Online Appendix

Does Precise Case Disclosure Limit Precautionary Behavior?

Evidence from COVID-19 in Singapore

August 13, 2021

Contents

1 Data Manipulation

1.1 Location Data Description

Lifesight provided our principal dataset, granular location information for individuals over long time periods. Each observation in this main dataset is an individual ID, unique to the phone; a timestamp; and a lat-long coordinate for that person at that time. Though they provided data for January through March 2020 with parts of 2019, our main analysis uses the 2020 data through March 17$^{\text{th}}$, the end of our period of study. Table 1 reports raw observation counts — after dropping duplicates — by month.

Table 1: **Basic Monthly Summary**

Time Period	Observations	ID-Day Count	Unique IDs
Jan 2019	311,021,056	6,909,862	1,295,054
Feb 2019	209,078,692	4,222,157	873,720
Oct 2019	195,046,083	8,789,171	1,894,819
Nov 2019	626,863	64,894	64,881
Dec 2019	171,326,775	7,131,877	1,722,650
Jan 2020	405,059,907	11,180,595	2,312,115
Feb 2020	512,084,761	12,502,090	2,492,687
Mar 2020	642,199,102	10,292,256	1,934,968

Note: We include the summary for data for months in which we have partial data, most extreme November 2019, though our analysis only includes the 2020 data for which we have complete information.

Our dataset is not balanced; individuals do not appear in our coverage period daily nor do they provide locations on a consistent basis. In our empirical analysis we assume that any data missing for an individual is not a function of their behavior. That is, the data available for an individual should be representative, if not completely, of their movement within the day. Table 2 provides a deeper look at observations available by ID (person), the first section on unique people per day for a given month and the second section on the number of observations per person-day. The first section makes clear that ID counts alone are not sufficient as measures of activity volume. There is significant variance both across months and within months in terms of people identified in the dataset. The second demonstrates the wide variance in observations per person-day; the data are highly skewed right, though there is not much variance in the number of times we see a person, conditional on that individual. This latter point supports our usage of models with individual fixed effects to capture inherent differences in observation frequency across individuals as well as focusing on day-level outcomes rather than putting significant weight on individual observations.

Table 2: **Summary Statistics on IDs**

	Unique IDs by Day		Daily Observations by ID		
Time Period	Average	Variance	Average	Median	Avg Deviation
Jan 2019	222,865	55,455	45.02	6	0.019
Feb 2019	150,778	31,214	49.52	6	−0.043
Oct 2019	283,492	101,562	22.19	2	−0.022
Nov 2019	2,163	11,823	9.66	1	−0.609
Dec 2019	230,042	151,142	24.02	2	−0.077
Jan 2020	360,639	74,252	36.23	3	−0.016
Feb 2020	431,047	66,220	40.97	3	0.016
Mar 2020	331,960	89,307	62.41	6	0.077

Note: The average "deviation" in daily observations by ID (person) is constructed by first calculating for each individual how many observations they produce per day deviates from their average (standard score) and averaging this score across all individuals for the month.

1.2 Data Cleaning

We take several passes at cleaning the raw dataset and remove observations that fit into one of several categories in the following order:

[1] Inaccurate longitude-latitude data

[2] People linked to devices with unrealistic travel behavior

Category 1 observations arise from issues with how the data is collected. One issue is that in the event that the precise location of a person cannot be determined, they might be "resolved" to a particular lat-long based on a guessed IP address. In the data this manifests as an unrealistic number of people in the exact same long-lat coordinates. For long-lat where we observe more than 500 unique people per week, we remove all observations associated with the long-lat for that particular week. Additionally, GPS data is occasionally collected with accuracy at a such low resolution it is useless for our purposes. We hence remove specific observations where the horizontal accuracy of a GPS reading is higher than 250m.

For category 2 observations we carry out simple operations on the data to calculate an individual's travel distance and speed between successive pings. Travel distances, like speeds, are not directly observable in the data. To calculate distance we take the aerial distance between two successive pings for an individual. To calculate speed between two pings, we take the distance and divide it by the elapsed time. This speed calculated is only a lower bound on speed and hence appropriate to check for unrealistically fast travel speeds.

What we classify as category 2 observations include people with excessive travel — more than 100 km per day — or unrealistic travel speeds, which we calculate as the distance traveled over the time elapsed between two location observations — over 140 kph, which is far greater than the highest speed limit within Singapore of 90 kph — in which case we remove observations for that person-day. This category also includes people with insufficient movement — individuals who do not change their long-lat position for the duration they are in the sample — in which case we remove that person entirely.

Table 3 illustrates how much data we drop through this procedure. Generally, about a third of observations are dropped per month, but the number of devices, or people, kept in the sample drops to a quarter. Most of these devices are dropped because they exhibited little movement of the life we observe them in the sample. Unfortunately, we cannot distinguish between collection errors or genuine lack of movement. Given the high bar we set for removing these observations — essentially no movement every single day — it seems reasonable that in the worst case that the data collection is correct for some of these people, they do not represent a large fraction of the populace. Recall that we have data prior to knowledge of the pandemic so these true cases would not be individuals exhibiting extreme isolation. This change is more obvious in Table 4 where the average daily observations per device increases significantly for the months of our study.

Table 3: **Summary of Basic Cleaned Data**

	No Conditions			Cleaned Data		
Time Period	Observations	ID-Day Count	Unique IDs	Observations	ID-Day Count	Unique IDs
Jan 2019	311,021,056	6,909,862	1,295,054	210,125,557	4,431,761	493,989
Feb 2019	209,078,692	4,222,157	873,720	130,457,919	2,836,238	390,246
Oct 2019	195,046,083	8,789,171	1,894,819	141,109,196	2,990,717	490,593
Nov 2019	626,863	64,894	64,881	429,868	15,937	15,927
Dec 2019	171,326,775	7,131,877	1,722,650	118,353,840	2,285,094	411,184
Jan 2020	405,059,907	11,180,595	2,312,115	287,106,021	4,140,000	546,178
Feb 2020	512,084,761	12,502,090	2,492,687	329,464,674	4,762,227	569,803
Mar 2020	642,199,102	10,292,256	1,934,968	413,326,755	4,616,688	457,482

Table 4: **ID Statistics for Cleaned Data**

	No Conditions				Cleaned Data			
	Unique Daily IDs		Daily Observations by ID		Unique Daily IDs		Daily Observations by ID	
Time Period	Average	Variance	Average	Avg Deviation	Average	Variance	Average	Avg Deviation
Jan 2019	222,865	55,455	45.02	0.019	142,931	25,847	47.42	−0.009
Feb 2019	150,778	31,214	49.52	−0.043	101,283	17,156	46.00	−0.096
Oct 2019	283,492	101,562	22.19	−0.022	96,454	48,280	47.19	0.106
Nov 2019	2,163	11,823	9.66	−0.609	637	3,173	26.97	−0.666
Dec 2019	230,042	151,142	24.02	−0.077	73,706	62,134	51.80	−0.138
Jan 2020	360,639	74,252	36.23	−0.016	133,538	40,266	69.35	0.005
Feb 2020	431,047	66,220	40.97	0.016	164,172	41,679	69.20	0.030
Mar 2020	331,960	89,307	62.41	0.077	148,891	28,568	89.55	0.034

Note: The average "deviation" in daily observations by ID (person) is constructed by first calculating for each individual how many observations they produce per day deviates from their average (standard score) and averaging this score across all individuals for the month.

