THE LANCET
Digital Health

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *Lancet Digit Health* 2020; published online Oct 28. https://doi.org/10.1016/S2589-7500(20)30243-0.

SUPPLEMENTARY APPENDIX

Evaluating the impact of demographic, socioeconomic factors, and risk aversion on mobility during COVID-19 epidemic in France under lockdown: a population-based study

Giulia Pullano[1,2,†], Eugenio Valdano[1,†], Nicola Scarpa[1], Stefania Rubrichi[2,*], Vittoria Colizza[1,*,°]

1 INSERM, Sorbonne Université, Pierre Louis Institute of Epidemiology and Public Health, Paris, France.
2 Orange Labs, Sociology and Economics of Network and Services (SENSE), Chatillon, France.

† Co-first authors

* Contributed equally

° *vittoria.colizza@inserm.fr*

S1. Data

Mobile phones

The 1436 areas of mainland France are groups of municipalities defined according to the 2018 EPCI (Établissements Publics de Coopération Intercommunale) level[1]. The average distance between the centroids of two adjacent areas is 22 km. Computation of the origin-destination matrices is based on the on-the-fly processing of signaling messages exchanged between mobile phones and the mobile network, usually collected by mobile network operators to monitor and optimize the mobile network activities. Such messages contain information about the identifiers of the mobile subscriber and of the antenna handling the communication, the timestamp and the type of event to be recorded (e.g., voice call, SMS, handover, data connection, location update). Knowing the spatial localization of the antennas allows reconstructing the approximate position of the device in communication. This was then used to compute the total number of displacements, with no residual information tracing back to the individual users. A displacement (or trip) was defined as the movement between any two consecutive locations along the trajectory of the user, where the user spent at least 1 hour. That is, in-between locations were not considered if the user spent there less than 1 hour. The value of the time cutoff was set by the data provider considering i) spatial resolution: the finer the spatial resolution, the shorter the time cutoff; ii) the time cutoff should be longer than the average time needed to cross a location, to avoid splitting a single trip into shorter sub-trips. Conversely, an excessively long cutoff would cause distinct displacements to merge into a single one. The data provider carried out sensitivity tests on several durations, and used a subset of mobile devices as testers, a standard methodology to extract displacements from mobile phone data. Travel flows we used in the study were previously adjusted by the data owner (Orange) to be representative of the general population, using spatially stratified market share data, socioeconomic data from INSEE, data on mobile phone ownership from INSEE, and customer socio-demographic information provided upon subscription. Most of our analysis is on French SIM cards, and excludes foreign SIM cards. Visitors purchasing French SIM cards may be a potential source of bias. However, free roaming within the EU has dramatically decreased the need to purchase a local SIM card when traveling. In addition, data have been adjusted by the data owner using findings from the EVE tourism survey for foreign visitors, conducted by the DGE (French General Directorate for Enterprise) and the Bank of France[2].

Hospitalizations

Cumulative hospitalization data contain all hospitalizations, in each region, from the beginning to the epidemic, up until April 5, 2020. We chose this date as it was halfway through lockdown. It was far enough from the start (March 17, 2020) for mobility to stabilize at lockdown value in each region, possibly influenced by the information on hospitalization levels. It was also far enough from the end (May 11, 2020) not to be influenced by possible relaxed adherence to restrictions.

INSEE statistics and government data

Statistics on working age population can be found here[3].

Statistics on work-related commuting can be found here[4].

Statistics on school-related commuting can be found here[5].

School holidays mark the periods in which students stay at home from school. Four geographic zones (A, B, C, and D) exist, and each of them has a specific holiday calendar. Each French region belongs to one of these zones. French school calendars are accessible here[6].

S2. Timeline fit and prediction

Prophet uses a decomposable time series model whose components fit into three types: trend (non-periodic changes), seasonality (a technical term that refers to any periodic change), holidays (abruptly irregular patterns occurring over one or more days). The model is then statistically framed as a generalized additive model[7], and fitted using Markov Chain Monte Carlo (MCMC), implemented in STAN[8]. To fit our data, we allowed for a weekly-periodic signal in the timeline fit (termed “weekly seasonality” in Prophet), and used school holidays by region as additional (additive) regressors.

