



**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, Emanuele Del Fava, Daniela Perrotta, Marco Bonetti, Carlo Giovanni Camarda, Emilio Zagheni

### ► To cite this version:

Ugofilippo Basellini, Diego Alburez-Gutierrez, Emanuele del Fava, Emanuele Del Fava, Daniela Perrotta, et al.. Linking excess mortality to Google mobility data during the COVID-19 pandemic in England and Wales. 2020. hal-02899654v2

**HAL Id: hal-02899654**

**<https://hal.science/hal-02899654v2>**

Preprint submitted on 10 Feb 2021

**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 mobility data during the first wave of COVID-19 in England and Wales

Ugofilippo Basellini<sup>\*1,2</sup>, Diego Alburez-Gutierrez<sup>1</sup>, Emanuele Del Fava<sup>1</sup>, Daniela Perrotta<sup>1</sup>, Marco Bonetti<sup>3</sup>, Carlo G. Camarda<sup>2</sup>, and Emilio Zagheni<sup>1</sup>

<sup>1</sup>*Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany*

<sup>2</sup>*Institut national d'études démographiques (INED), Aubervilliers, France*

<sup>3</sup>*Carlo F. Dondena Centre & Covid Crisis Lab, Bocconi University, Milan, Italy*

February 9, 2021

## Abstract

Non-pharmaceutical interventions have been implemented worldwide to curb the spread of COVID-19. However, the effectiveness of such governmental measures in reducing the mortality burden remains a key question of scientific interest and public debate. In this study, we leverage digital mobility data to assess the effects of reduced human mobility on excess mortality, focusing on regional data in England and Wales between February and August 2020. We estimate a robust association between mobility reductions and lower excess mortality, after adjusting for time trends and regional differences in a mixed-effects regression framework and considering a five-week lag between the two measures. We predict that, in the absence of mobility reductions, the number of excess deaths could have more than doubled in England and Wales during this period, especially in the London area. The study is one of the first attempts to quantify the effects of mobility reductions on excess mortality during the COVID-19 pandemic.

**Keywords:** SARS-CoV-2 · non-pharmaceutical interventions · human mobility · digital trace data

## 1 Introduction

After the first cases of COVID-19 were identified in Wuhan City, China, in December 2019, the outbreak rapidly spread globally reaching pandemic proportions. As of December 27, 2020, over 79 million cases of infections and 1.75 million deaths have been reported worldwide (World Health Organization, 2021), although these values are likely underestimated due to cross-country differences in disease monitoring and reporting, asymptomatic cases, medically unattended cases, and deaths indirectly related to COVID-19 (Havers et al., 2020; Pullano et al., 2020).

Local and national governments across the globe implemented various non-pharmaceutical interventions (NPIs) aimed at reducing human mobility and close contacts in the population and, consequently, the probability of transmission of the SARS-CoV-2 virus. These measures include travel bans, cancellations of public gatherings, social distancing, school closures, recommendations

---

\*Corresponding author: [basellini@demogr.mpg.de](mailto:basellini@demogr.mpg.de)

Address: Konrad-Zuse-Str. 1, 18057 Rostock, Germany

Phone: +49 381 2081 264

to work from home and stay at home, and nationwide lockdowns (Brauner et al., 2020; Hale et al., 2020). In Europe, such interventions were introduced in the first half of 2020, with considerable cross-country differences in terms of strictness, timing and duration. Most of these measures were then lifted during the summer as the number of new infections and deaths declined, but also to alleviate their long-term socioeconomic costs on society and citizens. A second wave of infections and deaths prompted the re-introduction of NPIs after the summer, again with different strategies across Europe (Kupferschmidt, 2020).

A growing body of literature has documented the link between the introduction of NPIs and the reduction of the SARS-CoV-2 virus' transmission (Brauner et al., 2020; Davies et al., 2020; Dehning et al., 2020; Flaxman et al., 2020; Hsiang et al., 2020) and the human cost of letting the virus spread unchecked (Buss et al., 2021). However, it still remains unclear how to measure the effectiveness of such interventions in alleviating the mortality burden of COVID-19. Estimating the impact of NPIs on mortality reductions is critical for policy makers to make informed decisions, in the context of both current and future pandemics.

In this article, we study the relationship between human mobility and excess mortality at the sub-national level in England and Wales during the first wave of the COVID-19 pandemic, specifically from February 15 to August 14, 2020. Note that at the time of writing (February 2021), the second wave of infections and deaths is still unfolding in the UK and the inclusion of these data could lead to misleading findings. Approximately 51,500 COVID-19 deaths were registered during this period of time (Office for National Statistics, 2020b), which includes almost one month before any NPIs were put in place. The British government applied various mobility restrictions strategies between March 12 and 24, 2020, including encouragement of social distancing, closure of schools, ban of public events, and total lockdown (Cameron-Blake et al., 2020; Flaxman et al., 2020). The goal of this paper is to estimate the impact of these governmental decisions, and consequent reduction in human movements, on excess all-cause mortality.

Human mobility plays a key role in the spread of infectious diseases (Riley, 2007; Tatem et al., 2006; Wilson, 1995). In our contemporary societies, 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 larger scale, than ever before. Population movement 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 for informing public health interventions, but generally not available or easily accessible, and the ongoing COVID-19 pandemic has once again amplified this long-standing issue. Recent work has highlighted the potential benefits of harnessing geo-located smartphone data to inform policy makers (Oliver et al., 2020) and to assess the impact of mobility restrictions on social distancing in near real-time (Badr et al., 2020; Davies et al., 2020; Pepe et al., 2020; Schlosser et al., 2020).

In this study, we exploit a publicly available human mobility dataset, the Google COVID-19 Mobility Reports (GCMR) (Google LLC, 2021). The GCMR reports changes in mobility of Google Maps users across different categories (e.g. supermarkets and pharmacies, workplaces, residential areas) with respect to the start of 2020. This dataset has been leveraged to monitor national mobility in the United Kingdom following the implementation of NPIs (Drake et al., 2020), and it has been employed (alongside social contacts data) to estimate the effect of tiered restrictions in England and the lockdown in Wales implemented in October 2020 (Davies et al., 2020).

We model the association between the reduction in human mobility and the excess all-cause mortality during the first wave of the COVID-19 pandemic. Given the high uncertainty surrounding the number of infections and deaths, we choose to estimate excess mortality, which indicates the number of deaths above what would be expected in a non-crisis period. This measure overcomes potential issues of incorrect death classification and registration, and is largely considered the best indicator of the pandemic's impact on mortality (Kontis et al., 2020; National Academies of Sciences, Engineering, and Medicine, 2020). Given that changes in mobility are not immediately reflected in changes in mortality, we analyse different time lags between the two measures to

calibrate the length of the expected delay in the effect of mobility restrictions on mortality. We expect reduction in mobility to first decrease the number of close social contacts in the population, leading to fewer new infections and, eventually, fewer deaths.

## 2 Material and methods

### 2.1 Data

We employ two data sources in our study: the UK Office for National Statistics (ONS) for mortality and population data, and the Google COVID-19 Community Mobility Reports (GCMR) for mobility data.

The ONS reports the weekly number of deaths registered in England and Wales, broken down by age group, sex and region of usual residence (Office for National Statistics, 2021). In the ONS classification, weekly data do not refer to calendar weeks, but to rolling 7-day periods, from Saturday to Friday, and we keep this notation throughout the article. For our purposes, we retrieve the weekly number of deaths by region for the years 2015–2020. We consider a total of ten regions based on the NUTS-1 (Nomenclature of Territorial Units for Statistics) subdivision, namely Wales plus nine regions in England (North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East, London, South East and South West). We focus our analysis on the first wave of the COVID-19 pandemic, hence we use data until week 2020-33 (i.e. until August 14, 2020). We obtain data on total population in each region for the years 2015–2019 from the ONS (Office for National Statistics, 2020a).

The GCMR reports daily mobility data in six categories of location: residential, workplaces, supermarket and pharmacy (grocery), transit, retail, and parks (Google LLC, 2021). Data are provided as percentage variations in number of visits or time spent in each category, relative to a pre-COVID-19 baseline period, defined from January 3 to February 6, 2020. 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. To protect users’ privacy, absolute mobility values are not available.

We obtain mobility data for each category of the GCMR for 108 sub-national regions (the GCMR’s *sub\_region\_1* variable), covering a geographic area inhabited by 99% of the population of England and Wales, from February 15 (the first available date in the dataset) to August 14, 2020. We aggregate the GCMR data by week (from Saturday to Friday, for consistency with mortality data) and region (taking the weighted average across all counties belonging to a given region, with weights equal to their population sizes) 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 future relative changes in visits to grocery stores. 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 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 on March 12, 2020.

### 2.2 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 week-specific coefficients for the death counts.

For a given region, let  $\mathbf{D} = (d_{w,t})$  denote a matrix containing the registered number of deaths in week  $w = 1, \dots, 53$  and year  $t = 2015, \dots, 2020$ , i.e., weekly registered deaths from week 2015-01 to week 2020-33. In alternative to the conventional Poisson distribution, we assume that the  $d_{w,t}$

are realisations of a random variable  $D_{w,t}$  that follows a Negative Binomial distribution. We thus allow for the overdispersion typically displayed in mortality data, i.e., the variance associated with the process is expected to be larger than what would be implied by the expected value in a Poisson model. 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  $\mu_{w,t} = \mathbb{E}(D_{w,t})$  denote the expected value of the Negative Binomial process. We model the expected number of deaths using a Generalized Additive Model (GAM) with logarithmic link, exposures as offset, and time-specific covariates:

$$\ln(\mu_{w,t}) = \ln(e_{w,t}) + f(x_{w,t}) + \alpha_0 + \alpha_w, \quad (1)$$

where  $e_{w,t}$  denotes the exposure to the risk of death, the function  $f(\cdot)$  is a smooth function over the observations  $x_{w,t} = 1, \dots, m$ , where  $m$  denotes the total number of observations,  $\alpha_0$  is an intercept and  $\alpha_w$  are week-specific coefficients. To ensure identifiability of the  $\alpha_w$  coefficients, the first week is taken as the reference group (i.e.  $\alpha_1 = 1$ ). Exposures over weeks and years are obtained by linear interpolation of mid-year populations in years  $t = 2015, \dots, 2019$  and linear extrapolation for the year 2020.

Whereas the coefficients  $\alpha_w$  capture the typical pattern of mortality in registered week  $w$  (including cyclical behavior of weekly mortality data, bank holidays effects, etc.), the smooth time component aims to describe long-term mortality trends. Given that mortality data are recorded by registration rather than occurrence, such a specification provides a better description of the data than including trigonometric functions for the seasonal pattern of deaths. 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, corresponding to a total of  $m = 267$  observations) and predict the expected number of deaths  $\hat{d}_w$  for the COVID-19 period (from week 2020-07 to week 2020-33). We use week 2020-06 as last estimated week before prediction 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  $\hat{d}_{w,2020}$  for the predicted weeks could be interpreted as the number of deaths that would have occurred in 2020 in the absence of the COVID-19 pandemic.

This approach allows us to readily evaluate the excess number of deaths  $\hat{\delta}_w$  for the weeks 07–33 in 2020 as the difference between the observed and the expected ones, i.e.  $\hat{\delta}_w = d_w - \hat{d}_w$ . In order to account for regional differences in population, we then compute the excess mortality rate, or per capita excess mortality,  $y_w = \hat{\delta}_w/e_w$ , dividing excess deaths by the (extrapolated) region-specific exposures in week  $w$  of 2020.

