



HAL
open science

Linking excess mortality to Google mobility data during the COVID-19 pandemic in England and Wales

Ugofilippo Basellini, Diego Alburez-Gutierrez, Emanuele del Fava, Daniela Perrotta, Marco Bonetti, Carlo Giovanni Camarda, Emilio Zagheni

► **To cite this version:**

Ugofilippo Basellini, Diego Alburez-Gutierrez, Emanuele del Fava, Daniela Perrotta, Marco Bonetti, et al.. Linking excess mortality to Google mobility data during the COVID-19 pandemic in England and Wales. 2020. hal-02899654v1

HAL Id: hal-02899654

<https://hal.science/hal-02899654v1>

Preprint submitted on 15 Jul 2020 (v1), last revised 10 Feb 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Linking excess mortality to Google mobility data during the COVID-19 pandemic in England and Wales

Ugo Filippo Basellini^{1,2}, Diego Alburez-Gutierrez¹, Emanuele Del Fava¹, Daniela Perrotta¹, Marco Bonetti³, Carlo G. Camarda², and Emilio Zagheni¹

¹*Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany*

²*Institut national d'études démographiques (INED), Aubervilliers, France*

³*Carlo F. Dondena Centre for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy*

July 7, 2020

Abstract

Following the outbreak of COVID-19, a number of non-pharmaceutical interventions have been implemented to contain the spread of the pandemic. Despite the recent reduction in the number of infections and deaths in Europe, it is still unclear to which extent these governmental actions have contained the spread of the disease and reduced mortality. In this article, we estimate the effects of reduced human mobility on excess mortality using digital mobility data at the regional level in England and Wales. Specifically, we employ the Google COVID-19 Community Mobility Reports, which offer an approximation to the changes in mobility due to different social distancing measures. Considering that changes in mobility would require some time before having an effect on mortality, we analyse the relationship between excess mortality and lagged indicators of human mobility. We find a negative association between excess mortality and time spent at home, as well as a positive association with changes in outdoor mobility, after controlling for the time trend of the pandemic and regional differences. We estimate that almost 130,000 excess deaths have been averted as a result of the increased time spent at home. In addition to addressing a key scientific question, our results have important policy implications for future pandemics and a potential second wave of COVID-19.

1 Introduction and Background

The COVID-19 pandemic started in December 2019 and rapidly spread globally. As of July 1, 2020, over 10 million cases of infections and 500,000 deaths have been reported worldwide ([World Health Organization, 2020](#)). This is likely an underestimate of the burden of the disease because of several reasons, including under-reporting related to cross-country differences in disease monitoring and classification, asymptomatic cases, medically unattended cases, and deaths indirectly related to COVID-19.

Due to the absence of immediate effective treatment or a vaccine, local and national governments have implemented non-pharmaceutical interventions (NPIs) as a strategy to curb the spread of COVID-19 during the first half of 2020. These measures include travel bans, cancellations of public gatherings, social distancing, school closures, recommendations to work from home and to stay at home, up to nationwide lockdowns ([Hale et al., 2020](#)). In several countries, particularly in Europe, some of these public health interventions have been gradually lifted – at the time of writing – following the decline in the number of new infections and deaths, and in order to ease the economic impact of the epidemic. However, it remains unclear how to assess the effectiveness of the various NPIs in reducing infections and saving lives.

In this paper, we examine the relationship between excess mortality and reduced human mobility in England and Wales during the COVID-19 pandemic, specifically from February 15 to

June 19, 2020. The British government applied various control strategies by first encouraging social distancing on March 16, 2020, then ordering the closure of schools on March 21, 2020, and finally banning public events and ordering a total lockdown on March 24, 2020. The goal of this paper is to assess the impact of these governmental decisions, and consequent reduction in human movements, on mortality during the pandemic. Indeed, human mobility plays a key role in the spread of infectious diseases (Wilson, 1995; Tatem et al., 2006; Riley, 2007). In our contemporary society, where millions of people travel and commute every day within and across cities and regions, infectious diseases have the opportunity to spread more rapidly, and on a large scale, than ever before. Human mobility can in fact increase the disease prevalence by introducing new pathogens into susceptible populations or by increasing social contacts between susceptible and infected individuals (Wesolowski et al., 2016). Timely, accurate, and comparative data on human mobility are therefore critical, but generally not available or easily accessible. The ongoing COVID-19 pandemic has highlighted the potential benefit of geo-located smartphone data to inform public health (Oliver et al., 2020) and assess the impact of mobility restrictions on social distancing in near real-time (Pepe et al., 2020; Badr et al., 2020).

In this study, we use the publicly available data provided by Google COVID-19 Mobility Reports (GCMR) (Google LLC, 2020a), which is one of the best data sources to assess changes in mobility or the lack thereof (i.e., time spent at home) since the beginning of the pandemic (Drake et al., 2020). The GCMR reports percentage changes in mobility of Google Maps users compared to a baseline period before the pandemic (from January 3 to February 6, 2020) in various settings (e.g. supermarkets and pharmacies, public transports, workplaces, residential areas). We focus on the ‘residential’ category of the GCMR, a measure of the relative change in the time that users spent at their home addresses, as provided to or estimated by Google Maps.

We link the reduction in human mobility to all-cause mortality during the COVID-19 pandemic. Given the high uncertainty surrounding the number of infections and deaths, we chose to estimate the excess mortality rate, which indicates the number of deaths above what would be expected in a non-crisis period, controlling for the size of the population. This choice is also in line with the growing general consensus in the scientific community that excess mortality is the best indicator to assess the impact of the pandemic (National Academies of Sciences, Engineering, and Medicine, 2020).

We anticipate that a potential effect of changes in mobility on mortality would not be immediate, but mediated by a time lag. One would expect mobility reductions to first affect the number of social contacts in the population; the decrease in contacts should then reduce the number of newly infected individuals, and eventually the number of deaths from the disease. As such, in this paper we analyse the association between excess mortality and past values of human mobility, considering different time lags between the two measures. Accounting for a time lag of five or more weeks, we find a positive correlation between increased mobility and excess mortality, and a negative correlation between time spent at home and excess mortality. These relationships are significant within a mixed-effects regression setting that controls for the time trend of the pandemic and different regional effects. We estimate that almost 130,000 excess deaths have been averted as a result of the increased time spent at home following the NPIs implemented in England and Wales. As such, we argue that NPIs have been effective to limit the potential negative effects of the pandemic on mortality. Our findings not only address an ongoing debate on the actual benefit of NPIs as a strategy to control the spread of COVID-19, but also reveal important implications for public health decision-making in the event of a second wave of infections or future pandemics. Moreover, we hope that our methodologically-informed reflections will help generate a broader conversation on the type of information that private companies could provide to help the scientific community improve the understanding of the course of the epidemic, while also taking into account the ethical, privacy, and commercial constraints that private companies have to consider.

In the following sections, we describe the data used in the analysis and the methodological

approach we use to measure the effect of reduced mobility on excess mortality during the COVID-19 pandemic. Then we present our findings and close with a discussion and an outlook for future work.

