Socio-Economic Planning Sciences 78 (2021) 101083

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps

# The socio-economic determinants of COVID-19: A spatial analysis of German county level data

Andree Ehlert

*Harz University of Applied Sciences, Friedrichstr. 57-59, 38855 Wernigerode, Germany*

A R T I C L E I N F O

*Keywords:*
COVID-19
Germany
Socio-economic
Social determinants
Spillover
Spatial analysis
County level

A B S T R A C T

The study explores the association of socioeconomic, demographic, and health-related variables at the regional level with COVID-19 related cases and deaths in Germany during the so-called first wave through mid-June 2020. Multivariate spatial models include the 401 counties in Germany to account for regional interrelations and possible spillover effects. The case and death numbers are, for example, significantly positively associated with early cases from the beginning of the epidemic, the average age, the population density and the share of people employed in elderly care. By contrast, they are significantly negatively associated with the share of schoolchildren and children in day care as well as physician density. In addition, significant spillover effects on the case numbers of neighbouring regions were identified for certain variables, with a different sign than the overall effects, giving rise to further future analyses of the regional mechanisms of action of COVID-19 infection. The results complement the knowledge about COVID-19 infection beyond the clinical risk factors discussed so far by a socio-economic perspective at the ecological level.

## 1. Introduction

Since its outbreak in December 2019 in Wuhan Province, China, the respiratory disease COVID-19 has developed in just a few weeks into a pandemic with currently (12 February 2021) about 107 million infections and about 2.4 million deaths worldwide. Many countries have responded with one or more successive lockdowns (including curfews and school closures), in particular to avoid overloading the health system. The initial exponential growth in case numbers has now slowed in many countries or, in many cases, resulted in multiple separate waves of infection.

In Germany the first confirmed cases were reported at the end of January 2020. The number of new infections per day rose to more than 100 at the beginning of March 2020 and reached a preliminary maximum of about 7000 at the end of March 2020 (what defined the peak of the first wave). The initial cases were mainly attributed to imported clusters spread at major events (e.g. sports, carnival).

The social and economic consequences of the lockdown measures taken in almost all countries to contain the pandemic are serious and their long-term development is not yet foreseeable. Currently, with the exception of a few flagship countries or regions, there is still no general vaccination (worldwide approx. 160 million vaccinated, 12.02.2021) and thus no effective herd immunity. This increases the pressure for empirical evaluation of prevailing contact restrictions and, hence, for the development and implementation of targeted and efficient future political measures. While clinical and epidemiological research is currently discussing person-specific risk factors for infection or probability of survival, which often have a strong ad hoc influence on government lockdown measures, there are so far few results on the question of which socio-economic and region-specific factors (after controlling the initial case numbers of a region) are associated with the spread of COVID-19 at an ecological level.

In order to understand the ecological study approach chosen here and to distinguish it quite clearly from (clinical) studies based on individual data, the following should be emphasized: This study's units of analysis are groups (401 German counties) with the corresponding data aggregated at the group level. Hence, the conclusions derived also apply only at the group level. Therefore, no direct conclusions can be derived at the individual level (this can lead to so-called ecological fallacies). In this regard, there are numerous examples in the literature of factors having a different influence at the group level than at the individual level, see, e.g. Ref. [8] and the literature cited therein.

Nevertheless, ecological studies do have a high relevance and justification, especially around (political) measures that also refer to groups instead of individuals (as is the case with many collective COVID-19 measures). Furthermore, they serve to generate hypotheses about

*E-mail address:* aehlert@hs-harz.de.

https://doi.org/10.1016/j.seps.2021.101083
Received 22 July 2020; Received in revised form 3 May 2021; Accepted 10 May 2021
Available online 14 May 2021

possible mechanisms of action in the sense of exploratory studies before individual-level studies are even available, as is the case worldwide at the onset of the COVID-19 pandemic. To address a classic example of such mechanisms of action: It can be assumed that persons of low socio-economic status are more severely affected by COVID-19. The possible interrelationships are manifold. For example [11], highlights the coincidence of multiple (and thus mutually reinforcing) health risk factors (such as living in crowded conditions, poorer hygienic conditions, jobs that are unfavorable to health with little opportunity for home office, therefore increased use of public transport, etc.) associated with low social status. For comparable ecological studies in influenza and COVID-19, see, e.g. Refs. [1,35,40].

The current hypotheses concerning socio-economic factors influencing COVID-19 at an ecological level are manifold. For example, the average age of the population, regional vaccination rates against tuberculosis, climatic factors, or the way schools and childcare facilities are managed, have recently been discussed in the daily press in terms of their impact on the spread of COVID-19. With this ecological analysis, further socio-economic factors at the level of the 401 counties in Germany (whose infection numbers vary considerably, see Fig. 1) will be analysed with the help of spatial statistical methods.

Besides the explorative character of this analysis (at the current state of research), three questions in particular are to be answered empirically: First, what possible influence do schoolchildren and children in day care have on the occurrence of infection? Second, can the thesis of the spread of COVID-19 via institutions for care of the elderly be empirically confirmed at regional level? Third, what influence do spillover effects, i.e. the interlinked spread of COVID-19 between neighbouring regions, have on infections? Other factors which have not yet been quantitatively investigated, but which can be discussed based on the results of this study include, for example, the question of the influence of the size of the dwelling (regarding the question of frequency of contacts), tourism (supra-regional spread) or the sectoral structure (contacts in the office or on the building site).

To answer these questions, the case numbers and deaths at the district level provided for Germany by the Robert Koch-Institut (RKI) are associated with socio-economic, demographic and health-related data at the same regional level. The analysis period is Germany's first wave of COVID-19 cases and deaths until mid-June 2020. The data are analysed

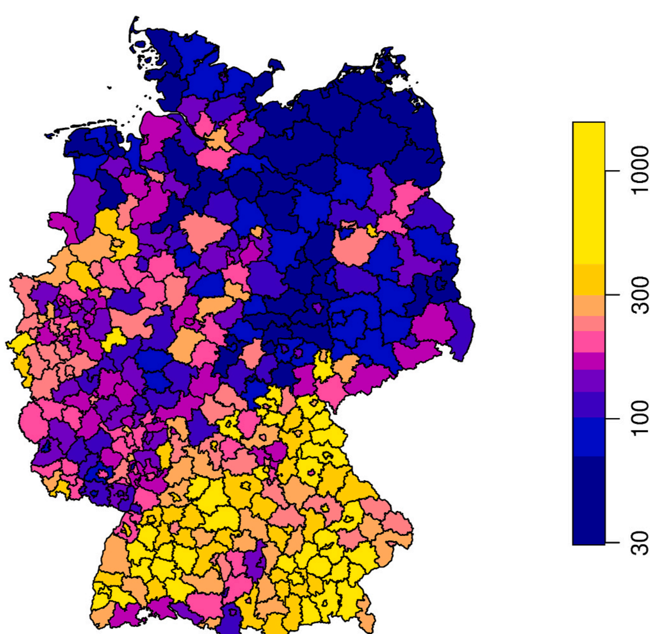


**Fig. 1.** Cumulative infection numbers per 100,000 population from the start of the COVID-19 pandemic to June 15, 2020 (data source: RKI).

by three types of spatial models, the spatial autoregressive model (SAR), the spatial error model (SEM) and the spatial autoregressive combined model (SAC). The spatial position of the regions relative to each other is represented by a weighting matrix. These models can reduce possible bias of an ordinary least squares (OLS) estimation in case of spatial spillovers and/or increase the efficiency of the estimation.