In addition to fitting the GAM, we perform a sensitivity analysis on the computation of the excess mortality rate. In particular, we derive another estimate of  $y_w$  by computing, for each week, the expected number of deaths as the average of the observed deaths between 2015 and 2019.

### 2.3 Combining Google categories into a single index

For each region analysed in the paper, we aim to extract the most relevant signals from the different categories of the GCMR by merging them into a single ‘‘Combined Google index’’. Ideally, all three dimensions of the GCMR (Google categories, regions and time) should be considered simultaneously. This rules out region-independent principal component analysis (PCA) of the mobility data over time. A solution is instead given by a multilinear principal component analysis (MPCA) (Lu et al., 2008). Whereas standard PCA reduces the dimensionality of a two-dimensional data set, MPCA allows to extract features of a multidimensional object such as the GCMR. In both approaches, the goal is to retain as much as possible the variation present in the original data set.

To do so, we construct a tensor object (a multidimensional array) containing the three types of data, and we extract the first component of the MPCA for the dimensions time and region. We

thus obtain a “combined Google index” for each region over time retaining most of the information regarding the mobility during the period analysed. In R, this can be achieved by using the `rTensor` package (Li et al., 2018).

## 2.4 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 mobility exists, and its magnitude. The spread of COVID-19 mainly occurs through contacts between infectious and susceptible individuals (Zhang et al., 2020). Hence, a reduction in mobility 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 of  $x$  weeks in the mobility data, which corresponds to the amount of time necessary for the change in mobility to have an impact on mortality. In other words, we analyse the relationship between excess mortality and changes in human mobility that occurred  $x$  weeks before.

Moreover, we work with rates that vary over weeks and for different regions. While the mortality trend will be assumed to remain constant in space, we need to account for the regional heterogeneity in excess mortality and response to mobility changes, 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 across regions.

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

$$y_{r,w} = \beta_0 + u_r + (\gamma_0 + \gamma_r) g_{r,w-x} + \sum_{s=1}^n \beta_s B_s(w) + \varepsilon_{r,w}, \quad (2)$$

where  $\beta_0$  is the common intercept and  $u_r$  are the region-specific random intercepts, added to capture average regional differences;  $\gamma_0$  is the common mobility coefficient, which can be interpreted as the average change in per capita excess mortality for a unit change in the mobility indicator (with respect to the baseline period), and  $\gamma_r$  are the region-specific random slopes that modify the effect of mobility change for region  $r$  during the (lagged) week  $w - x$ , i.e.,  $g_{r,w-x}$ . Random intercepts  $u_r$  and random slopes  $\gamma_r$  are normally distributed with mean zero and variance  $\sigma_u^2$  and  $\sigma_\gamma^2$ , respectively, with their dependence captured by the covariance term  $\sigma_{u,\gamma}$ , which allows the computation of a model-based correlation coefficient. Evidence for  $\sigma_u^2$  and  $\sigma_\gamma^2$  being greater than zero implies the existence of regional differences in the baseline levels of mortality and heterogeneity in the responses to mobility changes, respectively. The baseline mortality time trend is modelled in a flexible way using a non-parametric approach based on  $B$ -spline bases  $B_s$ , with  $\beta_s$  denoting the associated coefficients. Finally,  $\varepsilon_{r,w}$  is the vector of the residuals, distributed as  $\varepsilon_{r,w} \sim N(0, \sigma^2)$  and assumed to be independent of the random effects  $u_r$  and  $\gamma_r$ .

To account for uncertainty in the estimates, related to both the computation of excess mortality and the mixed-effects regression in our estimates, we employ a bootstrap approach to derive 95% pointwise confidence intervals for the model results. Specifically, we generate 1,000 simulated death counts from the deviance residuals of the GAM; for each simulation, we compute the regression in Equation (2) and derive fitted excess mortality rates. 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).

Finally, all data and source codes necessary for the reproducibility of our results are available at the following [blind for peer-review] repository: [https://osf.io/4pfb7/?view\\_only=3155acef790e4e5f8ac96f4c2c6ede71](https://osf.io/4pfb7/?view_only=3155acef790e4e5f8ac96f4c2c6ede71)

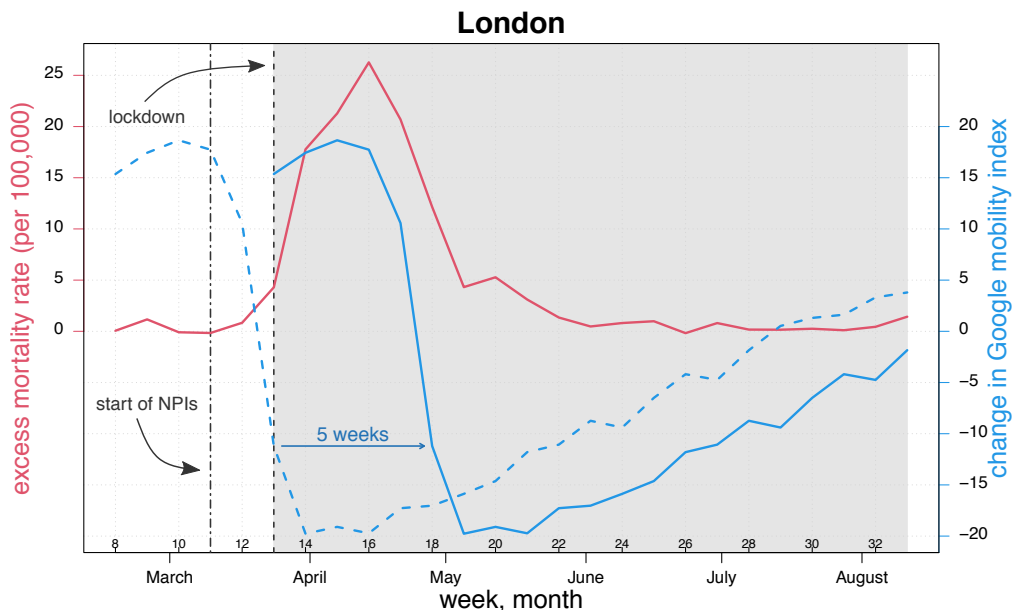
### 3 Results

#### 3.1 Time series of excess mortality rate and Google mobility

We compute the excess mortality rate and the combined “Google mobility index” at the regional level in England and Wales during weeks 8–33 of 2020. Figure 1 shows the time series of the two variables in the region of London, plus: (i) a forward lag of five weeks for the mobility data, (ii) the time period that we will analyse in our regression framework (weeks 13–33, highlighted in the grey shaded area), and (iii) the start of NPIs on March 12 (week 11) and the enforcement of the lockdown on March 24 (week 13).

The figure shows the drastic and sustained mobility reduction since week 11, resulting from the various NPIs implemented between March 12 and 24, 2020. The graph also exhibits the sharp increase in the excess mortality rate during weeks 13–22, which peaked at week 16 before returning to levels close to zero. Finally, the five-week forward shift of the mobility data highlights the correspondence between the decrease in (forward) mobility and the reduction in excess mortality starting from week 17 onwards.

We provide a series of additional graphs related to these two variables in Appendix A. Specifically, the estimation of the excess mortality rate using the GAM is exemplified for two regions in Figure A.1. Figure A.2 shows the time series of the six categories of the GCMR across the ten regions analysed, as well as their aggregation into the combined Google index. Moreover, Figures A.3 and A.4 report the same information shown in Figure 1 for the other nine regions under study.



**Figure 1.** Time series of excess mortality rate per 100,000 individuals (red line) and change in Google mobility index at week  $t$  (dashed blue line) and with a five-week forward shift (solid blue line) in the region of London during weeks 8–33 of 2020. Solid lines in the grey shaded area correspond to values analysed as described in the “Statistical analysis” section. Vertical lines indicate the start of NPIs on March 12 (week 11) and the lockdown ordered on March 24 (week 13), respectively. The Google index was multiplied by 10 for illustration purposes. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

#### 3.2 Statistical analysis

We investigate the relationship between excess mortality and mobility using a mixed-effects regression approach. To account for the delay between the two phenomena, we analyse excess mortality

with respect to mobility changes that occurred five weeks in the past. We found this lag to be the shortest one displaying a positive relationship between mortality and mobility (see Table A.1 as well as Figures A.5 and A.6 in Appendix A). We perform sensitivity analysis on the lag selection, and find that our results do not change for lags of 6 or 7 weeks (see Table A.1 in Appendix A).

We find a strong and significant association between mobility reduction and excess mortality after five weeks, after controlling in the regression model for the pandemic time trend and for regional differences. Table 1 reports the results of the mixed-effects regression models, considering the combined Google index and the six categories of the GCMR independently. The models include a smooth function of time (using 5  $B$ -splines), as well as random intercepts and random slopes for each region. We standardise the mobility data to aid the interpretation and comparison of the estimated coefficients from the various models.

**Table 1.** Estimated coefficients and 95% confidence intervals from the linear mixed-effects regression between excess mortality rate (per 100,000 individuals) and changes in mobility occurred five weeks before, measured separately for each model with the combined Google index and the six categories of the GCMR: grocery, workplaces, residential, transit, retail and parks. For the parks category, we considered only random intercepts since the the model with both random intercepts and slopes did not converge. Estimation is performed using restricted maximum likelihood. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

| Dependent variable: excess mortality rate (per 100,000) |                     |                      |                     |                         |                      |                      |                       |
|---|---------------------|----------------------|---------------------|-------------------------|----------------------|----------------------|-----------------------|
| Linear mixed-effects regression                         |                     |                      |                     |                         |                      |                      |                       |
|   | Google index        | grocery              | workplaces          | residential             | transit              | retail               | parks                 |
| <b>Fixed effects</b>                                    |                     |                      |                     |                         |                      |                      |                       |
| Mobility changes<br>5 weeks before<br>(95% CI)          | 3.75<br>(2.7, 4.81) | 2.56<br>(1.78, 3.35) | 3.11<br>(2.02, 4.2) | -3.53<br>(-4.61, -2.44) | 2.93<br>(1.86, 3.99) | 3.64<br>(2.46, 4.81) | 0.70<br>(-0.15, 1.54) |
| <b>Random effects<br/>(variance)</b>                    |                     |                      |                     |                         |                      |                      |                       |
| Region (intercept)                                      | 2.67                | 1.82                 | 2.31                | 3.11                    | 3.46                 | 2.47                 | 1.45                  |
| Mobility (slope)  | 0.66                | 0.83                 | 0.86                | 0.62                    | 0.73                 | 1.20                 | –                     |
| Residual  | 3.35                | 3.26                 | 3.56                | 3.56                    | 3.78                 | 3.47                 | 5.60                  |
| Observations  | 210                 | 210                  | 210                 | 210                     | 210                  | 210                  | 210                   |
| Groups  | 10                  | 10                   | 10                  | 10                      | 10                   | 10                   | 10                    |
| Log-Likelihood  | -437.42             | -434.56              | -443.69             | -443.03                 | -449.61              | -441.62              | -480.42               |
| AIC   | 896.84              | 891.12               | 909.37              | 908.06                  | 921.22               | 905.24               | 978.84                |
| BIC   | 933.66              | 927.94               | 946.19              | 944.87                  | 958.04               | 942.06               | 1008.97               |