S3. Spatial smoothing

We used a standard gaussian kernel with fixed characteristic distance[9], and weighed locations by their population. Let x_i be the value in location i of the quantity we want to smooth. Let w_i be the population in i, and Δ_{ij} the geodesic distance between locations i, j. Then, the smoothed value is

$$x_i^{smoothed} = \frac{\sum_j w_j x_j e^{-\left(\frac{\Delta_{ij}}{\lambda}\right)^2}}{\sum_j w_j e^{-\left(\frac{\Delta_{ij}}{\lambda}\right)^2}}$$

Where λ is the characteristic distance parameter. The sum runs over all the locations. The numerical implementation of spatial smoothing was done in Python (v. 3.7), using standard libraries (numpy 1.18, scipy 1.4).

S4. Additional results

Univariate analysis for outgoing and internal mobility

Mobility reduction (week April 06-12)	**outgoing**		**internal**	
	Pearson coefficient	p-value	Pearson coefficient	p-value
Number of hospitalized per 100,000 inhabitants (April 05)	0.73	<0.01	0.55	0.053
Standard of living - 9th decile	0.63	0.02	0.72	<0.01
% population in age 24-59	0.91	<0.01	0.76	<0.01
% highly impacted workers	0.80	<0.01	0.64	0.02
Number of deaths per 100,000 inhabitants (April 05)	0.63	0.02	0.46	0.11

Table S1. Correlation coefficients. The table reports the correlation coefficients and their p-value for the indicators considered and internal and outgoing regional mobility.

Mobility reduction (week April 06-12)	**outgoing**		**internal**	
	Pearson coefficient	p-value	Pearson coefficient	p-value
Number of hospitalized per 100,000 inhabitants (April 05)	0.57	0.03	0.34	0.14
Standard of living - 9th decile	0.12	0.36	0.61	0.02
% population in age 24-59	0.85	<0.01	0.63	0.01
% highly impacted workers	0.61	0.02	0.42	0.08

Table S2. Correlation coefficients without Île-de-France. The table reports the correlation coefficients and their p-value for the indicators considered and internal and outgoing regional mobility, computed excluding Île-de-France.