2 Data Description

2.1 Mortality Data for England and Wales

The weekly number of deaths registered in England and Wales was obtained from the Office for National Statistics (ONS) (2020a). The ONS reports weekly mortality data broken down by age group, sex and region of usual residence. Note that, in the ONS classification, weekly data do not refer to calendar weeks, but to rolling 7-day periods, from Saturday to Friday, and we will keep this notation throughout the manuscript. For our purposes, we retrieved the weekly number of deaths by region for the years 2015–2020. Specifically, the total number of deaths in England and Wales is divided into ten territories: nine NUTS-1 regions for England, and a single region comprising Wales. The nine regions in England are: North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East, London, South East and South West. At the time of writing, the data were available until week 2020-25 (i.e. until June 19, 2020). Finally, data on the total population by region for the years 2015–2019 were obtained from the Office for National Statistics (2020b).

2.2 Google Mobility Data

We estimate human mobility from February 15 to June 19, 2020, using the freely-available Google COVID-19 Community Mobility Reports (GCMR) dataset (Google LLC, 2020a). The potential of this data has been recognized (Drake et al., 2020), but, to the best of our knowledge, it has not been linked to excess mortality outcomes yet. Specifically, the GCMR uses the data underlying the ‘Popular Times’ feature of Google Maps.¹ The GCMR reports mobility as percentage changes relative to a baseline period (from January 3 to February 6, 2020). Absolute mobility values are not available to protect users’ privacy. The data account for weekly seasonality of movement by estimating a set of seven baseline weekdays using the median value for each particular weekday during the 5-week baseline period. Daily relative change is estimated as the percentage change with respect to the corresponding baseline weekday for any given report date. In the UK, the GCMR data are aggregated by Local Authority, category of place (e.g. supermarket and pharmacy, workplaces, residential) and report date, but contains no individual-level information and no demographic characteristics of users such as sex, age, or educational attainment.

In this paper, we focus on the *residential* category of the GCMR, which is defined as the time users spent at home, using the home addresses provided to or estimated by Google Maps. Our focus on a variable related to time use and duration of events is consistent with the epidemiological literature on time use and the spread of close-contact infectious diseases (Zagheni et al., 2008; De Cao et al., 2014).

The time spent at home is measured as the relative change in the duration of stay in residential areas on a given day relative to the duration of stay in the corresponding weekday of the baseline period. We aggregate the GCMR data by week (from Saturday to Friday, for consistency with mortality data) and region (taking the average across all Local Authorities belonging to a given region) to make it comparable to the mortality data. Relying on a fixed baseline period ignores yearly seasonality of movement, which may be affected by weather patterns, national holidays, vacation periods, etc. Important bank holidays, extreme weather events, or other major events during the 5-week long baseline period can affect the estimates of the future relative change in

¹Including all Google Map users, irrespective of the type of device and operating system, as long as users have not opted-out of the “Location History” feature. Popular times data are only available for businesses that get “enough visits from users”: <https://support.google.com/business/answer/6263531?hl=en-GB>.

time spent at home. We have no access to the raw mobility data used to produce the GCMR, but we find no evidence of any major events in England or Wales, such as major bank holidays or extreme weather events, that could have systematically biased the Google mobility data during the baseline period. Furthermore, the baseline week is based on the median value, which would be largely unaffected by short-lived temporary fluctuations in absolute mobility values. Finally, the baseline period is not affected by restrictions on movement, which were first introduced in late March (Ganyani et al., 2020).

3 Methods

In this section, we introduce the methodology that we use in our article. Specifically, we first describe the model that we employ to estimate excess mortality in Subsection 3.1. Then, Subsection 3.2 illustrates the approach that we use to measure the effect of changes in mobility on excess mortality.

3.1 Computing excess mortality

Weekly mortality data generally show strong cyclical behavior. As such, we opt for a modulation model that accounts for seasonal patterns using sine and cosine functions.

For a given region, let $\mathbf{d} = (d_t)$ denote a vector containing the observed number of deaths over time, i.e., weekly registered deaths from week 2015-01 to week 2020-25 (until June 19, 2020). In alternative to the conventional Poisson distribution, we assume that d_t are realizations of a random variable D_t that follows a Negative Binomial distribution. We thus take into account overdispersion typically displayed in mortality data, i.e., the variance associated with the process is expected to be larger than what implied by the expected value. The main source of overdispersion in the data is likely the varying strength of the seasonal pattern, which is mainly due to variation in the yearly influenza epidemics.

Let $\mu_t = \mathbb{E}(D_t)$ denote the expected value of the Negative Binomial process. We model the expected number of deaths using a Generalized Additive Model with logarithmic link, exposures as offset, and time-specific covariates:

$$\ln(\mu_t) = \ln(e_t) + \alpha_0 + \alpha_1 \cos(\omega) + \alpha_2 \sin(\omega) + v_t, \quad (1)$$

where e_t denotes the exposure to the risk of death, v_t is a smooth function of time, $\omega = 2\pi/p$, and p is the period (e.g. $p = 52$ for normal years, $p = 53$ for leap years). Whereas sine and cosine functions capture the cyclical behavior of weekly mortality data, the smooth time component aims to describe eventual long-term mortality trends. A similar model with additional varying-coefficient components has been proposed by Eilers et al. (2008) to describe seasonal time series and incidence tables. Here, estimation of the model described in Equation (1) is performed in R (R Development Core Team, 2020) using the `mgcv` package (Wood, 2019).

For each region independently, we model data for the pre-pandemic period (from week 2015-01 to week 2020-06) and predict expected number of deaths \hat{d}_t for the COVID-19 period (from week 2020-07 to week 2020-25). In particular, we use week 2020-06 as a threshold for consistency with the Google mobility data, for which the baseline period runs from week 2020-02 to week 2020-06. Consequently, the effect of the pandemic is not captured by the time trend of the model, and \hat{d}_t for the predicted weeks could be interpreted as the number of deaths that would have occurred in 2020 in the absence of COVID-19.

This approach allows us to readily evaluate the excess number of deaths $\hat{\delta}_t$ for the weeks 7–25 in 2020 as the difference between the observed and the expected ones, i.e. $\hat{\delta}_t = d_t - \hat{d}_t$. In order to account for regional differences in population, we then compute the excess mortality rate, or per capita excess mortality, $y_t = \hat{\delta}_t/e_t$ by dividing excess deaths by the region-specific mid-year

population in 2019 (i.e., we use population data from the previous year, since data for 2020 are still not available).

3.2 Modelling the relationship between excess mortality and mobility

Given the estimated excess mortality rate for each week and region, we intend to assess whether an association with the change in physical mobility exists or not. The spread of COVID-19 mainly occurs through social contacts between infectious and susceptible individuals (Zhang et al., 2020). Hence, a reduction in mobility, as well as an increase in the time spent at home, should lead to a reduction in social contacts, then in the infection spread and, ultimately, in the COVID-related mortality. However, this process requires time, as we would expect the reduction in physical mobility observed today to possibly have an impact on the infection spread and the related mortality in the coming weeks. This calls for the introduction of a time lag in the mobility data, which corresponds to the amount of time necessary for the change in mobility to have an impact on mortality. In what follows, a time lag of x will indicate that we analyse the relationship between excess mortality and changes in human mobility that occurred x weeks in the past.

Moreover, we work with rates that vary over weeks and for different regions. Whereas the effect of mobility and the mortality trend will be assumed to remain constant over space, we need to account for the regional heterogeneity in excess mortality given that data within each region are likely correlated. This setting calls for a mixed-effects modelling approach, since we aim to know whether an association between excess mortality and human mobility over time still exists, after controlling for the variation in regions.