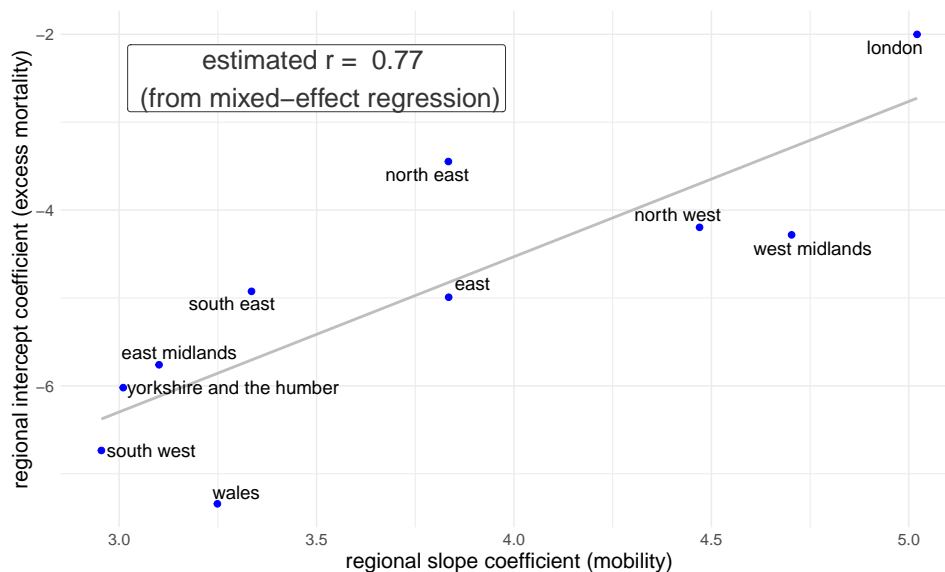
We estimate that a reduction of one standard deviation in the combined Google mobility index is associated with a reduction of 3.75 in the excess mortality rate per 100,000 individuals five weeks later. This is a strong effect, given that the Google index changed by almost 4 standard deviations across all regions following the introduction of the NPIs (see Figure A.2 in Appendix A). Moreover, five out of six categories of mobility indicators reported by Google display a similar robust relationship with excess mortality, while the parks category is the only one which does not display a significant association. This could be related to the greater volatility of the parks mobility time series with respect to the other ones.

These results are robust to a series of sensitivity analysis, which are reported in Appendix B. The estimated coefficients are robust to a change in the computation of the excess mortality rate. The historical-based approach to estimate excess mortality is shown and compared to the GAM in Figures B.1 and B.2, while Table B.1 reports the models’ results using this different computation of the excess mortality rate. Similarly, the number of  $B$ -splines employed to describe the time trend of the epidemic does not change the estimated mobility coefficients by a great extent. As a matter of fact, our choice of 5  $B$ -splines produces the most conservative estimate of the effect of mobility on excess mortality (see Table B.2). Finally, the exclusion of the region of London from



the analysis does not influence the magnitude and significance of the estimated coefficients (see Table B.3).

It is important to analyse cross-sectional differences across regions in terms of both excess mortality and response to mobility reductions. The regression model can account for such variations by means of region-specific random intercepts (corresponding to different levels of excess mortality) and random slopes (corresponding to different responses to mobility changes). Focusing on the model with the combined Google index (here and in the remainder of the section), Figure 2 shows that the intercept and the slope have a positive correlation of 0.77 across the regions, so that those with larger intercepts (i.e. higher excess mortality levels) also have larger slopes (i.e. larger effects of reduction in mobility).



**Figure 2.** Estimated region-specific intercepts and mobility slopes, as well as their estimated correlation  $r$  from the mixed-effects regression in England and Wales by region during weeks 13–33 of 2020. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

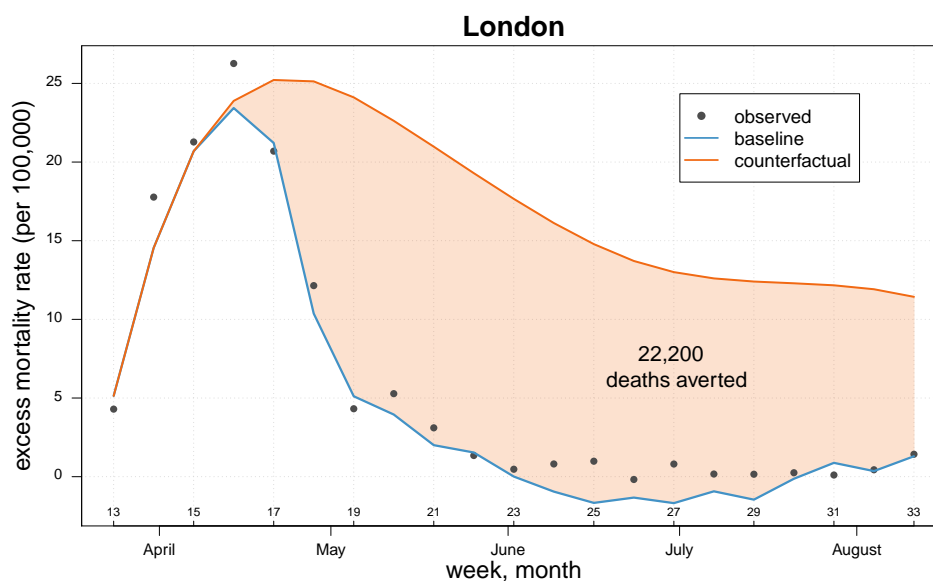
Finally, the estimated model allows us to estimate the number of deaths averted by mobility reductions. This is achieved by simulating a counterfactual worst-case scenario in which mobility is assumed not to have dropped following the introduction of the NPIs but rather remained at the levels observed before the implementation of such measures. Table 2 reports the results of this analysis, as well as the estimates of observed excess mortality during the period analysed. The fit of the model and the counterfactual analysis for the region of London are shown in Figure 3, and the graphs for the remaining nine regions are shown in Figure A.7 of Appendix A. We estimate that about 62,100 excess deaths occurred in England and Wales during the weeks 13–33 of 2020, and that an additional 93,700 excess deaths (95% confidence intervals 85,400–102,500) would have occurred if mobility had not reduced. These absolute estimates display expected significant regional variations, ranging from a minimum of 3,600 excess deaths averted in North East to a maximum of 22,200 excess deaths averted in London.

## 4 Discussion

Local and national governments around the world have implemented a variety of policies aimed at reducing social contacts to curb the transmission of and deaths from COVID-19. We investigated the relationship between excess mortality and changes in human mobility to assess the effectiveness of non-pharmaceutical interventions. Specifically, we leveraged digital data derived from the Google

**Table 2.** Population size, estimated number of excess deaths, and estimated number of deaths averted by the mobility reductions (counterfactual analysis) with 95% confidence intervals by region in England and Wales during weeks 13–33 of 2020. Estimates have been rounded to the nearest hundredth to avoid giving a false sense of precision in the presence of uncertainty (as in Kontis et al., 2020); as such, figures for the Total row may differ from the sum of the regions. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

| Region                   | Population (2019) | Estimated excess deaths | Deaths averted (counterfactual) | 95% CI             |
|--------------------------|-------------------|-------------------------|---------------------------------|--------------------|
| North East               | 2,669,941         | 4,100                   | 3,600                           | (3,300 ; 4,100)    |
| North West               | 7,341,196         | 9,000                   | 12,600                          | (11,600 ; 13,500)  |
| Yorkshire and The Humber | 5,502,967         | 5,100                   | 6,200                           | (5,600 ; 7,100)    |
| East Midlands            | 4,835,928         | 4,400                   | 5,900                           | (5,200 ; 6,700)    |
| West Midlands            | 5,934,037         | 7,200                   | 10,600                          | (9,700 ; 11,400)   |
| East                     | 6,236,072         | 6,300                   | 9,500                           | (8,600 ; 10,500)   |
| London                   | 8,961,989         | 11,000                  | 22,200                          | (20,900 ; 23,100)  |
| South East               | 9,180,135         | 9,300                   | 12,800                          | (11,700 ; 14,400)  |
| South West               | 5,624,696         | 3,800                   | 6,400                           | (5,700 ; 7,400)    |
| Wales                    | 3,152,879         | 1,900                   | 3,800                           | (3,300 ; 4,400)    |
| Total                    | 59,439,840        | 62,100                  | 93,700                          | (85,400 ; 102,500) |



**Figure 3.** Observed and estimated excess mortality rate (per 100,000 individuals) from the mixed-effects model in the baseline and counterfactual scenarios for the region of London during weeks 13–33 of 2020. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

COVID-19 Community Mobility Reports (Google LLC, 2021) to explore the association between mobility and excess mortality at the regional level in England and Wales.

We found a strong positive relationship between the mobility of Google Maps users and population-level excess mortality, which is considered to be the best indicator of the impact of the pandemic on mortality (National Academies of Sciences, Engineering, and Medicine, 2020). Our analysis determined that a time lag of at least five weeks is needed to reveal a positive association between mobility and mortality, while smaller lags display a negative relationship. A five-week

time period is consistent with preliminary estimates of the disease duration from infection to death, with the incubation period (i.e., from infection to symptom onset) 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 the course of disease (i.e., from symptom 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)). This is also consistent with preliminary studies in the US, where mobility reductions assessed via mobile phone data were found to anticipate the exponential decay of COVID-deaths by 6 weeks (Kogan et al., 2020). Finally, the period of time between infection and death is also related to the individual-level Susceptible-Exposed-Infectious-Recovered (SEIR) compartmental models that have been proposed to describe the spread of COVID-19 (Lin et al., 2020).

Furthermore, our findings are aligned with the existing evidence that mobility restrictions and stay-at-home measures are effective NPIs in the context of a global pandemic (Brauner et al., 2020; Davies et al., 2020; Dehning et al., 2020; Del Fava et al., 2020; Flaxman et al., 2020; Lai et al., 2020). The government’s handling of the COVID-19 pandemic in England and Wales has been heavily criticised, particularly given the delayed introduction of NPIs as compared to other countries in Europe (Hale et al., 2020). Nevertheless, the residents of England and Wales appeared to have followed the government guidelines to stay at home when they were issued, as evidenced by: (i) the drastic reduction in mobility of Google Maps users after mid-March, (ii) the large reduction in the daily number of face-to-face contacts per person after the introduction of the physical distancing guidelines (Del Fava et al., 2020; Jarvis et al., 2020), and (iii) the large adoption of social distancing measures and reduction in mobility following the implementation of NPIs as reported through online surveys (Perrotta et al., 2020).

We estimate that about 62,100 excess deaths occurred in England and Wales during weeks 13–33 of 2020 with respect to what was expected from previous years, in line with other findings (Aburto et al., 2021; Office for National Statistics, 2020b). Furthermore, we estimate that an additional 93,700 excess deaths were averted by the reduced mobility following the introduction of the NPIs. 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 in the absence of these interventions.

A limitation of our analysis is that, given the almost simultaneous introduction of different NPIs in March 2020 and the resulting reduction in mobility, we cannot disentangle the individual contribution of each intervention to the change in mobility and mortality. Moreover, the relationship between mobility data and excess mortality is necessarily based on location-based measures, and as such it may be prone to ecological fallacy. Nonetheless, our methodology allows us to isolate and estimate the effect of mobility reductions on excess mortality controlling for (i) the time trend of the epidemic and (ii) all other unobserved region-specific factors, which are allowed for through the random-effect specification of the model.