Finally, as described in the main text, part of the analysis is conducted on a subsample of the data for which we have estimates of a person's home. Lifesight created these estimates based on regular pings from a person's phone during off hours, like the early morning or late evening. Table 5 recreates the first two panels of the same table from the main paper using this slice of the data. The additional restriction maintains about 80 to 90% of the original sample, though on about two-thirds of individuals. While we have no reason to believe the home estimation selects individuals in a way that is biased, individuals who ping more will be more likely to have a home estimate. This is evident from the higher observations per day in this subsample. We also highlight that the principle travel statistics we share in Panel B are roughly the same across the two subsamples. Note that in Section 2.3 of this appendix we additionally test whether results from the main analysis are robust to this home definition or subsample use; we find that they are robust.

Table 5: **Data Summary**

	Jan 2020	Feb 2020	Mar 2020
Panel A: Cell Phone Data			
Person-Day Count	3,425,997	3,818,295	2,244,850
Unique People	332,256	362,912	264,570
Avg Obs Per Person-Day	78.04	81.71	104.84
	(131.96)	(157.51)	(145.62)
Panel B: Travel Statistics			
Avg KM Traveled Per Day	20.00	14.29	16.95
	(25.90)	(22.86)	(24.70)
Avg % Staying Home	22.85	27.81	26.43
	(0.18)	(0.15)	(0.10)
Avg Subzones Visited Per Day	2.95	2.09	2.81
	(2.92)	(1.97)	(2.76)

Note 1: Data for March 2020 only covers through the 17^{th}, the end of our period of study. The standard deviation for select averages are presented in parentheses.
Note 2: Panels A and B use the a subsample of the data with home estimates available.

1.3 Geography Data

This section covers the process of linking Lifesight longitude-latitude data to specific places in Singapore, typically known as reverse geocoding. Reverse geocoding is easily possible by mapping services like Google but is prohibitive for over hundreds of millions of observations, and does not produce results necessarily useful for our analysis. The same concerns apply to open source solutions like Nominatim through Open Street Maps.

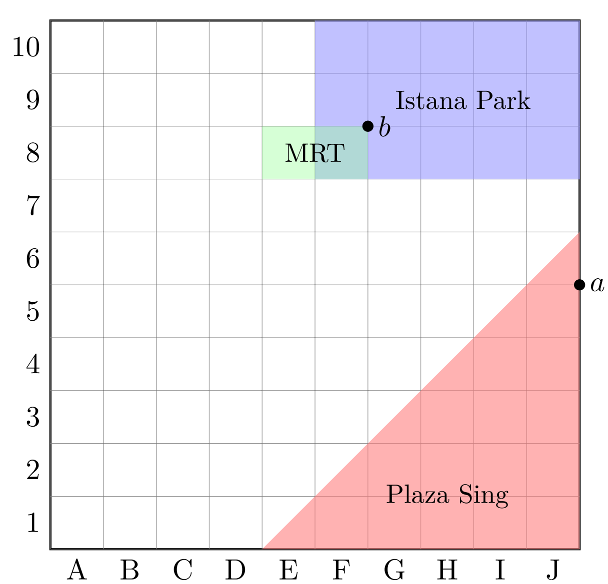
The methodology presented here instead uses more standard point-in-polygon identification using place and location data from Open Street Maps. The data provided lists building and land area common names and use classifications. We use the latter to determine the nature — for example, commercial, industrial, retail, or residential — of particular areas.

An issue that is not resolved with this methodology is that these places can overlap. In dense cities like Singapore there are many multi-use buildings, and horizontal GPS coordinates alone are not sufficient to identify what part of the building the individual uses. An additional trade off is that buildings and land-use records do not form a complete cover of Singapore.

Figure 1 depicts a highly stylized map of Singapore with some common challenges. The Istana Park, Plaza Singapura, and a mass rapid transit (MRT) station are included along with two sample location points, denoted a and b. Notice the Istana and MRT overlap in one square. White space is meant to depict areas without any specific place information available.

A simple geocoding operation would indicate location a is in Plaza Singapura. Location b cannot be assigned to a particular place, as it is contained within both Istana Park and the MRT station. In the main analysis of this paper we are somewhat agnostic in the specific locations of a and b in that we do not try to claim if b was in the park or the MRT station. As we describe in the next subsection, if the person is ever in a land area with a particular classification type, we claim that person visited that classification type that type. Location a would be tagged as retail. Location b would be tagged as both transit and recreational, though we do not use either of the latter classifications in the main analysis of the paper.

Figure 1: **Sample Geography**



1.4 Generating Outcome Data

Our outcome variable for individual i at time t with a home location in subzone j of area k are gathered in the set a_{it}. They include travel distance in meters ($TravelDist_{ijkt}$); a dummy which takes the value one if i stays within the subzone of their home ($StayHome_{ijkt}$); a dummy which takes the value one if i visits a subregion with an industrial-, commercial-, or retail-use classification ($IndRetCom_{ijkt}$); a dummy which takes the value one if i visits an with a residential-use classification ($Residential_{ijkt}$) outside the own home. In the following we provide a brief description on the construction of the summary variables:

1. $TravelDist_{ijkt}$: For each individual and day we order the device signals according to time. We proceed to calculate the aerial distance. Finally, $TravelDist_{ijkt}$ is the sum of these distances during a day.

2. $StayHome_{ijkt}$: For each individual and day we consider if two conditions are met. First, the individual needs to send signals from only one subzone. Secondly, the signals have to be within the subzone of their home. If both conditions are fulfilled, $StayHome_{ijkt}$ takes the value one.

3. $IndRetCom_{ijkt}$: For each individual and day we consider if an individual has sent at least one signal during the day from a place with a land class (see Section 1.3 of the Appendix) that is defined as industrial, retail or commercial.

4. $Residential_{ijkt}$: For each individual and day we consider if an individual has sent at least one signal during the day from a place with a land class (see Section 1.3 of the Appendix) that is defined as residential. To avoid that we simply count individuals when they are at home, we exclude observations that are extremely close to their home location. Specifically we exclude observations that have a long-lat coordinate identical to their home's up to the fourth decimal place.

2 Additional Analysis

2.1 Full Regression Models

In the following section we present additional regression evidence for model 1, the outflow travel regressions, of the main paper. In detail, each of the following tables presents one outcome variable of the set a_{it}: travel distance in meters ($TravelDist_{ijkt}$); a dummy which takes the value one if i stays within the subzone of their home ($StayHome_{ijkt}$); a dummy which takes the value one if i visits a place with an industrial-, commercial-, or retail-use classification ($IndRetCom_{ijkt}$); a dummy which takes the value one if i visits a place with a residential-use classification ($Residential_{ijkt}$) outside the own home. For each outcome variable the alternative specifications in each table drop fixed effects and run a naive pooled regression in specification (1), include only date fixed effect in specification (2), only individual fixed effects in specification (3), and both in specification (4).

Table 6: **Estimation of Local and General Response, TravelDist**

	TravelDist			
	(1)	(2)	(3)	(4)
$LocalCases_{jt-1}$	456.873***	426.596***	$-$124.491***	$-$61.433***
	(24.014)	(23.718)	(14.374)	(14.429)
$NonLocalCases_{jt-1}$	$-$361.327***	$-$532.699***	$-$91.125***	$-$28.045***
	(10.660)	(13.691)	(5.713)	(6.776)
Individual FE	No	No	Yes	Yes
Date FE	No	Yes	No	Yes
N	9,482,376	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (1). One observation corresponds to an individual on a specific date. Each model specification corresponds to the outcome variable of $TravelDist$, travel distance in meters. $LocalCases$ are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. Standard errors are reported in parentheses and clustered at the individual level.