This paper complements the existing literature in at least three areas. Firstly, the literature on COVID-19 tends to focus on clinical studies to identify individual risk factors and develop curative treatments, see e.g. Ref. [43]. However, the contribution of this study lies precisely in the identification of key socio-economic and demographic factors and the possible regional mechanisms of action through which COVID-19 spreads at an ecological level. Examples from the literature on the analysis of socio-economic risk factors of health status are numerous. As a classic example [12], deals with the various, also socio-economic, determinants of individual health. For infectious diseases such as pneumonia and influenza [13], has investigated the influence of socio-economic factors in addition to individual risk. Socio-economic risk factors for HIV were analysed by Ref. [2], for pertussis by Ref. [17], for salmonella infections by Ref. [41] and for H1N1 pandemic mortality in the USA by Ref. [29], to name only some examples.

Secondly, socio-economic risk factors in relation to COVID-19 have so far been analysed rather for regions outside Germany or in an international comparison of countries. Examples are [24] for 20 European countries, which points to a positive influence of high social activity and high population density on COVID-19 infections, or [35,36] for a worldwide analysis of risk factors based on country data. Examples of regional analyses in Spain are [25], for an analysis at municipal level within Catalonia [23], for France [14], for Iran [32], for China [30], or [4] for Italy, to name a few. For the analysis at the district level within Germany [28], finds significantly negative effects especially on income and education. In a spatial regression discontinuity analysis [3] investigate the influence of the tuberculosis vaccination along the former inner-German border under consideration of numerous socio-economic factors at the district level. A time series approach taking into account regional factors such as age and population density was carried out by Ref. [22] and aims to explain the decline in infection rates in the later phase of the first COVID-19 wave.

Thirdly, however, in the above analyses (with the exception of [5]) no spatial econometric models (and hence no possibility for spatial spillover effects) have been considered. Outside Germany, similar approaches have been applied by Ref. [34] for spatial models (without spillover effects) at the European country level. For 31 regions in China [15] analyses the spatial spillover effects for the COVID-19 case numbers, but without considering socio-economic covariates.

The rest of the paper is structured as follows. Section 2 details the data and methods used. Results are presented in Section 3. Section 4 provides a discussion and concludes.

## 2. Methods

### 2.1. The data

Our analysis is a retrospective ecological study on the level of the 401 administrative districts and district-free cities (counties) in Germany, whose population ranges from about 34,000 (Zweibrücken) to about 3,664,000 (Berlin). It should be emphasized again that this is not a clinical study at the level of single individuals, so that accordingly statements can only be made at the population level. However, this study design is well established in the health sciences and provides very practice-oriented results, especially for questions in the areas of health services research, public health and epidemiology. See also [21] for a detailed discussion.

As outcome variables the case numbers and deaths at the district level, which are currently updated daily by the RKI, are used. The cut-off date for this analysis is 15. June 2020. The data are based on reports from the local health authorities to the RKI. As officially recorded COVID-19 cases, the figures are a lower limit of the actual number of cases present (dark figure).

The socio-economic, demographic and health-related covariables at the district level come from two public sources, namely the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR) and the Central Research Institute of Ambulatory Health Care in Germany (ZI). Table 1 provides an overview of the data.

In selecting the covariates, we were guided by existing ecological studies on respiratory diseases and the motivation for the relevance of the corresponding variables discussed there. For example, the influence of income and education variables as well as the region type on COVID-19 infections is discussed in [28]. While the motivation for the first two variables is widely accepted in the literature as classic determinants of health status, the region type in this context can provide further evidence on the accessibility or remoteness of a region and the associated transport flows, in addition to mere population density. Importantly, *region type* is classified according to the presence of cities and the proportion of the population living there, which will certainly have an impact on the spread of COVID-19 at the ecological level. For many regions in Germany, there are significant differences for this indicator compared to mere population density.[1] For a motivation of the employment level on pandemic events, see e.g. [20]. Demographic variables, population density and information on health status are discussed in Ref. [3]. The physician density is brought into play in Ref. [30] in connection with COVID-19 infections in China. Note also that a very similar selection and grouping of socio-economic factors (but at the level of nation states) on COVID-19 infections is used in [35].

In general, the actual mechanisms of action behind the socio-economic factors are complex. For example, they can act simultaneously at the individual and population level and their direction is often unclear a priori. In the case of the economic status of a region, for example, the better infrastructure available to bridge a lockdown or further contact restrictions can be seen as an effect that slows down the spread of infection. The same applies to better medical treatment. On the other hand, a region's higher economic output generally increases the networking and thus the frequency of contact between individuals, which can have an increasing effect on the number of infections. Also, with higher economic performance, more resources are available for testing for COVID-19, so that this alone can also result in a positive association.

However, some rather uncommon variables that we have included in Table 1 should be briefly motivated explicitly. In addition to the classical economic indicators, the share of *academics* serves as a proxy for higher education and an indication of the economic progressiveness and possibly openness of a region. The literature also pays attention to the faster diffusion of knowledge among academics, which could facilitate mitigation strategies in the wake of COVID-19. On the other hand, social (cross-regional) interaction among academics may also be assumed to be higher [38], which in turn accelerates an infection event. The direction of effect is unclear a priori, leaving the *academics* variable all the more interesting for our analysis.

The *industry share* of the workforce was included as a plausible indicator for the possibility of working from a home office, which is often lacking in industry, and which would then favor the occurrence of infection. A more international perspective on the same question is answered by the *immigration for humanitarian reasons* variable. In addition to the mere proportion of foreigners, this variable allows a distinction to be made between highly qualified (and therefore more likely to work flexibly and online) foreigners and foreigners who are in Germany for humanitarian reasons and either perform low-skilled work (under conditions that are unfavorable from a pandemic perspective) or are often housed in collective accommodations, which also facilitate the

