Landscape and Urban Planning

An exploration of factors characterising unusual spatial clusters of COVID-19 cases in the East Midlands region, UK: a geospatial analysis of ambulance 999 data.

--Manuscript Draft--

Manuscript Number:	LANDUP-D-21-00412R1
Article Type:	VSI: EIDs and Planning
Section/Category:	Undecided
Keywords:	COVID-19, vulnerability, bioecological model, exposure, underlying susceptibility, built environments
Corresponding Author:	Harriet Elizabeth Moore, PhD University of Lincoln College of Science Lincoln, Lincolnshire UNITED KINGDOM
First Author:	Harriet Elizabeth Moore, PhD
Order of Authors:	Harriet Elizabeth Moore, PhD
	Bartholomew Hill
	Mark Gussy
	Niro Siriwardena
	Frank Tanser
	Robert Spaight
	Graham Law
	Chris Thomas
Abstract:	Complex interactions between physical landscapes and social factors increase vulnerability to emerging infections and their sequelae. Relative vulnerability to severe illness and/or death (VSID) depends on risk and extent of exposure to a virus and underlying health susceptibility. Identifying vulnerable communities and the regions they inhabit in real time is essential for effective rapid response to a new pandemic, such as COVID-19. In the period between first confirmed cases and the introduction of widespread community testing, ambulance records of suspected severe illness from COVID-19 could be used to identify vulnerable communities and regions and rapidly appraise factors that may explain VSID. We analyse the spatial distribution of more than 10,000 suspected severe COVID-19 cases using records of provisional diagnoses made by trained paramedics attending medical emergencies. We identify 13 clusters of severe illness likely related to COVID-19 occurring in the East Midlands of the UK and present an in-depth analysis of those clusters, including urban and rural dynamics, the physical characteristics of landscapes, and socio-economic conditions. Our findings suggest that the dynamics of VSID vary depending on wider geographic location. Vulnerable communities and regions occur in more deprived urban centres as well as more affluent peri-urban and rural areas. This methodology could contribute to the development of a rapid national response to support vulnerable communities during emerging pandemics in real time to save lives.
Suggested Reviewers:	Alex Lechner ALechner@lincoln.ac.uk I don't know if our manuscript will go to the same/different reviewers as before, or if our suggestions were taken up. Alex is a geospatial expert.
	Gary Bosworth gary.bosworth@northumbria.ac.uk Rurality specialist
	Maxime Inghels MInghels@lincoln.ac.uk

	Diego Cuadros cuadrodo@ucmail.uc.edu
	Andrew Tomita atomita@ahri.org
Response to Reviewers:	

26th March 2021

Addressing the editor,

Thank you for considering our manuscript for publication in this special edition of *Landscapes and Urban Planning.* The submitted manuscript is revised as instructed following the feedback we received earlier in the year (LANDUP-D-20-00215). The editor (Joan Iverson Nassauer) requested that we re-submit including a cover letter outlining our intention to resubmit to the special edition. Our submission includes an in-depth table of revisions addressing each concern raised by the three reviewers and Guest Editor (Jim Spencer). The two main concerns raised were related to the reliability of the data and the theoretical framing of the problem. We feel that both have been addressed adequately. We have included much more detail about the validity of the data, including supplementary material demonstrating a strong correlation between suspected daily cases and confirmed daily cases for our region over the period of our research. Regarding framing, we have restructured the research around bioecological models.

In addition to the current manuscript, we previously submitted a perspective piece which is complimentary to the research presented here. In light of the changes made to the current manuscript, we will also need to make some revisions to the perspective piece. If the current manuscript is selected for publication, we would be happy to revise the perspective piece for the special edition, or for a later edition. The second manuscript is brief and could be ready for resubmission shortly if desired.

Kind regards,

Dr Harriet Elizabeth Moore,

On behalf of Professor's Siriwardena, Law, Thomas, Gussy and Tanser, Bartholomew Hill, and Robert Spaight, and the EDGE Consortium.

Response to reviewers for LANDUP-D-21-00412

Addressing the Co-Chief-in-Editor and Special Edition Guest Editor,

Thank you for the opportunity to resubmit our manuscript LANDUP-D-21-00412. We have considered all comments from the Guest Editor and two Reviewers carefully and made further substantial changes to the manuscript which we feel improves the overall quality and clarity. The table below details the changes we have made and, in some cases, our explanation for why requested changes have not been made. As instructed, we have prioritised the comments and suggestions of the Guest Editor and endeavoured to address these as clearly as possibly in our revisions.

Kindly,

Dr Harriet Moore, on behalf of the research team.

Reviewer Comment	**Response**	**Section**
	Guest Editor	
Thank you for resubmitting the manuscript for review. i believe that you have a made substantive changes and explained your thinking in several places that clarifies the intents of the paper. In general, i think that the paper is publishable with some revisions based mostly on Reviewer #1. While Reviewer #2 brings up some interesting points, I believe that you have framed this paper's contributions as related to a methodology for targeting public/medical efforts rather than as a set of established "scientific" findings. In this way, the simple techniques of geocoding 999 calls is something of possible use to health professionals - is this the first time used? What kinds of policies might it imply?	We thank the Guest Editor for the recommendation to publish with revisions. We agree that the contribution of the manuscript is in part methodological. However, we feel the study also sheds light on vulnerable regions and factors that may explain vulnerability in relation to COVID-19. Suspected cases of respiratory virus have been used elsewhere to consider early patterns of vulnerability during the H1N1 outbreak in London (Balasegaram et al., 2012). Balasegaram et al. (2012) drew links between vulnerability and deprivation using clinical reports of suspected H1N1. This study linked resident postcodes to IMD scores to explore the relationship between poverty and severe illness. Our study uses the clinical reports of paramedics to identify vulnerable regions in real time during a pandemic, as well as to explore landscape scale factors, including deprivation, that may explain vulnerability. As the Guest Editor suggests, our methodology is a novel approach that health professionals could use in the early months of a pandemic to identify communities who may be vulnerable to severe illness and require additional support from community medical services. As suggested, we have included more specific mention of the novelty of the approach in the introduction section: *To our knowledge, ambulance data have not previously been utilised to identify communities that may be vulnerable to severe illness from COVID-19 or to investigate social and environmental factors that may influence vulnerability. Our novel methodology presents an opportunity for health professionals to identify and support vulnerable communities who are likely to*	Section 1

	experience severe illness from a new virus in the early phase of a pandemic.	
? If the contribution is the method, then I do think, however, that the paper needs to reinforce this basic contribution better and make it clearer, especially in the context of the landscapes that define these clusters (e.g. tobacconists, fast-food outlets, bars, etc)? What might a cluster-based intervention related to these landscapes look like. these are speculative suggestions, but based on preliminary analysis. Doing this will force you to develop a proposition associated with each cluster landscape variable.	Thank you, we agree and have clarified the landscape-related features that need to be considered in a cluster-based intervention. Our study uses the AHAHI which was developed drawing on the evidence which links landscape-related features and health outcomes. In the manuscript (Section 3.3) we specify: *According to Daras et al. (2019) healthy features of landscapes associated with more positive health outcomes include closer proximity to active and passive green space as well as health services, such as general practitioners and emergency departments. Hazardous features of landscapes associated with poorer health outcomes include poor air quality, further distance from healthy features of landscapes, and closer proximity to retail vendors like fast food outlets, tobacconists, off-license stores, pubs, bars, and clubs.* We have added a brief statement to this section outlining a proposition for the association between healthy features of landscapes and expected health outcomes related to severe cases of COVID-19, and the associated between hazardous features of landscapes and expected health outcomes related to severe cases of COVID-19: *The AHAHI is a validated metric that synthesises features of built environments that are commonly related to health outcomes in the wider health literature (Green et al., 2018)* *According to Daras et al. (2019) healthy features of landscapes associated with more positive health outcomes include closer proximity to active and passive green space as well as health services, such as general practitioners and emergency departments. This is because physical access to health services is associated with health service use and health maintenance. Thus, we might expect to observe lower rates of suspected severe COVID-19 located nearer to the health services and healthy features of physical environments included in our analysis, such as green and blue spaces (Table 1). We might also expect lower rates in areas with better air quality, including lower levels of Nitrogen Dioxide, Particulate Matter, Sulphur Dioxide.* *Hazardous features of landscapes associated with poorer health outcomes include poor air quality, further distance from healthy features of landscapes, and closer*	

	proximity to retail vendors like fast food outlets, tobacconists, off-license stores, pubs, bars, and clubs. This is because distance from hazardous retail environments is a proxy measure for individual behaivour; people who live closer to fast food outlets are more likely to consume fast food, and subsequently to experience underlying health conditions like diabetes and obesity (Green et al., 2018). On this basis, we could anticipate that unusual clusters of suspected COVID-19 are likely to occur closer to retail vendors, further from health services and healthy physical environments, and in areas with poorer air quality. We hope that these alterations clarify the relationship between landscape features and health outcomes. Regarding how we have interpreted the results of our analysis and how these results translate into recommendations, please see the below. Further, we emphasise that our analysis considers the cumulative effect of numerous health related behaviours rather than the specific effect of each on health outcomes. This is intentional and is the novelty of our approach. Underlying health problems related to severe COVID-19, such as diabetes and obesity, are associated with multiple lifestyle characteristics (e.g., not exercising, poor diet, smoking) rather than one specific element of lifestyles. These lifestyle characteristics tend to cluster together. Thus, from a policy perspective, it is very difficult to target recommendations at specific retail outlets. A more pragmatic approach is to reinforce the need for urban planning and economic incentives that encourage low-income families to lead healthier lifestyles. See below.	
Why might these businesses be relevant? Because there is a proprietor there managing people, or because it is where people are likely to get infected; if the latter, then the intervention might be a first aid kit for the businesses, or better instructions on public walls for what to do, while if the latter, then maybe a targeted approach for examining wastewater for COVID-19 viruses (something that is currently being done for hypothesized clusters).	Thank you for the request for clarification. The relationship between landscape features and health outcomes varies according to features of landscapes related to *exposure* and features of landscapes related to *underlying susceptibility*. In some cases, such as passive green space, it is appropriate to suggest interventions such as stricter enforcement of social distancing in parks and arboretums (see below). In other cases, such as access to tobacconists, fast food outlets, or off-licenses, health outcomes reflect individual choice and behaviours that may increase susceptibility to severe illness, which may be amenable to public health measures.	See changes to Section 6.2 and Section 7 below
Some other important revisions and considerations: 1) reviewer 2 notes that your model is able to detect only 42% of cluster membership, but you also state that it captures 99% of non-cluster cases. Rather	The clusters were identified through spatial analysis using SatScan software. The regression analysis is separate and explores features of landscapes that explain whether an individual case falls into a cluster or occurs outside of a cluster. The regression model does	Section 6.1

than focus on these numbers as facts, it can clarify what this means in practice for an emergency manager. E.g. they can be certain that a "non-cluster" case should be treated as a "one-off" kind of case and further surveillance is not needed in that area. This is the kind of policy interpretation needed to translate the numeric findings into recommended action, and what makes a methodology defined in real time utility useful in decision-making, even if it is scientifically imperfect.	not alter which cases occur in clusters or outside of clusters, it demonstrates how well the factors in the model predict cluster/non-cluster membership. The location of clusters remains the same regardless of the predictive accuracy of the regression model. Overall good model fit for binary logistic regression is around 70%. Our overall model fit is much higher which suggests that the factors in the model explain whether individual cases fall into clusters or occur outside of clusters well. The predictive accuracy for cases in clusters is acceptable even though it is lower than for cases outside of clusters. This is not uncommon when the number of cases in one condition (non-cluster) is substantially larger than the number of cases in the second condition (cluster). In the manuscript in a footnote: *The asymmetry in predictive accuracy for cases appearing in clusters compared to cases not appearing in clusters is a common phenomenon of highly unequal datasets (Calabrese, 2014) and reflects the true rarity of cases appearing in clusters.* Further, the predictive accuracy of a binary logistic model does not translate into the proportion of times policy makers should target a region for intervention (with reference to Reviewer Two comments). The model fit indicates the proportion of variance observed that is accurately predicted by the model. It is an indication of how likely the factors in the model are to explain group membership. For behavioural research, as compared to physical research (e.g., using environmental data) predictive accuracy above 50% is very unusual. This does not negate the value of the findings. We have added a comment to the discussion that addresses the varying predictive accuracy of the model for each condition (Section 6.1).	
2) in the introductory sections of the paper there are a number of typos, and in one case i am unsure if you mean "usual" or "unusual" clusters (an important distinction). beyond these, however, i think that greater and more consistent precision on what is meant can clarify a lot of issues that make it more confusing than need be. For example, in teh abstract: "exposure" = "geographic exposure"? "susceptibility" = "individual susceptibility"? "vulnerable communities"="vulnerable geographic	Thank you for these suggestions. We have made numerous changes to text to clarify these, and other, points, as advised. However, we are unclear on some of these suggestions. We have defined susceptibility as related to underlying health issues. Exposure is related to exposure to the COVID-19 virus. Exposure can occur for numerous reasons which are explored in the manuscript, such as urban density, crowded living arrangements, or nature of employment. These ideas cannot be summed up as one type of 'exposure'. These ideas are expounded in	Section 1

