



OPEN

Housing situations and local COVID-19 infection dynamics using small-area data

Diana Freise¹, Valentin Schiele¹ & Hendrik Schmitz^{1,2,3}✉

Low socio-economic status is associated with higher SARS-CoV-2 incidences. In this paper we study whether this is a result of differences in (1) the frequency, (2) intensity, and/or (3) duration of local SARS-CoV-2 outbreaks depending on the local housing situations. So far, there is not clear evidence which of the three factors dominates. Using small-scale data from neighborhoods in the German city Essen and a flexible estimation approach which does not require prior knowledge about specific transmission characteristics of SARS-CoV-2, behavioral responses or other potential model parameters, we find evidence for the last of the three hypotheses. Outbreaks do not happen more often in less well-off areas or are more severe (in terms of the number of cases), but they last longer. This indicates that the socio-economic gradient in infection levels is at least in parts a result of a more sustained spread of infections in neighborhoods with worse housing conditions after local outbreaks and suggests that in case of an epidemic allocating scarce resources in containment measures to areas with poor housing conditions might have the greatest benefit.

By the end of 2022, the World Health Organization (WHO) registered more than 656 million SARS-CoV-2 (severe acute respiratory syndrome Corona virus type 2) cases worldwide¹, with the number of confirmed cases representing only a fraction of the actual cases². Since the beginning of the COVID-19 pandemic, researchers of various disciplines investigate the spread of the virus as understanding this may help to develop strategies to cope with future epidemics. In this paper, we use small-scale data (see below for an explanation of the term small-scale) to study outbreak patterns of SARS-CoV-2. In particular, we focus on the role of the residential environment in order to learn what role it plays in the frequency, intensity and duration of local outbreaks and, thus, in the spread of the epidemic. This may help in the future to allocate scarce resources in containment measures to the areas where they will have the greatest benefit.

Several literature reviews address previous research on the spread of SARS-CoV-2 and contextualize the factors analyzed, the region and timing, and the methodology used. Most of the literature to date uses spatial and spatio-temporal methods in most cross-sectional and ecological studies, not only to investigate clusters in the spread or to simulate its evolution, but also to understand the role of socio-economic status and associated determinants in driving dispersal (see, for example, Franch-Pardo et al.^{3,4} and Nazia et al.⁵ for more information on the methodology used in this context). Alidadi and Sharifi⁶ review the literature examining the human (made) factors (for instance, density, transportation and mobility, demographic and socio-economic factors) responsible for the spread. They identify a total of seven important fields, one of which is housing conditions. Khanijahani⁷ focuses on literature examining in particular racial and ethnic as well as socioeconomic disparities in the pandemic and, thus, reviews parts of this literature strand. Since a full review of this literature is beyond the scope of this article, we summarize the main findings of these studies below, and refer the interested reader to the above-mentioned literature reviews for a more exhaustive overview.

Studies show differences in the first phase of the pandemic compared to following waves. In Germany, for instance, the first infections spread particularly in more affluent areas, where infections occurred through travelers from skiing holidays and other international travels. Socially more deprived districts were less affected at this time. This changed in the second phase, where districts with higher educated people had lower incidence rates⁸. This is confirmed by Rohleder et al.⁹ who also find higher deprivation as well as non-nationality to be positively correlated to a higher incidence risk, except for the first wave. Similar relations are also observed in the United States¹⁰, and cities like Barcelona¹¹, Helsinki¹² or Rio de Janeiro¹³. In addition, Hoebel et al.¹⁴ observe higher mortality for socio-economically deprived German districts. These findings already suggest housing conditions to play a role in the infection process, as deprivation, housing and health are strongly linked to each other and

¹Paderborn University, Paderborn, Germany. ²RWI Essen, Essen, Germany. ³Leibniz Science Campus Ruhr, Essen, Germany. ✉email: hendrik.schmitz@uni-paderborn.de

are not trivial to be disentangled¹⁵. While both urban density and connectivity within or between cities contribute to the spread, the latter one is better indicator¹⁶. Therefore, with the need for social distancing especially in the first phase of the pandemic, this is probably easier for people in better housing situations allowing also for home office, isolation of infected individuals and more testing opportunities. The role of housing characteristics in its variations is therefore of particular interest in the understanding of COVID-19. Ahmad et al.¹⁷ find higher incidence and mortality rates for US counties characterized by a higher percentage of households with severe housing problems, including overcrowding, high housing cost burden, incomplete kitchen facilities, and incomplete plumbing facilities. Besides overcrowding (see also Lee et al.¹⁸), areas with a higher proportion of multigenerational households are associated with higher incidence rates¹⁹. Overcrowded homes are not only associated with higher incidence, but also mortality²⁰.

We contribute to this literature in the following way. While the focus of many of these studies lies on the differences in levels of infection rates and deaths by socioeconomic status, we study *the evolution of local SARS-CoV-2 outbreaks* on a rather disaggregate level. This might help to understand how the commonly observed differences in infection *levels* across neighborhoods are shaped. Specifically, we ask whether local outbreaks are (a) more frequent, (b) more severe, and (c) longer lasting in disadvantaged areas compared to more affluent areas. To study patterns of local SARS-CoV-2 outbreaks, we combine small-scale district-level infection data with small-scale housing quality data for the city of Essen, a major city in the west of Germany with nearly 600.000 inhabitants and 50 local districts (“Stadtteile”), and rely on event-study graphs allowing to visualize outbreak trajectories and potential differences by residential environment. Event study methods are increasingly used in the literature on COVID-19/SARS-CoV-2, e.g. to study the impact of policy interventions on the spread of the epidemic (see e.g. Askitas et al.²¹, Bárcena-Martín et al.²², Bertocchi and Dimico²³ and Dave et al.²⁴), due to their flexibility, simplicity and transparency. Their major advantage in studying outbreak trajectories, also compared to epidemiological models, is that their use does not require prior assumptions about transmission characteristics of SARS-CoV-2, the immune status and susceptibility, or behavioral responses of the population (all of which changed during the pandemic). Our results show that local outbreaks are not more frequent or severe in less affluent districts. However, while they tend to end more than 20 days after the onset of the outbreak in districts with better housing, cases are higher in less affluent districts even after at least 25 days, suggesting sustained transmission of infections in these districts also in the longer run. This finding may not only explain the socio-economic disparity in infection rates found in previous studies, but also calls for a greater focus of interventions such as local vaccination initiatives or information campaigns in neighborhoods with poor housing conditions.

Data

For our analysis we combine small-scale data from different data sources. The unit of observation is the local district (“Stadtteil”) in Essen, a large city in Western Germany with nearly 600.000 inhabitants and 50 local districts. Essen has an area of 210 square kilometers, thus, the average size of the 50 local districts is around four square kilometers (or two by two kilometers). This is typically called “small-scale” or “small-area” as it is much smaller than, for instance, the county level (“Kreis”) for which German infection data are typically available with the largest county having a size of 74 by 74 kilometers. SARS-CoV-2 incidence rates—which are the numbers of new cases recorded per 100,000 people in seven days—are provided to us by the city of Essen. These are also used to calculate the number of cumulative cases (per 1000 inhabitants). We use daily data from 1 March 2020 to 31 December 2021, that is, 50 times 671 data points on infection rates. Information on housing situation is provided by the FDZ Ruhr. The data is based on different sources: the Federal Agency for Cartography and Geodesy (information on the quality of the residential area), ImmoScout24—the leading online platform for residential and commercial real estate in Germany—(information on the rent index) and the Federal Statistical Office of Germany (information on household sizes), see²⁵.