**Table 1**
Basic sample characteristics for outcomes and county-level covariates.

| Indicator | Definition | Year | Mean | SD | N | Source |
|---|---|---|---|---|---|---|
| COVID-19 | | | | | | |
| Case numbers | As of 15. June 2020 | 2020 | 464.99 | 656.15 | 401 | RKI |
| Early cases | As of 08. March 2020 | 2020 | 2.63 | 15.22 | 401 | RKI |
| Death count | As of 15. June 2020 | 2020 | 21.92 | 30.29 | 401 | RKI |
| Economy | | | | | | |
| HH income | Average monthly household income in EUR per inhabitant | 2017 | 1872.56 | 215.76 | 401 | BBSR |
| Unempl. rate | | 2017 | 5.36 | 2.41 | 401 | BBSR |
| Pers. services | Share of employees in personal services | 2017 | 23.87 | 4.65 | 401 | BBSR |
| Pers. nursing | Personnel in nursing homes per 10,000 inhabitants | 2017 | 97.71 | 23.28 | 401 | BBSR |
| Academics | Share of employed academics | 2017 | 13.06 | 6.20 | 401 | BBSR |
| Industry share | Employees in industry per 100 working-age population | 2017 | 18.25 | 8.72 | 401 | BBSR |
| Service share | Employees in the service sector per 100 working-age population | 2017 | 39.24 | 14.84 | 401 | BBSR |
| Tourist beds | Beds in tourist accomodation per 1000 inhabitants | 2017 | 41.78 | 49.31 | 401 | BBSR |
| Old age pov. | Percentage of the population with basic old-age provision in the population 65 years and older in promille | 2017 | 22.37 | 14.63 | 401 | BBSR |
| Demography | | | | | | |
| Mean age | | 2017 | 44.54 | 1.96 | 401 | BBSR |
| Over 75 | Population 75 years and older | 2017 | 11.72 | 1.62 | 401 | BBSR |
| Women | Share of women | 2017 | 50.60 | 0.64 | 401 | BBSR |
| Health | | | | | | |
| Heart failure | Proportion of heart failure patients in relation to all patients (covered by statutory health insurance) | 2017 | 3.85 | 1.42 | 401 | ZI |
| COPD | Proportion of COPD (chronic obstructive pulmonary disease) patients in relation to all patients over 40 years (covered by statutory health insurance) | 2017 | 6.46 | 1.50 | 401 | ZI |
| Physicians | Physicians per 10,000 inhabitants | 2017 | 14.59 | 4.41 | 401 | BBSR |
| Hospital beds | Hospital beds per 1000 inhabitants | 2016 | 6.35 | 3.89 | 401 | BBSR |
| Pharmacies | Pharmacies per 100,000 inhabitants | 2017 | 27.00 | 4.90 | 401 | BBSR |
| Infant mortality | Share of deaths under 1 year per 1000 inhabitants under 1 year | 2017 | 3.31 | 1.76 | 401 | BBSR |
| Life expectancy | | 2017 | 80.66 | 1.01 | 401 | BBSR |
| Need of care | Persons in need of care per 10,000 inhabitants | 2017 | 428.13 | 106.03 | 401 | BBSR |
| Region | | | | | | |
| Pop. density | | 2017 | 533.75 | 702.71 | 401 | BBSR |
| Car density | Cars per 1000 inhabitants | 2017 | 579.16 | 70.98 | 401 | BBSR |
| Commuter balance | Commuter balance (in minus out) per 100 employees subject to social insurance at the place of work | 2017 | −10.36 | 29.72 | 401 | BBSR |
| Share of foreigners | | 2017 | 10.03 | 5.15 | 401 | BBSR |
| Immigr. human. Reasons | Proportion of humanitarian refugees in the population | 2017 | 1.88 | 1.14 | 401 | BBSR |
| Multi-fam. houses | Share of apartments in multi-family houses | 2017 | 45.83 | 19.39 | 401 | BBSR |
| Region type | 1 = urban regions. 2 = regions with urbanisation | 2017 | 2.05 | 0.79 | 401 | BBSR |

(*continued on next page*)

[1] See https://www.bbsr.bund.de for details.

**Table 1** (*continued*)

| Indicator | Definition | Year | Mean | SD | N | Source |
|---|---|---|---|---|---|---|
| | approaches. 3 = rural regions | | | | | |
| Pupils | Pupils per 100 inhabitants | 2017 | 10.12 | 1.50 | 401 | BBSR |
| Childcare | Share of children under 3 years of age in day-care facilities among children in the corresponding age group | 2017 | 32.27 | 12.08 | 401 | BBSR |

spread of the virus.

To check that the multiple indicators (especially in the health domain) do not lead to a pronounced multicollinearity in the model, we also examined the variables in Table 1 for correlations, which, however, turn out to be rather low overall. Only some variables show correlations greater than 0.8, namely *physicians* and *service share* (0.86), *commuter balance* and *service share* (0.83), and *over 75* and *mean age* (0.92). Among the health variables, only *hospital beds* and *physicians* have a value of a similar magnitude (0.79).

### *2.2. Econometric methods*

In contrast to a standard linear model, in our analysis the spatial distance of the observation units (counties) from each other will be taken into account. Spatial statistical models (see below) may then reflect the fact that outcomes in one area are affected by outcomes in neighbouring regions (spatial spillover effects) and/or a spatial autocorrelation of the residuals. Common to all spatial models is the description of the neighbourhood relations via a so-called spatial weighting matrix (i.e. a symmetrical $N \times N$ matrix). For this purpose, the geocodes (longitude and latitude of the centre of the districts) provided by the provider Opendatasoft (under the creative commons licence) were used.[2] The command spmatrix in Stata/MP 16.1 was used to create an inverse distance matrix from the coordinates, so that according to Tobler's law [39] regions closer to each other receive a higher weight.

There are then essentially two approaches to spatial autocorrelation in the dependent variable or error terms, namely SAR and SEM, as well as numerous combinations and variations thereof. The technical details shall be omitted here, with reference to the excellent presentation in Ref. [9]. See also [15] for an application of these models in the context of COVID-19. It is important, however, to note that in the presence of the SAR model the estimates of an OLS model may be biased and that the true effect of an independent variable is reflected by direct and indirect spillover effects (and, together, total effects). We will come back to this when discussing the results in Section 3. In a SEM model, there are no spatial spillover effects, but due to the spatial structure in the error terms, simple OLS estimation might be inefficient. We will consider a third model in our analysis, the so-called SAC, as a combination of spatial autocorrelation in the dependent variables and in the errors, and refer to Ref. [9] for the technical details.

## 3. Results

### *3.1. Case numbers*

The econometric analysis was performed with the command spregress in Stata/MP 16.1. As a first step, it should be checked whether significant spatial dependency exists at all compared to an OLS model and which of the three proposed models best describes the data. Therefore, as suggested in Ref. [9], a likelihood ratio (LR) test of the models against an OLS model is performed. The results are given in the penultimate line of Table 2. Apparently the OLS model can be rejected against all three spatial models to the 1% level. Further the SAR model can be rejected (10% level) against the SAC (which are nested models by design).

The individual independent variables are grouped in the left column of Table 2. The estimated values with the corresponding standard errors in brackets are shown in the following columns for the respective models. Dummies for the 16 federal states were included in the model (estimates not shown) to check for the potential influence of country-specific lockdown measures and their different time frames.[3]

The estimated coefficients in Table 2 show a rather uniform picture across the three spatial models with regard to their significance and sign. In fact, there are only two variables, *women* and *childcare*, for which there are differences in significance (at least at the 10% level) between the three models: *Childcare* is only (negatively) significant in the SAR model, while *women* is only (positively) significant in the SEM and SAC models. In detail, it can be seen that, as expected, the early case numbers turn out positively significant (1% level). With regard to the economic variables, in contrast to what is expected (according to the recent media discussion), no significant negative influence of the income variable and no significant positive influence of the unemployment rate can be found. Here, it is also important to point out once again that the results of an ecological study do not necessarily have to correspond with the process of infection at the individual level. However, similar results were reported in Refs. [23,35], for example, and (in contrast to the prevalence of long-term widespread diseases) were attributed to the fact that the spread of the virus is favoured by high economic activity.