communities"? "factors" = "geographic factors"? These qualifiers can clarify some of the issues that seem to have confused reviewer 2. Please look through the entire initial section and see where you can be more specific to keep the analysis clearer.	the body of the manuscript. The abstract is necessarily a summary of the key ideas. We have changed all references in the introduction of 'communities' to 'communities and regions' to clarify for Reviewer 2.	
3) the primary methodological weakness - and one that the paper seems not to address adequately - is the 999 call denominator. The model adequately controls for population density, but does it account for selection bias in all 999 calls? In other words, some areas may have no telecom access, so no 999 calls come in for anything (including but not limited to COVID calls). This does not mean that COVID-19 does not occur there, but it never show up in the data. some recognition of this and possible implications is needed, and i did not see it (though perhaps I missed it). adn again, linking it to policy interventions, the model shows that there are clusters of COVID cases, even though this methodological limitation may mean that there are disproportionately larger clusters in 999 call deserts, it can still be useful. The policy goal is the identify large numbers of cases to target effectively, while the scientific goal may be to optimize resource investment. The latter "perfect" should not be the enemy of the "good" former.	We agree, this is a common feature of health services data which captures people who present to and are able to access services rather than those who choose not to or are unable to use services, sometimes referred to as the 'iceberg' of illness (Hannay 1980). Our analysis focusses on 'severe illness' from COVID-19 where emergency services are required to transport patients to hospital rather than all such cases and have ensured this is a consistent narrative throughout the manuscript. The policy goal, in the context of our manuscript, is to identify regions with communities that are more vulnerable to severe illness leading to an ambulance call, rather than to identify regions with the highest number of community cases. Further, the cluster analysis conducted in SatScan scans for unusual clusters *per population*. Thus, regardless of whether some cases of suspected COVID-19 are missing from our data, the identified clusters are valid. In addition, any missing data that occurred at random would bias the findings toward the null hypothesis of complete spatial randomness, but could not account for the detection of a spurious cluster. In the current version of the manuscript, this issue is addressed in the limitations: *Secondly, factors beyond the scope and scale of this research may affect ambulance use. People within close proximity to hospitals with A&E services are more likely to access those services directly rather than calling an ambulance. Similarly, willingness to call an ambulance may vary between communities. Poor health literacy, including ability to recognize symptoms of illness, is often associated with deprivation (Niksic et al., 2015). Thus, qualitative community scale research is needed to ground truth the trends and associations reported here.* We have added the following:	Section 6.5

	As a result, it is likely that our data does not represent all severe cases of suspected COVID-19 in the study region.	
overall, I think you can tighten up the "scientific" parts of the analysis and identify the core weaknesses along this angle, but then emphasize how the method can be useful and open up new modes of action that might be low-impact and easily implemented. On this latter point, there is a forthcoming article in Journal of the American Planning Association (Spencer, Marasco, and Eichinger) titled: "Planning for Emerging Infectious Disease Pandemics: Upstream Causes and Proportional Responses during the case of Avian Influenza 2004-2005.," that identifies the pressing need to find a range of politically palatable policy options for controlling pandemics that go beyond vaccination. This is something you might use to suggest why understanding clusters in real time is important.	Thank you. We like the idea of providing more specific recommendations for potentially 'politically palatable policy options'. Unfortunately, we are unable to access the suggested forthcoming publication as it is not yet publicly available. In the original manuscript there was a focus on identifying regions for localised lockdowns and more targeted intervention. We removed much of this content in response to the previous reviews which suggested not to focus on transmission. The current manuscript refers to implications for more targeted policies for social distancing in the conclusion. In the revision, we have reintroduced further detail in the conclusion about how policy makers might apply our methodology to inform localised mitigation measures. Specifically, we emphasise the possibility of utilising our findings for early localised lockdowns, as well as longer-term approaches to reducing vulnerability, such as some options for promoting health behaviours in deprived communities and increasing access to health services in isolated communities.	Section 7
Your core argument is that real-time assessment of unusual clusters can suggest place-based interventions that few public health and medical officials might even think of, let alone have any tentative data to support action/intervention. Spencer et.al (2020) in Landscape and Urban Planning has some similar types of recommendations regarding geographically targeted surveillance related to the household built environment.	Our core argument is that real-time analysis can offer more targeted and timely intervention. We also reflect on some possible nuances that could be made to existing mitigation policies (see below). Underlying susceptibility, as well as exposure, are factors in severe illness. Underlying susceptibility is partly the product of long-term individual choices, behaviours etc. This component of vulnerability cannot be addressed as a pandemic unfolds. Rather, communities who are vulnerable in this way can be targeted for better support earlier on. Spencer et al. (2020) highlight problems with developing urban areas, such as inadequate water and waste removal infrastructure, that are common to developing countries. Sanitation is a huge factor in exposure in poorer regions. There are some opportunities for novel interventions informed by our analysis, such as our suggestion about monitoring green space use below. However, many of the factors known to influence illness severity are related to underlying susceptibility and individual health behaviours, which in turn are related to systemic issues like deprivation, education, and systemic inequality. Our	Section 6.2, Section 7, Section 6.4

	findings emphasise the need of the UK Government to 'level up health', rather than specific new interventions that can prevent severe illness from occurring in vulnerable communities and regions. Following the advice of the Guest Editor, in our revised manuscript we include some additional reflections and recommendations for mitigating the impact of systemic inequalities on vulnerable communities during a pandemic. Specifically, we have included the following: *Our findings about vulnerability in rural areas suggest some policy responses for future pandemics and phases of lockdown. News reporting during the first national phase of lockdown suggests that the public viewed rural areas as less vulnerable to contagion and mortality related to COVID-19 compared to urban areas (e.g., McCarthy, 2020).Further, rural communities reported the phenomenon of people from urban and peri-urban areas 'flocking' to rural regions for recreation during phases of lockdown when only essential travel was legally permitted (Asquith, 2020). In the event of future phases of lockdown, mitigating high rates of severe illness in rural areas with aging populations may require more stringent policing of travel between urban and rural areas.* (Section 6.4) *It is possible that proximity to Passive Green Space reflects social behaviour during the pandemic. In a perspective piece published in this Special Edition we examine the relationship between landscape features and the implications for COVID-19 exposure and underlying susceptibility in more depth. During extended phases of lockdown parks and arboretums became social hubs that were poorly monitored by local authorities. News reports documented continual violations of social distancing rules in public spaces like beaches and common green areas. Thus, improving the monitoring and enforcement of social distancing in these spaces may be a future avenue for reducing rates of severe COVID-19.* (Section 6.2) • *Identifying vulnerable communities in real-time could inform earlier localised lockdowns to mitigate transmission and reduce rates of severe illness. Targeting areas where contagion is likely to result in high rates of hospitalisation would also reduce burden on emergency medical services;*	

<table>
<tr><td></td><td>

Opportunities for mitigating transmission also include more effective monitoring and enforcement of social distancing rules in Passive Green Space, including parks, commons and arboretums;
The dynamics of vulnerability vary between urban centres and more peripheral or rural regions, and between more deprived compared to more affluent communities. The opportunities for minimising the impacts of a pandemic include reducing the underlying susceptibility of communities as well as minimising transmission. In part, this involves urban planning to enhance opportunities for health behaviours. Improving the safety of green spaces for cost-free exercise and increasing infrastructural and financial access to healthy food would promote healthier lifestyles in deprived communities. Further, improving access to health services in more affluent and isolated communities may help to mitigate the most severe outcomes of a pandemic. However, in both cases this requires top-down financial investment to encourage healthy retail outlets to locate in deprived neighbourhoods, and health services to locate in low-density neighbourhoods.

(Section 7)
We hope that these additions address the concerns of the Guest Editor.
</td><td></td></tr>
<tr><td colspan="3" align="center">Reviewer One</td></tr>
<tr><td>One of the previous recommendations was to better explain the relationship between communicable, infectious diseases and the urban form (population density, housing crowdedness, etc.). The authors have not explicitly explained this relationship. There is plenty of public health literature that explores this relationship, especially related to respiratory (i.e. Tuberculosis) and diarrheal diseases. With COVID being an infectious, communicable respiratory disease, the relationship between urban/community characteristics and communicable diseases need to be explicitly explained.</td><td>
Section 2 (Conceptualising the relationship between severe suspected COVID-19 cases and built environments) considers the relationship between features of built environments and severe illness form infectious disease, including the following:
In the case of COVID-19, the relationship between severe illness and characteristics of the built environment involves both direct and indirect pathways. On the one hand, environments can influence the direct exposure of individuals to communicable disease. On the other hand, landscape features can indirectly affect the underlying susceptibility of communities to severe symptoms, compared to experiencing mild symptoms or presenting as asymptomatic, by supporting or preventing healthy lifestyles. Features of neighbourhoods that can influence health behaviours like exercise include access to green space for passive recreation such as walking, and facilities for active exercise, such as sports grounds or
</td><td>Section 1</td></tr>
</table>

	leisure centres (Hartig et al., 2020). Further distance from these healthy landscape features is associated with lower levels of activity and higher risk of cardiovascular disease (Shen & Lung, 2016) and obesity (Lachowycz & Jones, 2011). However, landscape features can also reflect the social characteristics of wider living environments; high crime rates co-occur with poor physical infrastructure like housing in deprived communities. Crime can deter access to nearby outdoor spaces (Gomez et al., 2004) while poor housing indicates lower incomes and a greater likelihood of underlying chronic health conditions (Krieger & Higgins, 2002). The focus of our manuscript is not exclusively urban landscapes and the transmission of infectious disease; we focus specifically on *urban landscapes and severe illness from infectious disease*. The fact that COVID-19 is communicable is only one component of severity. However, we have added a statement to the introduction to make the relationship between urban space and infectious disease explicit: *Characteristics of urban landscapes that are typically associated with the transmission of infectious diseases include population and employment density (Hu et al., 2013), and housing crowdedness (Low et al., 2013; Neiderud et al., 2015). However, these relationships are rapidly changing and vary depending on region and specific location within urban areas. For example, extended urbanisation is shifting the dynamics of vulnerability; in some cases, communities on urban peripheries may be more vulnerable than those in denser urban centres with greater access to healthcare and social support (Connolly et al., 2021). Indeed, in the case of COVID-19, typical relationships between urban space and infectious disease do not consistently explain mortality, with high rates of severe cases occurring in less dense urban areas (Frank & Wali, 2021). Thus, there is a need to consider how urban landscapes influence the underlying susceptibility of communities to severe illness as well as exposure to infectious diseases.* (Section 1)	
The authors addressed the recommendation of including a framework to justify the selection of variables in the study. However, the authors' application of the Bronfenbrenner's conceptual models focus on the social contexts, not necessarily on the social and environmental contexts. The	McLeroy et al (1988) do not address physical environmental factors. The research (*An ecological perspective on health promotion programs*) refers exclusively to ‘social environmental factors’ including the following: intrapersonal factors (e.g., knowledge, skills), interpersonal processes (e.g., social networks), institutional factors (e.g., organisational arrangements),	N/A

McLeroy et al (1988) offers a socio-ecological framework that addresses the environmental context that influences health.	community factors (e.g., relationships between organisations), and public policy. There is no mention of the physical environment. The use of the term 'environment' in socio-ecological modelling typically refers to 'wider social context' rather than the physical environment. Our adaption of Bronfenbrenner's model introduces elements of the physical environment. This is a novel adaptation of the model.	
The introduction mentioned the risk of severe COVID cases vary by ethnicity, however the study does not include the racial/ethnic makeup of communities, specifically in the clusters. Also, are racial/ethnic data collected from 999 responses?	Ethnicity is mentioned as an example of demographic characteristics. We are clear in the manuscript that we are unable to include ethnicity in our analysis. Section 3.3 states: *While ethnicity is also commonly associated with severe symptoms (Sze et al., 2020), reliable data was unavailable in real-time.*	N/A
In the data collection section, it is unclear if the 999 data are geocoded by address then aggregated by postal code, or if the data were only aggregated by postal code.	Section 3.4 states: *The database of suspected COVID-19 cases was obtained from EMAS[1], including the date 999 calls were received, partial postcodes of ambulance attendance locations, sex, and age.* We have made this adjustment to be clearer: *The database of suspected COVID-19 cases was obtained from EMAS[2], including the date 999 calls were received, partial postcodes* ***(rather than full addresses)*** *of ambulance attendance locations, sex, and age.*	Section 3.4
The author used the term "suspected COVID" cases and "suspected severe COVID" cases. How is "severe" defined?	We have added this clarifying statement to Section 1: *In this context, 'severe illness' refers to patients presenting with severe symptoms that require the attendance of emergency medical services.*	Section 1
When using the SatScan, can you control for size in determining cluster area? What is the window area compared to the study area? I recommend describing the clusters in terms of size relative to the community with incidence rates.	Section 3.5.1 specifies: *The Poisson Model was purely spatial. The model parameters included unconstrained spatial cluster size, and the criteria for reporting hierarchical clusters was set to 'no cluster centres in other clusters'.* This approach is preferable to setting a specific cluster size because the size of the cluster is empirically determined from the data.	N/A

[1] This research, including use of patient data for statistical and spatial analysis, was approved by the NHS Health Research Authority, IRAS ID: 264573.