Let $y_{r,t}$ denote the excess mortality rate for a given region r in week t . We model this variable as follows:

$$y_{r,t} = \beta_0 + u_r + \gamma_1 g_{r,t-x} + \sum_{s=1}^n \beta_s B_s(w_t) + \varepsilon_{r,t}, \quad (2)$$

where u_r are the region-specific random intercepts, which are normally distributed with mean zero and variance σ_u^2 , β_0 is the common intercept, and γ_1 is the coefficient associated to the mobility data in each region r during the (lagged) week $t - x$, i.e., $g_{r,t-x}$. The time trend is modelled in a flexible way using the B -spline bases B_s , where β_s are the associated coefficients. Finally, $\varepsilon_{r,t}$ is the vector of the residuals, distributed as $\varepsilon_{r,t} \sim N(0, \sigma^2)$ and assumed to be independent of the random effects u_r .

In our model, the baseline mortality trend is assumed to be a non-parametric function of the series of weeks, which is modelled with a series of n B -splines. This choice gives us sufficient flexibility to capture the pattern of the pandemic over time. Mobility data from the GCMR are considered as an additional fixed effect, and the associated regression coefficient will be interpreted as the change in per capita excess mortality for a unit change in the mobility indicator, always as compared to the baseline period (cf. Section 2.2). A random intercept for each of the ten regions is added since we expect the mortality-mobility relationship within regions to be correlated. If the random effect variance σ_u^2 is significantly large, then some regions will have a lower or higher excess mortality trend to begin with, given their series of mobility data.

We fit the model in R using the `lme4` package (Bates et al., 2015), and we compute B -spline bases using the `splines` package (R Development Core Team, 2020).

4 Results

In this section, we show the results of the methods that we employ in this paper. Specifically, subsection 4.1 shows the results of the computation of excess mortality. In subsection 4.2, we investigate the correlation between excess mortality and time spent at home, considering differ-

ent lags for the latter variable. Finally, subsection 4.3 presents the results of the mixed-effects regression.

4.1 Estimating excess mortality

We first consider excess mortality, estimated using the Generalized Additive Model (GAM) of subsection 3.1. Figure 1 shows the results of this approach for two specific regions, namely, East Midlands and Wales. The grey shaded areas in the panels correspond to the COVID-19 pandemic period that we will analyse in the remainder of the paper, and for which corresponding death counts are not employed in the estimation of the GAM.

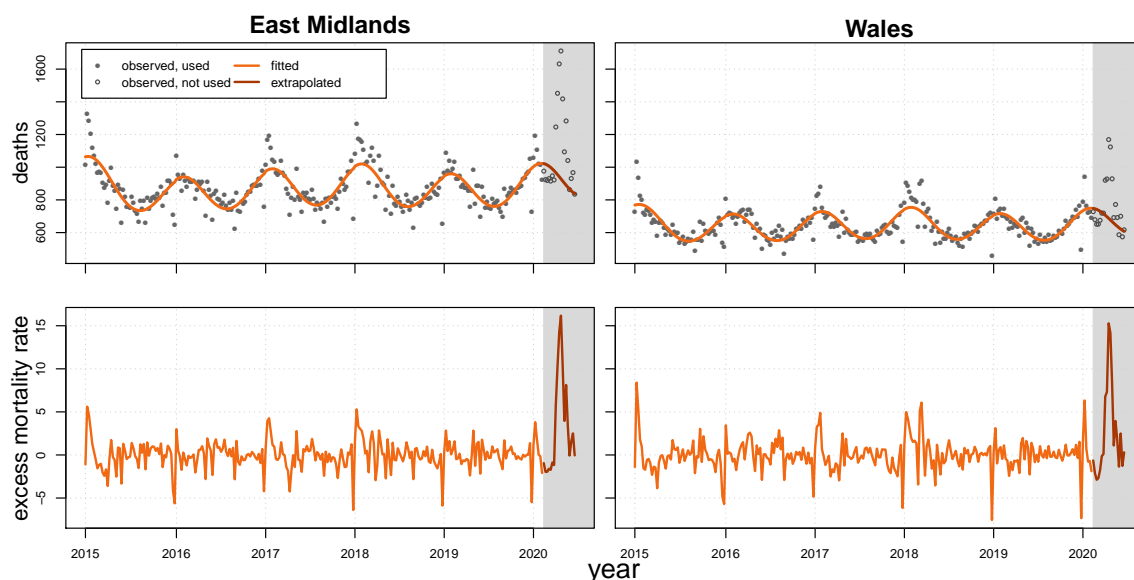


Figure 1. Observed (dots) and fitted (lines) weekly number of deaths (upper panels) and excess mortality rate (per 100,000 individuals, lower panels) in East Midlands and Wales for the years 2015–2020. The grey shaded area corresponds to the COVID-19 period analysed in the remainder of the paper. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#).

Figure 1 shows that the fitted model captures well the seasonal pattern of the mortality data, and that the smooth time trends allows for different amplitudes of the cyclical behaviour. The lower panels display the excess mortality rate for the two regions: despite being characterized by different absolute numbers of deaths, the regional excess rates have similar values due to different population sizes. Finally, the spike in number of deaths and excess mortality resulting from the COVID-19 pandemic is clearly visible in the last weeks of 2020 (grey shaded area).

In addition to employing the GAM, we perform a sensitivity analysis on the computation of excess mortality by estimating, for each week, the expected number of deaths as the average of the observed deaths between 2015 and 2019. Figure A.1 in the Appendix A shows the results of this approach; in particular, the estimated excess mortality rates (and the results in the next subsections) are very similar to those obtained with the GAM.

4.2 Correlation of excess mortality and time spent at home

Having estimated excess mortality, we can investigate its relationship with changes in time spent at home during the COVID-19 pandemic. Figure 2 shows the time series of the excess mortality rate and the residential category of Google data for the weeks 8–25 (the first week available for Google data and the last available data point, respectively) by region. In addition, the figure shows: (i) the forward lag of five weeks for the Google data, (ii) the weeks that will be analysed in the regression setting in Subsection 4.3 (grey shaded area), and (iii) the start of two non-pharmaceutical

interventions: the encouragement of social distancing (March 16) and the enforcement of the lockdown (March 24).