In our work, we exploit the potential of digital-trace 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 detail concerning the collection and processing of the mobility data. One limitation is that Google does not share absolute 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. Moreover, no information is provided on the population composition of Google Maps users, such as age-group or sex breakdowns, thus limiting our ability to assess the representativeness of the data. This lack of detail makes the data less informative than it could possibly be, if all raw measures were made available together with the description of the algorithms used to produce them.

Nonetheless, we believe that the Google mobility data provide a first and valuable approximation to the changes in human mobility occurred during the COVID-19 pandemic. Such data are necessarily affected by biases related to population sampling, which depends on the market share of the operator providing the data and the different usage across socio-demographic groups. However,

we are reassured by existing evidence on Internet penetration rates and Google Maps coverage in the UK, and by the high consistency in mobility estimates provided by Google. England and Wales have in fact 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 worldwide use an Android operating system, on which Google Maps is installed by default (StatCounter, 2020), and survey results indicate that approximately 60% of Android users in the UK have Google Location History reporting enabled (Ruktanonchai et al., 2018). In addition, existing evidence shows very high levels of consistency in mobility estimates obtained from Google and from call detail records (Ruktanonchai et al., 2020). Finally, we are reassured by the similar general patterns of the Google mobility data as compared to other digital sources, such as Apple and Facebook (which only report data for England and Wales as a whole and not at the sub-national level, see Figure A.8 in Appendix A). Thus, we are confident that this data source approximates well the overall mobility trends at the regional level.

Our study provides evidence on the positive impact of NPIs to mitigate the mortality burden 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. An increasing number of studies are showing the great potential of digital data for public health research (Coppersmith et al., 2018; Eichstaedt et al., 2018; Oliver et al., 2020; Reece and Danforth, 2017), and the current COVID-19 pandemic further highlights the need for an open conversation on how these data can be used ethically to help save lives.

## Acknowledgements

The authors would like to thank Vladimir Shkolnikov and Marília Nepomuceno for providing useful comments and discussions on an earlier version of this article.

## Author contributions

UB and EZ conceived the study. UB and DAG retrieved and elaborated data. UB, DAG, EDF 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.

## Competing interests

The authors declare no competing interests.

## References

- Aburto, J. M., Kashyap, R., Schöley, J., Angus, C., Ermisch, J., Mills, M. C., and Dowd, J. B. (2021). Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysis. *Journal of Epidemiology and Community Health*, pages jech-2020-215505.
- Apple (2020). COVID-19 - Mobility Trends Reports. Available at <https://covid19.apple.com/mobility>. Accessed on October 27, 2020.
- 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.

- Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A. B., Leech, G., Altman, G., Mikulik, V., Norman, A. J., Monrad, J. T., Besiroglu, T., Ge, H., Hartwick, M. A., Teh, Y. W., Chindelevitch, L., Gal, Y., and Kulveit, J. (2020). Inferring the effectiveness of government interventions against covid-19. *Science*.
- Buss, L. F., Prete, C. A., Abraham, C. M. M., Mendrone, A., Salomon, T., de Almeida-Neto, C., França, R. F. O., Belotti, M. C., Carvalho, M. P. S. S., Costa, A. G., Crispim, M. A. E., Ferreira, S. C., Fraiji, N. A., Gurzenda, S., Whittaker, C., Kamaura, L. T., Takecian, P. L., da Silva Peixoto, P., Oikawa, M. K., Nishiya, A. S., Rocha, V., Salles, N. A., de Souza Santos, A. A., da Silva, M. A., Custer, B., Parag, K. V., Barral-Netto, M., Kraemer, M. U. G., Pereira, R. H. M., Pybus, O. G., Busch, M. P., Castro, M. C., Dye, C., Nascimento, V. H., Faria, N. R., and Sabino, E. C. (2021). Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science*, 371(6526):288–292.
- Cameron-Blake, E., Tatlow, H., Wood, A., Hale, T., Kira, B., Petherick, A., and Phillips, T. (2020). Variation in the response to COVID-19 across the four nations of the United Kingdom. *Blavatnik School of Government Working Paper*. October 26, 2020. Available at: <https://www.bsg.ox.ac.uk/research/publications/variation-response-covid-19-across-four-nations-united-kingdom>.
- 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.
- Davies, N. G., Barnard, R. C., Jarvis, C. I., Russell, T. W., Semple, M. G., Jit, M., and Edmunds, W. J. (2020). Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in england: a modelling study. *The Lancet Infectious Diseases*.
- Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., and Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, 369(6500).
- 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*.
- 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*.
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoțiuc-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.
- Facebook (2020). Movement Range Maps. Available at <https://data.humdata.org/dataset/movement-range-maps>. Accessed on October 27, 2020.
- 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*.
- 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 (2021). Google COVID-19 Community Mobility Reports. Available at <https://www.google.com/covid19/mobility/>. Accessed on January 10, 2021.
- Hale, T., Angrist, N., Kira, B., Petherick, A., Phillips, T., and Webster, S. (2020). Variation in government responses to COVID-19. *Blavatnik School of Government Working Paper*. May 25, 2020. Available at: [www.bsg.ox.ac.uk/covidtracker](http://www.bsg.ox.ac.uk/covidtracker).
- Havers, F. P., Reed, C., Lim, T., Montgomery, J. M., Klena, J. D., Hall, A. J., Fry, A. M., Cannon, D. L., Chiang, C.-F., Gibbons, A., Krapivunaya, I., Morales-Betoulle, M., Roguski, K., Rasheed, M. A. U., Freeman, B., Lester, S., Mills, L., Carroll, D. S., Owen, S. M., Johnson, J. A., Semenova, V., Blackmore, C., Blog, D., Chai, S. J., Dunn, A., Hand, J., Jain, S., Lindquist, S., Lynfield, R., Pritchard, S., Sokol, T., Sosa, L., Turabelidze, G., Watkins, S. M., Wiesman, J., Williams, R. W., Yendell, S., Schiffer, J., and Thornburg, N. J. (2020). Seroprevalence of antibodies to SARS-CoV-2 in 10 sites in the united states, march 23-may 12, 2020. *JAMA Internal Medicine*, 180(12):1576.
- Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L. Y., Hultgren, A., Krasovich, E., Lau, P., Lee, J., Rolf, E., Tseng, J., and Wu, T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, 584(7820):262–267.
- Jarvis, C. I., Van Zandvoort, K., Gimma, A., Prem, K., working group, C. C.-., Klepac, P., Rubin, G. J., and Edmunds, W. J. (2020). Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Medicine*, 18.
- Kogan, N. E., Clemente, L., Liautaud, P., Kaashoek, J., Link, N. B., Nguyen, A. T., Lu, F. S., Huybers, P., Resch, B., Havas, C., et al. (2020). An Early Warning Approach to Monitor COVID-19 Activity with Multiple Digital Traces in Near Real-Time. *arXiv preprint arXiv:2007.00756*. Available at <https://arxiv.org/abs/2007.00756>.
- Kontis, V., Bennett, J. E., Rashid, T., Parks, R. M., Pearson-Stuttard, J., Guillot, M., Asaria, P., Zhou, B., Battaglini, M., Corsetti, G., McKee, M., Cesare, M. D., Mathers, C. D., and Ezzati, M. (2020). Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nature Medicine*, 26(12):1919–1928.
- Kupferschmidt, K. (2020). Europe is locking down again—but its strategy is unclear. *Science*, 370(6517):644–645.
- 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.
- Li, J., Bien, J., and Wells, M. T. (2018). rTensor: An R package for multidimensional array (tensor) unfolding, multiplication, and decomposition. *Journal of Statistical Software*, 87(10):1–31.
- Lin, Q., Zhao, S., Gao, D., Lou, Y., Yang, S., Musa, S. S., Wang, M. H., Cai, Y., Wang, W., Yang, L., and He, D. (2020). A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases*, 93:211–216.

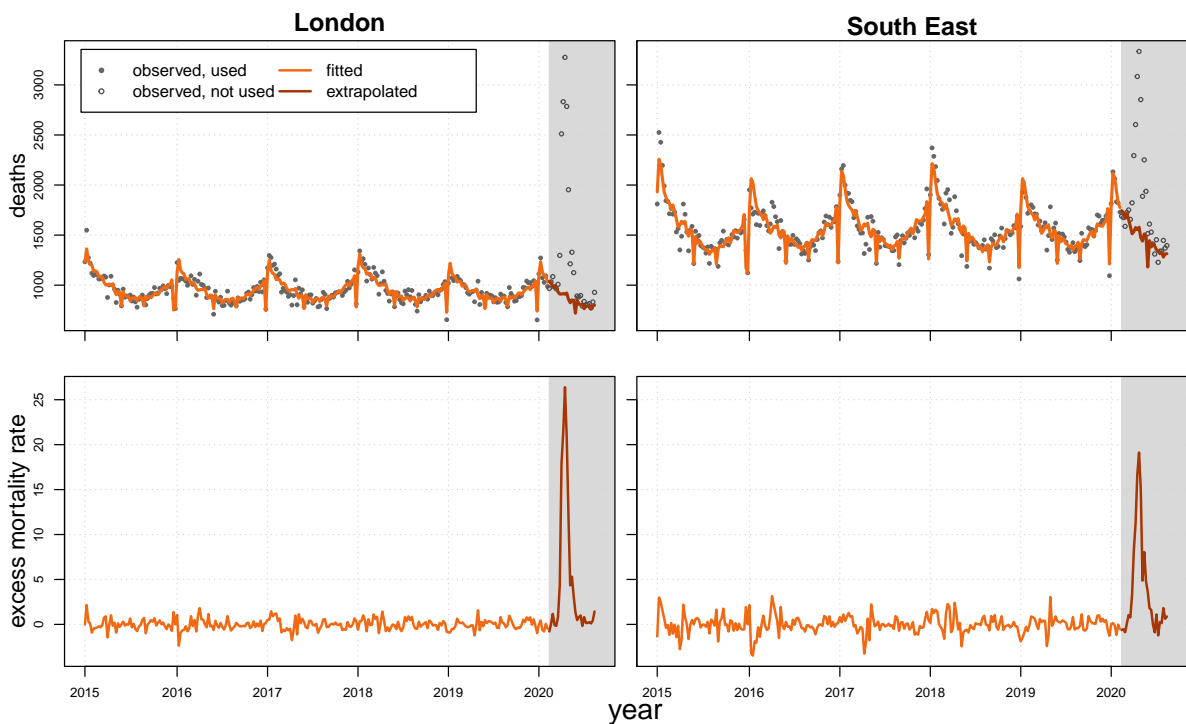
- Lu, H., Plataniotis, K., and Venetsanopoulos, A. (2008). MPCA: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, 19(1):18–39.
- National Academies of Sciences, Engineering, and Medicine (2020). *Evaluating Data Types: A Guide for Decision Makers using Data to Understand the Extent and Spread of COVID-19*. The National Academies Press, Washington, DC.
- Office for National Statistics (2020a). 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 2, 2020.
- Office for National Statistics (2020b). Statistical bulletin: Deaths registered weekly in England and Wales, provisional: week ending 14 August 2020. Available at <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending14august2020>. Accessed on January 21, 2021.
- Office for National Statistics (2021). Deaths registered weekly in England and Wales, provisional. Available at <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales>. Accessed on January 12, 2021.
- 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 dataset to assess mobility changes in Italy following national lockdown. *Scientific Data*, 7(1):230.
- Perrotta, D., Grow, A., Rampazzo, F., Cimentada, J., Del Fava, E., Gil-Clavel, S., and Zagheni, E. (2020). Behaviors and attitudes in response to the COVID-19 pandemic: Insights from a cross-national Facebook survey. *medRxiv*.
- Pullano, G., Domenico, L. D., Sabbatini, C. E., Valdano, E., Turbelin, C., Debin, M., Guerrisi, C., Kengne-Kuetche, C., Souty, C., Hanslik, T., Blanchon, T., Boëlle, P.-Y., Fignon, J., Vaux, S., Campèse, C., Bernard-Stoecklin, S., and Colizza, V. (2020). Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature*.
- 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.
- Ruktanonchai, N. W., Floyd, J., Lai, S., Ruktanonchai, C. W., Sadilek, A., Rente-Lourenco, P., Ben, X., Carioli, A., Gwinn, J., Steele, J., et al. (2020). Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science*.