Interesting with regard to our second research question (cf. Section 1) is especially the positive influence of employees in the nursing professions across all three models (5% level). Apparently, even after controlling for the number of people in need of care (see further down in the table, not significant), infection numbers seem to be influenced by the number of people employed in care, which would confirm the recent media discussion empirically consistent. On the other hand, the tourist bed sector cannot be significantly associated with infections in our models, although numerous channels of action are conceivable even after the lockdown measures have come into effect. It is possible, however, that these have already manifested themselves in the early case numbers and are therefore largely covered by this variable.

With regard to the employment structure, it is striking that, in contrast to the industry and service share, only the share of academics has a significant (1% level) positive association with infections (across all three models). This may indicate higher economic activity in the respective counties (although the income variable itself remains insignificant) or also increased social contacts favoring the spread of the virus, see also [38]. The result is in line with [35], who reports medium evidence for social connectedness and economic development as determinants of COVID-19 infections at the country level.

The demographic variables reflect the fact that a rising average age has a significantly positive (at least 10% level) influence on the number of cases, whereas the proportion of persons over 75 years seems to have a significantly negative (5% level) influence. This seems to contradict the media presentation that especially elderly people are affected by COVID-19 infections (which seems to apply even to the death figures with regard to Table 3, see below). However, this result can be empirically confirmed for example by Ref. [23] for Catalonia. As a possible explanation, it can again be cited that a high proportion of older persons tends to indicate a regionally lower level of economic and social activity, thus limiting the spread of the virus.

In the area of health indicators it is noticeable that the prevalence of

[2] Called up on 11.06.2020 at https://public.opendatasoft.com/explore/?sort=modified.

[3] See e.g. https://www.bundesregierung.de/breg-de/themen/coronavirus/corona-bundeslaender-1745198.

**Table 2**
COVID-19 case numbers: Coefficient estimates for the SAR, SEM and SAC models discussed in Section 2.2.

| Variable | SAR | | SEM | | SAC | |
|---|---|---|---|---|---|---|
| | Coeff. | (Std. Err.) | Coeff. | (Std. Err.) | Coeff. | (Std. Err.) |
| Early cases | 6.78** | (1.24) | 6.58** | (1.23) | 6.50** | (1.22) |
| Economy | | | | | | |
| HH income | 0.22 | (0.17) | 0.14 | (0.17) | 0.18 | (0.17) |
| Unempl. rate | −22.35 | (22.91) | −30.24 | (22.84) | −28.42 | (22.81) |
| Pers. services | −2.96 | (7.90) | −6.65 | (7.96) | −6.44 | (8.00) |
| Pers. nursing | 2.97* | (1.45) | 2.91* | (1.43) | 3.03* | (1.42) |
| Academics | 28.77** | (9.14) | 27.72** | (9.12) | 28.11** | (9.09) |
| Industry share | −0.27 | (3.95) | 0.37 | (3.88) | 0.06 | (3.87) |
| Service share | 3.85 | (4.17) | 2.31 | (4.07) | 2.40 | (4.05) |
| Tourist beds | −0.67 | (0.56) | −0.57 | (0.56) | −0.64 | (0.56) |
| Old age pov. | 4.12 | (3.78) | 4.81 | (3.74) | 4.49 | (3.74) |
| Demography | | | | | | |
| Mean age | 94.82† | (54.85) | 101.74† | (53.93) | 106.05* | (53.83) |
| Over 75 | −117.55* | (52.88) | −127.96* | (52.65) | −129.65* | (52.49) |
| Women | 70.64 | (45.65) | 85.73† | (45.36) | 83.25† | (45.24) |
| Health | | | | | | |
| Heart failure | 15.29 | (24.67) | 10.60 | (24.79) | 10.21 | (24.74) |
| COPD | −21.46 | (20.98) | −13.58 | (21.68) | −14.02 | (21.67) |
| Physicians | −33.73** | (12.92) | −37.47** | (12.83) | −36.45** | (12.81) |
| Hospital beds | −7.82 | (9.50) | −3.69 | (9.31) | −4.28 | (9.32) |
| Pharmacies | 2.98 | (6.33) | 6.46 | (6.48) | 5.09 | (6.58) |
| Infant mortality | −4.36 | (11.06) | −3.57 | (10.84) | −3.25 | (10.80) |
| Life expectancy | 140.42** | (40.79) | 118.54** | (40.92) | 125.50** | (41.20) |
| Need of care | 0.44 | (0.42) | 0.39 | (0.43) | 0.39 | (0.43) |
| Region | | | | | | |
| Pop. density | 0.34** | (0.07) | 0.36** | (0.07) | 0.37** | (0.07) |
| Car density | −0.42 | (0.61) | −0.35 | (0.60) | −0.37 | (0.60) |
| Commuter balance | 0.52 | (1.69) | 0.37 | (1.70) | 0.30 | (1.71) |
| Share of foreigners | −0.57 | (10.88) | −5.75 | (10.85) | −2.85 | (11.01) |
| Immigr. human. reasons | 8.59 | (21.25) | 8.64 | (20.76) | 8.20 | (20.68) |
| Multi-fam. houses | −2.98 | (3.68) | −1.38 | (3.72) | −1.61 | (3.72) |
| Regions urban. appr. | −13.50 | (63.12) | −49.40 | (66.35) | −50.08 | (66.31) |
| Rural regions | 28.28 | (78.19) | −4.35 | (81.83) | −8.42† | (81.93) |
| Pupils | −36.05 | (19.50) | −39.19* | (18.67) | −35.43† | (18.78) |
| Childcare | −9.04† | (4.84) | −6.69 | (5.03) | −7.65 | (5.10) |
| Interactions (centered) | | | | | | |
| Childcare × mean age | 3.35** | (1.30) | 2.97* | (1.30) | 3.03* | (1.30) |
| Pupils × mean age | 19.04* | (8.93) | 18.57* | (8.53) | 18.15* | (8.50) |
| Nursing × mean age | 0.57 | (0.47) | 0.42 | (0.47) | 0.47 | (0.47) |
| Constant | −16728.64** | (4347.27) | −15734.84** | (4307.13) | −16324.19** | (4323.41) |
| lambda | −1.04** | (0.33) | | | −0.69 | (0.49) |
| rho | | | 1.03** | (0.03) | 1.06** | (0.06) |
| sigma2 | 121389.73** | (8594.28) | 119734.46** | (8466.83) | 118921.29** | (8417.15) |
| Log likelihood | −2917.81 | | −2917.28 | | −2916.17 | |
| LR chi2 (OLS) | 9.51** | | 10.56** | | 12.78** | |
| LR chi2 (SAC) | 3.27† | | 2.23 | | | |

The significance level symbols are † for 10%, ∗ for 5%, and ∗∗ for 1%.

widespread diseases such as COPD and heart failure has no significant influence on the number of cases on an ecological level. The fact that this ecological result cannot, of course, be transferred to the individual level in individual cases is only repeated here, cf [42]. for first results of a significant positive significant association of a severe course of COVID-19 with COPD at individual level. However, a high regional physician density seems to have a significantly reducing effect on the number of cases. This may indicate, for example, an early implementation of quarantine and hygiene measures favoured by the presence of physicians. This correlation apparently does not apply to hospital beds and pharmacies.