Table 7: **Estimation of Local and General Response, HomeStay**

	HomeStay			
	(1)	(2)	(3)	(4)
$LocalCases_{jt-1}$	0.021^{***}	0.018^{***}	0.003^{***}	0.001^{***}
	(0.001)	(0.001)	(0.0003)	(0.0003)
$NonLocalCases_{jt-1}$	-0.004^{***}	-0.013^{***}	0.004^{***}	0.0001
	(0.0002)	(0.0003)	(0.0001)	(0.0002)
Individual FE	No	No	Yes	Yes
Date FE	No	Yes	No	Yes
N	9,482,376	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (1). One observation corresponds to an individual on a specific date. Each model specification corresponds to the outcome variable $HomeStay$, a dummy variable that takes the value one if an individual remains at their home subzone for an entire day. $LocalCases$ are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. Standard errors are reported in parentheses and clustered at the individual level.

Table 8: **Estimation of Local and General Response, IndComRet**

	IndComRet			
	(1)	(2)	(3)	(4)
$LocalCases_{jt-1}$	0.709^{***}	0.614^{***}	-0.166^{***}	-0.117^{***}
	(0.053)	(0.052)	(0.035)	(0.035)
$NonLocalCases_{jt-1}$	-0.429^{***}	-0.609^{***}	-0.099^{***}	-0.083^{***}
	(0.022)	(0.029)	(0.013)	(0.016)
Individual FE	No	No	Yes	Yes
Date FE	No	Yes	No	Yes
N	9,482,376	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (1). One observation corresponds to an individual on a specific date. Each model specification corresponds to the outcome variable $IndComRet$, a dummy variable that takes the value one if an individual enters at least one industrial, commercial or retail place. $LocalCases$ are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. Standard errors are reported in parentheses and clustered at the individual level.

Table 9: **Estimation of Local and General Response, Residential**

	Residential			
	(1)	(2)	(3)	(4)
$LocalCases_{jt-1}$	3.462^{***}	3.116^{***}	-0.097^{***}	-0.055^{**}
	(0.044)	(0.044)	(0.026)	(0.026)
$NonLocalCases_{jt-1}$	-1.208^{***}	-2.617^{***}	0.325^{***}	-0.029^{**}
	(0.022)	(0.028)	(0.011)	(0.013)
Individual FE	No	No	Yes	Yes
Date FE	No	Yes	No	Yes
N	9,482,376	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (1). One observation corresponds to an individual on a specific date. Each model specification corresponds to the outcome variable *Residential*, a dummy variable that takes the value one if an individual enters a residential place except their own residence. *LocalCases* are the number of local cases in a subregion announced in the evening of $t-1$. *NonLocalCases* are the cases of the region k announced. Standard errors are reported in parentheses and clustered on the individual level.

2.2 Excluding Dates

As we observe data anomalies on the 4th of February, this section shows results from our outflow regression analysis without observations from that date. Table 10 replicates the outward flow analysis summary from the main analysis. In comparison those results here we observe a slight change in the coefficients but not substantial enough to alter our interpretations of the results.

2.3 Definition of Local Cases

In our main analysis we study the travel behavior of individuals by considering COVID-19 cases close to an individual's home location, estimated by Lifesight. While we trust the residence estimates, the analysis necessarily drops individuals for whom home location estimates are not available. In addition, as we argue in the paper, individuals may not necessarily consider geographical distance from announced cases as much as the risk of potential contact with infected individuals. Thus, individuals may change their behavior not only for cases closes to home but also if they have visited areas outside home where a positive COVID-19

Table 10: **Estimation of Local and General Response without the 4th of February**

	TravelDist	StayHome	IndComRet	Residential
	(1)	(2)	(3)	(4)
$LocalCases_{jt-1}$	-59.918^{***}	0.132^{***}	-0.114^{***}	-0.047^{*}
	(14.445)	(0.034)	(0.035)	(0.026)
$NonLocalCases_{jt-1}$	-32.771^{***}	0.019	-0.088^{***}	-0.037^{***}
	(6.792)	(0.015)	(0.016)	(0.013)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Mean Outcome	13901	25.72	29.05	78.72
Mean Local Effect in Percent	-0.43	0.51	-0.39	-0.06
Mean Aggregate Effect in Percent	-0.24	0.07	-0.3	-0.05
N	9,306,704	9,306,704	9,306,704	9,306,704

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents results of regression model (1) excluding observations from the 4th of February. One observation corresponds to an individual on a specific date. Each model specification corresponds to a different outcome variable. *TravelDist* is the travel distance in meters, *StayHome* is a dummy variable that takes the value one if an individual remains at their home subzone for an entire day. *IndComRet* is a dummy that takes the value one if an individual enters at least one industrial, commercial, or retail place. *Residential* is a dummy that takes the value one if an individual enters a residential place except their own residence. Note, that we multiply outcome variables *StayHome*, *IndComRet* and *Residential* by 100 so that the coefficients are interpreted as percentage points. *LocalCases* are the number of local cases in a subregion announced in the evening of $t-1$. *NonLocalCases* are the cases of the region k announced. For all models we include individual and date fixed effects. We calculate the mean local effect and mean aggregate effect as the percentage difference from the average outcome. Standard errors are reported in parentheses and clustered at the individual level.

case has been announced.

Within this section we show the robustness of our results to specifications taking into account both of these considerations. We include all observations, independent of their home estimate, and further build a new measure of local cases. We specifically consider cases announced in subzones where an individual has travelled during the last five days. We do not restrict that the subzone is within a specific region. As individual travel may span over different regions and as we use date fixed effects to control for national trends in travel behavior, we are not able to identify local- and national-case responses separately. Nevertheless, this approach offers the possibility to demonstrate the robustness of our local-case estimates. The regression model is comparable to equation (1) in the main article:

$$\mathbf{a}_{ijt} = \beta_1 LocalCases_{jt-1} + \gamma_i + \rho_t + \varepsilon_{ijt}, \qquad (1)$$

where a_{it} is a set of outcome variables: travel distance in meters ($TravelDist_{ijt}$); a dummy which takes the value one if i visits a place with an industrial-, commercial-, or retail-use classification ($IndRetCom_{ijt}$); a dummy which takes the value one if i visits a place with a residential-use classification ($Residential_{ijt}$). Note that we exclude the outcome $StayHome_{ijkt}$ since we consider the entire sample, including individuals without a home estimate. $LocalCases_{jt-1}$ are the sum of the number of cases of those subzones j announced in the evening of $t-1$ that an individual has visited during the last five days.

We present results of this alternative model in Table 11. We find results quantitatively similar to those in the main paper for the effect of case announcements on travel distance; the probability of entering at least one industrial, retail, or commercial place in a day; and on the probability of entering a residential place.

2.4 Endogeneity Concern: Reverse Causality

Our identification is based on the assumption that new cases are exogenous to previous travel behavior of individuals. In particular we require that past behavior is not correlated with the number of cases. As COVID-19 is a contagious disease, the aggregate of previous behavior

Table 11: **Estimation of Local Response, New Definition**

	TravelDist	IndComRet	Residential
	(1)	(2)	(3)
$LocalCases_{jt-1}$	-61.747^{***}	-0.090^{***}	-0.072^{***}
	(5.281)	(0.011)	(0.009)
Individual FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
Mean Local Effect in Percent	-0.5	-0.33	-0.1
N	11,294,646	11,072,665	11,072,665

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model without considering the home location estimate as well as a new definition of local cases. One observation corresponds to an individual on a specific date. Each model specification corresponds to a different outcome variable. $TravelDist$ is the travel distance in meters, $IndComRet$ is a dummy that takes the value one if an individual enters at least one industrial, commercial or retail area. $Residential$ is a dummy that takes the value one if an individual enters a residential place except their own residence. Note, that we multiply outcome variables $IndComRet$ and $Residential$ by 100 such that the coefficients are interpreted in percentage points. $LocalCases$ are the number of local cases in a subregion announced in the evening of $t-1$ which an individual has visited during the last five days. For all models we include individual and date fixed effects. We calculate the mean local effect as the percentage difference from the average outcome. Standard errors are reported in parentheses and clustered at the individual level.

is certainly a key determinant in its spread. We believe, however, that in the time period of limited spread we examine, the link between individual behavior and new cases is tenuous. Part of this argument is that around 15% of cases in this time period were imported, and thus not a function of domestic behavior. We further provide regression evidence that past individual behavior is not correlated to increases in local cases.