Figure 1 illustrates the regional variation of cumulative cases per 1000 inhabitants within the city and over the course of the pandemic until the end of 2021. With the mutation to the Omicron variant and less severe progression of Covid infection, official data that only count positively tested individuals became less reliable when less and less people with infections decided to get tested. Therefore, we only analyze outbreaks up to this point in time. The classification of the waves is based on the incidence development in Essen. Figure S1 in the Supplementary Information shows the incidence rate in Essen until 31 December 2021. According to our classification, we declare a wave to be over as soon as it reaches a low point after flattening out. In Fig. 1, it is noticeable that the north of Essen has more cases, with the exception of the first wave where the first cases occurred predominantly in the wealthier districts.

The regional differences in Essen—a city with a strong North–South-divide—also become apparent in the consideration of residential factors in Fig. 2. The left map shows the regional variation in the rent index (in euros per square meter) which reflects the local rent. More expensive neighborhoods are prevalent in the south.

To map the location quality of the residential environment, we look at the share of addresses that are classified as good residential locations. Good residential locations are characterized by open and green spaces, a well-kept street scene and very good transport connections, good shopping possibilities and a good image. The middle map shows the regional variation in the share of addresses in good residential location, ranging from less than 13 percent in northern local districts to more than 55 percent in some southern local districts. We provide additional information on how this measure is constructed in Sect. S2 of the Supplementary Information.

Less crowded households may be less prone to Corona infections. We calculate the share of households with less than five persons—in this sense less crowded households—and see in the right map of Fig. 2, that more than 96 percent of households consist of less than five persons in most districts in the south of Essen. This share is slightly lower in the northern local districts.

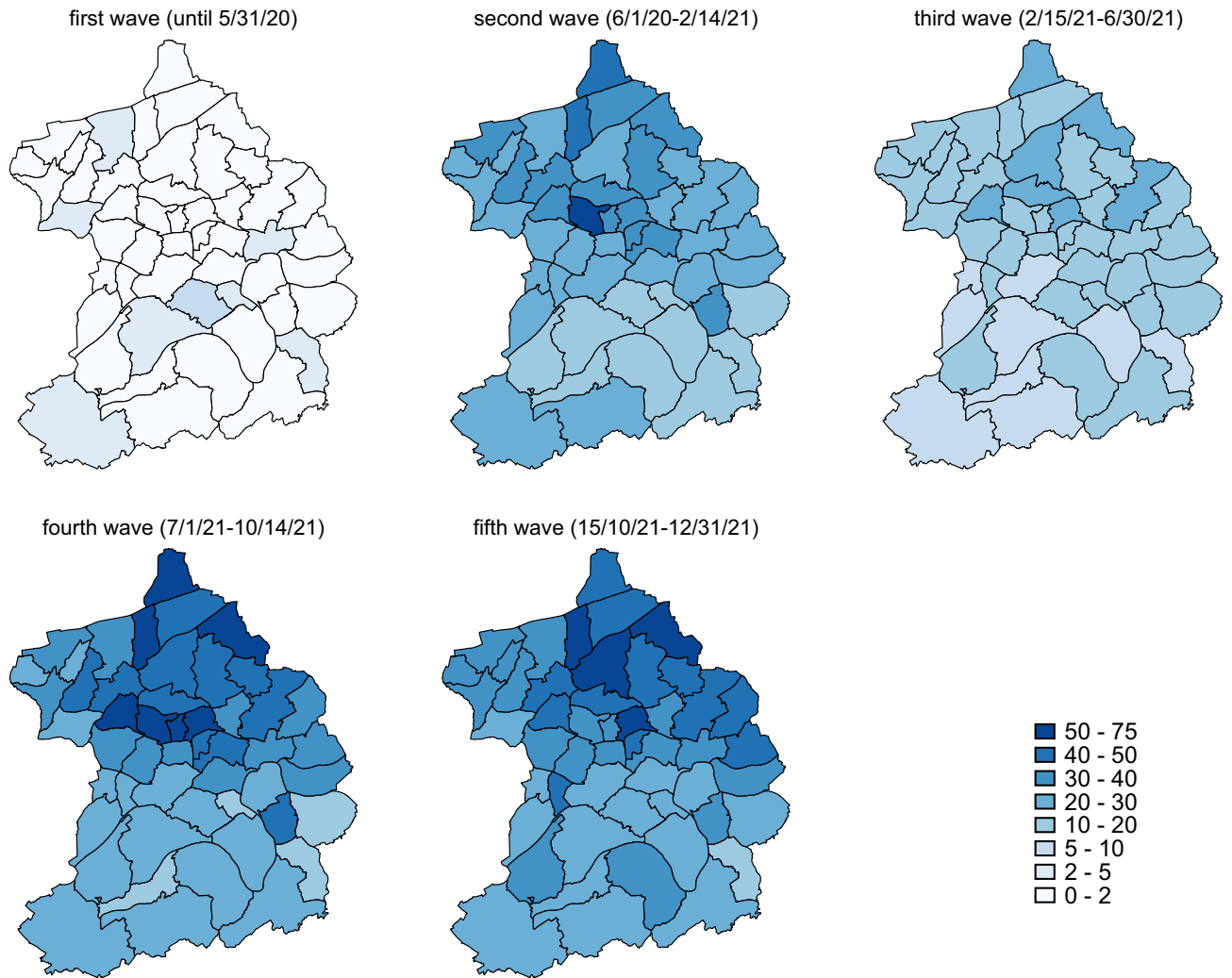


Figure 1. Illustration of the regional variation of cumulative cases between local districts. Each map shows cumulative cases per 1000 inhabitants for a specific time frame. The figure was created using Stata 17 (<https://www.stata.com/>).