	Table 2 describes the clusters relative to population size and includes relative risk which is the same as incidence rate.	
In line 207 page 6, "ethnicity is also commonly associated with severe symptoms". What ethnic groups are you referring to?	Ethnicity refers to the ethnicity of the individual, not to a specific ethnic group. All ethnicity is associated with severe symptoms. 'Associated' means 'less likely' as well as 'more likely'. Some ethnic groups are more/less likely to experience severe symptoms. The variable 'ethnicity' is related to severity.	N/A
There are several typos throughout the document. Different Font size on p. 2 line 76. What is GDPR (in line p. 2 line 84)? Lines 253-254 needs clarity. Line 311- Both needs a lowercase b. Line 314- raster's and polygons. Figure 3, please list what the values mean for each cluster? This is explained later, but needs to be explained earlier. Line 13 page 12- AHAHI? There are more typos further in the document.	Thank you for picking up these errors. We have made most of the suggested changes. We are unsure what values the reviewer suggests including in Figure 3. The figure is intended to show location rather than the characteristics of each cluster which is detailed elsewhere. There are no visible indictors of these characteristics in the figure. We are unsure what the error is regarding AHAHI. We have introduced the Index earlier, using the acronym in the remainder of the document is appropriate. We have thoroughly checked the remainder of the document for typos and made numerous changes.	Throughout
The use of acronyms and spelling these out are inconsistent throughout the document.	The study involves numerous acronyms. We have attempted to remind the reader of the meaning of long acronyms (such as AHAHI) at appropriate points in the manuscript. We have added the acronym after every full use. Full titles are acronyms are now included in each figure caption.	Throughout
The first 3 sections need better organization	Reviewer Two and the Guest Editor have not raised this as an issue.	N/A
	Reviewer Two	
Thank you for submitting your revised manuscript, which is better organized and describes more clearly the method and caveats of the study.	Thank you.	N/A
Unfortunately I am still struggling to match the claims of the paper ("these analyses offer a real-time approach for identifying and protecting vulnerable communities in the critical early stages between the first confirmed case of a new EID and widespread community testing") and the strength of the model itself. The model predicts less than 42% of cases in "unusual clusters," and the unusual clusters themselves tell very different stories. The authors conclude that "Taken together, these observations suggest that the relative contribution of demographic, socio-	Please see responses to the Guest Editor above. Regarding contribution, our analysis elucidates complex relationships; the factors that may explain vulnerability *vary spatially*. Deprivation is important in some contexts and less important in others. By comparing clusters in rural/urban and inland/coaster locations, our findings go beyond identifying individual determinants, and instead speak to the geographic surroundings of communities. For example, urban epidemiology research tends to suggest that high density urban areas are most at risk of infectious diseases. Our analysis suggests that the relationship between population and severe illness is	

economic, and environmental factors to vulnerability varies depending on wider geographic location." I am struggling to see how these findings advance the original claim of being able to "identify and protect vulnerable communities." The model results do not tell a particularly clear story (see Table 6), nor seem to advance urban plannings' understanding of pandemic contagion beyond a general and uncontroversial finding that physical AND social factors matter. The interpretation of findings from the perspective of urban and regional planning or urban spatial theory is weak, presenting the results of the model list-wise but not providing much meaningful interpretation or generalization.	more complex; landscapes interact with people in multiple ways, influencing their underlying health as well as their exposure to disease. Together these factors influence whether or not a virus like COVID-19 will result in severe illness requiring costly emergency services.	
Returning to the central claim of the paper, I am struggling to see how a public health professional would use these findings to meaningfully improve surveillance of a fast-moving pandemic - given the low power of the model and highly site- and context-specific nature of the findings.	See response to the Guest Editor above.	N/A

References:

Balasegaram, S., Ogilvie, F., Glasswell, A., Anderson, C., Cleary, V., Turbitt, D. and McCloskey, B., 2012. Patterns of early transmission of pandemic influenza in London–link with deprivation. *Influenza and Other Respiratory Viruses*, 6(3), pp.e35-e41.

Daras, K., Davies, A., Green, M. and Singleton, A., 2018. Developing indicators for measuring health-related features of neighbourhoods. *Consumer Data Research*, pp.167-77.

Daras, Konstantinos., Green, Mark A., Davies, Alec., Singleton, Alex., Barr, Benjamin., 2019. Access to Healthy Assets and Hazards (AHAH). https://doi.org/10.6084/m9.figshare.8295842.v1, viewed 5 August 2020.

Hannay, D.R., 1980. The 'iceberg' of illness and 'trivial' consultations. *The Journal of the Royal College of General Practitioners*, *30*(218), pp.551-554.

Highlights:

- Air quality and features of urban landscapes are risk factors for COVID-19.
- Deprived areas face different challenges for mitigating contagion compared to affluent areas.
- Identifying clusters of COVID-19 transmission could be used to inform 'isolate, test, trace'.
- Ambulance calls reflect acute cases of COVID-19 and could be used for pre-hospital triage.
- Factors that are associated with clusters are highly location specific.

An exploration of factors characterising unusual spatial clusters of COVID-19 cases in the East Midlands region, UK: a geospatial analysis of ambulance 999 data.

[1]Harriet Elizabeth Moore
[2]Bartholomew Hill
[2]Niro Siriwardena
[2]Graham Law
[2]Chris Thomas
[2]Mark Gussy
[2]Robert Spaight
[2]Frank Tanser

[1]DIRE Research Group Lead
[1,2]EDGE Consortium affiliates

An exploration of factors characterising unusual spatial clusters of COVID-19 cases in the East Midlands region, UK: a geospatial analysis of ambulance 999 data.

Abstract:

Complex interactions between physical landscapes and social factors increase vulnerability to emerging infections and their sequelae. Relative vulnerability to severe illness and/or death (VSID) depends on risk and extent of exposure to a virus and underlying health susceptibility. Identifying vulnerable communities and the regions they inhabit in real time is essential for effective rapid response to a new pandemic, such as COVID-19. In the period between first confirmed cases and the introduction of widespread community testing, ambulance records of suspected severe illness from COVID-19 could be used to identify vulnerable communities and regions and rapidly appraise factors that may explain VSID. We analyse the spatial distribution of more than 10,000 suspected severe COVID-19 cases using records of provisional diagnoses made by trained paramedics attending medical emergencies. We identify 13 clusters of severe illness likely related to COVID-19 occurring in the East Midlands of the UK and present an in-depth analysis of those clusters, including urban and rural dynamics, the physical characteristics of landscapes, and socio-economic conditions. Our findings suggest that the dynamics of VSID vary depending on wider geographic location. Vulnerable communities and regions occur in more deprived urban centres as well as more affluent peri-urban and rural areas. This methodology could contribute to the development of a rapid national response to support vulnerable communities during emerging pandemics in real time to save lives.

1. Introduction:

There is growing recognition in the fields of epidemiology (Lofus, 2004; Miller et al., 2012; Norris et al., 2020; Viegi et al., 2006), and urban planning (Durand et al., 2010; Northridge et al., 2003; Seo et al., 2019; Spence et al., 2020) that complex interactions between physical landscapes and social factors drive vulnerability to contagion and severe illness, and thus, understanding these mechanisms should be the focus of disease prevention and management. Urban landscapes can simultaneously influence chronic conditions associated with *susceptibility* to severe pathogenic illness, such as obesity by facilitating access to fast food vendors (Daras et al., 2018), and increase *exposure* to disease through high density urbanism (Goryakin et al., 2017; Wu et al., 2016).

On March 11th, 2020 the World Health Organization declared the novel coronavirus disease 2019 (COVID-19) a global pandemic. Over the subsequent months, the research community has prioritised understanding contagion and transmission pathways (Park et al., 2020) as well as identifying and supporting vulnerable communities and regions (Daras et al., 2021; Khalatbari-Soltani et al., 2020; Patel et al., 2020). Vulnerable communities include those with pre-existing chronic conditions that are known to increase susceptibility to severe COVID-19 illness, such as diabetes (Peric & Stulnig, 2020) and overweight or obesity (Steinberg et al., 2020). Medical and public health research has led to widespread appreciation that the impact of the pandemic has, and continues to be, heterogenous; some communities and regions appear to be more vulnerable to severe illness than others (Marmot et al., 2020; Patel et al., 2020).

One common observation is that urban and peri-urban areas tend to be sites of high rates of infection and mortality compared to more dispersed rural areas (Stier et al., 2020). High rates of infection suggest that urban landscapes are more exposed to transmission, and, given that many cases of infection are asymptomatic or involve very mild symptoms (Kim et al., 2020), high rates of mortality may indicate that communities within those landscapes are also more susceptible to severe illness (Guilmoto et al., 2020). Characteristics of urban landscapes that are typically associated with the transmission of infectious diseases include population and employment density (Hu et al., 2013), and housing crowdedness (Low et al., 2013; Neiderud et al., 2015). However, these relationships are rapidly changing and vary depending on region and specific location within urban areas. For example, 'extended urbanisation' is shifting the dynamics of vulnerability; in some cases, communities on urban peripheries may be more vulnerable than those in denser urban centres with greater access to healthcare and social support (Connolly et al., 2021). Indeed, in the case of COVID-19, typical relationships between urban space and infectious disease do not consistently explain mortality, with high rates of severe cases occurring in less dense urban areas (Frank & Wali, 2021). Thus, there is a need to consider how urban landscapes influence the underlying susceptibility of communities to severe illness as well as exposure to infectious diseases.

Surprisingly, of more than 40,000 papers using clinical diagnoses of COVID-19 published in 2020, fewer than 150 have considered the implications for urban planning, and those that do tend to focus on the impact that 'lockdown' has had on urban landscapes, such as pollutant rates, rather than the impact of urban landscapes on health outcomes (Sharif & Khavarian-Garmsir, 2020). In this study, we consider the relationship between built environments, exposure to emerging infectious diseases (EIDs), and susceptibility to severe COVID-19 illness provisionally diagnosed by medically trained professionals. In this context, 'severe illness' refers to patients presenting with severe symptoms that require the attendance of emergency medical services.

Efforts to shield the most vulnerable communities and regions in society are more likely to be effective if they happen rapidly and in real-time (Kasda et al., 2020). Compared to other common disasters like flooding, obtaining geographically accurate data to evaluate the spatial dimensions of a pandemic, and to support vulnerable communities and regions, faces unique challenges. In the UK, the use of contact tracing and testing to identify community cases of COVID-19 commenced after the first confirmed case on January 31st, 2020. However, community testing ceased in early March as cases rose rapidly and the virus was classified as a category 3 pathogen, confining testing to level 3 laboratories. On April 2nd the UK government outlined a five-pillar strategy for expanding testing capacity including the introduction of community testing in early May, with laboratory capacity expanding rapidly over the next two months (The Health Foundation, 2020).

In the interim before laboratory capacity increased, identification of severe illness and vulnerable communities was based on individual's self-reporting of symptoms, such as via the NHS Test and Trace App, and the clinical judgement of medical professionals. Routine medical data collected within the first hour of emergency department (ED) admittance for severe illness has since been demonstrated to predict positive COVID-19 cases with a high degree of accuracy, including self-reported olfactory and taste dysfunction (OTD) (Wee, 2020; Patterson, 2020; Printza & Constantinidis, 2020) and blood oxygen levels (Soltan et al., 2021). These measures, guided by Public Health England's case definition symptom criteria (PHE, 2020) are also used by paramedic clinicians attending ambulance call outs for suspected COVID-19 cases.