With regard to regional characteristics, as expected, population density appears to have a significant (1% level) impact on infections, which means empirical confirmation of common epidemiological models of virus spread and also corresponds to public perception. Note that the region type itself is insignificant (with urban regions as base category). Here, no further significant association seems to emerge when looking at the housing situation in multi-family houses. Nevertheless, the result is important, as it seems to exclude a possible channel, at least in this empirical snapshot, at the ecological level. The same applies to car density (also as a proxy for air quality), commuter balance and the proportion of foreigners. The central result in this area for our research question No. 1, however, are the significantly negative coefficients for the share of schoolchildren as well as the childcare rate for infants (the latter is significant only in the SAR model). In this ecological study (in contrast to the predominant public discussion), the density of pupils seems to have not only no positive, but even a negative influence on infections.

In order to analyze in more detail a possible fallacy regarding the average age factor in relation to the above result for schools and day care (in the sense that more children are simply negatively associated with a region's average age), interaction terms were included in the analysis in the lower part of Table 2. These are centered and should be interpreted as follows (referring to the excellent presentation in Ref. [27] for details): For regions with a mean average age (i.e., a mean with respect to the 401 regions), the just discussed values for childcare and pupils given in Table 2 reflect the marginal effects. However, for regions with an above mean average age, significant interactions emerge, i.e. pupils and childcare have a different marginal effect depending on the average age of the region. For example, for each additional year that the region is

**Table 3**
COVID-19 death count: Coefficient estimates for the SAR, SEM and SAC models discussed in Section 2.2. The significance level symbols are as before.

| | SAR | | SEM | | SAC | |
|---|---|---|---|---|---|---|
| Variable | Coeff. | (Std. Err.) | Coeff. | (Std. Err.) | Coeff. | (Std. Err.) |
| Early cases | 0.29** | (0.07) | 0.27** | (0.07) | 0.27** | (0.07) |
| Economy | | | | | | |
| HH income | −0.00 | (0.01) | −0.00 | (0.01) | 0.00 | (0.01) |
| Unempl. rate | −0.28 | (1.39) | −0.57 | (1.38) | −0.51 | (1.37) |
| Pers. services | −0.31 | (0.48) | −0.51 | (0.48) | −0.41 | (0.47) |
| Pers. nursing | 0.28** | (0.09) | 0.30** | (0.09) | 0.30** | (0.09) |
| Academics | 1.13* | (0.55) | 1.14* | (0.55) | 1.07† | (0.55) |
| Industry share | 0.00 | (0.24) | −0.01 | (0.23) | −0.01 | (0.23) |
| Service share | 0.07 | (0.25) | 0.04 | (0.25) | −0.00 | (0.24) |
| Tourist beds | −0.03 | (0.03) | −0.03 | (0.03) | −0.03 | (0.03) |
| Old age pov. | 0.32 | (0.23) | 0.35 | (0.23) | 0.35 | (0.23) |
| Demography | | | | | | |
| Mean age | 5.70† | (3.33) | 5.44† | (3.26) | 6.16† | (3.23) |
| Over 75 | −6.71* | (3.21) | −6.64* | (3.18) | −6.94* | (3.17) |
| Women | 4.10 | (2.77) | 4.82† | (2.74) | 4.91† | (2.72) |
| Health | | | | | | |
| Heart failure | 0.31 | (1.50) | −0.05 | (1.50) | −0.52 | (1.52) |
| COPD | −0.26 | (1.27) | 0.04 | (1.31) | −0.05 | (1.33) |
| Physicians | −1.65* | (0.78) | −2.02** | (0.78) | −2.06** | (0.78) |
| Hospital beds | −0.72 | (0.57) | −0.40 | (0.56) | −0.43 | (0.54) |
| Pharmacies | −0.08 | (0.38) | 0.00 | (0.39) | −0.17 | (0.39) |
| Infant mortality | 0.53 | (0.67) | 0.52 | (0.66) | 0.50 | (0.64) |
| Life expectancy | 7.11** | (2.47) | 6.07* | (2.47) | 6.79** | (2.46) |
| Need of care | −0.00 | (0.03) | −0.00 | (0.03) | 0.00 | (0.03) |
| Region | | | | | | |
| Pop. density | 0.01 | (0.00) | 0.01† | (0.00) | 0.01† | (0.00) |
| Car density | 0.02 | (0.04) | 0.02 | (0.04) | 0.02 | (0.04) |
| Commuter balance | 0.08 | (0.10) | 0.05 | (0.10) | 0.08 | (0.10) |
| Share of foreigners | −0.14 | (0.66) | −0.24 | (0.65) | −0.05 | (0.66) |
| Immigr. human. reasons | −0.44 | (1.29) | −0.40 | (1.26) | −0.42 | (1.23) |
| Multi-fam. houses | 0.19 | (0.22) | 0.25 | (0.22) | 0.26 | (0.23) |
| Regions urban. appr. | 1.62 | (3.82) | 0.24 | (4.00) | 0.56 | (4.10) |
| Rural regions | 3.20 | (4.72) | 0.82 | (4.94) | 1.53 | (5.02) |
| Pupils | −2.13† | (1.18) | −2.30* | (1.13) | −2.02† | (1.11) |
| Childcare | −0.51† | (0.29) | −0.44 | (0.30) | −0.54† | (0.31) |
| Interactions (centered) | | | | | | |
| Childcare × mean age | 0.03 | (0.08) | 0.02 | (0.08) | 0.02 | (0.08) |
| Pupils × mean age | 0.07 | (0.54) | 0.08 | (0.52) | 0.08 | (0.50) |
| Nursing × mean age | −0.01 | (0.03) | −0.02 | (0.03) | −0.02 | (0.03) |
| Constant | −910.67** | (263.65) | −840.43** | (260.34) | −931.62** | (258.98) |
| lambda | −0.66† | (0.39) | | | −0.99** | (0.42) |
| rho | | | 0.96**. | (0.04) | 1.71** | (0.29) |
| sigma2 | 446.41** | (31.57) | 437.17** | (30.93) | 429.97** | (30.71) |
| Log likelihood | −1792.95 | | −1791.47 | | −1789.73 | |
| LR chi2 (OLS) | 2.84† | | 5.79 | | 9.27**. | |
| LR chi2 (SAC) | 6.44* | | 3.48† | | | |

above the Germany-wide average age, the marginal effect of childcare increases significantly by 3.35 in the SAR model, for example (and by a similar magnitude for the other models). An analogous (significant) interaction effect emerges for pupils. Of course, this means that the negative marginal effect of child care and schoolchildren may even be reversed for regions with a high average age. Thus, childcare and pupils can have a different influence on the incidence of infection depending on a region's age. This could, for example, be due to the increased probability of infection of population groups above the average age, but cannot be conclusively answered based on region-specific data alone. Nevertheless, this key finding complements and qualifies previous general statements on the influence of kindergartens and schools on regional infections. It suggests that in future research the influencing factors should not only be analysed independently of each other, but also their possible interactions (moderator effects).

Finally, we also included interaction effects for nursing to check whether the positive influence of nursing varies with the average age of the region. It could be that particularly old regions are more sensitive to more nursing than younger regions because infections behave differently due to the high average age. This cannot be confirmed, however, as interactions are insignificant.

