsa = d_working[d_working['is_suspicious'] ==0]


def find_power(loc, sims):
    """
    for a given numeric location, return the power of the set
    """
    d_test = d_tsa[d_tsa['numeric_location']  == loc]
    treat_group = d_test[d_test['treatment'] == 1]['amount_used']
    ctrl_group = d_test[d_test['treatment'] == 0]['amount_used']
    n_per_group = int(np.ceil(np.mean([treat_group.shape[0],ctrl_group.shape[0]])))
    treat_mean = treat_group.mean()
    ctrl_mean = ctrl_group.mean()
    group_means = [treat_mean, ctrl_mean]
    treat_sigma = treat_group.std()
    ctrl_sigma = ctrl_group.std()
    group_sigmas = [treat_sigma, ctrl_sigma]
    ### number of simulations
    n_sims = sims
    ###store p val for each simulation
    sim_p = np.empty(n_sims)
    sim_p.fill(np.nan)

    for i_sim in range(n_sims):

        data = np.empty([n_per_group, n_groups])
        data.fill(np.nan)

        # simulate the data for this 'experiment'
        for i_group in range(n_groups):


              data[:, i_group] = np.random.normal(
                loc=group_means[i_group],
                scale=group_sigmas[i_group],
                size=n_per_group
            )

        result = scipy.stats.ttest_ind(data[:, 0], data[:, 1])

        sim_p[i_sim] = result[1]

    # number of simulations where the null was rejected
    n_rej = np.sum(sim_p < 0.05)
```

```
    prop_rej = n_rej / float(n_sims)

    print("Power for %s location: %s\n \
           \ttreatment mean, sigma: %s, %s\n \
           \tcontrol mean, sigma: %s, %s\n " \
          % (d_test['location'].iloc[0],prop_rej, round(treat_mean,2),
round(treat_sigma,2), round(ctrl_mean,2), round(ctrl_sigma,2)))
    return


for i in range(len(loc_set)):

    find_power(i, 10000)

def find_sample_required(location, sims, desired_power):
    """ find the number of days required for each location to
    see a statistically significant difference in control vs mean"""

    d_test = d_tsa[d_tsa['numeric_location']  == location]
    treat_group = d_test[d_test['treatment'] == 1]['amount_used']
    ctrl_group = d_test[d_test['treatment'] == 0]['amount_used']

    treat_mean = treat_group.mean()
    ctrl_mean = ctrl_group.mean()
    group_means = [treat_mean, ctrl_mean]
    treat_sigma = treat_group.std()
    ctrl_sigma = ctrl_group.std()
    group_sigmas = [treat_sigma, ctrl_sigma]

    # start at 30 days each
    n_per_group = 30



    n_groups = len(group_means)

    # number of simulations
    n_sims = sims

    # power level that we would like to reach
    desired_power = desired_power

    # initialise the power for the current sample size to a small value
    current_power = 0.0
    res_statement = []
    # keep iterating until desired power is obtained
    while current_power < desired_power:

        data = np.empty([n_sims, n_per_group, n_groups])
        data.fill(np.nan)

        for i_group in range(n_groups):

            data[:, :, i_group] = np.random.normal(
                loc=group_means[i_group],
                scale=group_sigmas[i_group],
                size=[n_sims, n_per_group])
```

```
        result = scipy.stats.ttest_ind(
            data[:, :, 0],
            data[:, :, 1],
            axis=1)

        sim_p = result[1]

        # number of simulations where the null was rejected
        n_rej = np.sum(sim_p < 0.05)

        prop_rej = n_rej / float(n_sims)

        current_power = prop_rej

        res_statement.append("With {n:d} days per group in {s} location, power =
{p:.3f}".format(
            n=n_per_group,
            p=current_power,
            s=d_test['location'].iloc[0]
        ))

        # increase the number of samples by one for the next iteration of the loop
        n_per_group+=1
    return print(res_statement[-1])

for i in range(len(loc_set)):

    find_sample_required(i, 5000, 0.8)
```

d. Model Evolution - previous versions of model

Please visit the modeling folder of our Github repo[11] for previous versions of the model.

e. Third Party Research[12]

Google Trends data: Collected on April 17, 2018 for the previous 90 days and limited to the United States. Table below shows the data from the days the experiment was in progress.

Day	"drug resistant"	"flu second wave"
3/4/18	15	0
3/5/18	37	0

[11] https://github.com/mathaholic/sanitization_habits_statistical_analysis/tree/master/modeling
[12] Add data sourced from Google Trends: https://trends.google.com/trends/explore?q=drug%20resistant and https://trends.google.com/trends/explore?q=flu%20second%20wave

3/6/18	50	0
3/7/18	0	0
3/8/18	13	0
3/9/18	39	0
3/10/18	0	0
3/11/18	44	0
3/12/18	0	0
3/13/18	62	0
3/14/18	12	0
3/15/18	25	0
3/16/18	65	0
3/17/18	15	0
3/18/18	15	0
3/19/18	62	0
3/20/18	38	0
3/21/18	0	0
3/22/18	13	0
3/23/18	66	0
3/24/18	0	0
3/25/18	30	0
3/26/18	38	46
3/27/18	52	93
3/28/18	26	63
3/29/18	93	64
3/30/18	55	100

3/31/18	46	0
4/1/18	0	38
4/2/18	39	31
4/3/18	89	0
4/4/18	64	15
4/5/18	0	0
4/6/18	79	0
4/7/18	30	0
4/8/18	15	0
4/9/18	38	0
4/10/18	50	0
4/11/18	51	0
4/12/18	38	0
4/13/18	80	16