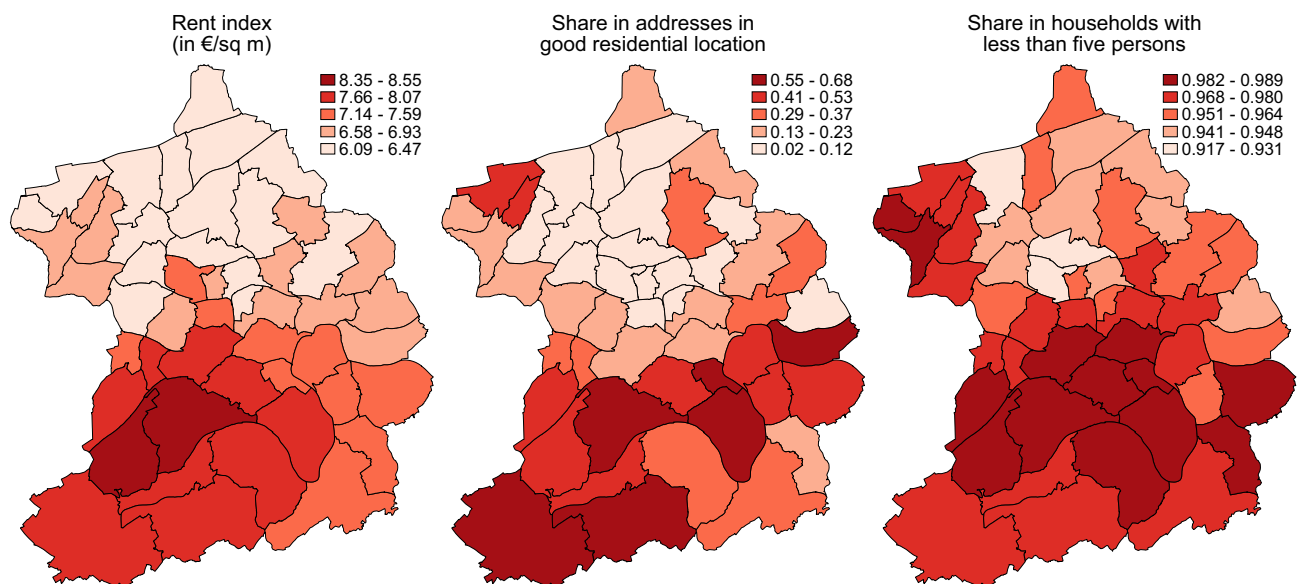


Figure 2. Illustration of the regional variation in housing characteristics between local districts. Classification is based on k-means partition cluster analysis. The figure was created using Stata 17 (<https://www.stata.com/>).

For the main analysis, we concentrate on these characteristics and divide each into terciles. Thus, the first tercile of the rent index contains a third of the districts with the lowest rent level. The first tercile of good residential location contains a third of the districts with the lowest share of addresses in good location and the first tercile considering the household size contains a third of the districts with the lowest share in households with less than five persons. Table 1 summarizes descriptive statistics for these housing characteristics where especially the mean differences in terciles for good residential location are apparent. The mean rent index ranges from around 6–8.5 euros per square meter. The average rent in a district from the group of 33.33% of districts with the lowest rent level, is 6.3 euros per square meter whereas in the group of 33.33% of districts with the highest rent level, it is 7.82 euros per square meter on average. The share in addresses in good location ranges from around 1.5 percent to over 67 percent. The difference in mean shares of the first and third tercile are in this respect more obvious. In districts with the lowest proportion of good location, on average every fifth address is in good quality, whereas in districts with the highest proportion of good location it is one in two addresses. The differences in shares in households with less than five persons ranges from around 92–99 percent.

When comparing Figs. 1 and 2, infections and housing situations seem to be linked. To quantify the graphical impression that there is a relationship between Corona infections and housing situation and to estimate the relative significance of the individual indicators, we run bivariate OLS regressions of cumulative cases per 1000 inhabitants on each characteristic, separately for each identified Corona wave in Essen. These regressions are intended to replicate the correlations from the existing literature. Although the data does not contain the necessary information to obtain causal estimates by including additional variables, the selected measures rather serve as indicators to reflect more general differences in housing situations. Table 2 shows standardized coefficient estimates. Except for the first wave (column (1)), the relationship between cumulative cases and housing characteristics is negative which is also in line with the previous discussed literature. One standard deviation increase in the rent index (and respectively addresses in good location or households with less than 5 persons) goes along with a 0.305 (and respectively 0.245 or 0.312) standard deviations increase in cumulative cases at the end of wave 1. With reference to columns (2)–(5), the relationships are throughout negative and range between 0.4 and 0.8 standard deviations decrease in cumulative cases. All in all, except for wave 1, local districts with higher rents, higher shares of addresses in good location and less crowded households seem to be less prone to Corona infections. The relationship is very similar in waves 2–4 but somewhat smaller in wave 5. A speculative reason for the less strong relationship in wave 5 is that at that time immunization rates of individuals in worse housing conditions had become better due to higher infection rates before. Yet, it is beyond the scope of the paper to scrutinize this claim.

Methods and results

Methods. The analysis so far shows a clear association between housing situation and SARS-CoV-2 incidence. In all waves except for the first, districts with a higher proportion of good housing conditions have significantly fewer cases than neighborhoods characterized by low rents, a low proportion of households in good residential location, and large household sizes. This suggests that the frequency, intensity, and duration of local outbreaks might depend on the residential environment. A potential explanation for this finding is that more

	Mean	SD	Min	1st terc.	2nd terc.	3rd terc.	Max
Rent index (in €/sq m)	7.01	0.67	6.09	6.30	6.91	7.82	8.55
Addresses in good location (in %)	26.64	20.04	1.53	5.68	24.48	51.19	67.68
Households < 5 (in %)	96.62	1.82	91.69	94.46	97.06	98.46	98.86

Table 1. Descriptive statistics for housing characteristics.

Dep. var.: Cumulative cases (per 1000 inhabitants) from current wave	(1)	(2)	(3)	(4)	(5)
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Rent index (in euros/sq m)	0.305** (0.137)	−0.562*** (0.119)	−0.789*** (0.089)	−0.684*** (0.105)	−0.416*** (0.131)
Addresses in good location (in %)	0.245* (0.140)	−0.567*** (0.119)	−0.518*** (0.123)	−0.638*** (0.111)	−0.400*** (0.132)
Households < 5 (in %)	0.312** (0.137)	−0.761*** (0.094)	−0.697*** (0.104)	−0.735*** (0.098)	−0.470*** (0.127)
<i>N</i>	50	50	50	50	50

Table 2. Bivariate regression results of cumulative cases on housing characteristics. Coefficients estimates are standardized. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Figure S6 in the Supplementary Information shows the joint distribution of cumulative cases and housing conditions by wave and for each of the three housing measures.

affluent individuals may find it easier to reduce their personal risk of infection following a Corona outbreak in their community through social distancing and hygiene measures (home office, isolation of infected individuals in the household, regular testing, wearing KN95/FFP2 masks, etc.) than individuals living in more precarious residential areas. This, in turn, could contribute to a rapid stagnation or even decline of incidence in one neighborhood after an outbreak, whereas it may rise for a longer period of time in another and only decline later. We now change the time dimension in our analysis from *wave to day* and exploit the daily information on 7-day Corona incidence. Using event studies, we will test the hypothesis that the trajectories of local outbreaks differ by the residential environment. This requires to define what constitutes a local outbreak and its onset in a first step.