In our initial test we evaluate if historic individual travel is correlated with local case counts by estimating the following regression model with day-individual level observations:

$$LocalCases_{jt} = \sum_{s=0}^{14} \beta_{t-s} a_{ijt-s} + \gamma_i + \rho_t + \varepsilon_{it}, \tag{2}$$

where $LocalCases_{jt}$ are the number of cases in subzone j announced at t. a_{ijt-s} is the set of outcome variables. We consider two outcome variables: the travel distance (in meters) of individual i with home location in j in $t-s$ ($TravelDist_{ijt-s}$) and a dummy which takes the value one if i stays within the subzone of their home j in $t-s$ ($StayHome_{ijt-s}$). γ_i and ρ_t are individual and day fixed effects, respectively.

The regression evaluates if the behavior of individual i on day t and the 14 preceding days are associated with local case announcements on t. Because a key assumption is that previous behavior is not correlated to local cases, we test if the coefficients of β_{t-s} are not significantly different from zero. However, if a typical individual's behavior can impact the spread of COVID-19, we would expect a positive correlation. Incorporating that the incubation period of COVID-19 is up to two weeks, we would see multiple positive coefficients for the fortnight before a case announcement. In the regression on home-stay behavior, we would expect negative coefficients if there is a significant linkage between behavior and spread.

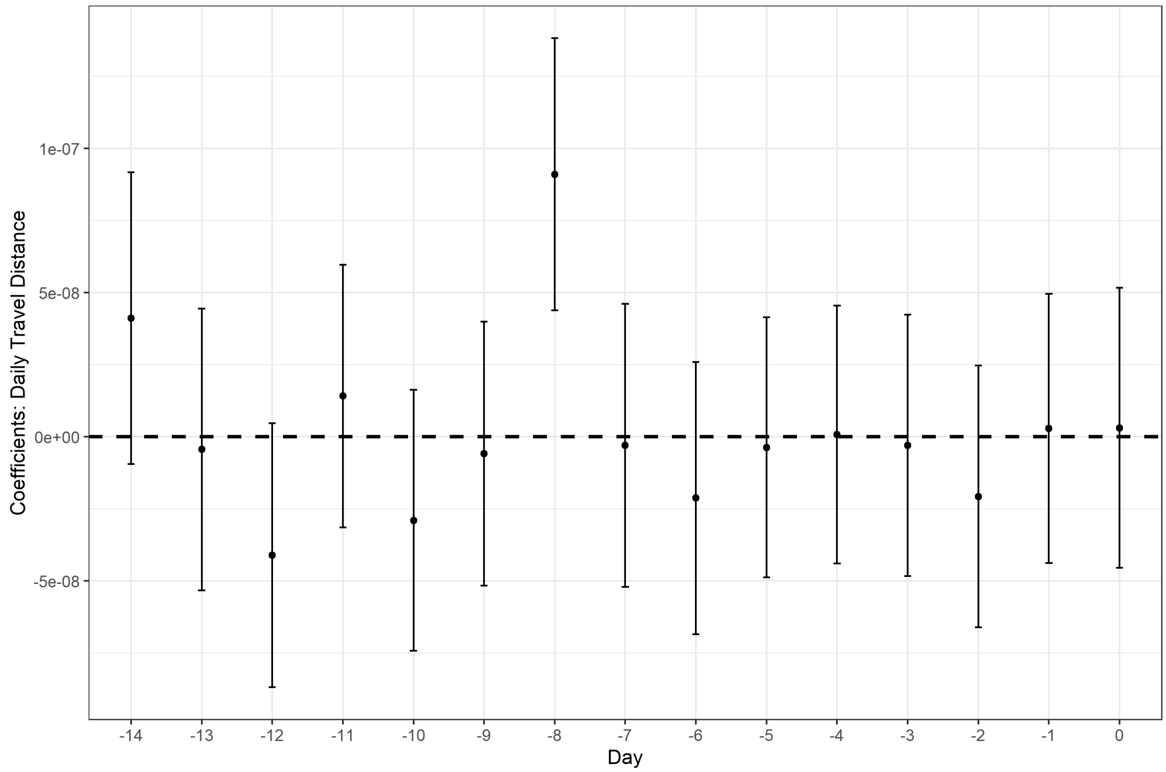
We show our results for travel and home-stay behavior in Figures 2 and 3, respectively. Both regression results generally support that there is no strong linkage between individual past behavior and local cases. For prior travel distance all coefficients, except the for 8 days before the announcement of a local case, are not significantly different from zero. Further, we do not observe any systematic pattern in these coefficients over the lagged times, a result which leads us to believe that the 8$^{\text{th}}$ day coefficient is an anomaly. The results are similar when considering the impact of an individual staying within a home subzone. Overall, these tests do not contradict our assumption that past individual behavior is not strongly correlated with local case announcements.

In a second set of tests, we evaluate the correlation between local cases and a higher inflow of individuals during the preceding fortnight. We use subzone-date level observations to estimate the following model:

$$LocalCases_{jt} = \sum_{s=0}^{14} \beta_{t-s} Visits_{jt-s} + \gamma_j + \rho_t + \varepsilon_{jt}, \tag{3}$$

where $LocalCases_{jt}$ are the number of cases in subzone j announced at t. $Visits_{jt-s}$ are the number of unique visitors in subzone j at time $t-s$. γ_j are the subzone and ρ_t time fixed effects. The regression shows if total visits in t and the 14 preceding days are correlated to the local cases announced in t. If activity is unrelated to the spread, we expect that the coefficients for all days are not significantly different from zero.