- Ruktanonchai, N. W., Ruktanonchai, C. W., Floyd, J. R., and Tatem, A. J. (2018). Using Google Location History data to quantify fine-scale human mobility. *International Journal of Health Geographics*, 17(1):28.
- Schlosser, F., Maier, B. F., Jack, O., Hinrichs, D., Zachariae, A., and Brockmann, D. (2020). COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences*, 117(52):32883–32890.
- StatCounter (2020). Mobile operating systems’ market share worldwide from January 2012 to December 2019. *Statista*. Retrieved July 24, 2020. Available at <https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/>.
- 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., Elstrand, 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 (2021). Coronavirus disease 2019 (COVID-19): Weekly epidemiological update, 29 December 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>. Accessed on January 21, 2021.
- 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 Additional Results

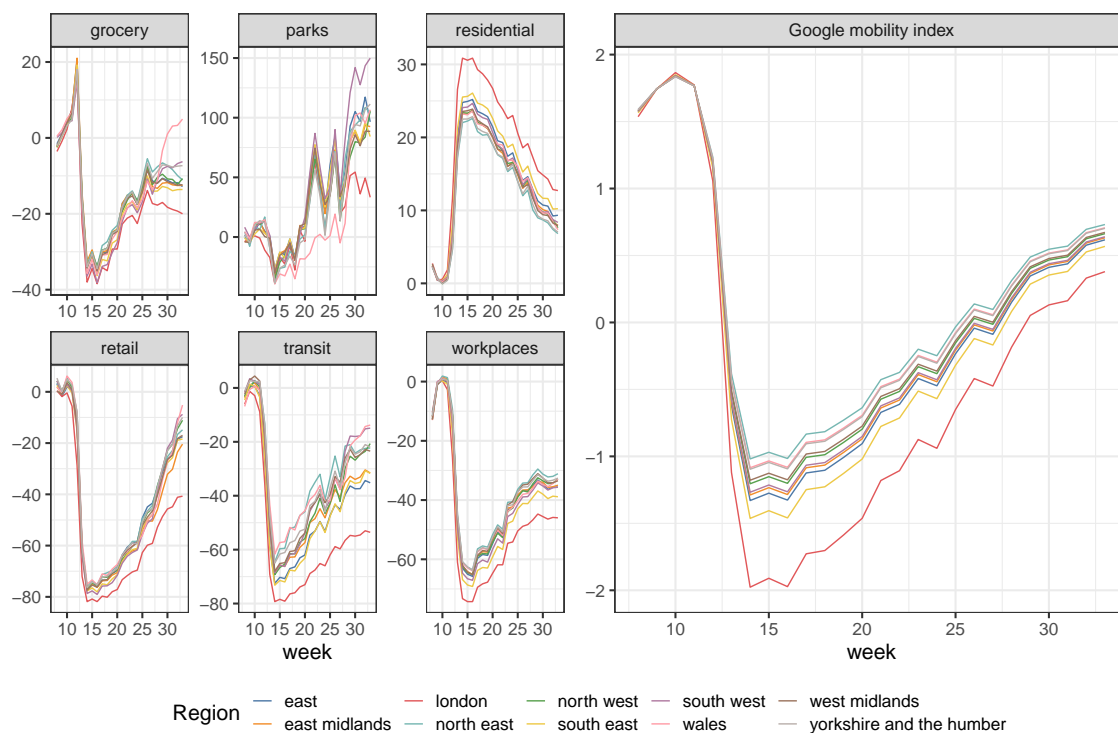
In this Appendix, we report additional results of our analysis. We start by presenting the results of computing the excess mortality rate using the GAM. Figure A.1 shows the results of this approach for two specific regions, London and South East. The grey shaded areas in the panels correspond to the first wave of the COVID-19 pandemic analysed in the paper, and for which corresponding death counts are not employed in the estimation of the GAM. The figure shows that the fitted model captures well the seasonal pattern of the mortality data, as well as the peculiarities of the registration of deaths in some specific weeks (i.e. the lower number of registered deaths during holidays, such as bank holidays and the last week of the year). The lower panels display the excess mortality rate for the two regions: despite being characterised 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 period analysed (grey shaded area).



**Figure A.1.** Observed (dots) and fitted (lines) weekly number of deaths (upper panels) and excess mortality rate (per 100,000 individuals, lower panels) in the regions of London and South East for the years 2015–2020. The grey shaded area corresponds to the COVID-19 period analysed in the paper. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021).

Figure A.2 shows the six categories of the Google COVID-19 Mobility Reports (Google LLC, 2021) (GCMR) – residential, workplaces, supermarket and pharmacy (grocery), transit, retail, and parks – and the combined Google mobility index for the ten regions of England and Wales during weeks 8–33 of 2020. In order to derive the combined indicator, we employ multilinear principal component analysis (MPCA) on five categories of the GCMR. We do not include the parks category in the MPCA for several reasons: (i) its pattern is more volatile than those of other categories, (ii) the category is subject to a higher proportion of missing data (see Figure A.9), and (iii) its inclusion decreases the amount of explained Frobenius norm (a measure comparable to the overall variance). The right panel of the figure shows that the combined index captures most of the peculiarities found in the Google mobility data, as the explained Frobenius norm amounts to 86.6%. For example, mobility in the region of London decreased considerably more than in other

regions, and this is well captured by the Google index.

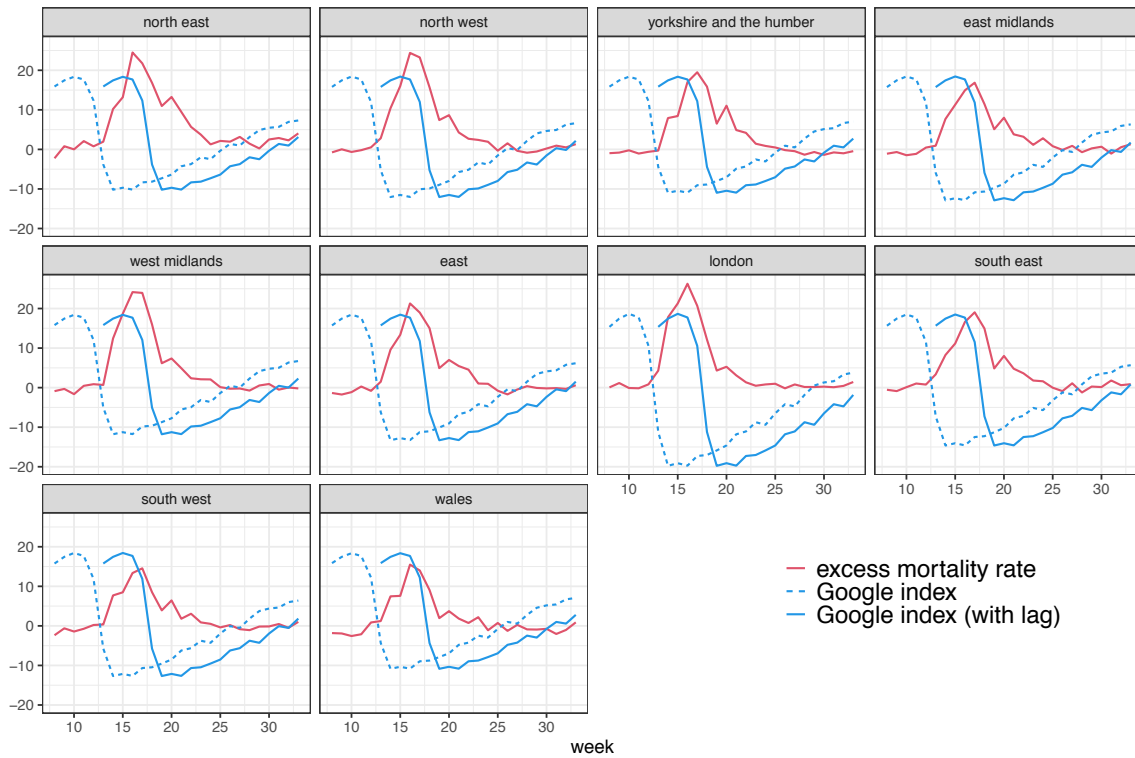


**Figure A.2.** Six categories of the GCMR and their combination into the Google mobility index in England and Wales by region during weeks 8–33 of 2020. *Source:* Authors’ own elaboration based on data from Google LLC (2021).

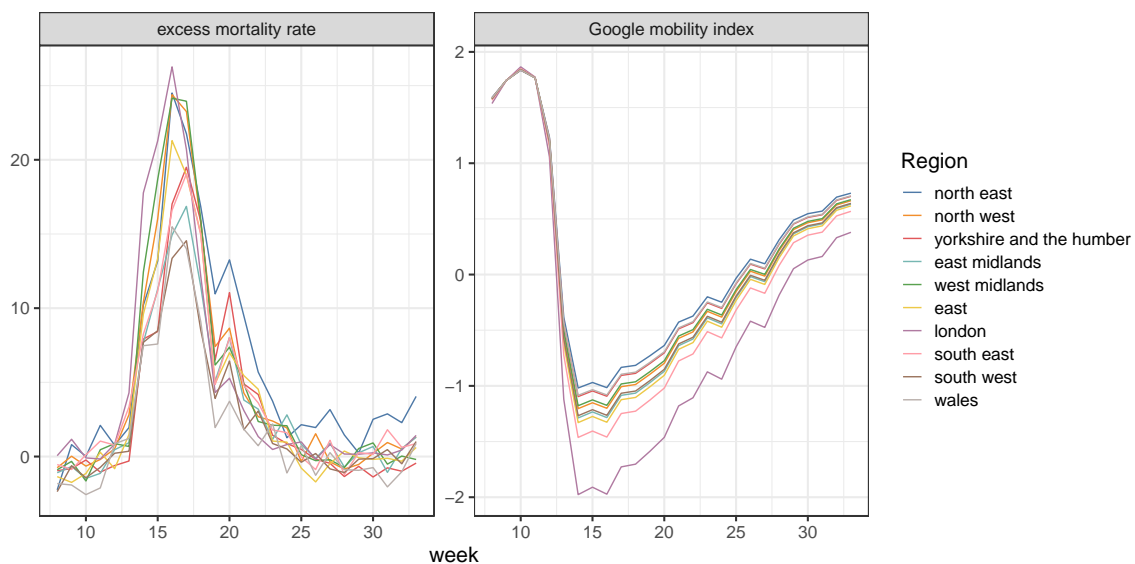
Next, we present the graphs corresponding to Figure 1 for all the regions analysed in our study. Figure A.3 shows the time series of the excess mortality rate (per 100,000 individuals) and the change in the Google mobility index for the ten regions of England and Wales during the weeks 8–33 of 2020. In addition, Figure A.4 shows the time series of the two variables for each of the ten regions side-by-side. Common patterns and regional variations clearly appear from the figure.