Identification of local Corona outbreaks. One way to define the onset of a Corona outbreak, which seems natural at first glance, would be to use the occurrence of a first case after a predefined period of time without active cases in the respective district. However, in view of the wide spread of SARS-CoV-2 from autumn 2020 at the latest and the associated permanent transmission of individual infections from one area to the other, such an approach is not very purposeful. Therefore, and since there is no generally accepted definition of local Corona outbreaks, we use a data-driven approach in this paper to identify (the onset of) local outbreaks. Broadly speaking, an outbreak is defined here as a sharp increase in the Corona incidence over a period of at least 7 days after a stagnation of the incidence over a period of at least 7 days. In our preferred specification, stagnation corresponds to an average change in daily Corona incidence in the range of ± 2 cases per 100,000 inhabitants. A sharp increase in incidence then corresponds to an increase in active cases by an average of at least 15 cases per 100,000 inhabitants per day and this for a period of at least 7 days. Technically, we identify (the onset of) outbreaks by running a series of linear regressions. For each district i and day k we fit the following two linear regressions to the data

$$y_{it} = \alpha_{1ik} + \beta_{1ik}t + \varepsilon_{1it} \quad \text{for } k - 7 < t < k \quad (1)$$

$$y_{it} = \alpha_{2ik} + \beta_{2ik}t + \varepsilon_{2it} \quad \text{for } k < t < k + 7 \quad (2)$$

where y_{it} denotes the 7-day Corona incidence (the sum of new Corona infections in the previous seven days) in district i on day t ; α_{1ik} and α_{2ik} are intercepts and β_{1ik} and β_{2ik} are the slope parameters of the of the rolling window regressions. This yields $k = 1, \dots, 657 (= 671 - 7 - 7)$ estimates for β_{1ik} and β_{2ik} for each of the $i = 1, \dots, 50$ districts. Whether day $t = k$ marks the starting point of a Corona outbreak in district i then follows from a comparison of the slope parameters belonging to this day with the corresponding threshold values. Day k only marks the onset of an outbreak if $-2 < \beta_{1ik} < 2$ and $15 < \beta_{2ik}$, i.e. if the incidences in the 7 days before day k have not changed by more than 2 cases per day and 100,000 inhabitants on average and the incidences in the 7 days following day k have increased on average by at least 15 cases per day and 100,000 inhabitants.

Figure 3 provides a stylized example that illustrates our approach to identify local Corona outbreaks. It shows simulated daily incidences for the arbitrarily chosen days $t = 1, \dots, 19$ in a hypothetical district, along with the fitted regression lines for varying values of k . In the upper left panel, the center of the rolling estimation window is day 8, i.e. $k = 8$ and accordingly the first regression line is fitted through the data for days 1–7 and the second through the data for days 9–15. In the next panel, the center of the estimation window is moved one day ahead, and the regression lines are fitted through the data for the days 2–8 and 10–16, respectively, and so forth. By comparing the slope parameter of the regression lines to the respective thresholds (i.e. $-2 < \beta_{1ik} < 2$ and $\beta_{2ik} > 15$), we are then able to identify the onset of an outbreak. In Fig. 3, we observe a slope parameter in all regression lines left of the center of the rolling window smaller than -2 and a slope parameter smaller than 15 in all regression lines right of the center for all panels except for the one in the lower right. Only on day 11, we observe a clear stagnation of cases in the 7 days before and a clear increase (>15 cases per day and 100,000 inhabitants) during the following 7 days. Thus, based on our approach for identification of local outbreaks, we would conclude that day 11 marks the start of the local outbreak in our hypothetical district.

In Fig. S2 in the Supplementary Information, we show a selection of outbreaks using the actual data from Essen, both from more and less affluent districts and at different points in time, along with the estimated regression lines that identify the start date of the respective outbreak. Table S1 in the Supplementary Information gives an overview of the number of local outbreaks identified in the data using this approach. In total, we identify 167 local outbreaks in the 50 districts of the city of Essen during the observation period. Most of these outbreaks occur in the second and fifth wave, several in waves 3 and 4, and only a negligible number in the first wave. Given that the thresholds for our outbreak definition are chosen somewhat arbitrarily, we also provide results for two alternative definitions where we use other thresholds and show that the results of our analysis are not sensitive to the exact definition of outbreaks. Table S1 also reports the thresholds and number of outbreaks identified using these alternative outbreak definitions.

Event study estimations. Based on the definition of Corona outbreaks and all local outbreaks identified in the data, we can analyze and visualize the courses of the outbreaks in order to investigate whether they differ depending on the residential environment. For this purpose, we use event-study graphs in the following, which plot the development of incidences as a function of the temporal distance to the onset of the local Corona outbreak.

We define the event time r as the time (in days) between a day and the day at which a local outbreak starts in district i :