Our research involves analysing records of provisional diagnoses of COVID-19 made by trained paramedics attending medical emergencies in the East Midlands of the UK. In S-1 we present a preliminary analysis comparing daily rates of suspected severe COVID-19 cases from our data set obtained from the East Midlands Ambulance NHS Trust (EMAS) to retrospective records of daily rates of cases confirmed by polymerase chain reaction (PCR) testing for the same region. The results indicate a very strong correlation between daily rates of suspected severe cases and confirmed cases for the East Midlands region for the period examined in the current research, $r=.96$, $p<.01$, $N=71$. This is unsurprising given that testing in this early phase of the pandemic was largely confined to cases of severe illness, such as those patients conveyed by ambulances. Thus, ambulance records may be a reliable measure of severe COVID-19 related illness in real-time. To our knowledge, ambulance data have not previously been utilised to identify communities and regions that may be vulnerable to severe illness from COVID-19 or to investigate the social and environmental factors that may influence vulnerability. Our novel methodology presents an opportunity for health professionals to identify and support vulnerable communities who are likely to experience severe illness from a new virus in the early phase of a pandemic.

Ambulance records of provisional diagnosis hold several important advantages over hospital admittance records and laboratory records of confirmed cases for identifying vulnerable communities affected by severe illness. Hospitals and laboratories are required under General Data Protection Regulation (GDPR) to aggregate patient information for reporting. Others have explored socio-economic predictors of aggregate confirmed cases in the UK at less granular scales (eg., Daras et al., 2021). However, without costly and time-consuming data linkage via NHS Digital, aggregated patient data does not allow for meaningful analysis of spatial patterns or characteristics of physical and social environments that explain geographical trends of severe illness. Ambulance data, including postcode region[1], allows a more granular analysis of factors that predict severe illness from COVID-19 infection in real-time.

Our study presents a novel methodology for identifying communities and regions that are vulnerable to severe illness during the early phase of a pandemic before laboratory testing is widespread. This involves considering both *risk of exposure* to a contagious disease as well as *underlying susceptibility* to severe illness. We identify unusual clusters of provisionally diagnosed severe COVID-19 cases in real-time using medical data collated by EMAS. Provisional diagnosis of suspected COVID-19 is determined by paramedics based on observed signs, such as patient acuity, self-reported symptoms, such as OTD, and objective medical measures, such as blood oxygen levels. Our analysis explores the characteristics of communities and regions within built environments where unusual clusters occur, including landscape features and socio-economic dynamics. In this context, ‘unusual clusters’ refers to high numbers of cases occurring within spatial proximity that are unlikely to have occurred by chance. Taken together, these analyses offer a real-time approach for identifying and protecting vulnerable communities in the critical early stages between the first confirmed case of a new EID and widespread community testing, as well as for identifying the characteristics of those communities most affected by severe illness over the course of a pandemic.

2. Conceptualising the relationship between severe suspected COVID-19 cases and built environments:

Theories about individual health and wider environments emerged in the 1980s as a critical response to medical and epidemiological paradigms; traditional medical models conceptualise health in terms of the presence or absence of biological disease, and the outcome of exposures that occur entirely at the level of the individual (Barbour, 1997). More contemporary ‘social’ perspectives suggest that health transcends the individual and recognize the important role that social networks (Smith &

[1] UK Postcodes include two components, for example LN6 7TS. Region postcodes include the first component and the first letter of the second component; LN6 7.

Christakis, 2008) and wider social environments, including deprivation, play in health outcomes (Marmot, 1998).

Efforts to consider interactions *between* individual or biological factors and social factors often draw on Bronfenbrenner's bioecological theories (Eriksson et al., 2018), including his original Ecological Social Model, as well as the more recent Process-Person-Context-Time model (Rosa & Tudge, 2013). These frameworks advanced the field of public health by introducing a way to conceptualise the multi-level social interactions that influence health and wellbeing. Bronfenbrenner's models divide the social world of an individual into four 'systems'; the Microsystem, including the most immediate elements of the social world, such as family, the Mesosystem, including extended social networks, the Exosystem, including wider community services, and the Macrosystem including commonly shared cultural and social beliefs and values (Bronfenbrenner, 1979).

While progressive, bioecological theories focus almost exclusively on the social world and social vulnerabilities and rarely consider landscape features that are also understood to influence health outcomes (Campbell & Wiesen, 2009; Cervero & Duncan, 2003; Williams, 2016). In contrast, in the field of urban planning, vulnerability is often conceptualised in relation to hazards and risks in the landscape, such as exposure to direct communicable disease, as well as more distal relationships, or 'teleconnections' (Seto et al., 2012) between landscape features, like access to green space (De Vries et al., 2003; Markevych et al., 2017) and underlying health conditions, such as obesity (Daras et al., 2018). Importantly, exposure to a virus does not necessarily precipitate a medical emergency, rather, severe symptoms requiring emergency medical attention reflect the cumulative effect of exposure *and* underlying susceptibility. Thus, vulnerability is multifaceted, incorporating components of the physical landscape and components of the social world. While both bioecological and urban risk theories have advanced ways of thinking about health outcomes and pathways of vulnerability, a holistic approach is needed that considers the range of factors in built environments that precipitate severe illness and death from COVID-19.

In the case of COVID-19, the relationship between severe illness and characteristics of the built environment involves both direct and indirect pathways. On the one hand, environments can influence the direct exposure of individuals to communicable disease. On the other hand, landscape features can indirectly affect the underlying susceptibility of communities to severe symptoms, compared to experiencing mild symptoms or presenting as asymptomatic, by supporting or preventing healthy lifestyles. Features of neighbourhoods that can influence health behaviours like exercise include access to green space for passive recreation such as walking, and facilities for active exercise, such as sports grounds or leisure centres (Hartig et al., 2020). Further distance from these healthy landscape features is associated with lower levels of activity and higher risk of cardiovascular disease (Shen & Lung, 2016) and obesity (Lachowycz & Jones, 2011). However, landscape features can also reflect the social characteristics of wider living environments; high crime rates co-occur with poor physical infrastructure like housing in deprived communities. Crime can deter access to nearby outdoor spaces (Gomez et al., 2004) while poor housing indicates lower incomes and a greater likelihood of underlying chronic health conditions (Krieger & Higgins, 2002).

While deprivation broadly is associated with susceptibility to severe illness, the socio-economic characteristics of patients with severe symptoms of COVID-19 have often been overlooked. Thus, Khalatbari-Soltani et al (2020) call for the systematic recording of these dynamics for identifying vulnerable groups in the early stages of a pandemic. Patel et al (2020) suggest that deprivation is likely to be associated with increased VSID from COVID-19 in three ways. Firstly, more deprived neighbourhoods often experience overcrowding which results in increased risk of infection compared to less densely populated areas. Secondly, poorer people are more likely to be employed in roles without opportunities to work from home which also increases risk of exposure. Finally, poverty is a risk factor for chronic comorbidities that in turn predict severe illness and hospitalisation from COVID-19, such as cardiovascular disease (Mehra et al., 2020), diabetes (Peric & Stulnig, 2020) and obesity (Steinberg et al., 2020).

Others have investigated the relationship between severe COVID-19 related illness and individual features of social worlds, such as deprivation (e.g., Patel et al., 2020) and physical landscapes, such as air pollution (Travaglio et al., 2021). We consider the *cumulative* impact of factors across demographic, socio-economic and environmental domains to explore the characteristics of vulnerable communities (Kiaghandi et al., 2020) identified spatially by unusual clusters. This approach resounds well with the underlying philosophy of bioecological modelling; health outcomes are the culmination of interactions between and within domains that make up the built environment, and across individual and neighbourhood scales. Thus, in addition to social interactions, we include features of physical landscapes in our analysis to consider vulnerability across scales in Bronfenbrenner's socio-ecological landscape.

3. Methods:

3.1 Site and location

The East Midlands is located in the Central Eastern part of England and spans an area of 15,627km2 (Figure 2). The estimated total population of the region is 4.8 million including the most populous urban areas of Derby, Leicester, Lincoln, Northampton and Nottingham (ONS, 2020a). The proportion of the population identifying as other than 'White UK' in the East Midlands is low (14.6%) compared to the national average (20.2%) (ONS, 2020b), although some regions, including Leicester, have a much higher proportion of non-white population. In 2016, 18.5% of people in the region lived in the most deprived quintile (Public Health England, 2018). Nottingham, Derby and Leicester are the economic core of this region, with around 48% of businesses, and 50% of the population located in these cities (European Commission, 2020). The East Midlands is also the 3rd most rural region in England (European Commission, 2020).

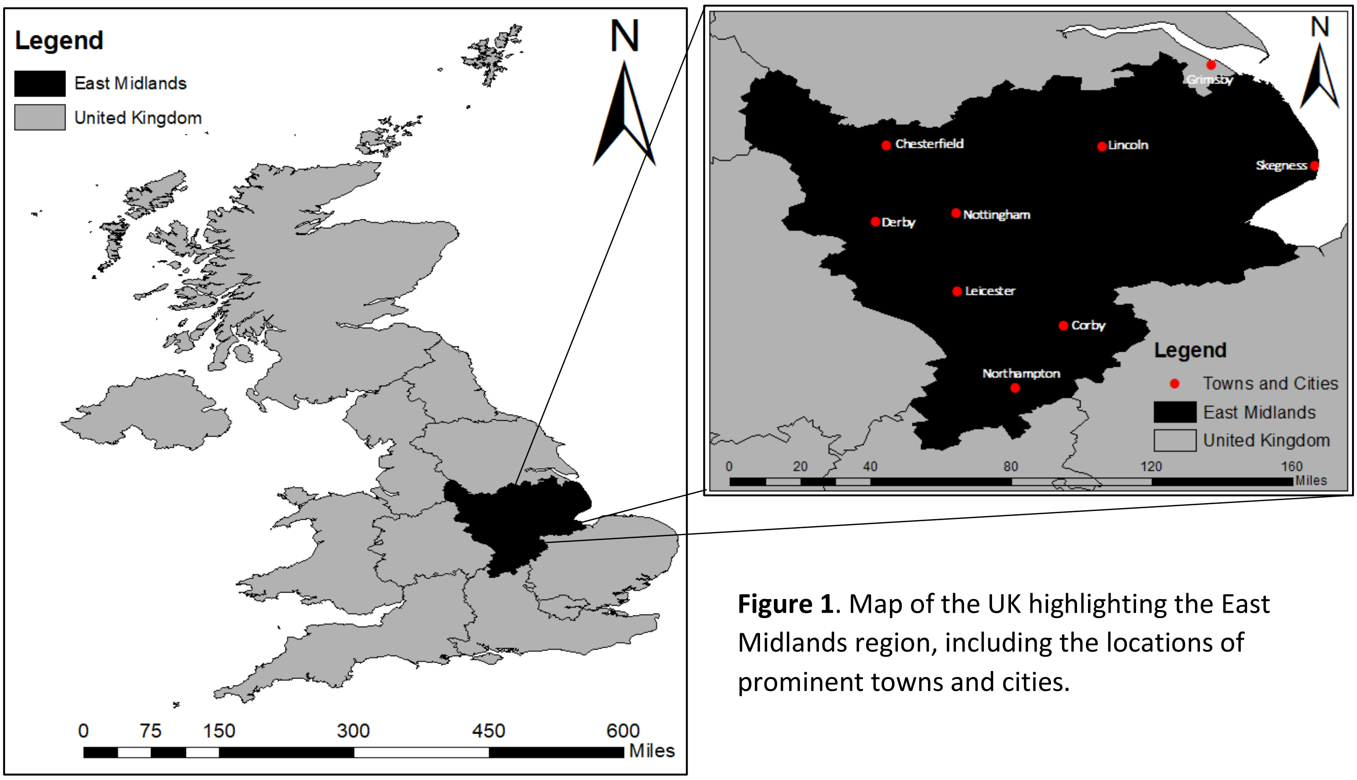


Figure 1. Map of the UK highlighting the East Midlands region, including the locations of prominent towns and cities.

3.2 Research aims and questions

The first aim of the research was to identify unusual clusters of suspected COVID-19 cases in the East Midlands of the UK using more than 10,000 records of provisional diagnoses for severe COVID-19 collated by EMAS during the first 'wave' of the pandemic between March 2nd and May 11th (Kontopantelis et al., 2021). This was achieved using a Kuldorff spatial scan statistic implemented in the geospatial software SatScan^{TM} which compares the actual distribution of cases to the predicted distribution based on population density. The null hypothesis tested is that cases are randomly distributed rather than occurring in unusual clusters. The second aim of the research was to explore factors that predict cluster membership. This analysis involved computing a binary logistic regression with variables including measures of patient demographics, deprivation, and landscape features (Section 3.3). The third aim was to elucidate the individual characteristics of each unusual COVID-19 cluster, using geospatial analysis and mapping to determine the strongest predictors of cluster membership.