Furthermore, we analyse the relationship between the excess mortality and changes in mobility, considering a wide range of possible lags for the latter variable. Figures A.5 and A.6 clearly show that the relationship between excess mortality and mobility varies 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 negative, i.e., a decrease in mobility is associated with a (future) higher excess mortality. However, for longer time lags (greater or equal to five weeks), we can observe a positive relationship between the two variables, i.e., a decrease in mobility is associated with a (future) lower excess mortality. This holds true for the combined Google mobility index as well as for four categories of the GCMR: grocery, retail, transit and workplaces. For the residential category, the relationship is reversed because time spent at home increased following the implementation of the NPIs (see Figure A.2); for the parks category, the relationship is negative for all lags of time.

Next, we analyse the role played by different time lags in the relationship between excess mortality and mobility within our regression approach. Table A.1 shows the mixed-effects regression models between excess mortality rate and changes in the Google mobility index occurred 3, 4, 5, 6 and 7 weeks prior. The relationship between excess mortality and mobility changes according to the time lag considered: for shorter time lags (3 and 4 weeks), the two variables have a negative or null association. For time lags of 5 or more weeks, the relationship becomes positive and significant at the 95% confidence level, and the magnitude of the estimated coefficient reduces as the lag increases.

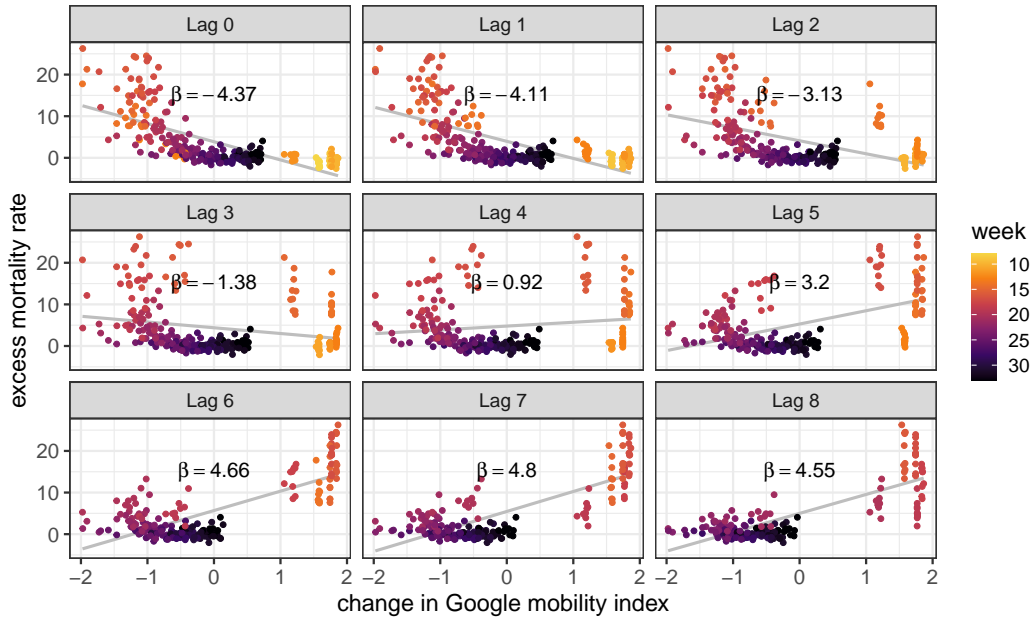


**Figure A.3.** Time series of excess mortality rate per 100,000 individuals (red lines) and change in Google mobility index in week  $t$  (dashed blue lines) and with a five-week forward shift (solid blue lines) in England and Wales by region during weeks 8–33 of 2020. The Google index was multiplied by 10 for illustration purposes. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

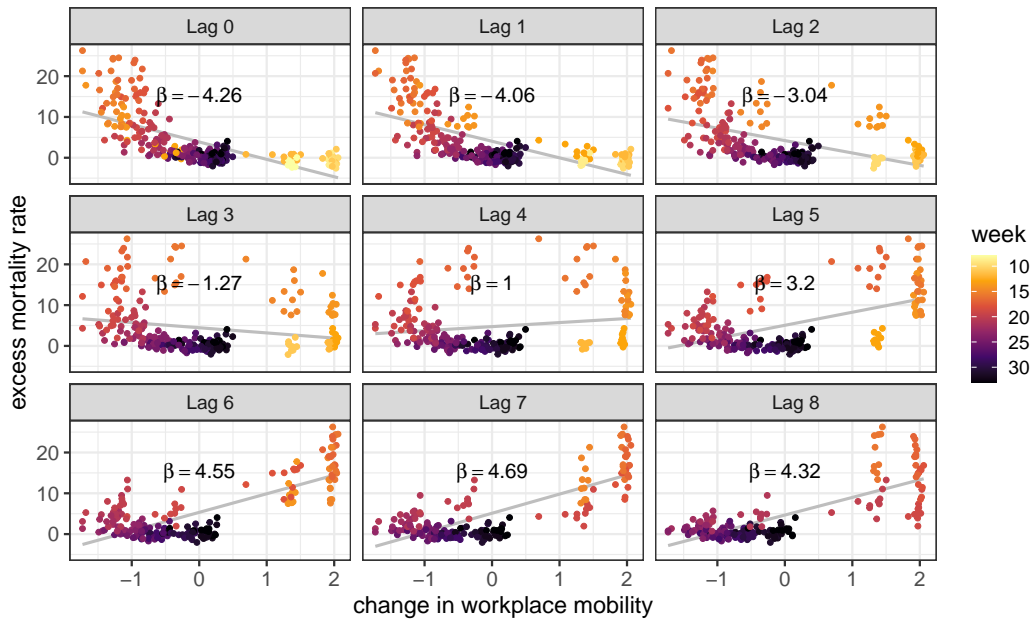


**Figure A.4.** Time series of excess mortality rate per 100,000 individuals and change in Google mobility index in England and Wales by region during weeks 8–33 of 2020. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

Figure A.7 shows, for each region analysed, the fit of the estimated mixed-effects regression model as well as the counterfactual analysis which assumes that mobility did not reduce after the introduction of the NPIs.



**Figure A.5.** Linear relationship (with slope equal to  $\beta$ ) between excess mortality rate (per 100,000 individuals) and change in the Google mobility index in ten regions of England and Wales during weeks 8–33 of 2020, considering different lags of time for mobility data. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

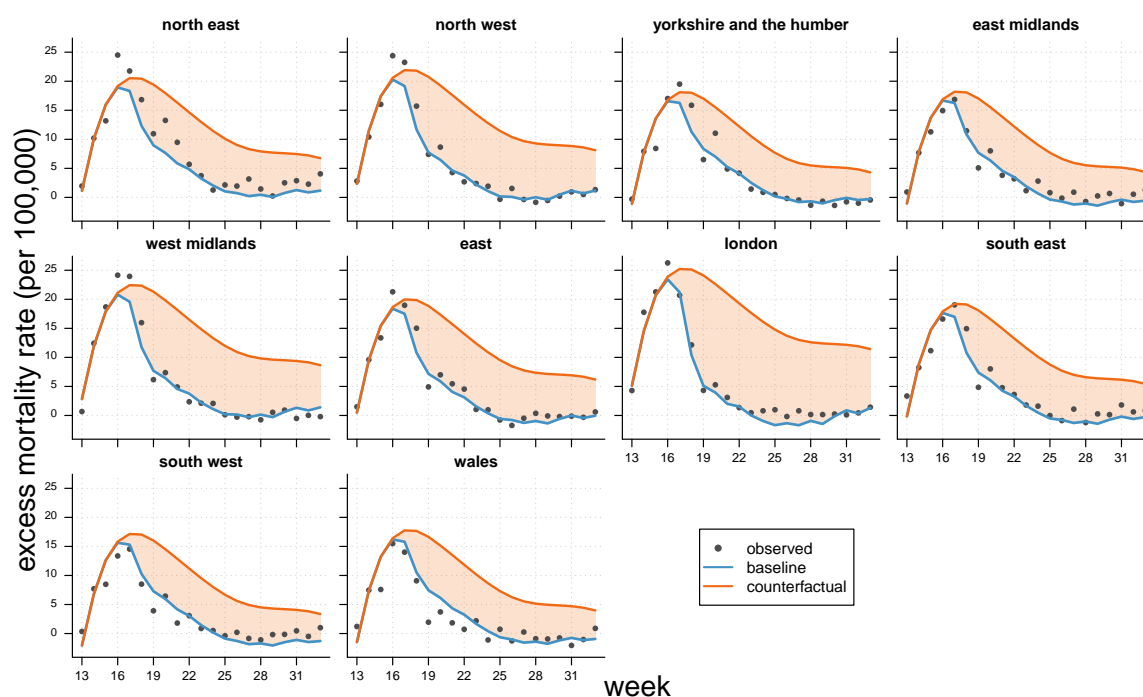


**Figure A.6.** Linear relationship (with slope equal to  $\beta$ ) between excess mortality rate (per 100,000 individuals) and (scaled) change in workplace mobility in ten regions of England and Wales during weeks 8–33 of 2020, considering different time lags for mobility data. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

Figure A.8 compares the GCMR with the mobility data provided by Apple (2020) and Facebook (2020). For the latter two, only country-level data are available for England and Wales (i.e. not at the sub-national level). The figure shows that the six Google categories of mobility indicators are aligned with those of other providers, hence capturing the general mobility patterns throughout

**Table A.1.** Estimated coefficients and 95% confidence intervals of linear mixed-effects regression between excess mortality rate (per 100,000 individuals) and changes in Google mobility index occurred  $x$  weeks before, using five choices of time lags between the two time series. *Note:* AIC and BIC values should not be compared here due to different number of observations, and they have been reported for completeness only. *Source:* Authors' own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