### 3.2. Death count

Results for the death counts in relation to COVID-19 are summarized in Table 3. In comparison with the results for the case numbers in Table 2, it is noticeable that the signs and significances for the estimated coefficients are virtually identical. To explain this, it should be taken into account that the death numbers usually represent a percentage (mortality) of the case numbers with a certain time lag. For Germany, this mortality is approx. 4.7% (as of June 15, 2020). However, a separate consideration of death numbers as outcome variables is nevertheless justified, since mortality (especially at the beginning of an epidemic) is by no means a time-constant variable and therefore the structural similarity of the socio-economic influencing variables compared to the case numbers is not known a priori.

It is particularly interesting that the interaction effects are all insignificant, unlike in Table 2. This means that the (significant) negative marginal effect of childcare and pupils on the number of deaths does not change as a function of the average age of the region. A possible explanation (and an interesting avenue for future research) could be that an infection event in kindergartens and schools does not spill over to populations with increased COVID-19-related mortality (e.g., aged over

80) because children have fewer contacts with this population.

### 3.3. Spillover effects

Regarding our question No. 3 (cf. Section 1) the LR tests implemented in Section 3.1 already support the thesis of a spatial dependency in the data. However, the question of actual regional spillover effects remains unanswered so far. In Ref. [9] it is proposed to analyze these spatial spillover effects by calculating the so-called (average) indirect effects (where the coefficients reported in Tables 2 and 3 correspond to the so-called direct effects). The indirect effects reflect how a change in the $k$-th independent variable of region $j$ affects cases (or deaths) in region $i$, as the result of an infinite sequence of spatial feedback effects. Since these effects naturally vary by region, summary measures have been proposed as quasi-average spillover effects and are reported in what follows. For technical details see e.g. Ref. [19].

To keep the presentation clear, we restrict ourselves in Table 4 to the significant indirect effects (where we also report the corresponding total effects reflecting the sum of the direct an indirect effects). Note that the SEM model has no spillover effects and is therefore omitted in Table 4.

In terms of interpretation, it should be noted that the significant negative spillover effects of *early cases* may indicate that early lockdown measures in affected regions and increased awareness and thus caution about COVID-19 were successful in preventing the spread of cases to neighbouring regions. However, some of the results in Table 4 should be interpreted with caution and in a more hypothesis-generating manner at the current state of (ecological) research. Given the significant negative spillover effects of *mean age* (meaning, in other words, that areas with a higher mean age have a protective effect on surrounding areas with respect to COVID-19 infections), it could be argued that age-related reduced travel to neighbouring regions leads to a containment effect on the number of cases there. In contrast, the significant positive spillover effects related to *over 75* could indicate that economic activity in the corresponding region is low due to age (hence the negative direct effect), but mobility of dependents (visitors) across county borders is rather high (hence the positive spillover effects). Further, the significant negative spillover effects of the *nursing* variable could be due to the fact that a concentration of nursing homes (other than a mere high percentage of the population over 75 as above) has a protective effect on surrounding counties, as nursing homes in Germany were closed off at an early stage of the pandemic by strict lockdown measures.

On the other hand, the density of doctors seems to lead to significant positive spillover effects, since here, in contrast to the nursing homes, increased patient traffic across regional borders is to be expected. The same argument (i.e., exporting infections to surrounding regions) seems plausible as an explanation for the significant positive spillover effects of pupils, since the location of schools often differs from their place of residence. Finally, it is interesting to note that the number of beds in tourist establishments does not seem to lead to significant spillover effects in neighbouring regions (not significant, not included in the table). Here, for example, the early restrictions on tourist travel in the context of the imposed contact restrictions could play a role.

Regarding the results for death count (not shown), only the variables *nursing* (negative), *physicians* (positive), and *life expectancy* (negative) show significant spillover effects. Since the signs correspond to those of the spillover effects for the case numbers, arguments similar to those above can be referred to for explanation.

**Table 4**
COVID-19 case numbers: Direct and indirect effects for the SAR model. The significance level symbols are as before.

| Variable | Coeff. | (Std. Err.) |
|---|---|---|
| Indirect (spillover) effects | | |
| Early cases | −3.37**. | (0.79) |
| Pers. nursing | −1.48† | (0.78) |
| Academics | −14.29** | (5.22) |
| Mean age | −47.10† | (28.67) |
| Over 75 | 58.39* | (28.55) |
| Physicians | 16.76* | (6.98) |
| fe expectancy | −69.75** | (23.45) |
| Pop. density | −0.17** | (0.05) |
| Pupils | 17.91† | (9.87) |
| Childcare | 4.49† | (2.52) |
| Total effects | | |
| Early cases | 3.46** | (0.87) |
| Pers. nursing | 1.52* | (0.75) |
| Academics | 14.71** | (5.09) |
| Mean age | 48.46† | (28.69) |
| Over 75 | −60.07* | (28.03) |
| Physicians | −17.24* | (7.12) |
| Life expectancy | 71.76** | (23.46) |
| Pop. density | 0.17** | (0.04) |
| Pupils | −18.42† | (10.60) |
| Childcare | −4.62† | (2.57) |

To highlight a key result, we have found that the density of pupils and the care rate of small children (kindergartens) seems to have a rather dampening (i.e. negative) effect on infections. However, the average age at the district level turned out to be a moderator effect, i.e. for regions with older population, this negative effect is mitigated (and can even be turned into positive). This result underlines that a general statement on the effect of kindergartens and schools on infection is not possible, but is at least qualified by an age variable. The task of future research can be to confirm these effects (possibly also on the basis of individual data) and possibly to uncover further significant moderator variables not discussed in this study.[4]

A possible limitation to this conclusion regarding schools and daycare centers can be discussed in view of the testing strategy followed in Germany, which does not guarantee that children are tested as regularly as adults. Thus, it could be argued that schools and day care centers stand as a mere indicator for a high number of undetected cases. As background information, it should be mentioned that, according to the national testing strategy, testing only follows the appearance of typical COVID-19 symptoms, whereas children are known to have a rather asymptomatic course. Unfortunately, no systematically collected data of the performed tests by age group and region are available for the period of the study. However, a (not regionally disaggregated) sample of voluntarily participating laboratories (about 30% of all testing laboratories), which continuously report their test results to the RKI, can be used as a first indication [33]. The results show that the proportion of tests performed among 0–4 year olds (at about 0.9% of the population) is indeed lower than for the overall average (about 1.8% of the population). However, the positive rate for 0–4 year olds (1.8%) is also more than half lower than the overall positive rate (about 4.4%) for all tests conducted by mid-June 2020. This consideration can at least partially consolidate our conclusion on kindergartens and schools (even if it cannot rigorously prove it): After all, equalizing the testing rate for children based on the above figures would tend to reduce the overall

## 4. Discussion

This ecological study addresses the association of socio-economic variables with COVID-19 infections and deaths. Data at the district level and spatial econometric models are applied to capture indirect spillover effects between the individual districts. In particular, we asked three research questions about the role of schools and child care, the role of nursing homes, and spatial spillover effects on infections and deaths.