$$r_{it} = t - e_i$$

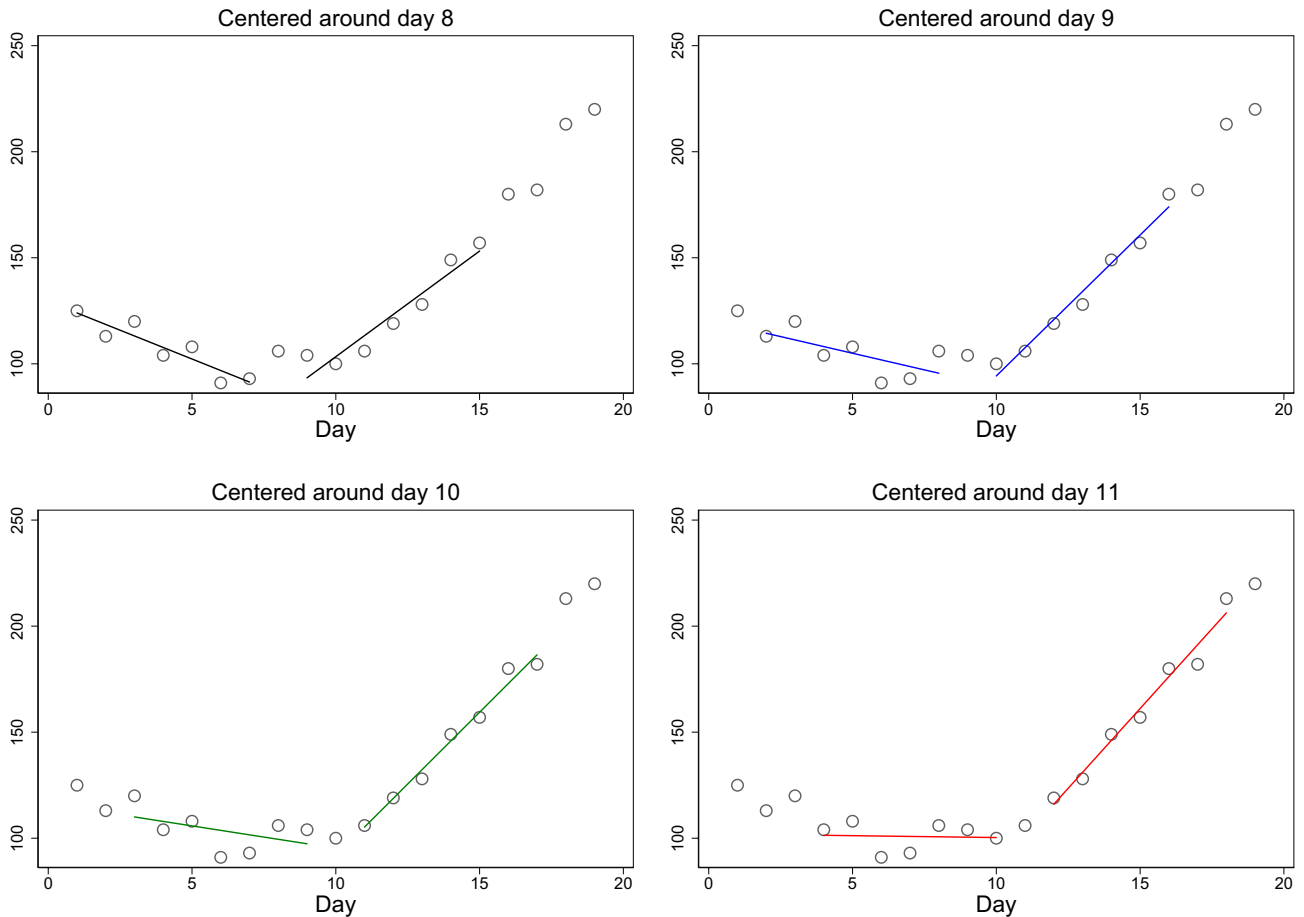


Figure 3. Stylized example of local Corona outbreak identification for one district. The vertical axis measures the 7-day incidence. The figure was created using Stata 17 (<https://www.stata.com/>).

where t denotes calendar time in days, i is the district, and e_i denotes the day that has been identified as the day when a local outbreak starts in district i . Therefore, $r = -1$ is the last day before the onset of the outbreak, $r = 0$ is the day of the onset of the outbreak and $r = 1$ is the first day after the onset of the outbreak and so on. Based on this definition of relative time, we run the following baseline fixed effects regression:

$$y_{it} = \sum_{\substack{j = -10 \\ j \neq -1}}^{24} \mu_j 1[r = j] + \alpha_i + \lambda_m + \tau_w + \varepsilon_{it} \tag{3}$$

where y_{it} denotes the 7-day incidence on day t in district i , that is, the sum of new Corona infections in the previous seven days. The indicator function $1[r = j]$ takes on the value 1 if $r = j$ and 0 otherwise. That is, we include a full set of binary variables that non-parametrically account for the relative time r (i.e. of the event time). μ_j (where $j = -10, \dots, 24$) are the 35 coefficients of the indicator variables. We restrict $\mu_{-1} = 0$ and trim the sample to the left at $j < -10$. For $j > 0$ we include all event-time indicators and report them up to day 24 after the start of an outbreak. α_i , λ_m , and τ_w are district, calendar month and wave fixed effects. The 7-day Corona incidence is—other than the reported new cases in Germany - not sensitive to the day of the week. Therefore, day of the week fixed effects are not included. By plotting the resulting event time coefficients against the relative distance to the start of an outbreak, we can empirically evaluate how cases evolve over the course of a local outbreak without making prior assumptions about the transmission properties of SARS-CoV-2 or the susceptibility of the population. Splitting the sample by residential characteristics then allows to study potential differences in outbreak patterns by housing conditions.

Results. Before turning to the event study results, Fig. 4 shows the geographical distribution of outbreaks by wave to assess whether districts with bad housing conditions experience Corona outbreaks more often. Local outbreaks can be observed in almost all districts at some point in time. In wave 3 and 4, it seems that local outbreaks happen more frequent in bad residential locations in the north while this is not the case in waves 1, 2 and 5. Overall, incidence of Corona outbreaks appear to be across all areas.

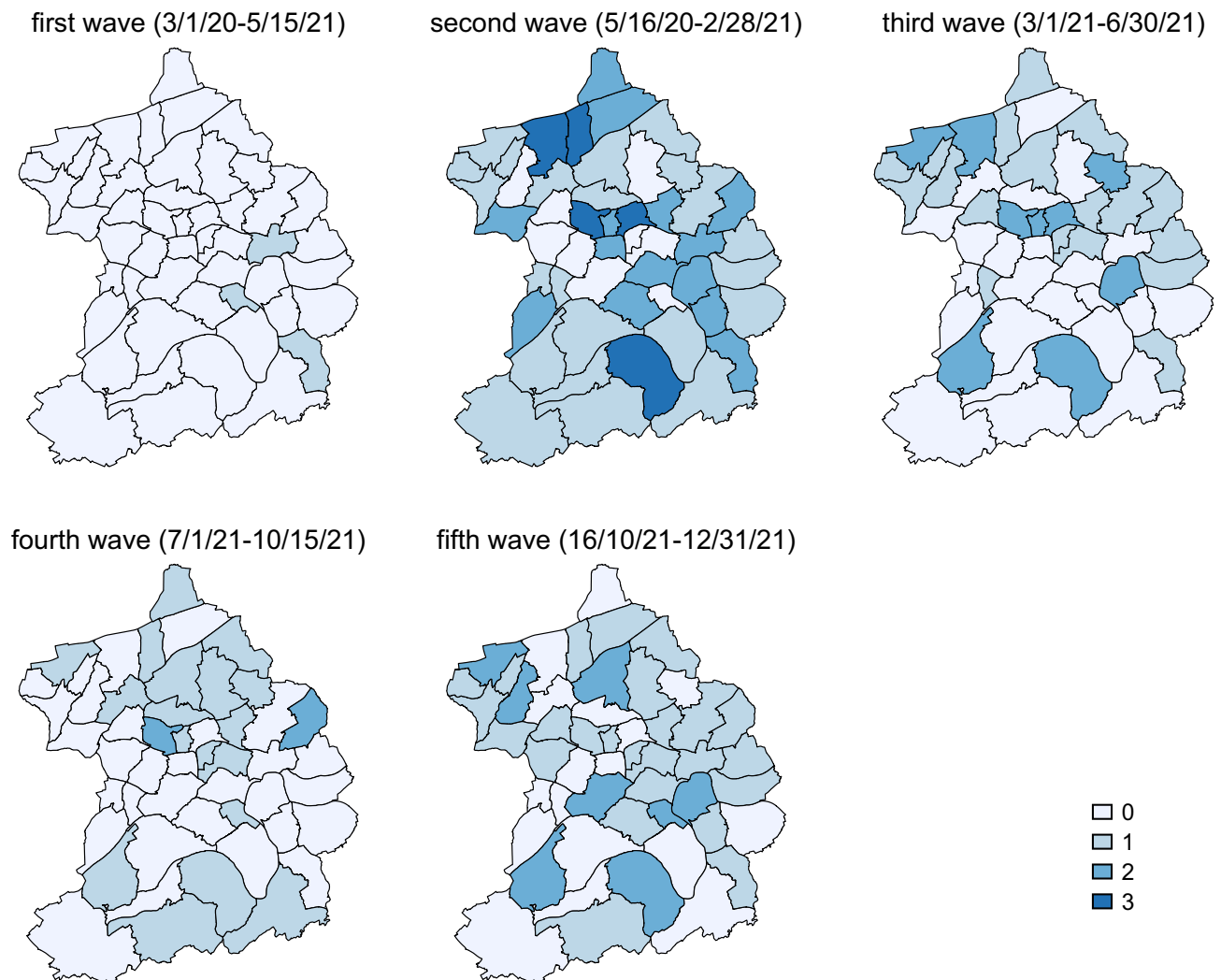


Figure 4. Number of outbreaks in neighborhoods. The figure was created using Stata 17 (<https://www.stata.com/>).