Figure 2: Regression Coefficients, Daily Travel Distance



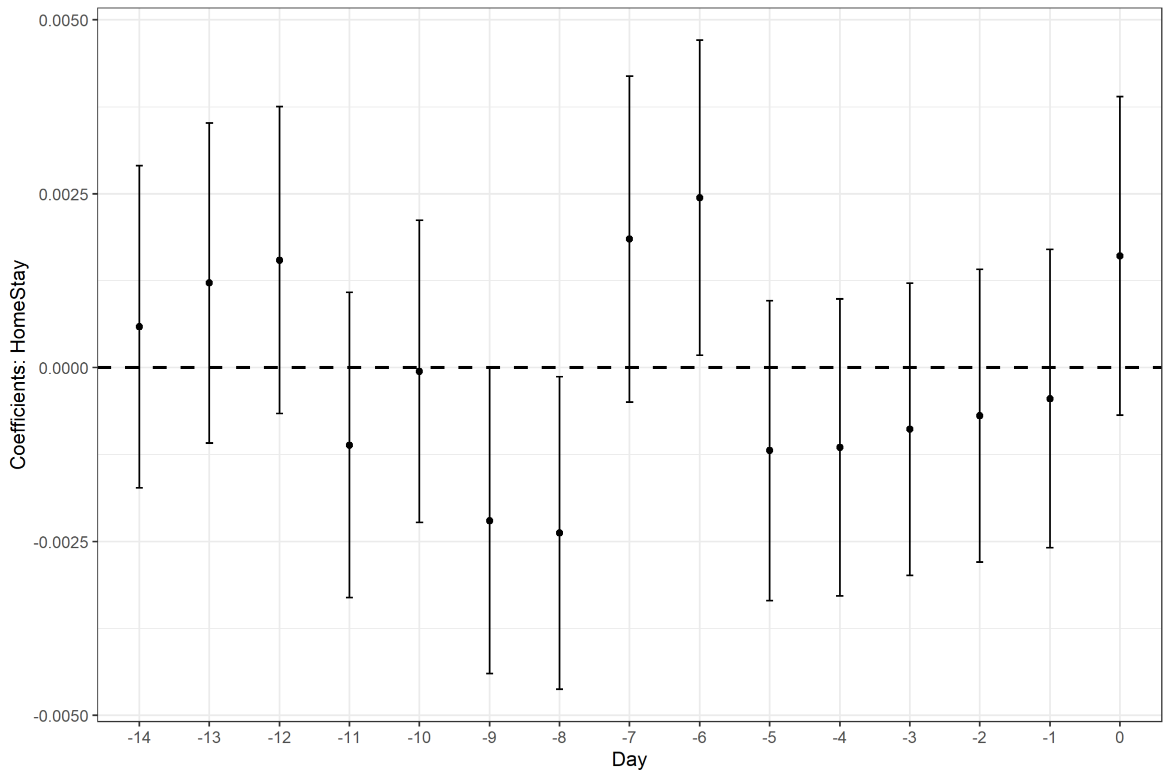
Notes: The figure reports regression coefficients of the number of local cases announced in $t = 0$ on the past daily travel distances. The coefficients correspond to the estimates of $\sum_{s=0}^{14} \beta_{t-s}$ in regression model 2. The regression includes individual and time fixed effects. The error bars are the 95% confidence interval.

Figure 4 shows the results of model 3. Similar to the previous tests we do not observe any specific pattern. Only two coefficients are significant at a five-percent significance level, and neither is associated with a consistent pattern across the preceding or succeeding lags. Overall, the test does not show any significant correlation between the total inflow of visitors and local cases. Taken together with the previous test, we are comfortable that our weak exogeneity assumption holds during the time period studied.

2.5 Identification Concern: Work-from-home Arrangements

When estimating the effect of local case announcements on individual's home location outflow and a location's inflow traffic, we interpret the effect as a voluntary reduction in an individual's travelled distance or visits. One potential concern is that work places use local

Figure 3: Regression Coefficients, Individuals Staying Home

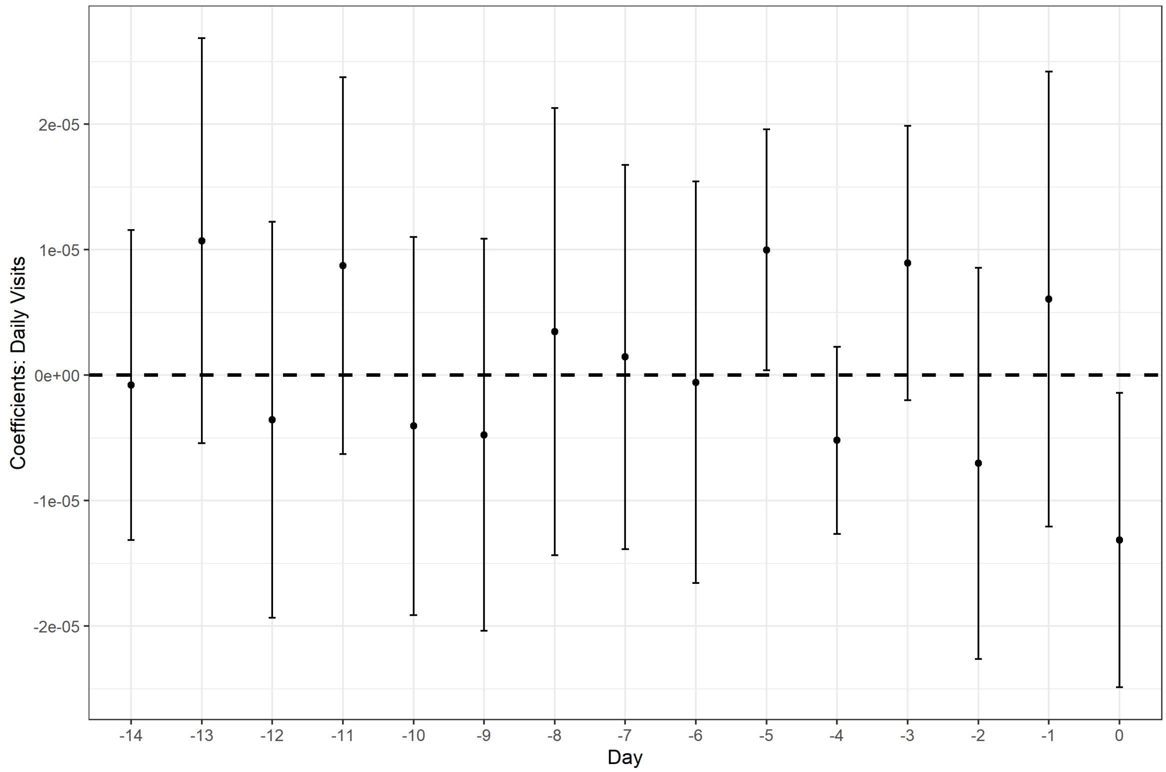


Notes: The figure reports regression coefficients of the number of local cases announced in $t = 0$ on the past stay of individuals in within their home subzone. The coefficients correspond to the estimates of $\sum_{s=0}^{14} \beta_{t-s}$ in regression model 2. The regression includes individual and time fixed effects. The error bars are the 95% confidence interval.

cases as a reason to switch from on-site work to work-from-home arrangements. In this case, we may not be estimating the precautionary behavior of individuals but rather involuntary reactions to their firms' work-from-home arrangements.

We first argue that it is unlikely that firms respond to local cases but rather to national trends in Singapore as well as government recommendations. Nevertheless, we show in the following robustness check that neither for our outflow nor in our inflow analysis we find evidence that work-from-home arrangements are a concern.

Figure 4: Regression Coefficients, Number of Visits



Notes: The figure reports regression coefficients of the number of local cases in $t = 0$ in subzone j on the number of visits in subzone j during the previous 14 days. The coefficients correspond to the estimates of $\sum_{s=0}^{14} \beta_{t-s}$ in regression model 3. The regression includes individual and time fixed effects. The error bars are the 95% confidence interval.

2.5.1 Outflow Analysis

We start with the outflow analysis. We use the original cellphone data to estimate an individual's work place. The individual-level variation allows us to control for cases at person's specific work place.

Our robustness exercises require identifying the workplace of most individual's in our sample. Using our cell phone data, we determine an individual's work place w as the second most-visited location (subregion) on record. While we cannot directly verify these workplace estimates, unlike with home data, we can show that our measurements are a decent approximation for this robustness exercise. With our estimates we construct the share of people working in workplace region j conditional on living in region i.[1] We then compare this number to the comparable shares provided by the 2020 Singapore Census.[2] To demonstrate the usefulness of our cellphone-based measurement, we report the normalized mean squared error from our workplace estimator. It obtains a value of 0.10, in comparison to 0.17 from an estimator that randomly assigns workplaces to individuals based on the overall distribution of employees across workplace regions.