3.3 Measures

Table 1 summarizes the datasets and measures included in the research. Data collated by and obtained from EMAS includes provisional diagnoses of suspected COVID-19 by medically trained clinicians, age, and sex. More severe COVID-19 symptoms tend to be associated with older age (Romero et al., 2020), and mortality is nearly twice as high in males compared to females (Ortolan et al., 2020). While ethnicity is also commonly associated with severe symptoms (Sze et al., 2020), reliable data was unavailable in real-time. The diagnostic algorithm employed by medically trained clinicians is guided by Public Health England's case definition criteria (PHE, 2020), including observations of illness, self-reported symptoms like OTD, and objective medical measures like blood oxygen levels.

The Index of Multiple Deprivation (IMD) is an aggregate measure of socio-economic indictors. Low scores indicate greater deprivation while higher scores indicate greater affluence. Decile values of IMD were used for both spatial and statistical analysis. The Access to Healthy Assets and Hazardous Index (AHAHI) includes neighbourhood measures of physical landscape features, such as distance (km) from medical services and retail outlets, as well as environmental measures of air pollution which are often associated with built-up areas like dense housing, transport infrastructure and power stations (Beevers et al., 2012; Pannullo et al., 2017). The AHAHI is a validated metric that synthesises features of built environments that are commonly related to health outcomes in the wider health literature (Green et al., 2018).

According to Daras et al. (2019) healthy features of landscapes associated with more positive health outcomes include closer proximity to active and passive green space as well as health services, such as general practitioners and emergency departments. This is because physical access to health services is associated with health service use and health maintenance. Thus, we might expect to observe lower rates of suspected severe COVID-19 located nearer to the health services and healthy features of physical environments included in our analysis, such as green and blue spaces (Table 1).

Hazardous features of landscapes associated with poorer health outcomes include poor air quality, further distance from healthy features of landscapes, and closer proximity to retail vendors like fast food outlets, tobacconists, off-license stores, pubs, bars, and clubs. This is because distance from hazardous retail environments is a proxy measure for individual behaviour; people who live closer to fast food outlets are more likely to consume fast food, and subsequently to experience underlying health conditions like diabetes and obesity (Green et al., 2018). On this basis, we could anticipate that unusual clusters of suspected COVID-19 are likely to occur closer to retail vendors, further from health services and healthy physical environments, and in areas with poorer air quality. We chose to include individual measures of each of these landscape features rather than the final aggregate AHAHI scores in order to examine the effect of specific landscape scale environmental variables on COVID-19 clusters.

Raw data for AHAHI input domains (Daras et al., 2019) were used to compute the binary logistic regression analysis. Decile values were used for the purpose of geospatial analysis and mapping. Given that high mortality rates have been associated with urbanity compared to rurality (Stier et al., 2020), we also included measures from the UK Rural and Urban Categories (RUC) scale to supplement our analysis of neighbourhood environments.

Table 1. Datasets, measures and sources

Dataset*	Measure		Source
EMAS COVID-19 2020	Suspected cases of COVID-19 (March 2nd-May 11th), sex, age		East Midlands Ambulance NHS Trust
IMD 2019	IMD Decile		https://hub.arcgis.com/datasets/communities::lower-super-output-area-lsoa-imd-2019-osgb1936
RUC 2011	Categorical scale 1 (most urban) to 10 (most rural)**		https://hub.arcgis.com/datasets/ons::rural-urban-classification-2011-of-lower-layer-super-output-areas-in-england-and-wales
AHAHI 2019	Retail Environment (distance in km)	Gambling, fast food, pubs/clubs/bars, off license, tobacconists	https://data.cdrc.ac.uk/dataset/access-healthy-assets-hazards-ahah
	Health services (distance in km)	GPs, A&E, dentists, pharmacies, leisure	

Physical environment (distance in km)[2]	Green Space (passive), Green Space (active), Blue Space
Air pollution (levels)[3]	Nitrogen Dioxide, Particulate Matter, Sulphur Dioxide

*all data scales at Lower Super Output Area

**only 8 categories were present in the East Midlands dataset; provisional diagnoses of COVID-19 requiring ambulance attendance in the East Midlands were not recorded in Urban-Major Conurbations, Villages, and Small Town and Fringe areas.

3.4 Data handling and cleaning

The database of suspected COVID-19 cases was obtained from EMAS[4], including the date 999 calls were received, partial postcodes (rather than full addresses) of ambulance attendance locations, sex, and age. In total, 10,438 records were received, however, 93 records were removed because they contained errors, were missing geospatial information, or were unable to link to postcode population. All records of suspected COVID-19 cases were successfully linked to IMD, AHAHI, and RUC values. Thus, the final dataset contained 10,345 geospatial points. Only call outs for provisionally diagnosed COVID-19 were included in the dataset.

3.5 Statistical and spatial data analysis

Data analysis was conducted in three steps. Step one involved identifying unusual clusters of suspected severe COVID-19 cases by using population data as a baseline for the expected distribution of cases. For this analysis data were represented at the postcode region scale. The output included the location of statistically significant clusters, and a binary dataset distinguishing all cases that fell within clusters from all cases that fell outside of clusters. Step two involved converting the postcode region data to Lower Super Output Area scale for the purpose of linking the COVID-19 dataset with existing national datasets, including IMD, the AHAHI, and RUC. The output was a unique linked database combining clinical and landscape scale data. In step three, statistical analyses were conducted to identify demographic, socio-economic, and environmental factors that predicted cluster and non-cluster membership, and geospatial analysis was used to characterize each individual cluster.

3.5.1 Identifying unusually high clusters of suspected COVID-19 cases

We applied a Kulldorff spatial scan statistic (Discrete Poisson model) implemented in SatScan™ software version 9.6.1 to perform the spatial analysis scanning to detect unusual clusters of COVID-19 cases across the surveillance area. A spatial scan statistic is a cluster detection test that detects the location of clusters and evaluates their statistical significance (Kulldorff et al., 2005; Kulldorff, 1997). This was done by gradually scanning a window across the study area, noting the number of observed and expected observations, based on population (ONS, 2011b), inside the window at each location using a Discrete Poisson model. For any given position of the centre, the radius of the circle changes continuously so that it can take any value. For each circle, the spatial scan statistic calculates the likelihood of the observed number of cases occurring inside and outside of the circle. The circle with the maximum likelihood is the most likely cluster, and thus the least likely to have

[2] Passive Green Space includes parks, gardens, golf courses, and allotments. Active Green Space includes sporting areas such as playing fields and tennis courts.

[3] PM, NO_2 and SO2 measures are annual µg m-3, micrograms per cubic meter of air.

[4] This research, including use of patient data for statistical and spatial analysis, was approved by the NHS Health Research Authority, IRAS ID: 264573.

occurred by chance. This method tests the null hypothesis that cases are randomly distributed. Statistically significance suggests that unusual spatial clustering is unlikely to have occurred by chance. The isotopic circular scan method employed by the software has previously been validated for identifying clusters of other infectious disease, such as malaria (Coleman et al., 2009), HIV (Namosha et al., 2013; Tanser et al., 2017), tuberculosis (Smith et al., 2018), and various chronic diseases (Cuadros et al., 2019; Tomita et al., 2020).

The Poisson Model was purely spatial. The model parameters included unconstrained spatial cluster size, and the criteria for reporting hierarchical clusters was set to ‘no cluster centres in other clusters’.

3.5.2 Data conversion to LSOA and database compilation

To compile the LSOA dataset, IMD, RUC and AHAHI scores were merged using the join tool in ArcGIS Pro 2.6.0. The join used Lower Super Output Area codes (LSOA11CD) as these identifiers are consistent between the EMAS COVID-19 database and the remaining datasets. These processes are visualised in Figure 2.

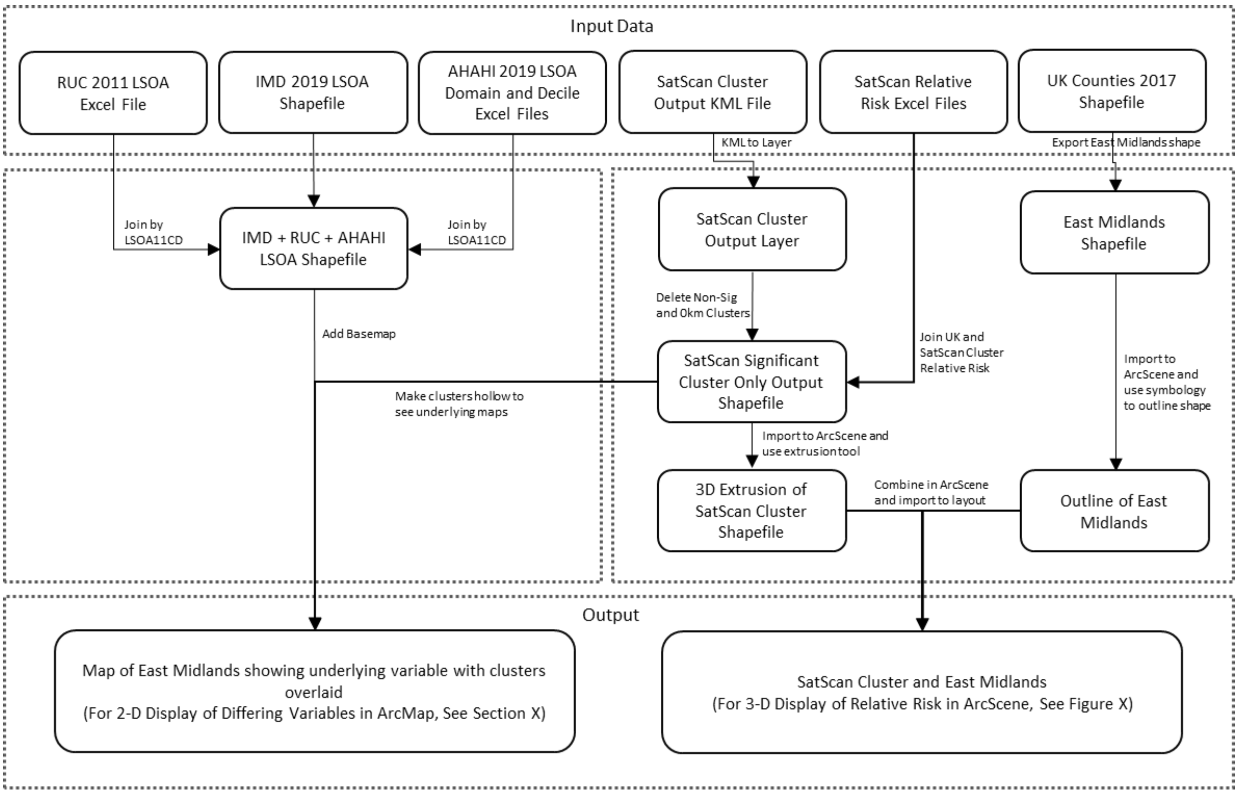


Figure 2. Schematic of database compilation and spatial analysis including data joining, and data display as 2-D and 3-D maps using ArcGIS Pro 2.6.0.

Geospatial analysis was also used to identify which cases fall into specific clusters compared to cases that are randomly distributed. In one instance, two clusters were found to overlap. However, for the purpose of characterizing clusters it was necessary to assign all cases to a single cluster. Thus, these cases (N = 54) were assigned to clusters based on their location from a centre line of intersection between the overlapping clusters. The output, a novel database, was used for regression analysis to identify factors that predict cluster membership (Section 5.2), and for geospatial analysis to produce maps representing teleconnections (Section 5.3).

3.5.3 Statistical analysis and spatial representation of significant clusters

Binary logistic regression analysis was used to identify factors that predict whether individual cases of suspected severe COVID-19 occur in unusual clusters. All measures reported in Table 3 and Table 4 were included in the regression model. While the IMD was included in the binary regression, we also conducted an ANOVA to determine whether mean differences in deprivation and affluence occur between areas with clusters and areas characterized by random distribution. In the UK, deprivation is often associated with early transmission patterns (Balasegaram et al., 2012), and high rates of contagion (Rushton et al., 2007). Thus, it is possible that deprivation is a common denominator for all areas with suspected cases of COVID-19, rather than a distinguishing feature of cluster membership. ANOVA was computed to explore more nuanced spatial differences between each cluster, and areas with cases that do not occur in clusters.