[4] Note that a single study should not include an arbitrary number of model variations in order to still ensure (against the background of a multiple comparisons problem) a statistically meaningful interpretation of the significance levels.

positive rate in the population.[5]

However, even if the empirical arguments just mentioned (due to lack of data and regional context) should not be considered rigorous and conclusive, there is also a similar tendency in the current body of research to suggest that schools and daycare centers are not seen as drivers of COVID-19 caseloads, which also supports our conclusion, see Refs. [6,26,37] for Germany and [7,16,18,31] for an international perspective. These studies also provide some plausible (albeit yet to be further explored) mechanisms of action that could help explain our ecological conclusion regarding schools. These include, for example, the organized and more amenable to public hygiene measures environment in daycare centers and schools versus the largely uncontrollable situation in children's homes. In addition, the presence in daycare centers and schools provides reliable daily access to up-to-date information on COVID-19 for children and parents. Added to this is the role of daycare centers and schools as learning sites for COVID-19-compliant behavior and active practice of hygiene concepts. However, positive spillover effects of pupil density to neighbouring regions could be observed, so this cross-regional export or import of infections due to schools and childcare should be carefully analysed also in future individual-level studies.

Turning to our second question, namely the role of care for the elderly and the employees in this area, first significant indications could be found. The results show that the sole quota of persons in need of care has no influence on infections and deaths, whereas the influence of the employment density in this sector on both outcomes turns out to be significantly positive. These results are all the more interesting because they also support the anecdotal ongoing evidence from the media about nursing homes as hotspots of COVID-19 dissemination and even differentiate between the different outcomes for residents and employees. However, it must be mentioned as a limitation that our data do not distinguish between day care at home and nursing homes.

We were also able to find clear evidence regarding our third question of spillover effects between neighbouring districts, which, to our knowledge, has not been investigated empirically for COVID-19 in Germany before. On the one hand, the suitability of spatial models compared to simple OLS approaches is statistically supported. On the other hand, indications of negative effects (i.e. a containment effect on neighbouring regions) could be found especially for the number of early cases of infection or the average age, whereas spillover effects for doctor and pupil density turn out to be positive. Since this means an export of infections across regional borders, these questions should be analysed in future studies in greater detail and, if possible, on an even smaller regional scale, in order to expand the knowledge about the ecological pathways of COVID-19 transmission.

The practical relevance of the results lies, among other things, in the derivation of new or adaptation of existing political measures for the containment of COVID-19, which by their very nature also do not function at the individual level, but at the population level. When interpreting these ecological results, different mechanisms of action, namely the numerous interdependent infection channels at population level, than in clinical or biological studies at individual level must be taken into account. To give one example: Certainly, a high social status, which can be sufficiently measured e.g. by income, at least statistically protects against infection or even death from COVID-19 at the individual level. Well-known mechanisms of action from the literature can include higher education associated with income, better access to medical care or an information advantage. At the population level, however, this effect can be reversed (especially in the initial phase of the epidemic). A high level of economic activity is often based on networking (including physical networking), travel and social contacts such as working with colleagues in teams or open-plan offices. In other words, both results can be empirically substantiated and are therefore not contradictory.

Our results have proven to be stable in numerous robustness checks: First, the three model types applied (SAR, SEM, SAC) show very comparable results with respect to our core results. In addition, the calculation of the weighting matrix was varied (using the normalize option in Stata) and the robustness of the age (±5 years) and date (±2 weeks) cutoffs was checked. Also, the possible influence of other health variables, such as specialist density or development of hospital care over the last 5 years, had practically no impact on our results. From a statistical perspective, there are further avenues for future studies, e.g., in the area of modeling the temporal dynamics of case numbers, which were explicitly not the subject of this analysis, cf. the generalized additive mixed model approach in Ref. [22] or the dynamic spatial Durbin panel models discussed in Ref. [10]. Finally, there is also the question of comparing the results with corresponding regional analyses for other countries.

[5] At least under the weak assumption that the unobserved positive rate among children who have not been tested so far because they are asymptomatic does not exceed that of those who have been tested, which can be considered plausible.

## Declaration of competing interest

The author declares that he has no competing interests nor financial relationships with any organizations that might have an interest in the submitted work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

[1] Amato M, Werba JP, Frigerio B, Coggi D, Sansaro D, Ravani A, Ferrante P, Veglia F, Tremoli E, Baldassarre D. Relationship between influenza vaccination coverage rate and COVID-19 outbreak: an Italian ecological study. Vaccines 2020;8(3):1–11.
[2] Bärnighausen T, Hosegood V, Timaeus IM, Newell ML. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. AIDS 2007;21(SUPPL. 7):1–16.
[3] Bluhm R, Pinkovskiy M. The spread of COVID-19 and the bcg vaccine: a natural experiment in reunified Germany. 2020. Available: https://ideas.repec.org/p/fip/fednsr/87993.html.
[4] Buja A, Paganini M, Cocchio S, Scioni M, Rebba V, Baldo V. Demographic and socio-economic factors, and healthcare resource indicators associated with the rapid spread of COVID-19 in Northern Italy: an ecological study. PloS One 2020;15(12):e0244535. Available: https://dx.plos.org/10.1371/journal.pone.0244535.
[5] Crighton EJ, Elliott SJ, Moineddin R, Kanaroglou P, Upshur R. A spatial analysis of the determinants of pneumonia and influenza hospitalizations in Ontario (1992-2001). Soc Sci Med 2007;64(8):1636–50.
[6] Jugendinstitut Deutsches, Robert Koch Institut. Quartalsbericht der Corona-KiTa-Studie. 2. Quartalsbericht (IV/2020) [Corona-KiTa study quarterly report. 2nd quarterly report (IV/2020)]. 2020. Available: https://corona-kita-studie.de/quartalsberichte-der-corona-kita-studie.
[7] Edmunds WJ. Finding a path to reopen schools during the COVID-19 pandemic. The Lancet Child and Adolescent Health 2020;4(11):796–7.
[8] Elford J, Ben-Shlomo Y. Geography and migration with special reference to cardiovascular disease. In: Kuh D, Ben-Shlomo Y, editors. A life course approach to chronic disease epidemiology. Oxford: Oxford University Press; 2004. p. 144–64.
[9] Elhorst J. Spatial econometrics. From cross-sectional data to spatial panels. Heidelberg: Springer; 2014.
[10] Elhorst JP, Gross M, Tereanu E. Spillovers in space and time: where spatial econometrics and Global VAR models meet. ECB Working Paper, no. 2134. 2018. Available: https://ssrn.com/abstract=3134525.
[11] Evans GW, Kim P. Multiple risk exposure as a potential explanatory mechanism for the socioeconomic status-health gradient. Ann N Y Acad Sci 2010;1186:174–89.
[12] Evans RG, Stoddart GL. Producing health, consuming health care. Soc Sci Med 1990;31(12):1347–63.
[13] Farr B, Bartlett C, Wadsworth J, Miller D. Risk factors for community-acquired pneumonia diagnosed upon hospital admission. Respir Med 2000;94(10):954–63.
[14] Ginsburgh V, Magerman G, Natali I. COVID-19 and the role of economic conditions in French regional departments. 2020. Available: https://ideas.repec.org/p/eca/wpaper/2013-305805.html.
[15] Guliyev H. Determining the spatial effects of COVID-19 using the spatial panel data model. Spatial Statistics 2020;38. Available: https://linkinghub.elsevier.com/retrieve/pii/S2211675320300373.
[16] Heavey L, Casey G, Kelly C, Kelly D, McDarby G. No evidence of secondary transmission of COVID-19 from children attending school in Ireland, 2020. Euro Surveill 2020;25(21).
[17] Huang X, Lambert S, Lau C, Soares Magalhaes RJ, Marquess J, Rajmokan M, Milinovich G, Hu W. Assessing the social and environmental determinants of pertussis epidemics in Queensland, Australia: a Bayesian spatio-temporal analysis. Epidemiol Infect 2017;145(6):1221–30.