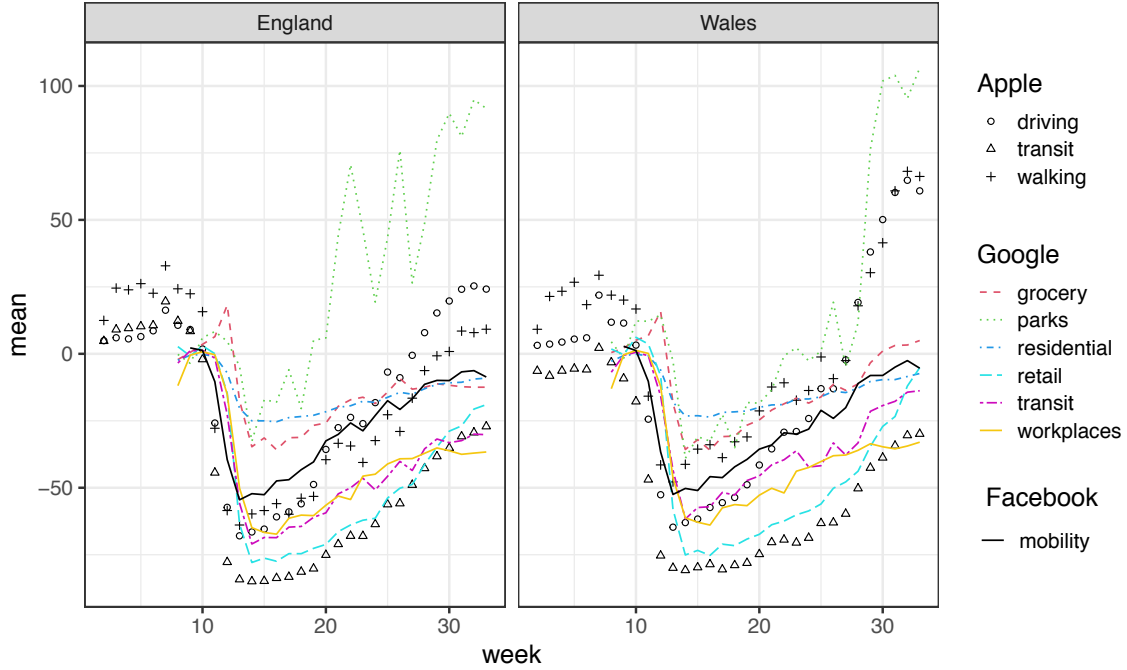
| Dependent variable: excess mortality rate (per 100,000) |                         |                        |                      |                      |                      |
|---|-------------------------|------------------------|----------------------|----------------------|----------------------|
| Linear mixed-effects regression                         |                         |                        |                      |                      |                      |
|   | 3-week lag              | 4-week lag             | 5-week lag           | 6-week lag           | 7-week lag           |
| Fixed effects   |                         |                        |                      |                      |                      |
| Mobility changes<br>$x$ weeks before<br>(95% CI)        | -6.44<br>(-7.80, -5.08) | -1.26<br>(-2.64, 0.12) | 3.75<br>(2.70, 4.81) | 2.73<br>(1.52, 3.94) | 1.26<br>(0.31, 2.20) |
| Random effects<br>(variance)                            |                         |                        |                      |                      |                      |
| Region (intercept)                                      | 1.95                    | 1.46                   | 2.67                 | 2.44                 | 1.48                 |
| Mobility (slope)  | 1.05                    | 0.66                   | 0.66                 | 0.95                 | –                    |
| Residual  | 7.26                    | 7.58                   | 3.35                 | 3.82                 | 7.55                 |
| Observations  | 230                     | 220                    | 210                  | 200                  | 190                  |
| Groups  | 10                      | 10                     | 10                   | 10                   | 10                   |
| Log-Likelihood  | -559.88                 | -539.16                | -437.42              | -428.07              | -462.03              |
| AIC   | 1141.76                 | 1100.32                | 896.84               | 878.15               | 940.05               |
| BIC   | 1179.58                 | 1137.65                | 933.66               | 914.43               | 966.03               |



**Figure A.7.** Observed and estimated excess mortality rate (per 100,000 individuals) from the mixed-effects model in the baseline and counterfactual scenarios for the ten regions in England and Wales during weeks 13–33 of 2020. *Source:* Authors' own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

England and Wales.

Finally, Figure A.9 shows the share of the missing data, computed as percentage of the regional



**Figure A.8.** Comparison of mobility indicators provided by Apple (categories *driving*, *transit* and *walking*), Facebook (category *mobility*) and Google (categories *residential*, *workplaces*, *grocery*, *transit*, *retail* and *parks*) for England and Wales in weeks 1–33 of 2020. *Note:* the sign of the *residential* category of Google is reversed for illustrative purposes. *Source:* Authors’ own elaboration based on data from Apple (2020), Google LLC (2021) and Facebook (2020).

population, for the six categories of the Google mobility data by region and week. For most weeks, regions and categories, missing data are relatively low. Only the parks and residential categories, particularly in the Welsh territory, are characterised by some degree of missing data.

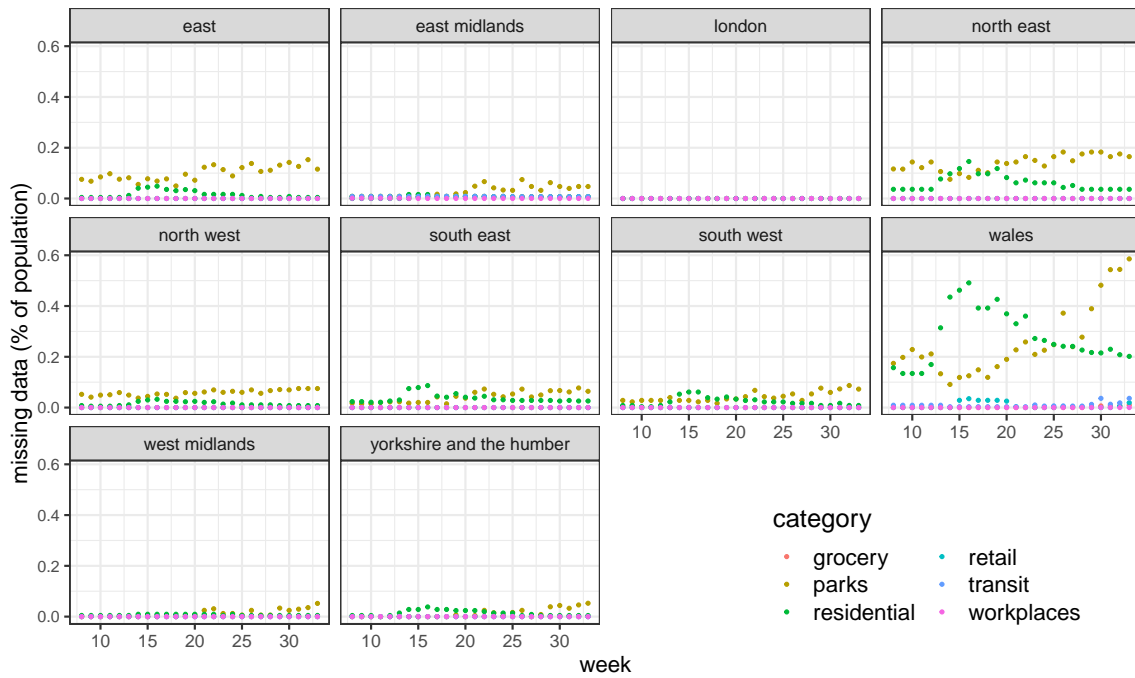
## B Sensitivity Analysis

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

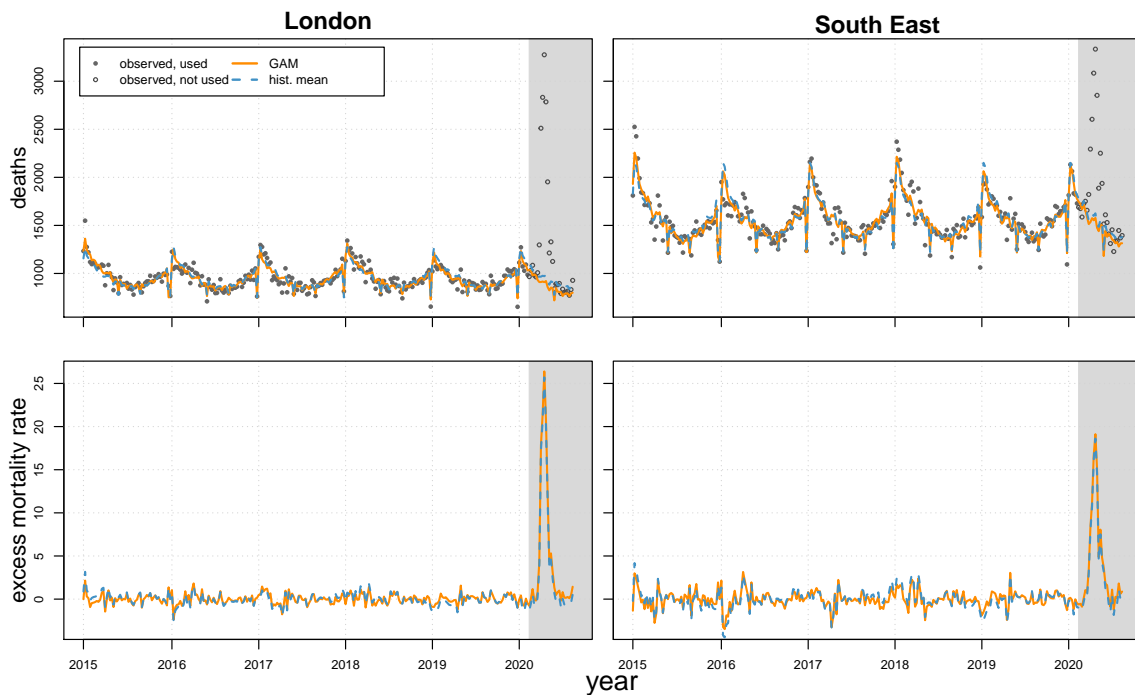
### B.1 Different estimation of excess mortality

The results shown in our paper are robust to the computation of expected deaths  $\hat{d}_t$  derived from the GAM. 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 B.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. In both cases, the decreasing time trend estimated by the GAM predicts a lower number of deaths in the period under study (grey area) as compared to the historical average, resulting in slightly higher excess mortality. Figure B.2 shows the excess mortality rate obtained with the two approaches for all ten regions of England and Wales during the period analysed.

Furthermore, Table B.1 reports the corresponding results of the mixed-effect regression models obtained from using this different computation of the excess mortality rate. Given that the estimates of the excess mortality rates are very similar to those of the GAM, it is not surprising that the results do not change by a great extent when using this alternative computation of excess mortality.

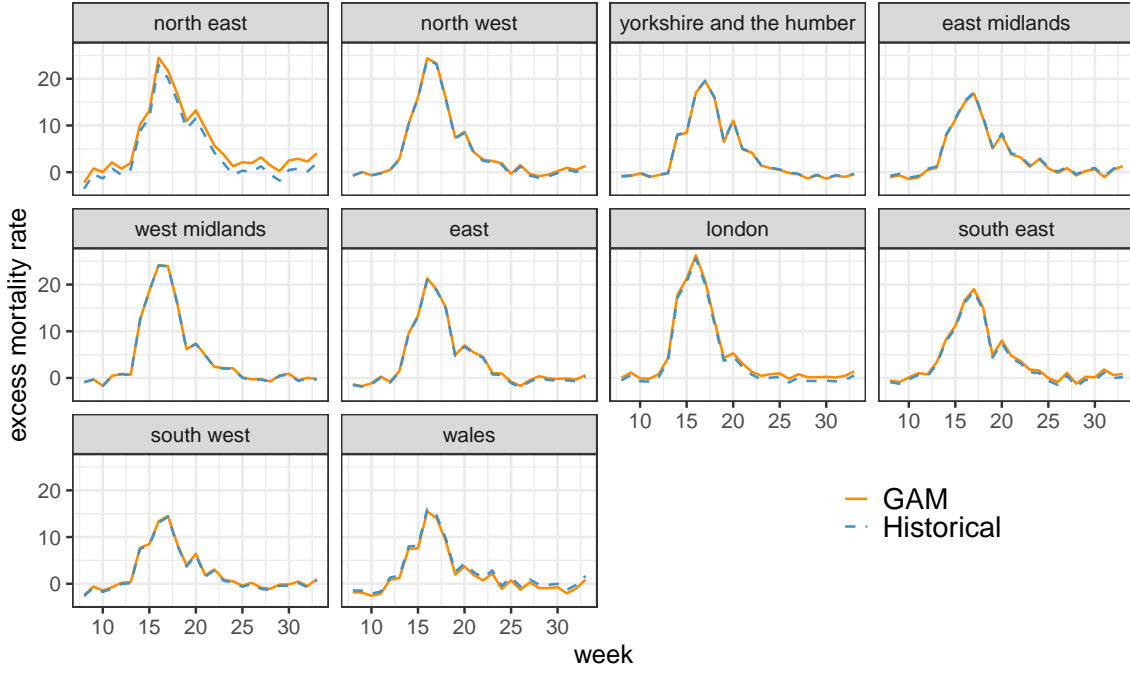


**Figure A.9.** Share of missing population in the GCMR by region, week and category (*residential*, *workplaces*, *grocery*, *transit*, *retail* and *parks*) for ten regions in England and Wales in weeks 8–33 of 2020. *Source:* Authors’ own elaboration based on data from Google LLC (2021).