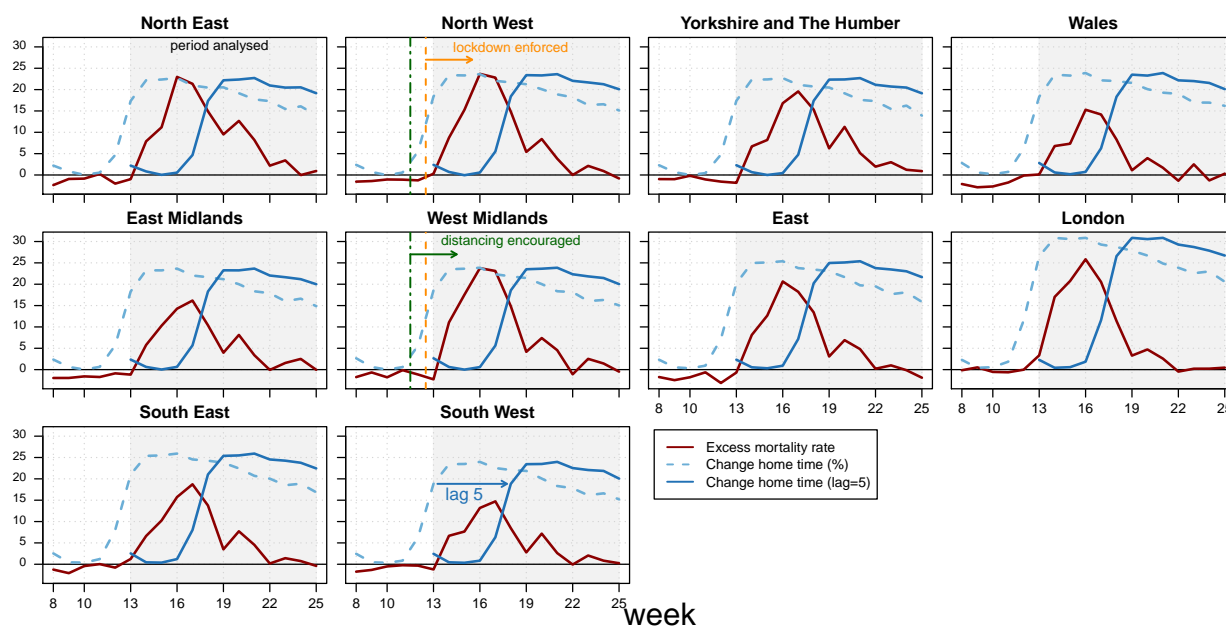


Figure 2. Time series of excess mortality rate (per 100,000 individuals, in red), change in time spent at home at time t (dashed lines, light blue) and with a five-week lag (solid lines, dark blue) by region in England & Wales for the weeks 8–25 in 2020. Solid lines correspond to values that will be analysed in the regression setting of subsection 4.3. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#) and [Google LLC \(2020a\)](#).

The combination of several non-pharmaceutical interventions between March 12 and 24 (nationwide school closure, advice to avoid public places and to self-isolate in presence of symptoms, in addition to distancing and lockdown) led to a significant increase of time spent in residence, which peaked around week 15 and remained relatively high in subsequent weeks. A similar increase is observable for the excess mortality rate across regions, which occurred about two weeks after the increase in residential permanence. Moreover, excess mortality peaked in weeks 16–17, returning to low levels at the end of the observation period.

We can now turn to analyse the correlation between excess mortality rate and changes in time spent at home, considering a wide range of possible lags for the latter variable. Figure 3 clearly shows that the relationship between excess mortality and residential permanence changes according to the length of time lags. When we consider none or short time lags (between one and three weeks), the relationship between the two variables is positive, i.e., an increase in residential permanence is associated with a (future) increase in excess mortality. However, for longer time lags (greater or equal to five weeks), we can observe a negative correlation between the two variables, i.e., an increase in time spent at home is associated with a (future) reduction of excess mortality.

Very similar correlation patterns, albeit with the reversed sign, are observable for the other categories of Google data. Figures A.2 and A.3 in Appendix A show the correlations between excess mortality rate and the grocery and workplaces categories reported by Google. For these and the remaining categories, we observe positive correlations between excess mortality and lagged values of outdoor mobility, i.e., a decrease in mobility is associated with a (future) reduction of excess mortality.

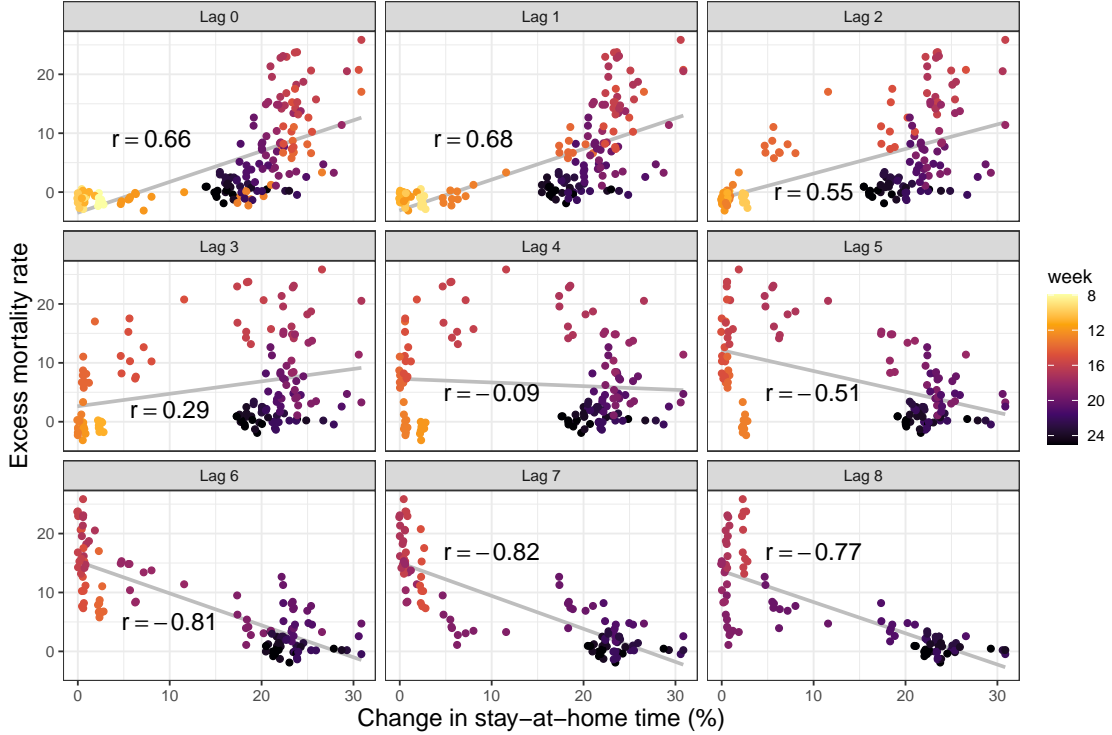


Figure 3. Linear relationship and Pearson correlation (r) between excess mortality rate (per 100,000 individuals) and change in time spent at home in ten regions of England & Wales during weeks 8–25 in 2020, considering different lags of time for residential permanence. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#) and [Google LLC \(2020a\)](#).

4.3 Model results

In this subsection, we investigate whether the negative relationship between excess mortality and lagged changes in time spent at home is statistically significant or not. To do so, we analyse the two variables within a regression setting, using the mixed-effects modeling approach introduced in Subsection 3.2. We consider a lag of five weeks for the changes in mobility as reported by the Google data. We select this lag because it is the first one to display a stronger negative correlation with excess mortality (cf. Figure 3), and it allows us to analyse a longer period of time with respect to choosing greater lags. Moreover, it is consistent with preliminary estimates of the incubation period (from infection to the onset of symptoms) up to almost two weeks (Lauer et al., 2020) and of the course of the disease (from the onset of symptoms to death) up to almost three weeks (Verity et al., 2020). As such, we analyse the weeks from 13 to 25, because Google data starts from week 8, and we bring the mobility data five weeks forward (cf. Figure 2). Nonetheless, the regression results are robust to different lags choices, given that the lag is greater than 5 weeks. Table 1 shows the results of this analysis; in addition, we also run a plain linear regression with the same covariates, without taking into account the random effects for the different regions.