To shed light on the trajectories of local outbreaks and address the question whether local Corona outbreaks are also more intense and longer in districts with bad housing conditions, we now turn to the event study results. Figure 5 plots the point estimates for μ_j , derived from estimating Eq. 3, along with 95% confidence intervals against relative time r for four specifications. The upper left panel shows baseline results for the entire sample, the following panels provide separate results by housing condition. Here, we distinguish between districts with high (in the top tercile) versus low (in the bottom tercile) shares of good housing, using rent levels as an indicator in the top right panel, housing locations in the bottom left panel, and household size in the bottom right panel (see “Data” section for details on definition of variables). That is, here we perform separate subsample analyses by districts with bad/good locations as defined by the three different measures of housing condition we use. The baseline results show that after a stagnation or slight decline in Corona incidence, cases initially rise sharply with the onset of a local outbreak. While this is not surprising given the way we define and identify outbreaks in the data, the results are still interesting. They show that the increase in case numbers accelerates during the first three to four days, but then quickly levels off. By the eighth or ninth day, the local outbreak has already peaked. Incidences then average at about 120 cases per 100,000 inhabitant above the level before the outbreak began. In the following approximately ten days, incidences fall again by about half but then tend to stagnate until at least day 24 after the start of the outbreak. On the one hand, this indicates that local outbreaks end relatively quickly, but on the other hand, it also shows that these outbreaks seem to lead to persistently higher levels of transmission of infection and thus sustain the spread of SARS-CoV-2. Interestingly, this seems to be especially true for less affluent residential areas.

The remaining panels of Fig. 5 show, that the patterns of local Corona outbreaks in less affluent residential areas are initially almost identical to those in better areas. In both less affluent and better residential areas, local outbreaks reach their peak at the same time. Up to this point, there is also no difference in the level of incidence growth. After the case numbers have reached their peak, they initially decline uniformly in both the less affluent and the better residential areas, but then diverge from around day 15 after the start of the outbreak. While the incidences in better districts seem to continue to decline on the following days, they stagnate from day 15

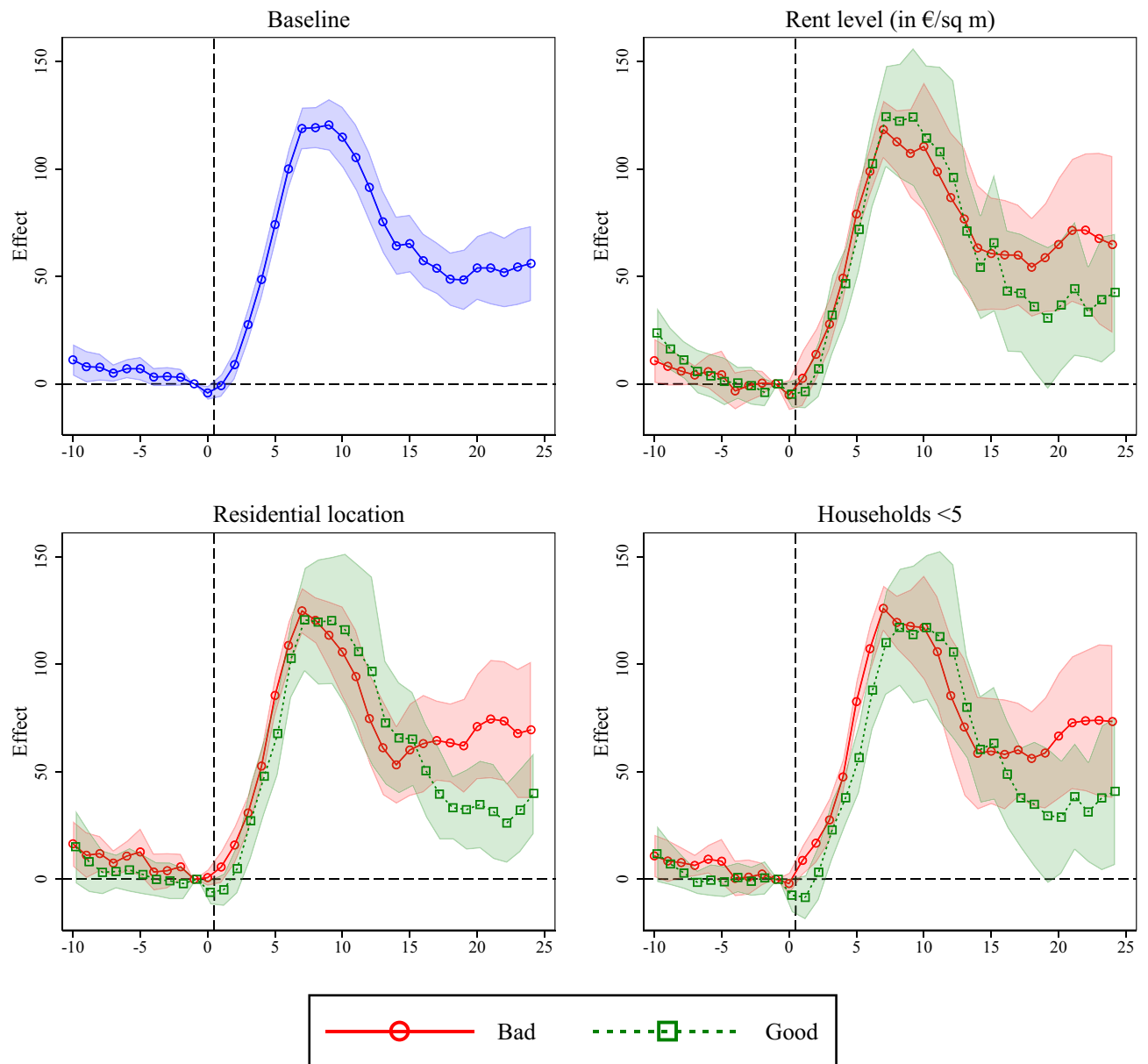


Figure 5. Event study results. Coefficients corresponding to μ_j in Eq. 3. μ_{-1} is restricted to zero. 95% confidence intervals reported. Standard errors clustered on district level. The figure was created using Stata 17 (<https://www.stata.com/>).

onwards on a higher level in less affluent districts. Although this difference is not particularly large and not statistically significant, it shows up consistently in all analyses.

While our approach of using event studies to identify and depict outbreak patterns requires few assumptions, it can be criticized for leaving many degrees of freedom as to the definition of a local outbreak. We thus show in Figs. S3 and S4 in the Supplementary Information that the results are largely robust to the use of alternative definitions of outbreaks that are described in Table S1.