**Figure B.1.** Observed (dots) and fitted (lines) weekly number of deaths (upper panels) and excess mortality rate (per 100,000 individuals, lower panels) using two different approaches (GAM in orange, weekly historical mean in blue) in the regions of London and South East for the years 2015–2020. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021).





**Figure B.2.** Excess mortality rate (per 100,000 individuals) computed with two different approaches (GAM in orange, weekly historical mean in blue) in the ten regions of England and Wales during the weeks 8–41 of 2020. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021).

**Table B.1.** Estimated coefficients and 95% confidence intervals of linear mixed-effects regression between excess mortality rate (per 100,000 individuals) and changes in mobility occurred five weeks before, measured separately for each model with the combined Google index and the six categories of the GCMR: grocery, workplaces, residential, transit, retail and parks. *Note:* the excess mortality rate (per 100,000 individuals) is computed from the historical average of weekly deaths instead of the GAM employed in Table 1. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

| Dependent variable: excess mortality rate (per 100,000) |                      |                      |                      |                         |                     |                      |                       |
|---|----------------------|----------------------|----------------------|-------------------------|---------------------|----------------------|-----------------------|
| Linear mixed-effects regression                         |                      |                      |                      |                         |                     |                      |                       |
| Fixed effects   | Google index         | grocery              | workplaces           | residential             | transit             | retail               | parks                 |
| Mobility changes<br>5 weeks before<br>(95% CI)          | 3.53<br>(2.53, 4.54) | 2.56<br>(1.94, 3.18) | 3.00<br>(1.97, 4.02) | -3.23<br>(-4.24, -2.23) | 2.62<br>(1.64, 3.6) | 3.63<br>(2.47, 4.79) | 0.62<br>(-0.21, 1.45) |
| Random effects<br>(variance)                            |                      |                      |                      |                         |                     |                      |                       |
| Region (intercept)                                      | 1.47                 | 0.49                 | 1.16                 | 1.82                    | 2.09                | 1.33                 | 0.52                  |
| Mobility (slope)  | 0.66                 | –                    | 0.84                 | 0.60                    | 0.73                | 1.20                 | –                     |
| Residual  | 3.40                 | 4.31                 | 3.61                 | 3.61                    | 3.85                | 3.52                 | 5.66                  |
| Observations  | 210                  | 210                  | 210                  | 210                     | 210                 | 210                  | 210                   |
| Groups  | 10                   | 10                   | 10                   | 10                      | 10                  | 10                   | 10                    |
| Log-Likelihood  | -434.45              | -451.02              | -439.35              | -440.55                 | -447.46             | -437.24              | -477.9                |
| AIC   | 890.89               | 920.05               | 900.71               | 903.09                  | 916.91              | 896.48               | 973.81                |
| BIC   | 927.71               | 950.17               | 937.52               | 939.91                  | 953.73              | 933.3                | 1003.93               |

## B.2 Other sensitivity analyses

The results shown in our paper are further robust with respect to the number of  $B$ -splines employed to describe the time trend of the epidemic. Table B.2 shows the mixed-effects regression models



between excess mortality rate and changes in the Google mobility index occurred 5 weeks prior for five different choice of  $B$ -spline bases. From the table, it emerges that the estimated mobility coefficient does not change to a great extent, taking values between 3.75 and 5.77 according to different number of bases. Moreover, the estimated coefficient is always statistically significant at the 95% confidence level.

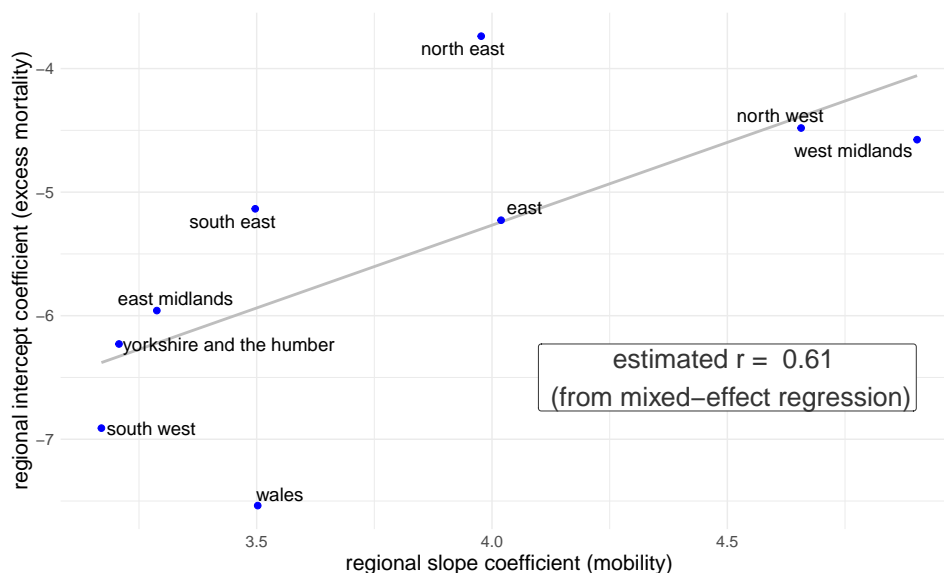
**Table B.2.** Estimated coefficients and 95% confidence intervals of linear mixed-effects regression between excess mortality rate (per 100,000 individuals) and changes in Google mobility index occurred five weeks before, using five different choices of  $B$ -spline bases for describing the time series of the epidemic. *Source:* Authors' own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

| Dependent variable: excess mortality rate (per 100,000) |                      |                      |                     |                      |                     |
|---|----------------------|----------------------|---------------------|----------------------|---------------------|
| Linear mixed-effects regression                         |                      |                      |                     |                      |                     |
|   | 3 $B$ -splines       | 4 $B$ -splines       | 5 $B$ -splines      | 6 $B$ -splines       | 7 $B$ -splines      |
| Fixed effects   |                      |                      |                     |                      |                     |
| Mobility changes<br>5 weeks before<br>(95% CI)          | 5.77<br>(5.03, 6.52) | 4.22<br>(3.48, 4.97) | 3.75<br>(2.7, 4.81) | 5.46<br>(4.36, 6.55) | 4.79<br>(3.77, 5.8) |
| Random effects<br>(variance)                            |                      |                      |                     |                      |                     |
| Region (intercept)                                      | 3.18                 | 2.81                 | 2.67                | 3.20                 | 2.99                |
| Mobility (slope)  | 0.48                 | 0.64                 | 0.66                | 0.60                 | 0.63                |
| Residual  | 5.13                 | 3.17                 | 3.35                | 2.98                 | 2.92                |
| Observations  | 210                  | 210                  | 210                 | 210                  | 210                 |
| Groups  | 10                   | 10                   | 10                  | 10                   | 10                  |
| Log-Likelihood  | -481.06              | -433.97              | -437.42             | -426.27              | -422.71             |
| AIC   | 980.12               | 887.93               | 896.84              | 876.54               | 871.42              |
| BIC   | 1010.24              | 921.4                | 933.66              | 916.71               | 914.94              |

Finally, our results are also robust to the exclusion of the region of London from the analysis. Analysing Figure A.4, London appears as an outlier compared to the other regions since its level of excess mortality and mobility decreased considerably more than in other regions. For this reason, we re-run all our analysis excluding London from the observations employed in our study. Table B.3 shows that the estimated mobility coefficients vary marginally with respect to those estimated in the presence of London, remaining significant at the 95% confidence level. Finally, Figure B.3 shows that the positive correlation between random slopes and random intercepts in the mixed-effects model reduces only slightly when excluding London from the analysis.

**Table B.3.** Estimated coefficients and 95% confidence intervals of linear mixed-effects regression between excess mortality rate (per 100,000 individuals) and changes in mobility occurred five weeks before, measured separately for each model with the combined Google index and the six categories of the GCMR: grocery, workplaces, residential, transit, retail and parks. *Note:* the region of London was removed from the analysis. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).

|  |                      | Dependent variable: excess mortality rate (per 100,000) |                      |                         |                      |                      |                       |       |
|--|----------------------|---|----------------------|-------------------------|----------------------|----------------------|-----------------------|-------|
|  |                      | Linear mixed-effects regression                         |                      |                         |                      |                      |                       |       |
| Fixed effects                                  |                      | Google index  | grocery              | workplaces              | residential          | transit              | retail                | parks |
| Mobility changes<br>5 weeks before<br>(95% CI) | 3.80<br>(2.64, 4.97) | 2.64<br>(1.91, 3.38)                                    | 3.05<br>(1.91, 4.20) | -3.67<br>(-4.91, -2.43) | 3.09<br>(1.93, 4.24) | 3.51<br>(2.36, 4.66) | 0.83<br>(-0.01, 1.66) |       |
| Random effects<br>(variance)                   |                      |   |                      |                         |                      |                      |                       |       |
| Region (intercept)                             | 1.66                 | 1.47  | 1.48                 | 1.52                    | 1.79                 | 1.7                  | 1.55                  |       |
| Mobility (slope)                               | 0.51                 | 0.51  | 0.51                 | 0.48                    | 0.37                 | 0.55                 | –                     |       |
| Residual                                       | 3.53                 | 3.08  | 3.78                 | 3.75                    | 3.91                 | 3.61                 | 4.92                  |       |
| Observations                                   | 189                  | 189   | 189                  | 189                     | 189                  | 189                  | 189                   |       |
| Groups   | 9                    | 9   | 9                    | 9                       | 9                    | 9                    | 9                     |       |
| Log-Likelihood                                 | -395.41              | -383.96   | -401.41              | -399.74                 | -403.92              | -397.42              | -420.07               |       |
| AIC  | 812.82               | 789.93  | 824.83               | 821.47                  | 829.85               | 816.85               | 858.13                |       |
| BIC  | 848.48               | 825.59  | 860.49               | 857.13                  | 865.50               | 852.51               | 887.31                |       |



**Figure B.3.** Estimated region-specific intercepts and mobility coefficients, as well as their estimated correlation  $r$  from the mixed-effects regression in nine regions of England and Wales (excluding London) during weeks 13–41 of 2020. *Source:* Authors’ own elaboration based on data from Office for National Statistics (2020a, 2021) and Google LLC (2021).