Regression output and cluster output from SatScan™ was used to display the relationship between determinants of clusters visually. The cluster output from SatScan™ was converted to a layer ('cluster shapefile') within ArcGIS Pro 2.6.0. Of 41 clusters identified, 13 were statistically significant ($P < 0.05$). All non-significant clusters were removed from the dataset. A polygon representing the East Midlands was extracted from the UK Counties 2017 shapefile ('UK shapefile') to create a background in ArcScene. Relative Risk values were assigned to each cluster within the cluster shapefile. Both the cluster shapefile and UK shapefile were converted to rasters and combined to create a unique raster displaying the Relative Risk of clusters in 3-D. The unique raster was then converted to a TIN in order to be represented clearly in ArcScene. This step addressed a display problem due to the rasters and polygons merging and warping the slope of elevation in the image. Displaying clusters involved using a scale of graduated colours from green to red that were manually selected based off the spread of the data.

4. Results:

4.1 Identifying unusually high clusters of suspected severe COVID-19 cases

SatScan™ Poisson Modelling identified 13 statistically significant ($P<.05$) unusually high clusters of suspected COVID-19 cases, displayed in Figure 3. Per 100,000 population, the number of observed cases range from 951 West of Skegness to 3,417 East of Rugby. By comparison, the range of cases occurring per 100,000 outside of clusters was between 8 and 660. Figure 4 demonstrates the relative risk of each cluster, meaning the likelihood of contracting severe illness in an area compared to regions where cases are randomly distributed. The spatial characteristics of each cluster, including approximate location, radius, expected and observed number of cases, P-Values, specific relative risk ratios, and the number of cases in each cluster per 100,000 population are reported in Table 2.

Figure 3. The geographic location of 13 statistically significant (P<.05) clusters of COVID-19, identified using a Kulldorff spatial scan statistic. Further details of clusters are given in Table 2.

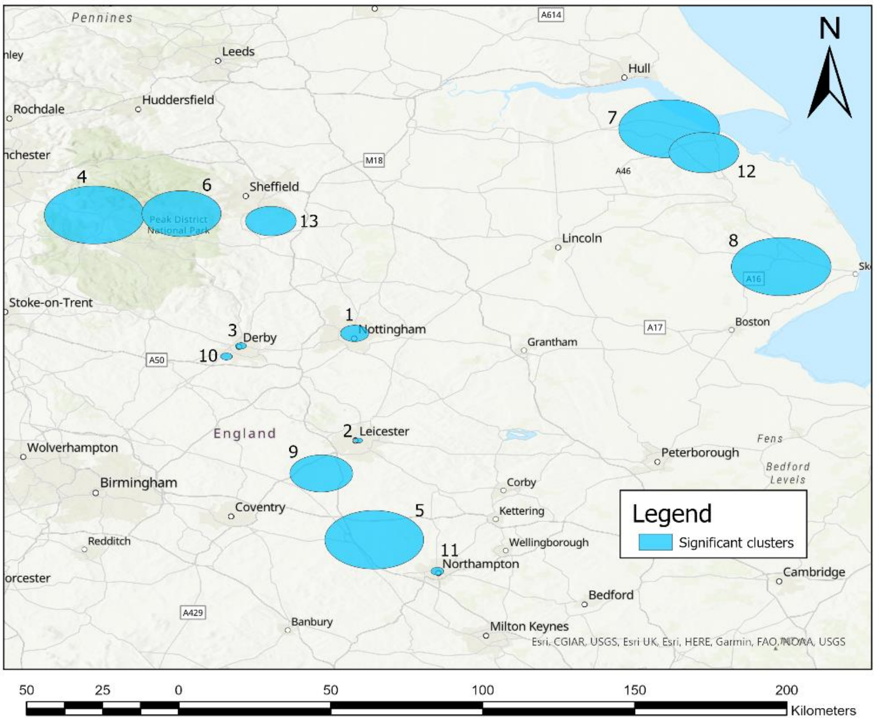


Figure 4. Spatial representation of relative risk of suspected cases of COVID-19 in the East Midlands of the UK between March 2nd and May 11th 2020. Taller clusters, and clusters closer to red on the colour gradient reflect greater risk of contracting COVID-19.

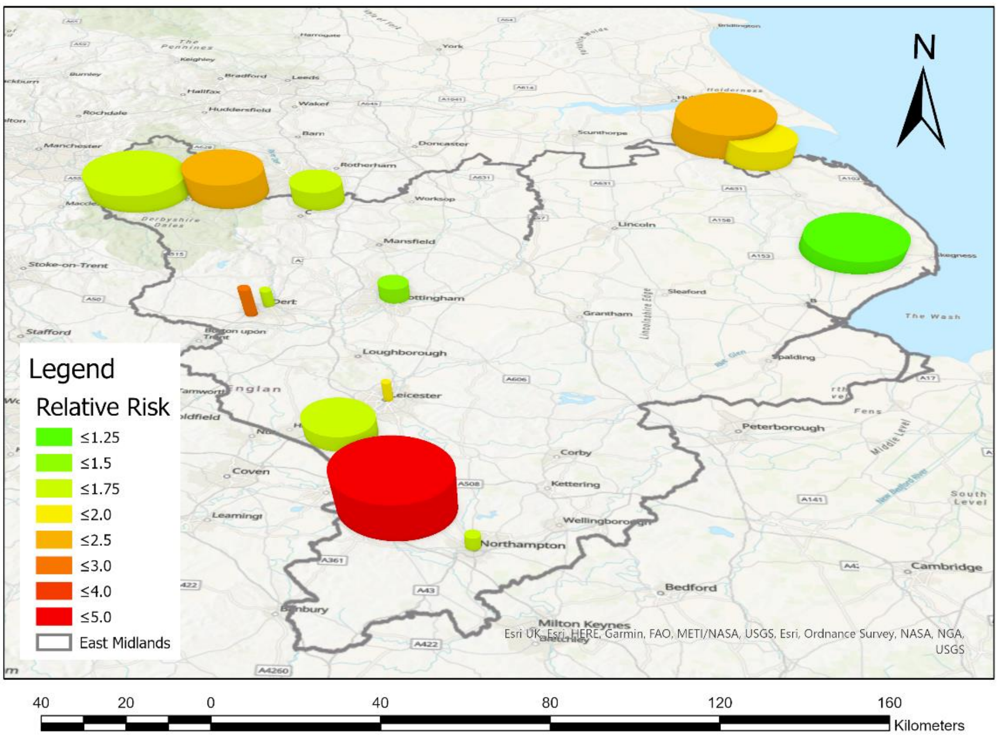


Table 2. Spatial characteristics of unusual clusters of suspected COVID-19 cases presented in Map 1, extracted from SatScan output, including population, number of cases, expected cases, log likelihood, P-value, relative risk, cases per 100,000 population and approximate location of clusters. Population has been determined at the regional postcode scale.

Cluster	Radius (km)	Population	Number of Cases	Expected Cases	Log Likelihood Ratio Cases	P – Value	Relative Risk	Cases per 100,000 population	Location
1	49.21	82,653	911	652.93	48.82	<0.00	1.43	1102	Nottingham
2	20.67	14,120	210	111.55	34.88	<0.00	1.90	1487	Leicester
3	2.78	32,220	379	254.53	27.19	<0.00	1.51	1176	Derby
4	33.98	18,430	233	145.59	22.53	<0.00	1.61	1264	West Peak District
5	.84	907	31	7.16	21.61	<0.00	4.34	3417	East of Rugby
6	1.08	4,331	77	34.22	19.76	<0.00	2.26	1777	East Peak District
7	9.92	3,690	65	29.15	16.34	<0.00	2.24	1761	West Grimsby
8	42.15	87,897	836	694.36	14.60	0.00	1.22	951	West of Skegness
9	11.16	9,235	121	72.96	13.29	0.00	1.67	1310	Southwest of Leicester
10	1.17	1,543	32	12.19	11.10	0.00	2.63	2073	Southwest of Derby
11	1.3	12,443	148	98.3	10.98	0.00	1.51	1189	Northampton
12	4.99	5,285	74	41.75	10.15	0.01	1.78	1400	East Grimsby
13	13.33	6,029	81	47.63	9.70	0.02	1.71	1343	North of Chesterfield

4.2 Factors that predict cases of COVID-19 falling into unusual clusters compared to randomly distributed cases

4.2.1 Descriptive statistics

In total, 10,345 cases of suspected severe COVID-19 with sufficient information to include in analysis were reported and recorded by EMAS between March 2nd and May 11th, 2020. Of all cases, 1,123 fell into unusual clusters compared to population, while the remaining 9,222 cases were distributed randomly. The mean (M) and standard deviation (SD) for measures of IMD and AHAHI included in our analysis are presented in Table 3. The proportion of cases in unusual clusters compared to randomly distributed by sex and RUC categories are presented in Table 4.

Table 3. Descriptive statistics for measures of Index of Multiple Deprivation (IMD), Access to Healthy Assets and Hazardous Index (AHAHI) and age for cases of severe COVID-19 in unusual clusters (M_IN, SD_IN) compared to cases randomly distributed outside clusters (M_OUT, SD_OUT). Measures of IMD are decile values. Measures of AHAH include four domains: distance (km) from retail environments, health services, physical environments, and air quality.

Domain	Factor	M_In	SD_IN	M_Out	SD_Out
Retail environments	Gambling	2.02	2.63	2.50	2.87
	Fast food	1.85	2.65	2.18	2.48
	Pubs/clubs/bars	1.40	1.91	1.87	2.22
	Off License	4.00	5.50	4.87	6.62
	Tobacconists	3.26	3.861	3.63	3.41
Health services	GPs	1.44	1.47	1.67	1.55
	A&E	16.76	16.40	12.52	10.30
	Dentists	1.65	1.97	2.10	2.28
	Pharmacies	1.21	1.50	1.39	1.62
	Leisure	3.12	3.95	3.95	4.317
Physical environment	Green Space (passive)	.34	.25	.36	.48
	Green Space (active)	.54	.59	.58	.55
	Blue Space	2.24	1.79	2.57	2.13
Air pollution	Nitrogen Dioxide	12.59	2.31	11.77	1.81
	Particulate Matter	13.64	1.60	14.30	.80
	Sulphur Dioxide	1.40	.29	1.24	.23
	IMDDecil	4.38	2.84	5.04	2.875
	Age	48.97	25.86	50.69	26.09

Table 4. Proportion of cases in unusual clusters (IN(%)) compared to randomly distributed cases outside clusters (OUT (%)) by sex and Rural Urban Classification Categories (RUC).

		In (%)	Out (%)
RUC	Urban major conurbation	<1	<1
	Urban minor conurbation	34.9	16.5
	Urban city and town	49.1	62
	Urban city and town in sparse setting	<1	.6
	Rural town and fringe	10.3	12.8
	Rural town and fringe in sparse setting	<1	<1
	Rural village and dispersed	4.3	7.3
	Rural village and dispersed in sparse setting	1.3	.2
Sex	Female	54	53
	Male	45	46
	Missing	<1	<1

4.2.2 Regression analysis

A binary logistic regression analysis was conducted to investigate factors that are associated with cluster membership. Given the highly unequal distribution of cases by binary categories, the probability cutoff was set to .6, as distinct from the usual cutoff of .5 for randomly distributed binary data (Calabrese, 2014), and the model parameters were set to predict the log-odds of membership in the major category (randomly distributed cases) compared to the minor category (unusual clusters). We also performed bootstrap sampling to account for dependencies between cases in clusters. This analysis did not change the P-values or significant predictors in the regression model.

The results indicate that 12 of 16 variables that input to the Access to Healthy Assets and Hazardous Index (AHAHI), and 4 rural and urban categories are significant predictors of whether cases are distributed randomly or appear in unusual clusters by population (Chi-square = 2028.36, df = 26, P = .00). Age, sex, Index of Multiple Deprivation (IMD) Decile, the remaining 4 AHAHI variables (Accessibility to leisure centres, gambling districts, Green Space and dentists) and 3 rural and urban categories (Urban major conurbation, Urban city and town and Rural town and fringe) are not significant. The model correctly predicted 41.3% of cases that appear in clusters and 98.7% of cases that do not appear in clusters, giving an overall percentage correct prediction rate of 92.5%[5].

Table 5 displays the binary logistic regression results for the independent variables that were found to be associated with cluster membership. Compared to randomly distributed cases, cases in clusters are more likely to be located closer to pubs/bars/clubs, Blue Space, off licenses, Passive Green Space, as well as in areas with higher levels of Nitrogen Oxide, and in RUC categories 'urban minor conurbation', 'urban city and town in sparse', 'rural town and fringe', and 'rural village and dispersed'. Randomly distributed cases that do not occur in clusters are located closer to tobacconists, GP practices, A&E hospitals, and pharmacies, as well as in areas with higher levels of particulate matter and Sulphur Dioxide.