Having determined work location w for each individual i, we test if cases in an individual's work location influence our estimation of local home cases' impact on travel. We propose two different regression models that extend our main analysis. The first explores if cases living in or visiting workplace w affects our previous findings on the impact of local and non-local cases on individual i's travel behavior. In the second approach we extend the control by also interacting the cases of an individuals workplace with the fraction of employees at an individual's home location that work in occupations more amenable to work-from-home agreements. In both models we test the stability of the effect of local cases on individual behavior.

The following regression model shows the first approach:

$$
\begin{aligned}
\mathbf{a}_{ijkt} = &\beta_1 LocalCases_{jkt-1} + \beta_2 NonLocalCases_{kt-1} + \\
&\beta_3 CasesWorkplace_{wt-1} + \beta_4 VisitedCasesWorkplace_{wt-1} + \gamma_i + \rho_t + \varepsilon_{ijkt},
\end{aligned} \tag{4}
$$

which is an extension to equation 1 in the main manuscript with two additional regressors. First, we add the number of cases for people living in the workplace w of individual i, $CasesWorkplace_{wt-1}$. Second, we include the number of positive cases who visited the workplace of the individual, $VisitedCasesWorkplace_{wt-1}$. As we are testing whether work-at-home arrangements impact our results, our null expectation is that our estimates of β_1

[1]Technically we construct the share working in planning area j conditional on living in region i. This is the most granular data available in the public 2020 census data for Singapore.

[2]As the census was conducted throughout 2020, work from home would be more commonplace than in early 2020, the time period for our cell phone data. Note, however, that respondents were asked to report their principal work's official address.

and β_2 are not substantially different from those in the main paper.

In a second robustness check, we add a new interaction term for workplace cases with the fraction of employees at an individual's home location that work in occupations with a high likelihood to permit work-from-home arrangements (WFH_{jk}).[3]

$$\begin{aligned}
\mathbf{a}_{ijkt} = &\beta_1 LocalCases_{jkt-1} + \beta_2 NonLocalCases_{kt-1} + \\
&\beta_3 CasesWorkplace_{wt-1} + \beta_4 VisitedCasesWorkplace_{wt-1} + \\
&\beta_5 CasesWorkplace_{wt-1} * WFH_{jk} + \beta_6 VisitedCasesWorkplace_{wt-1} * WFH_{jk} + \\
&\gamma_i + \rho_t + \varepsilon_{ijkt},
\end{aligned} \tag{5}$$

Because of individual fixed effects γ_i we can only identify the interaction with the cases at workplaces, not a coefficient on the variable WFH_{jk} alone. Again, we would expect that work-from-home arrangements should not change the impact of local and non-local cases.

We report results in Tables 12 and 13. We observe that the point estimates for the impact of local (β_1) and non-local (β_2) cases are fairly stable with the exception of the results for visiting non-home residential areas. Introducing the additional variables only decreases the precision of the estimates for the last two outcome variables (*IndComRet* and *Residential*). We favor models without these additional variables, as we cannot reject that the added variables are jointly insignificant.[4]

Generally speaking, the stability of the key coefficients across most outcomes reassures us that our main coefficients are not threatened by the inclusion of additional workplace cases.

[3] Using the Singapore Census for 2020, we calculate the fraction of employees with a high likelihood to work from home. We limit this fraction to those employed in the occupations listed as legislator, senior officials and managers, professionals, and clerical support workers.

[4] Considering regression model 4 we test if β_3 and β_4 are both zero with the alternative that at least one is unequal to zero. For the outcome *IndComRet* we observe a p-value of 0.2321 while the outcome *Residential* shows a p-value of 0.081. Thus, we cannot reject either hypothesis that the variables are jointly insignificant on the 5-percent significance level. Results are stronger for model 5. Testing that β_2 to β_5 are zero we get p-values of 0.1457 for *IndComRet* and 0.1463 for *Residential*. Thus, for both models we cannot reject that the added variables are jointly significant.

Table 12: Estimation of Local and General Response, Controlling for Workplaces

	TravelDist	StayHome	IndComRet	Residential
	(1)	(2)	(3)	(4)
$LocalCases_{jt-1}$	-79.248^{***}	0.195^{***}	-0.104^{**}	-0.053^{*}
	(18.622)	(0.041)	(0.045)	(0.032)
$NonLocalCases_{jt-1}$	-35.950^{***}	0.005	-0.100^{***}	-0.051^{***}
	(8.182)	(0.017)	(0.019)	(0.015)
$CasesWorkplace_{wt-1}$	-8.578	0.006	-0.064	-0.066^{*}
	(19.134)	(0.044)	(0.048)	(0.036)
$VisitedCasesWorkplace_{wt-1}$		-0.027	-0.016	0.041^{*}
		(0.026)	(0.028)	(0.022)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
N	7,693,261	7,693,261	7,693,261	7,693,261

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (2). One observation corresponds to an individual on a specific date. Each model specification corresponds to a different outcome variable. $TravelDist$ is the travel distance in meters, $StayHome$ is a dummy variable that takes the value one if an individual remains at their home subzone for an entire day. *IndComRet* is a dummy that takes the value one if an individual enters at least one industrial, commercial or retail place. *Residential* is a dummy that takes the value one if an individual enters a residential place except their own residence. Note, that we multiply outcome variables $StayHome$, *IndComRet* and *Residential* by 100 such that the coefficients are interpreted in percentage points. $LocalCases$ are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. We control for the number of cases in an individual's workplaces w. $CasesWorkplace_{wt-1}$ are the number of positive-tested individuals living in the workplace w of i. Further, $VisitedCasesWorkplace_{wt-1}$ are the number of positive cases who visited the workplace w of i. For all models we include individual and date FE. Standard errors are reported in parentheses and clustered on the individual level.

Table 13: Estimation of Local and General Response, Controlling for Workplaces II

	TravelDist (1)	StayHome (2)	IndComRet (3)	Residential (4)
$LocalCases_{jt-1}$	-89.634^{***}	0.183^{***}	-0.083^{*}	-0.051
	(19.038)	(0.042)	(0.046)	(0.033)
$NonLocalCases_{jt-1}$	-38.358^{***}	0.011	-0.094^{***}	-0.051^{***}
	(8.466)	(0.018)	(0.020)	(0.015)
$CasesWorkplace_{wt-1}$	193.658^{*}	0.431^{*}	-0.381	0.064
	(104.661)	(0.225)	(0.251)	(0.219)
$VisitedCasesWorkplace_{wt-1}$	237.895^{***}	-0.674^{***}	0.033	0.155
	(58.337)	(0.130)	(0.142)	(0.129)
$CasesWorkplace_{wt-1} * WFH_{jk}$	-453.820^{**}	-0.815^{*}	0.601	-0.282
	(204.355)	(0.440)	(0.496)	(0.449)
$VisitedCasesWorkplace_{wt-1} * WFH_{jk}$	-455.122^{***}	1.342^{***}	-0.115	-0.250
	(111.426)	(0.253)	(0.278)	(0.267)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
N	7,325,218	7,325,218	7,325,218	7,325,218

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (3). One observation corresponds to an individual on a specific date. Each model specification corresponds to a different outcome variable. $TravelDist$ is the travel distance in meters, *StayHome* is a dummy variable that takes the value one if an individual remains at their home subzone for an entire day. $IndComRet$ is a dummy that takes the value one if an individual enters at least one industrial, commercial or retail place. *Residential* is a dummy that takes the value one if an individual enters a residential place except their own residence. Note, that we multiply outcome variables *StayHome*, *IndComRet* and *Residential* by 100 such that the coefficients are interpreted in percentage points. *LocalCases* are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. We control for the number of cases in an individual's workplaces w. $CasesWorkplace_{wt-1}$ are the number of positive-tested individuals living in the workplace w of i. Further, $VisitedCasesWorkplace_{wt-1}$ are the number of positive cases who visited the workplace w of i. WFH_{jk} measures the fraction of employees from subregion j that have a high likelihood to allow work from home arrangements. Standard errors are reported in parentheses and clustered on the individual level.