It is interesting to observe that the estimated coefficients for the time spent at home differ between the linear and the mixed-effect model. In addition to the theoretical considerations outlined in Subsection 3.2, statistical criteria such as the AIC or the BIC, as well as an ANOVA test, between the two models suggest that the mixed-model setting is a more appropriate approach in this context. In the mixed-effect model, a one percentage point increase in time spent at home is associated to a reduction of 1.1 in the excess mortality rate (per 100,000 individuals). This is a rather strong effect, given that the residential category of Google data changed by about 20-30 percentage points in all regions following the introduction of the NPIs. Furthermore, the intraclass correlation coefficient, calculated as the ratio between the variance of the region-specific random

	Dependent variable: excess mortality rate		
	linear regression	linear mixed-effects regression	
	ML	ML	REML
Fixed effects			
change in stay-at-home time (lag = 5)	-0.23 (0.12)	-1.10 (0.15)	-1.10 (0.16)
[95% conf. interval]	[-0.47, 0.01]	[-1.38, - 0.81]	[-1.40, - 0.79]
Random effects (variance)			
region (intercept)		5.75	6.46
residual		3.13	3.51
Observations	130	130	130
Log-Likelihood	-303.16	-274.70	-256.87
AIC	636.31	581.40	545.74
BIC	679.32	627.29	591.62

Table 1. Estimated coefficients (with associated standard errors, in parenthesis) and 95% confidence intervals of regressing the excess mortality rate (per 100,000 individuals) on changes in time spent at home five weeks before using linear regression and linear mixed-effects regression with maximum likelihood (ML) and restricted ML (REML) estimations. The models consider a smooth function of time (using 12 B -splines), and the mixed-effects model includes a random intercept for each region. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#) and [Google LLC \(2020a\)](#).

effects (σ_u^2) and the total variance ($\sigma_u^2 + \sigma^2$), shows that 64.8% of the whole data variability can be attributed to the variability between regions.

It is worth pointing out that the estimated coefficient for the stay-at-home time is robust to the choice of number of B -splines chosen to model the pandemic. In the regressions shown in Table 1, we selected 12 B -spline bases, as this choice minimizes the AIC and BIC values with respect to smaller number of B -splines. However, we performed a sensitivity analysis by using different numbers of B -splines in the mixed-effect model. Table A.2 in Appendix A shows the results of this analysis. A further sensitivity analysis considered the different categories of mobility data reported by Google. Table A.1 in Appendix A shows the results of regressing excess mortality rate on the different mobility indicators available.

Figure 4 shows the observed and fitted excess mortality rate from the mixed-effects model in Table 1 with 95% confidence intervals by region. The fitted curves capture well the time-series of excess mortality by regions, and investigation of the model residual does not display systematic biases.

Finally, the estimated model allows us to compute the number of deaths averted by the increase of time spent at home. Figure 4 further displays the estimated excess mortality in the counterfactual scenario in which residential permanence would have not changed during the period analysed. We estimate that an additional 128,609 excess deaths would have occurred in England and Wales in the absence of the observed increase in residential time. These estimates display significant regional variations, ranging from a minimum of 5,093 excess deaths in North East to a maximum of 24,382 excess deaths in London. Table 2 reports the additional cumulative excess mortality rate and total number of excess deaths avoided for this counterfactual scenario by region. Clearly, these estimates are highly dependent on the modeling approach as well as the data inputs, which are characterized by some degree of uncertainty. We are currently working to quantify the

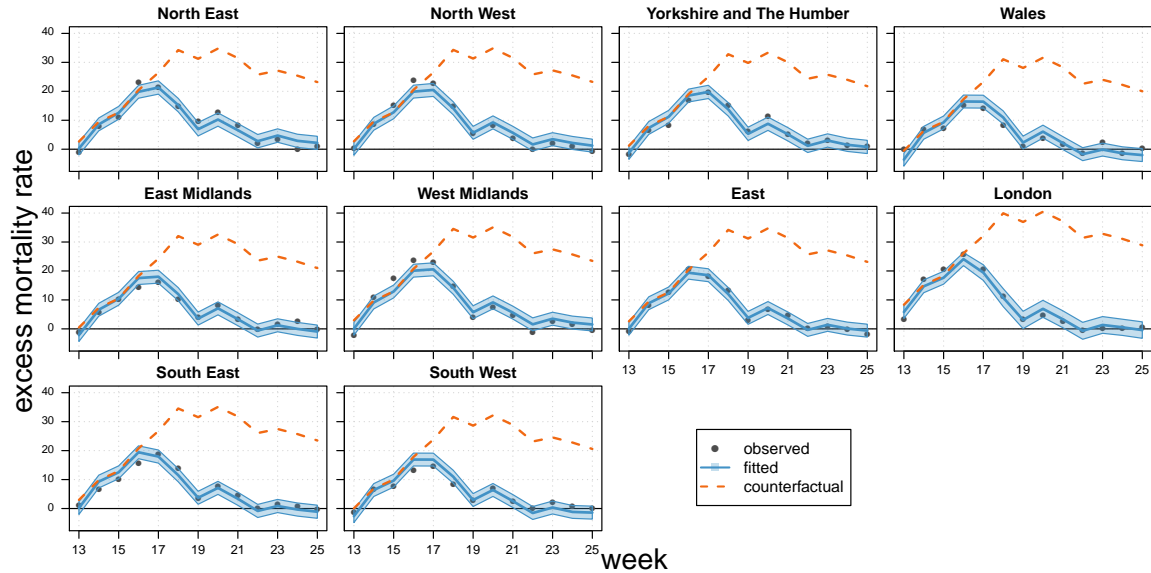


Figure 4. Observed and fitted excess mortality rate (per 100,000 individuals) from the mixed-effect model with 95% confidence intervals for the ten regions in England & Wales during the weeks 13–25. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#) and [Google LLC \(2020a\)](#).

uncertainty associated to these point estimates, in order to provide confidence intervals around them.

Region	Population (2019)	Cumulative excess mortality rate	Total number of excess deaths
North East	2,669,941	190.76	5,093.25
North West	7,341,196	200.53	14,721.26
Yorkshire and The Humber	5,502,967	191.19	10,521.39
Wales	3,152,879	203.21	6,406.85
East Midlands	4,835,928	200.27	9,684.88
West Midlands	5,934,037	202.50	12,016.54
East	6,236,072	218.02	13,595.78
London	8,961,989	272.06	24,382.42
South East	9,180,135	225.23	20,676.26
South West	5,624,696	204.64	11,510.18
Total	59,439,840	2108.41	128,608.80

Table 2. Counterfactual analysis showing the estimated additional excess mortality rate (per 100,000 individuals) and number of excess deaths that would have occurred if time spent at home would have not changed during the weeks 13–25 of 2020. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#) and [Google LLC \(2020a\)](#).

5 Discussion

Local and national governments across the world have implemented a variety of policies aimed at reducing social contacts to mitigate the transmission and deaths of COVID-19 during the first half of 2020. We investigated the relationship between excess mortality and changes in human mobility considering that the value and effectiveness of non-pharmaceutical interventions (NPIs) is still under debate. Specifically, we leveraged digital data derived from the Google COVID-19

Community Mobility Reports ([Google LLC, 2020a](#)) to explore the association between stay-at-home measures and excess mortality at the regional level in England and Wales during the COVID-19 pandemic.