Another potential concern with our approach is that it does not take into account the level of infections of SARS-CoV-2 from which the outbreaks start and ignores earlier outbreaks. Both the infection level at the start of a local outbreak and whether there have been previous outbreaks in the same district are likely to have an impact on the intensity and duration of outbreaks, as both influence the number of potentially susceptible individuals. If this number also differs between bad and good neighborhoods, it would be misleading to attribute differences in outbreak trajectories directly to differences in the residential environment. To deal with this concern, Fig. 6 repeats the analysis but now includes the cumulative number of cases since the beginning of the pandemic (time lag of 15 days) as an additional control. In fact, the differences in outbreak trajectories between districts with good and districts with poor housing environments reduce somewhat when controlling for previous infection activity, as Fig. 6 shows. Nevertheless, there are still differences towards the end of the outbreak, suggesting that

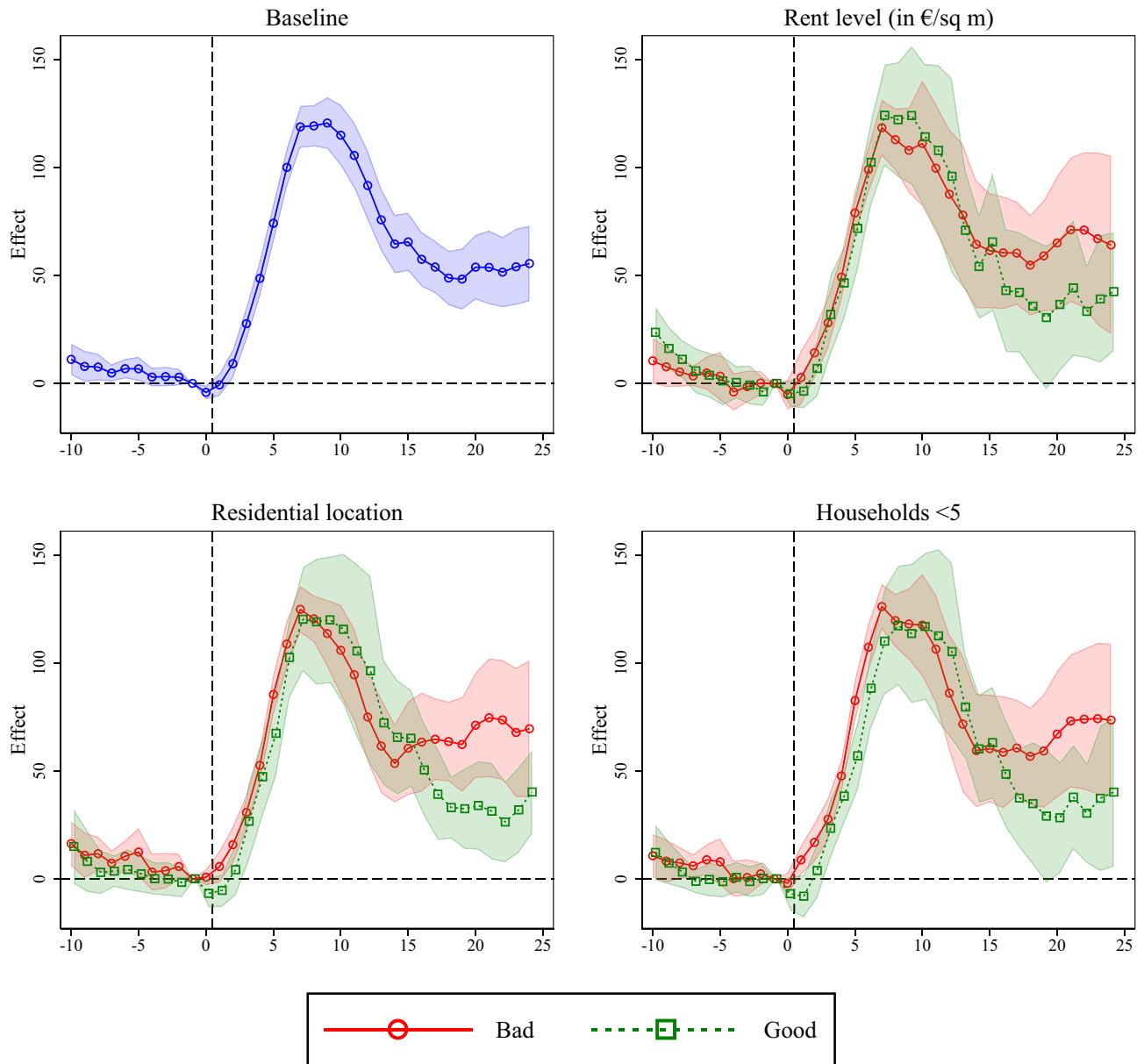


Figure 6. Event study results. Coefficients corresponding to μ_j in Eq. 3. μ_{-1} is restricted to zero. Regressions additionally account for (cumulative) cases up to day $t - 15$. 95% confidence intervals reported. Standard errors clustered on district level. The figure was created using Stata 17 (<https://www.stata.com/>).

local outbreaks in poor neighborhoods lead to somewhat higher incidences in the longer term than those in good neighborhoods.

Figure S5 shows event study results for alternative measures of crowding and affluence that support or interpretations of the main results. We provide descriptive statistics and explanations for these measures in Table S2.

Summary and conclusion

In this paper, we examine local outbreak patterns of SARS-CoV-2, focusing on the role of housing conditions for the frequency, intensity, and duration of local outbreaks. Using small-scale infection and housing quality data, we show that, in the city of Essen, Germany, local outbreaks did not occur significantly more frequently in areas with rather poor housing conditions. We also find almost no differences in the intensity of outbreaks (measured as the speed and magnitude of outbreaks) between poorer and better neighborhoods. However, our results suggest that outbreaks in more affluent neighborhoods decline further after reaching their peak than in less affluent neighborhoods, where they last longer on a higher level. This suggests that outbreaks lead to higher post-outbreak incidences and thus higher transmission rates in less affluent neighborhoods, which might explain, at least to some extent, the socioeconomic gradient in prevalence and mortality of SARS-CoV-2. While it has been found before that worse housing situations are correlated with higher incidences (see the cited studies in the introduction), we believe that it is (to the date of writing this paper) a new result in the literature that this

correlation seems to be due to longer durations of local outbreaks and not due to higher frequencies of outbreaks or higher intensities.

We are aware that this is the result of one case study in a German city only, but, taken at face value it has some implications. A common hypothesis for the well-known socio-economic gradient in SARS-CoV-2 infections is that people in disadvantaged areas have less possibilities for home-office, thus higher infection rates at the workplace. Yet, given that many individuals work in other districts and that frequencies of infection do not significantly vary by districts, this does not seem to be the most important reason. Instead, once the virus enters the district, it stays longer in disadvantaged areas, implying that transmissions are more likely in the private environment, partly induced by bad housing situations like overcrowded places. Yet, obviously this is just an interpretation that is consistent with the statistics and more evidence is needed to completely back that.

For policymakers it seems relevant to visit and address individuals in disadvantaged areas much more directly regarding, for instance, local vaccination initiatives and education of effectiveness of masks. Not because individuals in these districts are less willing to vaccinate or wear masks but because they have less other options to protect themselves.