2.5.2 Inflow Analysis

We now turn to the inflow analysis. In the following we show that locales with more occupied office space do not feature stronger inflow reductions following local case announcements. If white-collar work-from-home arrangements were a significant factor driving inflow reductions, we would expect a larger coefficient on cases for subregions with more offices.

We collect data on office space in each subzone j in 2019 from the REALIS database provided by the Urban Redevelopment Authority.[5] We then show regression evidence for the following model which corresponds to the model (2) in the main article:

$$\begin{aligned} Visit_{ijt} =& \beta_1 LocalCases_{jt-1} + \beta_2 InfectionVisit_{jt-1} + \\ & \beta_3 LocalCases_{jt-1} \cdot Offices_j + \beta_5 InfectionVisit_{jt-1} \cdot Offices_j \\ & \gamma_i \times \xi_j + \rho_t + \varepsilon_{ijt}, \end{aligned} \tag{6}$$

where the difference to the main regression is that we add interactions of $LocalCases_{jt-1}$ and $InfectionVisit_{jt-1}$ with the office space in hundred thousands of square meters, $Offices_j$. Negative coefficients β_3 or β_4 would indicate that the impact of locals cases or visiting cases is larger for areas with more offices, indicating that work-from-home arrangement could potentially matter. Our main results are more tenuous if β_1 and β_2 are insignificantly different from zero.

We show the results of the regression in Table 14. Models (1) to (3) are the specifications of the main paper. Model (4) includes the interaction with the office space in the hundred thousands. In comparison to model (3), the introduction of the interaction terms with the number of offices shrank the impact of local cases that are independent of the offices by 0.03 percentage points (38%). In subregion with more offices are indeed related to a lower visit if a local case is identified. However, we see that the number of offices are positively related to visits when infectious people entered such zones. Overall, we conclude that is possible that

[5] Additional information is available at https://www.ura.gov.sg/reis/index.

work-from-home arrangements have an effect on our observed effect. However, we do not find evidence that such arrangements are responsible for our estimated effect but in fact could shrink it to some extent.

Table 14: **Regression, Visiting Affected Areas, Offices**

	Visit	Visit	Visit	Visit
	(1)	(2)	(3)	(4)
$LocalCases_{jt-1}$	-0.31^{***}	-0.099^{***}	-0.081^{***}	-0.05^{***}
	(0.003)	(0.003)	(0.002)	(0.003)
$InfectionVisit_{jt-1}$	-0.276^{***}	-0.149^{***}	-0.017^{***}	-0.013^{***}
	(0.002)	(0.002)	(0.001)	(0.002)
$LocalCases_{jt-1} \times Office_j$				-0.047^{***}
				(0.001)
$InfectionVisit_{jt-1} \times Office_j$				0.009^{***}
				(0.001)
Subregion FE	Yes	Yes	No	No
Date FE	No	Yes	Yes	Yes
Subregion×Individual FE	No	No	Yes	Yes
N	477,903,426	477,903,426	477,903,426	477,903,426

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (2). One observation corresponds to a combination an individual, subregion, and specific date. We exclude observations from the sample that do not provide variation: (1) subregions that an individual has never visited and (2) home subregions of individuals. Each model specification corresponds to the outcome variable $Visit$, a dummy that takes the value one if the individual visits the subregion in t. Note, that we multiply outcome variable by 100 such that the coefficients are interpreted in percentage points. $LocalCases$ are the number of local cases in a subregion announced in the evening of $t-1$. $InfectionVisit$ are the number of newly announced cases that visited subregion j. Finally $Offices$ are the number of offices in hundred thousands. In Model (4) we interact the local cases as well as the cases of infectious persons visted subregion j with the number of offices. Note that we do not have an intercept of $Offices$ as they are not identified due to the fixed effects. Model specification (1) includes subregion fixed effects, specification (2) adds date fixed effects, and specifications (3) and (4) include date and subregion× individual fixed effects. Results are based on a bootstrapping procedure in which we draw 10% of the individuals in the full sample and repeat after replacement. We calculate the mean local effect and mean infection visit effect as percentage change from the average outcome. Standard errors are reported in parentheses and clustered on the individual level.

2.6 Identification Concern: Higher Local Incidence

In our main specification we argue that individuals respond to local COVID-19 cases. The argument stems from observing a stronger reaction to local cases. Another interpretation is that a local case announcement could indicate a higher incidence of COVID-19 compared to regional case announcements. In this case our estimate could pick up a reaction to perceived prevalence rather than to geographic proximity.

To test this alternative theory, we re-estimate our model with proxies for local and regional COVID incidence. Consider the following regression model in the fashion of our main specification:

$$\textbf{TravelDist}_{ijkt} = \beta_1 \frac{LocalCases_{jkt-1}}{pop_{jk}} + \beta_2 \frac{NonLocalCases_{kt-1}}{pop_k} + \gamma_i + \rho_t + \varepsilon_{ijkt}, \qquad (7)$$

where $TravelDist_{ijkt}$ is the travelled distance of an individual i with a home in subregion j, part of region k. $LocalCases_{jkt-1}$ and $NonLocalCases_{kt-1}$ are local and non-local cases respectively. The major difference from the main model is that we divide the announced cases with a proxy for population in subregion j or region k. As a result the regressor approximates the incidence of the virus, i.e. the cases in relation to the population announced in $t-1$. If β_2 is larger than β_1 we can rule out local incidence as driving observed behavior. We do not have a perfect estimate of pop_k and pop_{jk}. Thus we use two different proxies. First, we use data from the 2010 Singaporean census reporting the number of citizens and permanent residents within each subregion and region. One caveat of this data is that foreigners are not reported. As foreigners constitute to over 25% of the Singaporean population and tend to live in specific areas, this measurement likely introduces a measurement bias. As a second measure we calculate residential areas (square feet) in each subregion. This data also comes from the REALIS database.

Table 15 reports the results. We observe a clear difference to the main findings of the paper. Individuals do not respond stronger to local incidence than to non-local. Indeed the local response is not significantly different from zero in the specification using population and full fixed effects. Overall we conclude that we find no evidence that local incidence is driving force behind our result.

Table 15: **Estimation of Local and General Response to Prevalence of the Virus**

	TravelDist			
	Census Estimate		Residential Floor Area	
	(1)	(2)	(3)	(4)
$\frac{LocalCases_{jt-1}}{pop}$	128.532^{***}	18.416	-41.721^{***}	-34.404^{***}
	(26.684)	(26.726)	(12.644)	(12.668)
$\frac{NonLocalCases_{jt-1}}{pop}$	$-10{,}726.450^{***}$	$-3{,}490.480^{***}$	-890.897^{***}	-233.853^{***}
	(475.123)	(568.648)	(38.638)	(43.096)
Individual FE	No	Yes	No	Yes
Date FE	Yes	Yes	Yes	Yes
Mean Outcome	13808	13808.08	13808.08	13808.08
Mean Local Effect in Percent	0.93	0.13	-0.3	-0.25
Mean Aggregate Effect in Percent	-77.68	-25.28	-6.45	-1.69
N	9,482,376	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model 7. One observation corresponds to an individual on a specific date. Each model specification corresponds to the outcome variable of $TravelDist$, travel distance in meters. $\frac{LocalCases}{pop}$ are the number of local cases in a subregion announced in the evening of $t-1$ divided by the proxy of population in a subzone. $\frac{NonLocalCases}{pop}$ are the cases of the region k announced divided by the proxy of the region. Model (1) and (2) use the 2010 census estimate of citizens in 10 thousands individuals. Model (3) and (4) use sqft of residential properties. Standard errors are reported in parentheses and clustered at the individual level.