We found a strong negative correlation between the time that Google Map users spent at home and population-level excess mortality, which is considered the best indicator of the impact of the pandemic ([National Academies of Sciences, Engineering, and Medicine, 2020](#)). Our analysis determined that a time lag of at least five weeks was necessary for the association between mobility and mortality to materialize. We are unaware of previous scientific evidence regarding the time span that would be required for mobility changes to influence mortality outcomes during the COVID-19 pandemic. However, we found that this period of five weeks is consistent with preliminary estimates of the whole COVID-19 timeline from infection to death, namely, an incubation period (i.e., from infection to the onset of symptoms) that can last up to two weeks (11.5 days with a 95%CI of 8.2 to 15.6 days ([Lauer et al., 2020](#))), and a course of disease (i.e., from symptoms onset to death) that can last up to three weeks (17.8 days with a 95% CI of 16.9 to 19.2 days ([Verity et al., 2020](#))).

Using a mixed-effects regression approach, we found that an increase in time spent at home was associated with lower levels of excess mortality rate across all regions of England and Wales. These effects are strong, significant, and robust to different model specifications. Our findings are consistent with the existing evidence that stay-at-home measures are effective NPIs in the context of a global pandemic ([Lai et al., 2020](#); [Flaxman et al., 2020](#); [Del Fava et al., 2020](#)). The government’s handling of the COVID-19 pandemic in England and Wales has been heavily criticized, particularly given the delayed introduction of stay-at-home measures as compared to other countries in Europe ([Hale et al., 2020](#)). Nevertheless, by and large, residents of England and Wales appeared to have followed the government guidelines to stay at home when they were issued, as evidenced by: (i) a relative increase in the time that Google Map users spent at home after mid-March, and (ii) the large reduction in the number of face-to-face contacts per day per person after the introduction of the physical distancing guidelines, with respect to the pre-COVID-19 period ([Del Fava et al., 2020](#)).

We estimate that almost 130,000 excess deaths were averted by the increased time spent at home following the introduction of the NPIs in England and Wales. The number of lives that could have been saved if earlier and stricter measures had been put in place is unknown, but our results suggest that the potential number of excess deaths could have been much higher indeed in the absence of these interventions. Our estimates are sensitive to modeling choices as well as data inputs, which are characterized by some degree of uncertainty. Work is currently carried out to provide confidence intervals around the estimate by quantifying the uncertainty in our study.

We exploit the potential of Google data to estimate human mobility and explain excess mortality, but we are also aware of the shortcomings related to this data source. These are in particular due to the lack of details concerning the collection and processing of the mobility data. One limitation is that Google does not share absolute visitor numbers in their reports, but only relative changes with respect to the beginning of 2020. Back-engineering the underlying absolute measurements does not seem possible or desirable, given privacy concerns in sparsely populated areas. Furthermore, categories are not standard across regions and may contain different types of places (*parcs*, for example, may include castles, public gardens, observation decks, etc.). This lack of conceptual clarity also affects the ‘residential’ category used in our main analyses, which is only loosely defined in the GCMR documentation ([Google LLC, 2020b](#)). This lack of detail makes the data much less informative that it could possibly be, if all raw measures were made available together with the description of the algorithms used to produce them.

The lack of data and algorithmic transparency also limits our ability to assess the representativeness of the data. Nonetheless, there are reasons to believe that the Google mobility data provide a first and accurate approximation to the changes in human mobility occurred during the

COVID-19 pandemic. England and Wales have one of the highest rates of internet penetration in the world and one of the lowest degrees of gender inequality in internet access (Garcia et al., 2018). Furthermore, around three quarters of mobile phones in the UK use an Android operating system, in which Google Maps is installed by default.² Thus, we are confident that the aggregation of the data from daily Local Authorities inputs to regional weekly estimate approximates well the main mobility trends at the macro-region level. Finally, we are reassured by the strong and consistent relationships between some of the categories of mobility data and excess mortality, after controlling for time-specific trends of the epidemic and regional differences.

Our study provides important evidence on the positive impact of NPIs to mitigate the negative effects of the spread of COVID-19. We conclude by encouraging private companies such as Google to continue sharing data to foster academic research in areas of public interest. Further improvements include more transparency about the data-generating process and the sharing of raw data, keeping as a priority the privacy of users. The use of digital data clearly has great potential for public health research (see, e.g., Coppersmith et al., 2018; Reece and Danforth, 2017; Eichstaedt et al., 2018). We understand and share the privacy concerns related to the use of digital trace data in academic research. The current COVID-19 pandemic highlights the urgent need for an open conversation about how these data can be used ethically and, potentially, help save lives.

Authors' contributions

UB and EZ conceived the study. UB, DAG and EDF retrieved and elaborated data. UB and CGC performed analyses. All authors wrote the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript. All authors approved the final version of the manuscript.

References

- Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., and Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*, 0(0).
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Coppersmith, G., Leary, R., Crutchley, P., and Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10:1178222618792860.
- De Cao, E., Zagheni, E., Manfredi, P., and Melegaro, A. (2014). The relative importance of frequency of contacts and duration of exposure for the spread of directly transmitted infections. *Biostatistics*, 15(3):470–483.
- Del Fava, E., Cimentada, J., Perrotta, D., Grow, A., Rampazzo, F., Gil-Clavel, S., and Zagheni, E. (2020). The differential impact of physical distancing strategies on social contacts relevant for the spread of COVID-19. *medRxiv*, page 2020.05.15.20102657.
- Drake, T. M., Docherty, A. B., Weiser, T. G., Yule, S., Sheikh, A., and Harrison, E. M. (2020). The effects of physical distancing on population mobility during the COVID-19 pandemic in the UK. *The Lancet Digital Health*, page S2589750020301345.
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preotiuc-Pietro, D., Asch, D. A., and Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44):11203–11208.

²<https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/>

- Eilers, P. H. C., Gampe, J., Marx, B. D., and Rau, R. (2008). Modulation models for seasonal time series and incidence tables. *Statistics in Medicine*, 27(17):3430–3441.
- Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Imperial College COVID-19 Response Team, Ghani, A. C., Donnelly, C. A., Riley, S. M., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., and Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*.
- Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., and Hens, N. (2020). Estimating the generation interval for COVID-19 based on symptom onset data. preprint, Infectious Diseases (except HIV/AIDS).
- Garcia, D., Mitike Kassa, Y., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., and Cuevas, R. (2018). Analyzing gender inequality through large-scale Facebook advertising data. *Proceedings of the National Academy of Sciences*, 115(27):6958–6963. Publisher: National Academy of Sciences. eprint: <https://www.pnas.org/content/115/27/6958.full.pdf>.
- Google LLC (2020a). Google COVID-19 Community Mobility Reports. Available at <https://www.google.com/covid19/mobility/>. Accessed on June 30th, 2020.
- Google LLC (2020b). Google COVID-19 Community Mobility Reports: Understand the data. Available at https://support.google.com/covid19-mobility/answer/9825414?hl=en&ref_topic=9822927. Accessed on June 26th, 2020.
- Hale, T., Petherick, A., Phillips, T., and Webster, S. (2020). Variation in government responses to covid-19. *Blavatnik school of government working paper*, 31.
- Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., Wesolowski, A., Santillana, M., Zhang, C., Du, X., Yu, H., and Tatem, A. J. (2020). Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., and Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9):577–582.
- National Academies of Sciences, Engineering, and Medicine (2020). *Evaluating Data Types: A Guide for Decision Makers using Data to Understand the Extent and Spread of COVID-19*. The National Academies Press, Washington, DC.
- Office for National Statistics (2020a). Deaths registered weekly in England and Wales, provisional. Available at <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales>. Accessed on June 30th, 2020.
- Office for National Statistics (2020b). Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland. Available at <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland>. Accessed on June 2nd, 2020.
- Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., Nadai, M. D., Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., Cordes, N. d., Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., Saramäki, J., Scarpino, S. V., Tizzoni, M., Verhulst, S., and Vinck, P. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, 6:eabc0764.

- Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., and Tizzoni, M. (2020). Covid-19 outbreak response: a first assessment of mobility changes in Italy following national lockdown. *medRxiv*.
- R Development Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reece, A. G. and Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1):1–12.
- Riley, S. (2007). Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301.
- Tatem, A. J., Rogers, D. J., and Hay, S. I. (2006). Global transport networks and infectious disease spread. *Advances in parasitology*, 62:293–343.
- Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Griffin, J. T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., FitzJohn, R., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Laydon, D., Nedjati-Gilani, G., Riley, S., Elmsland, S. v., Volz, E., Wang, H., Wang, Y., Xi, X., Donnelly, C. A., Ghani, A. C., and Ferguson, N. M. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 20(6):669–677.
- Wesolowski, A., Buckee, C. O., Engø-Monsen, K., and Metcalf, C. J. E. (2016). Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *The Journal of infectious diseases*, 214(suppl_4):S414–S420.
- Wilson, M. E. (1995). Travel and the emergence of infectious diseases. *Emerging infectious diseases*, 1(2):39.
- Wood, S. N. (2019). *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation*. R package version 1.8-31.
- World Health Organization (2020). Coronavirus disease 2019 (COVID-19): Situation Report – 163. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200701-covid-19-sitrep-163.pdf?sfvrsn=c202f05b_2. (Accessed on July 3, 2020).
- Zagheni, E., Billari, F. C., Manfredi, P., Melegaro, A., Mossong, J., and Edmunds, W. J. (2008). Using time-use data to parameterize models for the spread of close-contact infectious diseases. *American journal of epidemiology*, 168(9):1082–1090.
- Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., Wu, Q., Merler, S., Viboud, C., Vespignani, A., Ajelli, M., and Yu, H. (2020). Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*.

A Sensitivity analysis

In this Appendix, we perform a sensitivity analysis of the results shown in the paper.

A.1 Different estimation of excess mortality

The results shown in our paper are robust to the computation of expected deaths \hat{d}_t derived from the Generalized Additive Model (GAM) of Equation (1) (cf. Subsection 3.1). Specifically, we re-run all our analysis using a different estimate of the excess mortality rate. Instead of using the \hat{d}_t predicted from the GAM, we computed \hat{d}_t for each week in 2020 as the average number of deaths observed in the corresponding weeks between the years 2015–2019. Figure A.1 shows the estimated numbers of expected deaths and excess mortality rate using the GAM and the historical weekly means of the observed death counts. The excess mortality rates estimated with the two approaches are very similar, especially in the period under study (grey area in 2020). As such, it is not surprising that the results shown in Section 4 do not change by a great extent when using this alternative computation of excess mortality.

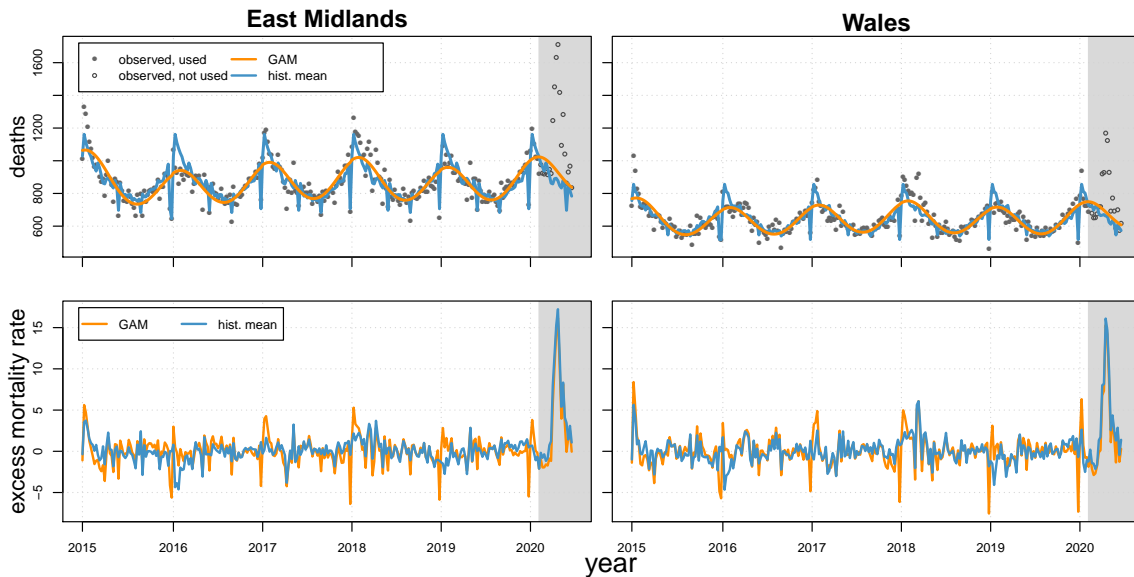


Figure A.1. Observed (dots) and fitted (lines) weekly number of deaths (upper panels) and excess mortality rate ($\hat{\delta}_t/e_t$ per 100,000 individuals, lower panels) using two different approaches (GAM in orange, weekly historical mean in blue) in East Midlands and Wales for the years 2015–2020. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#).

A.2 Exploring other Google mobility indicators

In this subsection, we investigate the relationship between excess mortality rate and other mobility indicators provided by Google LLC (2020a). Following the results shown in Section 4, we start from the correlation analysis between excess mortality and two additional categories: grocery and workplace mobility. Figures A.2 and A.3 shows these correlation, considering different lags of time for the mobility variables.

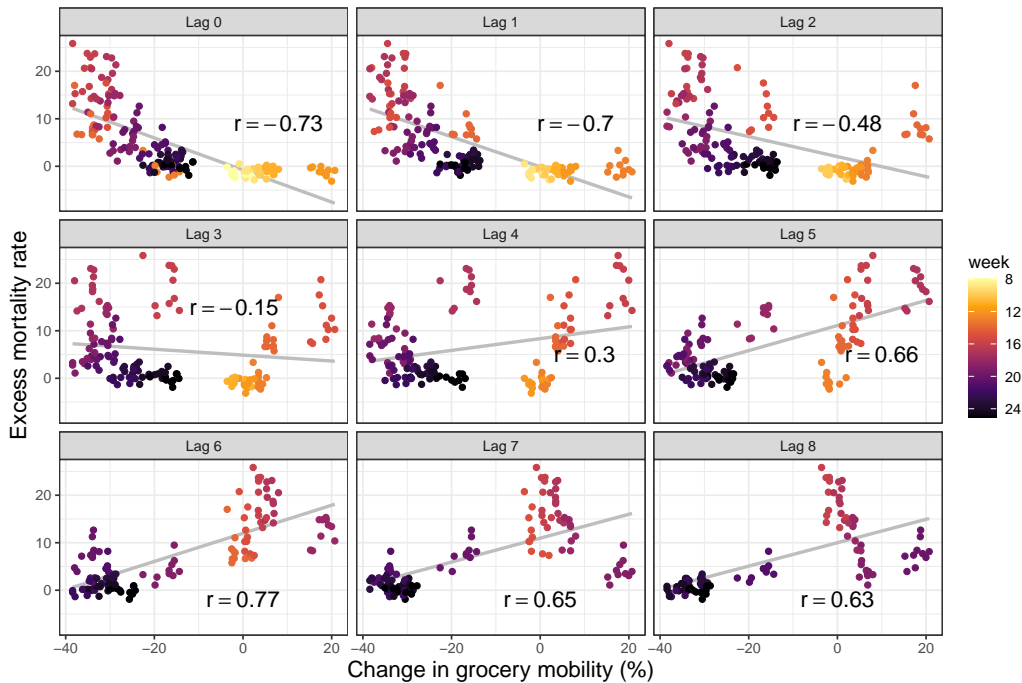


Figure A.2. Linear relationship and Pearson correlation coefficient (r) between excess mortality rate (per 100000) and change in grocery mobility in England & Wales for the weeks 8–25 in 2020, considering different time lags for grocery mobility. *Source:* Author’s elaborations on data from the Office for National Statistics (2020a,b) and Google LLC (2020a).