Data availability

Restrictions apply to the availability of the data that support the findings of this study, which were used under license for the current study, and so are not publicly available. The data are however available upon reasonable request for scientific purposes from the corresponding author Hendrik Schmitz (hendrik.schmitz@uni-paderborn.de).

Received: 15 March 2023; Accepted: 16 August 2023

Published online: 31 August 2023

References

1. World Health Organization. Daily cases and deaths by date reported to WHO. <https://covid19.who.int/WHO-COVID-19-global-data.csv> (2022).
2. Barber, R. M. *et al.* Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: A statistical analysis. *The Lancet* [https://doi.org/10.1016/S0140-6736\(22\)00484-6](https://doi.org/10.1016/S0140-6736(22)00484-6) (2021).
3. Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F. & Billa, L. Spatial analysis and GIS in the study of COVID-19. A review. *Sci. Total Environ.* **739**, 140033. <https://doi.org/10.1016/j.scitotenv.2020.140033> (2020).
4. Franch-Pardo, I., Desjardins, M. R., Barea-Navarro, I. & Cerdà, A. A review of GIS methodologies to analyze the dynamics of COVID-19 in the second half of 2020. *Trans. GIS: TG* **25**, 2191–2239. <https://doi.org/10.1111/tgis.12792> (2021).
5. Nazia, N. *et al.* Methods used in the spatial and spatiotemporal analysis of COVID-19 epidemiology: A systematic review. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph19148267> (2022).
6. Alidadi, M. & Sharifi, A. Effects of the built environment and human factors on the spread of COVID-19: A systematic literature review. *Sci. Total Environ.* **850**, 158056. <https://doi.org/10.1016/j.scitotenv.2022.158056> (2022).
7. Khanijahani, A. Racial, ethnic, and socioeconomic disparities in confirmed COVID-19 cases and deaths in the United States: A county-level analysis as of November 2020. *Ethn. Health* **26**, 22–35. <https://doi.org/10.1080/13557858.2020.1853067> (2021).
8. Plümper, T. & Neumayer, E. The pandemic predominantly hits poor neighbourhoods? SARS-CoV-2 infections and COVID-19 fatalities in German districts. *Eur. J. Public Health* **30**, 1176–1180. <https://doi.org/10.1093/eurpub/ckaa168> (2020).
9. Rohleder, S., Costa, D. & Bozorgmehr, K. Area-level socioeconomic deprivation, non-national residency, and Covid-19 incidence: A longitudinal spatiotemporal analysis in Germany. *EClinicalMedicine* **49**, 101485. <https://doi.org/10.1016/j.eclinm.2022.101485> (2022).
10. Abedi, V. *et al.* Racial, economic, and health inequality and COVID-19 infection in the United States. *J. Racial Ethn. Health Disparities* **8**, 732–742. <https://doi.org/10.1007/s40615-020-00833-4> (2021).
11. Mari-Dell'Olmo, M. *et al.* Socioeconomic inequalities in COVID-19 in a European urban area: Two waves, two patterns. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph18031256> (2021).
12. Siljander, M., Uusitalo, R., Pellikka, P., Isosomppi, S. & Vapalahti, O. Spatiotemporal clustering patterns and sociodemographic determinants of COVID-19 (SARS-CoV-2) infections in Helsinki, Finland. *Spat. Spatio-temporal Epidemiol.* **41**, 100493. <https://doi.org/10.1016/j.sste.2022.100493> (2022).
13. Silva, J. & Ribeiro-Alves, M. Social inequalities and the pandemic of COVID-19: The case of Rio de Janeiro. *J. Epidemiol. Community Health* **75**, 975–979. <https://doi.org/10.1136/jech-2020-214724> (2021).
14. Hoebel, J. *et al.* Emerging socio-economic disparities in COVID-19-related deaths during the second pandemic wave in Germany. *Int. J. Infect. Dis.* **113**, 344–346. <https://doi.org/10.1016/j.ijid.2021.10.037> (2021).
15. Thomson, H., Petticrew, M. & Morrison, D. Health effects of housing improvement: Systematic review of intervention studies. *BMJ* **323**, 187–190. <https://doi.org/10.1136/bmj.323.7306.187> (2001).
16. Jo, Y., Hong, A. & Sung, H. Density or connectivity: What are the main causes of the spatial proliferation of COVID-19 in Korea?. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph18105084> (2021).
17. Ahmad, K. *et al.* Association of poor housing conditions with COVID-19 incidence and mortality across US counties. *PLoS ONE* **15**, e0241327. <https://doi.org/10.1371/journal.pone.0241327> (2020).
18. Lee, W. *et al.* Urban environments and COVID-19 in three Eastern states of the United States. *Sci. Total Environ.* **779**, 146334. <https://doi.org/10.1016/j.scitotenv.2021.146334> (2021).
19. Ghosh, A. *et al.* Association between overcrowded households, multigenerational households, and COVID-19: A cohort study. *Public Health* **198**, 273–279. <https://doi.org/10.1016/j.puhe.2021.07.039> (2021).
20. Davies, B. *et al.* Community factors and excess mortality in first wave of the COVID-19 pandemic in England. *Nat. Commun.* **12**, 3755. <https://doi.org/10.1038/s41467-021-23935-x> (2021).
21. Askitas, N., Tatsiramos, K. & Verheyden, B. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. *Sci. Rep.* **11**, 1–13. <https://doi.org/10.1038/s41598-021-81442-x> (2021).
22. Bárcena-Martín, E., Molina, J., Muñoz-Fernández, A. & Pérez-Moreno, S. Vulnerability and COVID-19 infection rates: A changing relationship during the first year of the pandemic. *Econ. Hum. Biol.* **47**, 101177. <https://doi.org/10.1016/j.ehb.2022.101177> (2022).
23. Bertocchi, G. & Dimico, A. JUE insight: COVID-19, race, and gender. *J. Urban Econ.* <https://doi.org/10.1016/j.jue.2022.103484> (2022).

24. Dave, D., Friedson, A., Matsuzawa, K., Sabia, J. J. & Safford, S. JUE Insight: Were urban cowboys enough to control COVID-19? Local shelter-in-place orders and coronavirus case growth. *J. Urban Econ.* **127**, 103294. <https://doi.org/10.1016/j.jue.2020.103294> (2022) (**JUE Insights: COVID-19 and Cities**).
25. Hörnig, L. & Schaffner, S. *FDZ data description: Smaragd-survey*. Tech. Rep, RWI—Leibniz-Institut für Wirtschaftsforschung (2023).

Acknowledgements

Funding from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, project: COVID-19 als Beschleuniger sozialer Ungleichheit—Kleinräumige Analysen mit deutschen Nachbarschaftsdaten, 492351283) is gratefully acknowledged. We acknowledge support for the publication cost by the Open Access Publication Fund of Paderborn University.

Author contributions

D.E.: Data analysis, Writing. V.S.: Data analysis, Writing. H.S.: Conceptualization, Project administration, Data analysis, Funding acquisition, Writing.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-40734-0>.

Correspondence and requests for materials should be addressed to H.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023