2.7 Additional Fixed Effects

In the main manuscript we compare the impact of local and non-local cases while using date fixed effects to control for the national trend. In the following, to account for the possibility of region-wide mobility changes, we extend the main model for new fixed effects:

$$\mathbf{a}_{ijkt} = \beta_1 LocalCases_{jkt-1} + \beta_2 NonLocalCases_{kt-1} + \gamma_i + \rho_t + \xi^1_{jt} + \xi^2_{kt} + \varepsilon_{ijkt}, \quad (8)$$

where all variables correspond to regression model (1) in our main manuscript. In the specifications we test below, we selectively include subregion × date ($\xi^1_j t$) or region × date ($\xi^2_k t$) fixed effects in order to keep measure some effect from daily case counts. When including region × date or subregion × date fixed effects, the corresponding coefficient of local or non-local cases vanishes as the variation is captured by the fixed effects.

We present results for the four the outcome variables – travel distance; likelihood of staying at home; likelihood of visiting industrial, commercial, or retail areas; or residential areas – in Tables 16 to 19. Note that in each table the first specification includes the result with only individual and date fixed effects as a baseline reference. We find that the original results are robust to the inclusion of the new fixed effects controlling for local or non-local trends.

2.8 A Logit Specification

In our main model we use a linear probability model (LPM) to analyse binary outcomes. An alternative approach would be the use of a non-linear model such as a logit or probit specification. We found several problems while trying to implement a nonlinear model with our fixed effect structure and sample size. Beyond computational issues, a binary model in this case requires dropping a substantial number of observations because of individuals who do not change their behavior (measured by the relevant outcome variable) in the time period observed. For example, for our measurements of how many people visit other residential areas; nearly one third of observations — approximately 3 million of 9 million — are dropped. In the

Table 16: Locality x Date fixed effects, Outcome: TravelDist

	TravelDist		
	(1)	(2)	(3)
$LocalCases_{jt-1}$	-61.433^{***}	-52.861^{***}	
	(14.429)	(14.793)	
$NonLocalCases_{jt-1}$	-28.045^{***}		-30.124^{***}
	(6.776)		(6.995)
Individual FE	Yes	Yes	Yes
Date FE	Yes	No	No
NonLocal x Date FE	No	Yes	No
Local x Date FE	No	No	Yes
N	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (1). One observation corresponds to an individual on a specific date. The outcome variable $TravelDist$ is the travel distance in meters. $LocalCases$ are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. Each model includes different fixed effects. Standard errors are reported in parentheses and clustered on the individual level.

Table 17: Locality x Date fixed effects, Outcome: StayHome

	StayHome		
	(1)	(2)	(3)
$LocalCases_{jt-1}$	0.140^{***}	0.138^{***}	
	(0.034)	(0.034)	
$NonLocalCases_{jt-1}$	0.006		0.005
	(0.015)		(0.015)
Individual FE	Yes	Yes	Yes
Date FE	Yes	No	No
NonLocal x Date FE	No	Yes	No
Local x Date FE	No	No	Yes
N	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (1). One observation corresponds to an individual on a specific date. The outcome variable $StayHome$ is a dummy variable that takes the value one if an individual remains at their home subzone for an entire day. Note, that we multiply the outcome variable $StayHome$ by 100 such that the coefficients are interpreted in percentage points. $LocalCases$ are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. Each model includes different fixed effects. Standard errors are reported in parentheses and clustered on the individual level.

Table 18: Locality x Date fixed effects, Outcome: IndRetCom

	IndComRet		
	(1)	(2)	(3)
$LocalCases_{jt-1}$	-0.117^{***}	-0.098^{***}	
	(0.035)	(0.036)	
$NonLocalCases_{jt-1}$	-0.083^{***}		-0.084^{***}
	(0.016)		(0.016)
Individual FE	Yes	Yes	Yes
Date FE	Yes	No	No
NonLocal x Date FE	No	Yes	No
Local x Date FE	No	No	Yes
N	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (1). One observation corresponds to an individual on a specific date. The outcome variable *IndComRet* is a dummy that takes the value one if an individual enters at least one industrial, commercial or retail place. Note, that we multiply the outcome variable *IndComRet* by 100 such that the coefficients are interpreted in percentage points. *LocalCases* are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. Each model includes different fixed effects. Standard errors are reported in parentheses and clustered on the individual level.

Table 19: Locality x Date fixed effects, Outcome: Residential

	Residential		
	(1)	(2)	(3)
$LocalCases_{jt-1}$	-0.055^{**}	-0.063^{**}	
	(0.026)	(0.027)	
$NonLocalCases_{jt-1}$	-0.029^{**}		-0.034^{**}
	(0.013)		(0.013)
Individual FE	Yes	Yes	Yes
Date FE	Yes	No	No
NonLocal x Date FE	No	Yes	No
Local x Date FE	No	No	Yes
N	9,482,376	9,482,376	9,482,376

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents results of regression model (1). One observation corresponds to an individual on a specific date. *Residential* is a dummy that takes the value one if an individual enters a residential place except their own residence. Note, that we multiply the outcome variable *Residential* by 100 such that the coefficients are interpreted in percentage points. *LocalCases* are the number of local cases in a subregion announced in the evening of $t-1$. $NonLocalCases$ are the cases of the region k announced. Each model includes different fixed effects. Standard errors are reported in parentheses and clustered on the individual level.

case of our model we believe it is important to keep this larger sample. We require this full sample to estimate a population-wide average effect; dropping individuals whose behavior does not change from the range of cases experienced in the time period is not a random subsample of the population. While a standard fixed effects model should mute the impact of these individuals on coefficient estimates, in our two-way fixed effect model this is not the case. We found that even in the linear probability model, matching the censoring required for the logit model impacted our results. Hence the results in Table 20 are largely similar to our baseline results, but they more closely mirror the results from the linear probability models without the full sample. The results which are most substantially different, those on residential outcomes, are from regressions dropping the most observations.

Table 20: **Estimation of Local and General Response, Conditional Logit Model**

	StayHome	IndComRet	Residential
	(2)	(3)	(4)
$LocalCases_{jt-1}$	0.354^{**}	-0.547^{***}	0.010
	(0.186)	(0.168)	(0.068)
$NonLocalCases_{jt-1}$	0.068	-0.398^{***}	-0.194^{***}
	(0.084)	(0.074)	(0.029)
Individual FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
N	6,983,610	7,779,409	6,182,873

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table presents average marginal effects of a conditional logit regression model. One observation corresponds to an individual on a specific date. Each model specification corresponds to a different outcome variable. *StayHome* is a dummy variable that takes the value one if an individual remains at their home subzone for an entire day. *IndComRet* is a dummy that takes the value one if an individual enters at least one industrial, commercial or retail place. *Residential* is a dummy that takes the value one if an individual enters a residential place except their own residence. Note, that we multiply outcome variables *StayHome*, *IndComRet* and *Residential* by 100 such that the coefficients are interpreted in percentage points. *LocalCases* are the number of local cases in a subregion announced in the evening of $t-1$. *NonLocalCases* are the cases of the region k announced. For all models we include individual and date FE. Standard errors are reported in parentheses and clustered on the individual level.