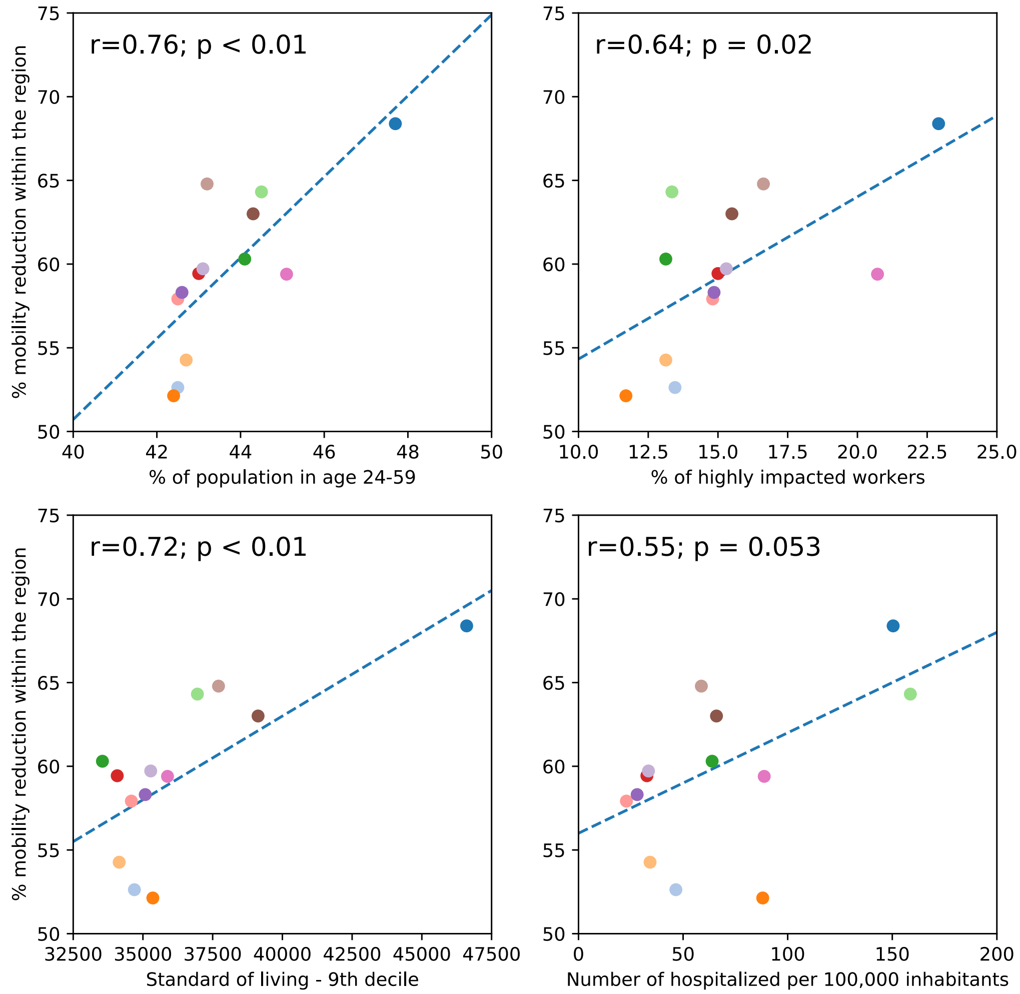

Figure S1. Reduction in internal mobility for the week April 6-12, 2020 vs. epidemic, socio-economic, and demographic indicators. The following plot is the equivalent of Figure 4 for internal mobility. Correlation is evaluated between reduction of internal traffic and the four considered indicators: a) the population in active age (24-59 years old), b) the fraction of employees in the sectors mostly affected by lockdown. c) the 90th percentile of the regional standard of living. Pearson correlation coefficients and their p-values are reported, d) the cumulated number of COVID-19 hospitalizations per 100,000 inhabitants on April 05, 2020.

Multivariate analysis

We focused on outgoing mobility, which already exhibited significant unadjusted correlations. Table S3 reports crude (unadjusted) as well as adjusted correlation coefficients, and their p-values. Multivariate correlation was performed by standardizing the variables and performing linear regression.

Mobility reduction (week April 06-12)	**crude (unadjusted)**		**adjusted**	
	Pearson coefficient	p-value	Pearson coefficient	p-value
Number of hospitalized per 100,000 inhabitants (April 05)	0.73	<0.01	0.51	0.01
Standard of living - 9th decile	0.63	0.02	-0.47	<0.01
% of population in age 24-59	0.91	<0.01	0.45	0.04
% of highly impacted workers	0.80	<0.01	0.53	<0.01
Number of deaths per 100,000 inhabitants (April 05)	0.63	0.02	0.47	<0.01

Table S3. Crude and adjusted correlation coefficients. The table reports the crude (unadjusted) and adjusted correlation coefficients and their p-value for the indicators considered and outgoing regional mobility.

The qualitative behavior is the same for both crude and adjusted coefficients (positive, significant associations), except for *Standard of living – 9th decile*. Its adjusted coefficient is still significantly different from zero, but it is negative, unlike the unadjusted one. To check that this effect is not due to the covariates being too correlated among themselves (multicollinearity), we computed Pearson correlation between standard of living and the other covariates and found it never goes above 0.44, which is the value of the correlation with the active population. We also computed the variance inflation factor (VIF), a measure of multicollinearity, and found it to be always below 5 except for the active population, for which it is however still slightly below 10. As additional safety checks, we performed the multivariate regression using LASSO to identify possible redundancy in the covariates, and found none. We also reran the multivariate regression without the active population, and found a similar result to what shown in the table, including the sign flip.

REFERENCES

1. Cartographie des EPCI à fiscalité propre. https://www.collectivites-locales.gouv.fr/cartographie-des-epci-a-fiscalite-propre. Accessed: April 2020.
2. Le Tourisme International En France. https://www.entreprises.gouv.fr/files/files/directions_services/etudes-et-statistiques/stats-tourisme/memento/2017/2017-11-MEMENTO-TOURISME-CHAP6-le-tourisme-international-en-france.pdf . Accessed: April 2020.
3. INSEE. Tableaux de l'économie française. https://www.insee.fr/fr/statistiques/3676623?sommaire=369693. Accessed: April 2020.
4. INSEE. Mobilités professionnelles en 2015. https://www.insee.fr/fr/statistiques/3566477. Accessed: April 2020.
5. INSEE. Mobilités scolaires en 2015. https://www.insee.fr/fr/statistiques/3566470. Accessed: April 2020.
6. education.gouv.fr. https://www.education.gouv.fr/calendrier-scolaire-100148. Accessed: April 2020.
7. Taylor, S. J. & Letham, B. *Forecasting at scale*. https://peerj.com/preprints/3190 (2017) doi:10.7287/peerj.preprints.3190v2.
8. Stan. *stan-dev.github.io* //mc-stan.org/.
9. Genebes, L., Renaud, A. & Sémércurbe, F. Spatial smoothing. Manuel d'analyse spatiale: http://www.epsilon.insee.fr/jspui/bitstream/1/85657/1/imethode131.pdf.