Results for outdoor mobility are consistent with those for the residential category (cf. Fig. 3), although with the opposite sign. Initially, changes in outdoor mobility are negatively correlated with excess mortality. However, when we consider lags greater or equal to five weeks, the relationship between the two variables becomes positive, i.e., an increase in outdoor mobility is associated with a (future) increase in excess mortality.

Next, we run additional mixed-model regressions between excess mortality and changes in mobility, considering all the remaining categories provided by Google. Table A.1 shows the results of the regressions. This analysis provides additional evidence that lagged reductions in outdoor mobility are associated with declines in excess mortality; specifically, the categories of workplaces, transit and grocery display stronger associations, while retail and recreation and parks do not show a significant association. It should be noted that we run these regressions with a single mobility indicator at a time due to the correlation between the different categories, which would likely introduce multicollinearity in regressions with more than one mobility covariate at a time.

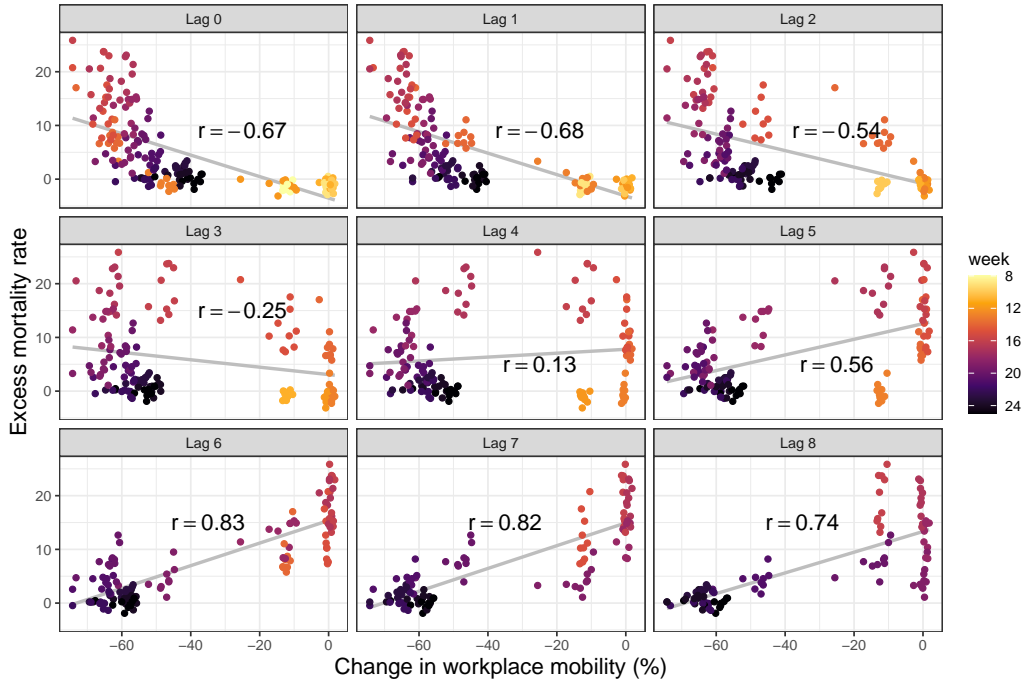


Figure A.3. Linear relationship and Pearson correlation coefficient (r) between excess mortality rate (per 100000) and change in workplace mobility in England & Wales for the weeks 8–25 in 2020, considering different time lags for workplace mobility. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#) and [Google LLC \(2020a\)](#).

	Dependent variable: excess mortality rate				
	Linear mixed-effects regression				
	workplaces	grocery	transit	retail	parks
Fixed effects					
change in mobility (lag = 5)	0.48 (0.10)	0.33 (0.15)	0.25 (0.06)	0.01 (0.12)	-0.09 (0.04)
[95% conf. interval]	[0.29, 0.67]	[0.04, 0.63]	[0.13, 0.37]	[-0.23, 0.24]	[-0.16, -0.01]
Random effects (variance)					
region (intercept)	4.39	1.81	4.86	1.90	2.01
residual	4.20	5.12	4.44	5.33	5.05
Observations	130	130	130	130	130
Log-Likelihood	-265.43	-272.32	-269.40	-274.95	-273.41
AIC	562.85	576.64	570.81	581.91	578.81
BIC	608.73	622.52	616.69	627.79	624.69

Table A.1. Estimated coefficients (with associated standard errors, in parenthesis) and 95% confidence intervals of linear mixed-effects regression between excess mortality rate (per 100,000 individuals) and changes in outdoor mobility five weeks before, measured with other categories of Google data: workplaces, grocery, transit, retail and parks. The models consider a smooth function of time (using 12 B -splines) and a random intercept for each region. Estimation is performed using restricted maximum likelihood. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#) and [Google LLC \(2020a\)](#).

A.3 Different number of B -spline bases

Here, we re-run the mixed-effect linear regression of Subection 4.3 using a different number B -spline bases. Table A.2 shows the results of this analysis. In particular, it is important to notice that the estimated mobility coefficients are in line with those found in the main analysis, ranging between -0.77 and -1.01 according to the number of B -splines. Similarly, the degree of random effect variance captured by the regional intercepts varies between 43.5% and 50.0%.

	Dependent variable: excess mortality rate			
	Linear mixed-effects regression			
	4 B -splines	6 B -splines	8 B -splines	10 B -splines
Fixed effects				
change in stay-at-home time (lag = 5)	-0.88 (0.11)	-1.01 (0.14)	-0.88 (0.14)	-0.77 (0.14)
[95% conf. interval]	[-1.08, - 0.67]	[-1.29, - 0.73]	[-1.16, - 0.61]	[-1.04, - 0.49]
Random effects (variance)				
region (intercept)	4.77	5.67	4.91	4.31
residual	6.19	5.68	5.29	4.52
Observations	130	130	130	130
Log-Likelihood	-308.39	-300.53	-292.12	-277.69
AIC	632.78	621.06	608.25	583.37
BIC	655.72	649.74	642.66	623.52

Table A.2. Estimated coefficients (with associated standard errors, in parenthesis) and 95% confidence intervals of regressing the excess mortality rate (per 100,000 individuals) on time spent at home five weeks before using linear mixed-effects regression with restricted maximum likelihood estimation. The models consider a smooth function of time with different numbers of B -splines and a random intercept for each region. *Source:* Author’s elaborations on data from the [Office for National Statistics \(2020a,b\)](#) and [Google LLC \(2020a\)](#).