[5] The asymmetry in predictive accuracy for cases appearing in clusters compared to cases not appearing in clusters is a common phenomenon of highly unequal datasets (Calabrese, 2014) and reflects the true rarity of cases appearing in clusters.

Table 5. Binary logistic regression analysis predicting cluster membership. Positive **B** values indicate an increased likelihood of random distribution and a decreased likelihood of cases occurring in clusters. Negative **B** values indicate a decreased likelihood of random distribution and an increased likelihood of cases occurring in clusters.

		B	SE	Wald	df	Exp(B)	95% CI
AHAHI	Accessibility to fast food outlets	-.16	.04	15.63	1	.85**	.78, .92
	Accessibility to pubs/bars/nightclubs	.2	.04	21.43	1	1.22*	1.12, 1.33
	Accessibility to Blue Space	.09	.02	14.5	1	1.1*	1.04, 1.14
	Accessibility to Off Licenses	.02	.01	5.18	1	1.02**	1, 1.04
	Accessibility to tobacconists	-.1	.02	17.73	1	.91*	.87, .95
	Passive Green Space (within 900m buffer)	.56	.1	33.26	1	1.75*	1.45, 2.11
	Accessibility to GP practices	-.14	.045	10.28	1	.87*	.92, 1.2
	Accessibility to A&E hospitals	-.12	.005	529.67	1	.9*	.89, .91
	Accessibility to pharmacies	-.11	.05	3.89	1	.9**	.81, 1.01
	Level of Nitrogen Dioxide (NO_2)	-1.12	.05	591.83	1	1.75*	.3, .4
	Level of Particulate Matter (PM10)	1.51	.06	662.64	1	4.53*	4.04, 5.9
	Level of Sulphur Dioxide (SO_2)	1.98	.28	48.26	1	7.22*	4.13, 12.6
RUC	Urban minor conurbation	-.92	.09	103.03	1	.4*	.33, .48
	Urban city and town in a sparse setting	-.54	.17	10.23	1	.58*	.48, .81
	Rural town and fringe	-3.01	1.26	5.77	1	.05**	.00, .58
	Rural village and dispersed	-3.9	.7	34.76	1	.02*	.00, .07

*Statistically significant at P < .01

**Statistically significant at P < .05

4.2.3 Index of Multiple Deprivation ANOVA

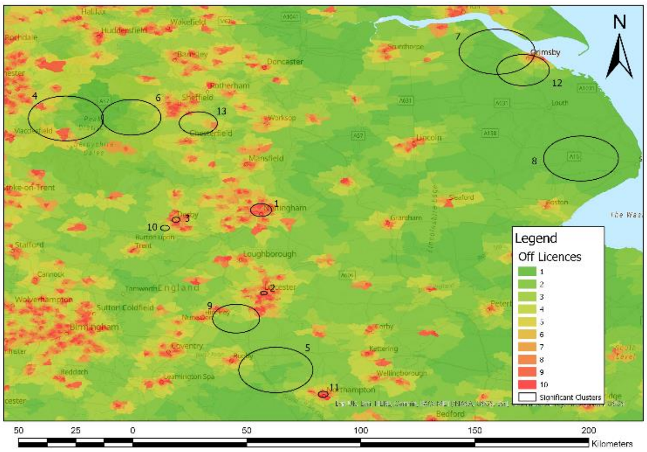
Regression analysis revealed that IMD deciles was not a significant predictor of whether provisionally diagnosed severe COVID-19 cases were distributed randomly or occurred in unusual clusters. ANOVA was also computed to identify whether IMD scores varied between each individual cluster, and areas with randomly distributed cases. There was a significant difference for IMD decile scores between Cluster 1 (M = 2.93, SD = 2), Cluster 2 (M = 2.42, SD = 1.21), Cluster 3 (M = 3.22, SD = 2), Cluster 4 (M = 6, SD = 2.33), Cluster 5 (M = 7.4, SD = 2), Cluster 6 (M = 9, SD = .6), Cluster 7 (M = 3.9, SD = 2.83), Cluster 8 (M = 2.86, SD = .83), Cluster 9 (M = 8.06, SD = 1.86), Cluster 10 (M = 7.84, SD = 1.9), Cluster 11 (M = 2.49, SD = 1.04), Cluster 12 (M = 4.68, SD = 3.18), Cluster 13 (M = 5.79, SD = 2.48), and cases that are evenly distributed (M = 5.04, SD = 2.88), $F(13, 10331)=40.96$, $p=.00$.

4.3 Characteristics of unusually high severe COVID-19 clusters

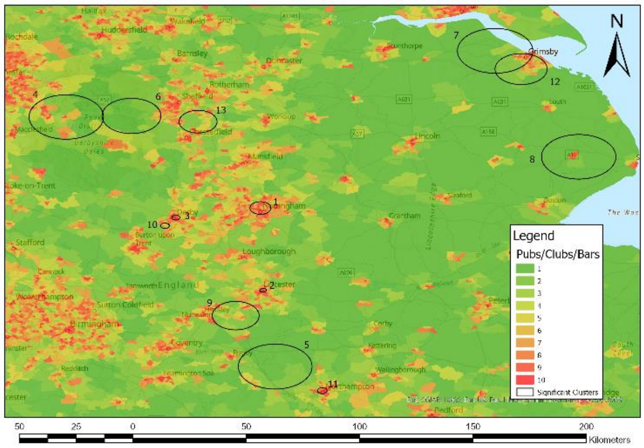
The statistical analysis presented in section 5.2 considers factors related to all the clusters of suspected severe COVID-19 cases in the East Midlands. The geospatial analysis presented below considers the characteristics of *individual* clusters. The following series of maps (Figures 6-9) display the distribution of factors related to retail environments, health services, physical environments (including RUC), air pollution, and IMD. With the exception of RUC, all other factors are represented as deciles values.

Figure 5. Maps depicting distance (km) from 'harmful' retail environments derived from the Access to Healthy Assets and Hazardous Index (AHAHI) that are associated with cluster membership, including off licenses, pubs/ bar/clubs, fast food outlets and tobacconists. The green spectrum indicates areas that are further away and the red spectrum indicates areas that are closer. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as black circles and numbered consistent with Table 2.

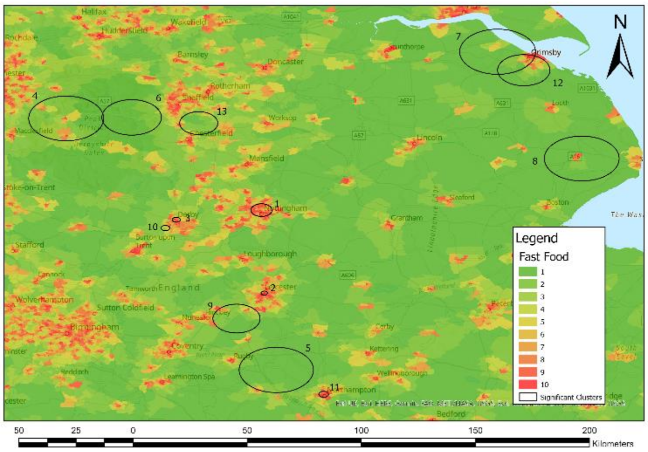
A. Off licenses



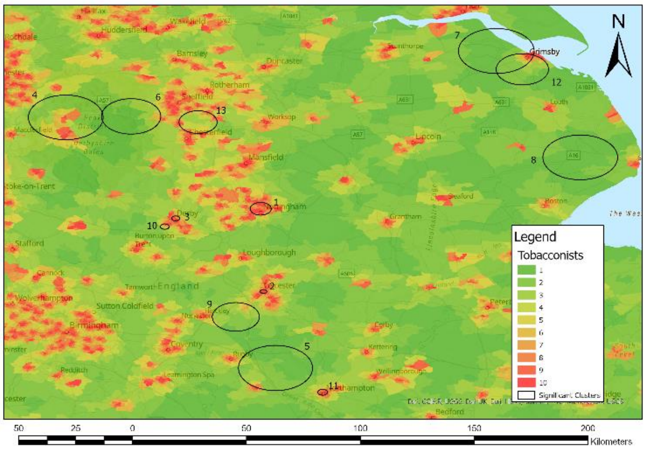
B. Pubs/bars/clubs



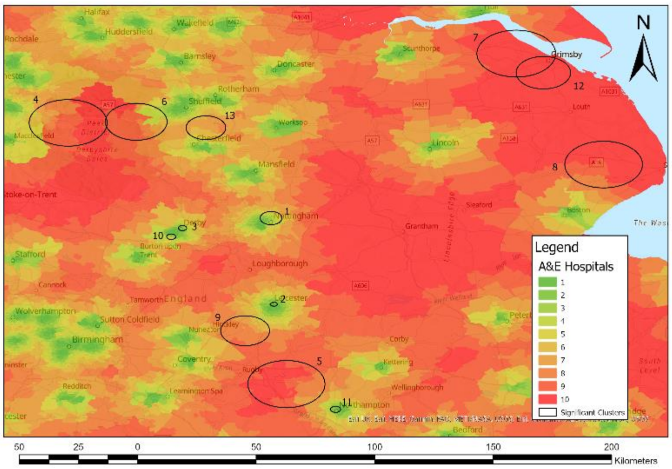
C. Fast food outlets



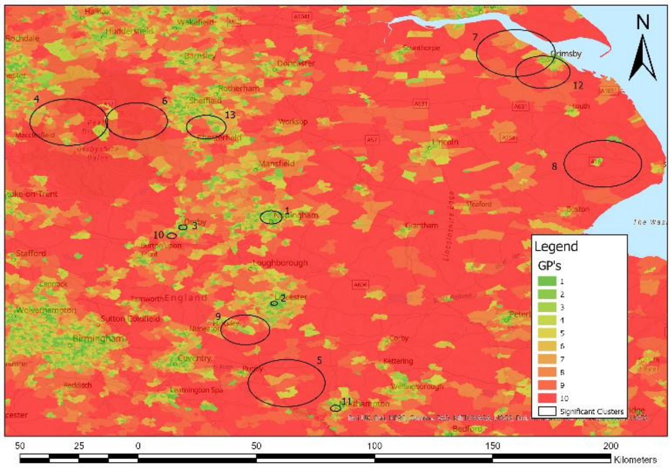
D. Tobacconists



E. A&E hospitals



F. GPs



G. Pharmacies

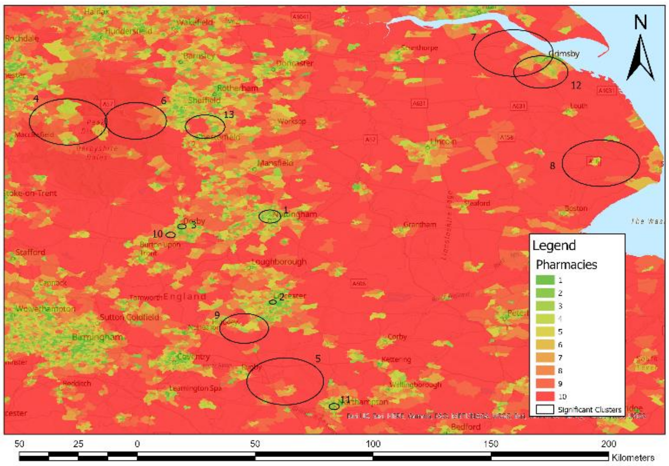
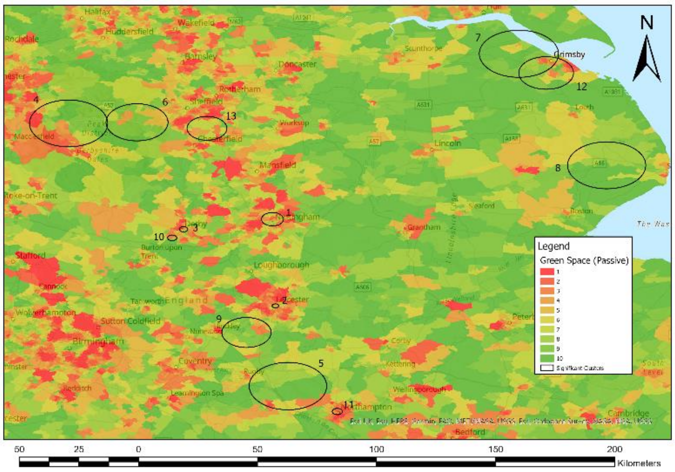
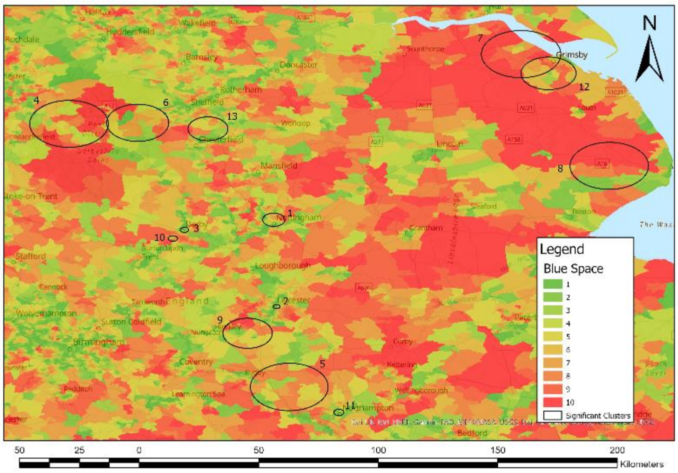


Figure 6. Maps depicting distance (km) from 'healthy' services derived from the Access to Healthy Assets and Hazardous Index (AHAHI) that are associated with cluster membership, including A&E hospitals, GPs, and pharmacies. The green spectrum indicates areas that are closer and the red spectrum indicates areas that are further away. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as black circles and numbered consistent with Table 2.