[18] Lee B, Raszka WV. COVID-19 transmission and children: the child is not to blame. e2020004879 Pediatrics 2020;146(2). Available: http://pediatrics.aappublications.org/lookup/doi/10.1542/peds.2020-004879.
[19] LeSage J. An introduction to spatial econometrics. Boca Raton: CRC Press; 2009.
[20] Lowcock EC, Rosella LC, Foisy J, McGeer A, Crowcroft N. The social determinants of health and pandemic h1n1 2009 influenza severity. Am J Publ Health 2012;102 (8):51–8.
[21] Macintyre S, Ellaway A. Ecological approaches: the rediscovery of the role of the physical and social environment. In: Berkman L, Kawachi I, editors. Social epidemiology. Oxford: Oxford University Press; 2000. p. 332–48.
[22] Mercker M, Betzin U, Wilken D. What influences COVID-19 infection rates: a statistical approach to identify promising factors applied to infection data from Germany. medRxiv; 2020. Available: https://www.medrxiv.org/content/medrxiv/early/2020/04/17/2020.04.14.20064501.full.pdf.
[23] Mogi R, Kato G, Annaka S. Socioeconomic inequality and COVID-19 prevalence across municipalities in Catalonia, Spain. 2020. https://doi.org/10.31235/osf.io/5jgzy. Available:.
[24] Mogi R, Spijker J. The influence of social and economic ties to the spread of COVID-19 in Europe. J Popul Res 2021. Available: http://link.springer.com/10.1007/s12546-021-09257-1.
[25] Oto-Peralías D. Regional correlations of COVID-19 in Spain. 2020. https://doi.org/10.31219/osf.io/tjdgw. Available:.
[26] Otte im Kampe E, Lehfeld A-S, Buda S, Buchholz U, Haas W. Surveillance of COVID-19 school outbreaks, Germany, March to august 2020. Euro Surveill 2020;25(38).
[27] Ozer Balli H, Sørensen BE. Interaction effects in econometrics. Empir Econ 2013;45 (1):583–603.
[28] Plümper T, Neumayer E. The pandemic predominantly hits poor neighbourhoods? SARS-CoV-2 infections and COVID-19 fatalities in German districts. Eur J Publ Health 2020;30(6):1176–80.
[29] Ponnambalam L, Samavedham L, Lee HR, Ho CS. Understanding the socioeconomic heterogeneity in healthcare in US counties: the effect of population density, education and poverty on H1N1 pandemic mortality. Epidemiol Infect 2012;140(5):803–13.
[30] Qiu Y, Chen X, Shi W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. J Popul Econ 2020;33(4):1127–72.
[31] Rajmil L. Role of children in the transmission of the COVID-19 pandemic: a rapid scoping review. BMJ Paediatrics Open jun 2020;4(1). e000722.
[32] Ramírez-Aldana R, Gomez-Verjan JC, Bello-Chavolla OY. Spatial analysis of COVID-19 spread in Iran: insights into geographical and structural transmission determinants at a province level. medRxiv; 2020. Available: http://medrxiv.org/content/early/2020/04/22/2020.04.19.20071605.abstract.
[33] Robert Koch Institut. Laborbasierte Surveillance von SARS-CoV-2. Wochenbericht vom 17.06.2020 [Laboratory-based surveillance of SARS-CoV-2. Weekly report, 06/17/2020]. Tech. Rep.. 2020. Available: http://ars.rki.de/Docs/SARS{_}CoV2/Archiv/Wochenberichte/20200617{_}Wochenbericht.pdf.
[34] Sannigrahi S, Pilla F, Basu B, Basu AS. The overall mortality caused by COVID-19 in the European region is highly associated with demographic composition: a spatial regression-based approach. arXiv 2020;62(July):102418. https://doi.org/10.1016/j.scs.2020.102418. Available:.
[35] Stojkoski V, Utkovski Z, Jolakoski P, Tevdovski D, Kocarev L. The socio-economic determinants of the coronavirus disease (COVID-19) pandemic. SSRN Electronic Journal 2020:1–26. Available: https://www.ssrn.com/abstract=3576037.
[36] Su D, Chen Y, He K, Zhang T, Tan M, Zhang Y, Zhang X. Influence of socio-ecological factors on COVID-19 risk: a cross-sectional study based on 178 countries/regions worldwide. 2020. https://doi.org/10.1101/2020.04.23.20077545. Available:.
[37] Thielecke M, Theuring S, van Loon W, Hommes F, Mall MA, Rosen A, Böhringer F, von Kalle C, Kirchberger V, Kurth T, Seybold J, Mockenhaupt FP, Chylla T, Linzbach E, von der Haar A, Körner J, Gertler M, Bernhard J, Rössig H, Kurzmann M, Peters F, Wiesmann C, Horn J, Steger J, Bethke N, Schmidergall T, Holz F, van den Berg A, Peña-Groth ML. SARS-CoV-2 infections in kindergartens and associated households at the start of the second wave in Berlin, Germany – a cross sectional study. medRxiv; 2020. https://doi.org/10.1101/2020.12.08.20245910. Available:.
[38] Tijhuis M. Social networks and health. 1994. Thesis, Utrecht.
[39] Tobler W. A computer movie simulating urban growth in the detroit region. Econ Geogr 1970;46:234–40.
[40] Trinh P, Jung TH, Keene D, Demmer RT, Perzanowski M, Lovasi G. Temporal and spatial associations between influenza and asthma hospitalisations in New York City from 2002 to 2012: a longitudinal ecological study. BMJ Open 2018;8(9):1–9.
[41] Varga C, Pearl DL, Mcewen SA, Sargeant JM, Pollari F, Guerin MT. Evaluating area-level spatial clustering of Salmonella Enteritidis infections and their socioeconomic determinants in the greater: a retrospective population-based ecological study. BMC Publ Health 2007;13:1–17.
[42] Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, Deng Y, Lin S. The impact of COPD and smoking history on the severity of COVID-19: a systemic review and meta-analysis. J Med Virol 2020. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25889.
[43] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020;395(10229):1054–62.

**Andree Ehlert** studied economics at the Universities of Hamburg and St Andrews, specializing in statistics and econometrics. He then obtained his PhD in the field of time series analysis and extreme value statistics at the Institute for Mathematical Stochastics at the Georg-August University of Göttingen. He has several years of industry experience in the statistical analysis of financial services and as a biometrician in pharmaceutical consulting. Dr. Ehlert has also worked as a post-doc at the Innovation Incubator, the University of Lüneburg and the Hamburg Institute of International Economics (HWWI) in the areas of regional and health economics, health services research, secondary data analysis, spatial statistics/econometrics, program evaluation and economic modeling. He currently holds the professorship for statistics and empirical social research at the Harz University of Applied Sciences.