H. Green Space (passive)



I. Blue Space



J. Rural Urban Categories (RUC)

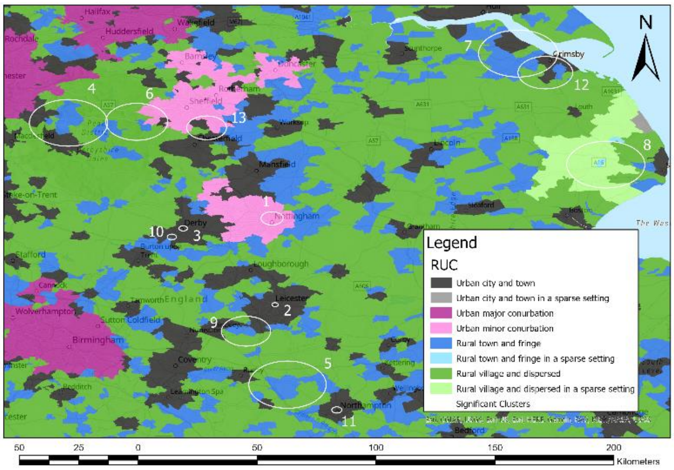
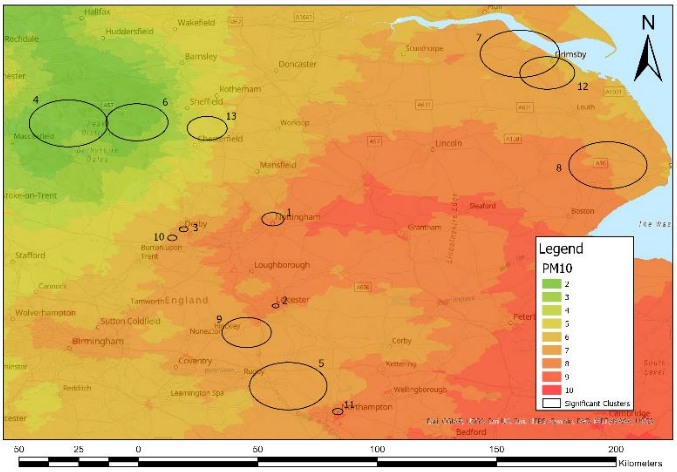
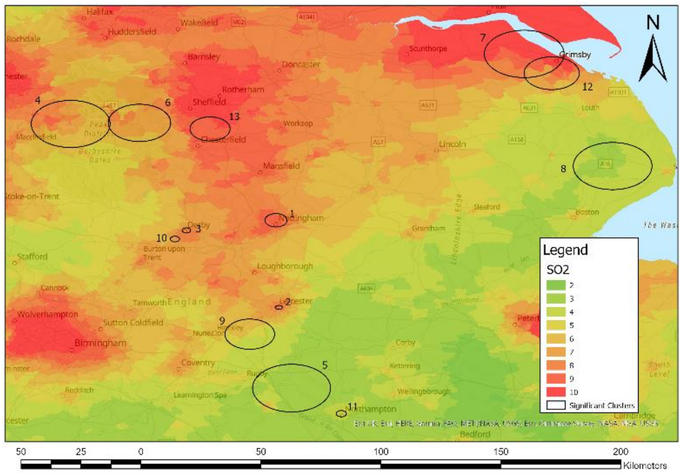


Figure 7. Maps depicting distance (km) from physical environments derived from the Access to Healthy Assets Hazardous Index (AHAHI) and degree of urbanization/rurality, that are associated with cluster membership, including Green Space (passive), Blue Space, and RUC categories. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as white circles and numbered consistent with Table 2.

K. Particulate Matter (PM10)



L. Sulphur Dioxide (SO_2)



M. Nitrogen Dioxide (NO_2)

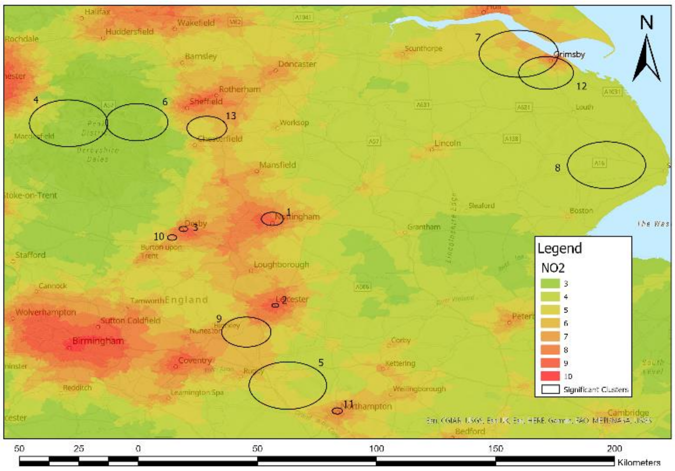
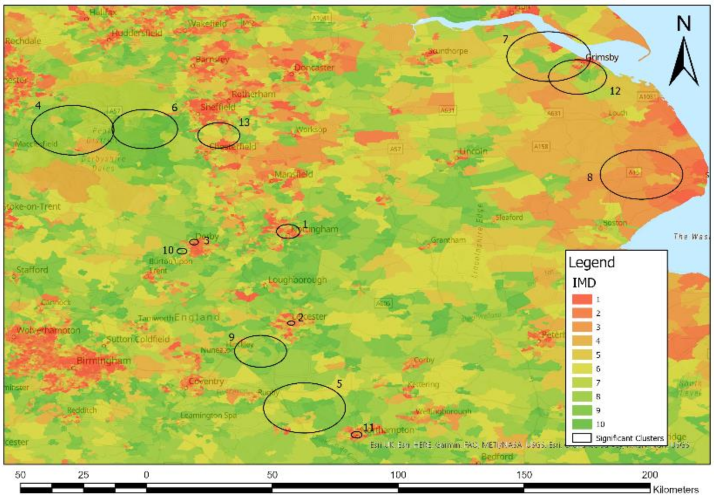


Figure 8. Maps depicting the level of pollutants derived from the Access to Healthy Assets Hazardous Index (AHAHI) that are associated with cluster membership, including Particulate Matter (PM10), Sulphur Dioxide (SO_2) and Nitrogen Dioxide (NO_2). The green spectrum indicates lower levels of pollutants and the red spectrum indicates higher levels. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as black circles and numbered consistent with Table 2.

Figure 9. Map of Index of Multiple Deprivation (IMD) distribution and unusual clusters of suspected COVID-19 cases. The green spectrum indicates greater affluence and the red spectrum indicates greater deprivation. The 13 clusters of high numbers of suspected COVID-19 cases (identified using a Kulldorff spatial scan statistic) are superimposed as black circles and numbered consistent with Table 2.



Importantly, clusters displayed on the maps reflect the radius within which individual cases of suspected COVID-19 occur. To preserve the anonymity of EMAS patients, we have not displayed the specific location of cases within clusters. Table 6 synthesizes the characteristics of each individual cluster compared to areas with cases that are randomly distributed. Average scores for RUC, IMD and AHAHI have also been deidentified [6]. Rather than exact values, Table 6 compares the characteristics of clusters to average values for all areas with randomly distributed cases. In some instances, the visual characteristics of a cluster may vary from the characteristics reported in Table 6. For example, Figure 9 displays the distribution of IMD scores within clusters. A cluster may appear to be predominately affluent (towards the green end of the colour scale) while most suspected COVID-19 cases fall within a small area that is severely deprived (towards the red end of the colour scale). Taken together, visual and statistical analyses represent cluster characteristics accurately while maintaining the anonymity of patient locations.

[6] Patient anonymity is a requirement of the approved IRAS. It may be possible to triangulate cluster information, such as radius, and specific values, such as IMD, to identify more specific locations. Our approach maintains anonymity and complies with the terms of ethical approval.

Table 6. Characteristics of individual clusters of unusually high suspected cases of COVID-19 compared to randomly distributed cases, including the proportion of cases in urban (U) and rural (R) areas (RUC), Index of Multiple Deprivation (IMD) Decile, and Access to Healthy Assets and Hazardous Index (AHAI) indictors (average distance (km) from retail environments, health services, and physical environments, as well as average level of air pollutants). For cases that are randomly distributed by population (Non-cluster), average values for each indicator, and the average score of aggregated indictors for each domain are reported. For each cluster, a '+' sign indicates when the average score for each indicator, or average aggregated domain score, is higher than the equivalent score for 'Non-cluster' cases. A '-' sign indicates when the average score is lower than the equivalent score for 'Non-cluster' cases. A score of '0' indicates no difference between cluster scores and non-cluster scores.

	Cluster	IMD	RUC (%)*		Retail*					Health*				Physical*			Pollution*			
		Decile	U	R	FF	PBC	OL	T	**X**	GP	A&E	P	**X**	B	G	**X**	PM	NO	SO	**X**
	Non-cluster	5.04	80	20	2.9	1.87	4.8	3.64	**3.3**	1.7	12.52	1.4	**5.21**	2.57	.37	**1.47**	7.25	6.16	6.61	**6.67**
1	Nottingham	-	+	-	-	-	-	-	**-**	-	-	-	**-**	-	0	**-**	+	+	+	**+**
2	Leicester	-	+	-	-	-	-	-	**-**	-	-	-	**-**	-	0	**-**	+	+	+	**+**
3	Derby	-	+	-	-	-	-	-	**-**	-	-	-	**-**	-	+	**-**	-	+	+	**+**
4	W. Peak	+	-	+	-	-	-	-	**-**	0	+	-	**+**	+	0	**+**	-	+	+	**-**
5	E. Rugby	+	-	+	+	+	+	+	**+**	+	+	+	**+**	-	0	**-**	-	-	-	**-**
6	E. Peak	+	-	+	+	+	+	+	**+**	+	+	+	**+**	-	-	**-**	-	+	+	**-**
7	W. Grimsby	-	+	-	-	+	+	+	**+**	0	+	0	**+**	+	0	**+**	-	+	+	**+**
8	W. Skeg	-	-	+	+	+	+	+	**+**	+	+	+	**+**	+	-	**+**	+	-	-	**-**
9	S.W. Leicester	+	-	+	+	-	-	+	**+**	0	+	+	**0**	+	-	**+**	0	-	-	**-**
10	S. W. Derby	+	+	-	-	-	+	-	**-**	-	-	-	**-**	-	0	**-**	-	+	+	**+**
11	Northampton	-	+	-	-	-	-	-	**-**	-	-	-	**-**	-	+	**-**	+	-	-	**0**
12	E. Grimsby	-	+	-	-	-	-	+	**-**	-	+	-	**+**	+	0	**+**	-	+	+	**+**
13	N. Chesterfield	+	+	-	+	-	-	-	**-**	-	-	+	**-**	-	-	**-**	-	+	+	**-**

*Rural and Urban Categories: Urban (U), Rural (R). Scores indicate % of sites in more urban and more rural areas
Retail Environment: Fast food (FF), Pubs/bars/clubs (PBC), Off license (OL), Tobacconists (T)
Health Services: General Practitioners (GPs), A&E Hospitals (A&E), Pharmacies (P)
Physical Environments: Blue Space (B), Green Space (passive) (G)
Air Pollution: Particulate Matter 10 (PM), Nitrous Oxide (NO), Sulphur Dioxide (SO)

6. Discussion:

One year into the COVID-19 pandemic research related to public health, epidemiology, and urban planning has advanced knowledge and understanding about the groups in society that are particularly vulnerable to severe illness from COVID-19. Most research about vulnerable communities and regions considers the association between COVID-19 cases and individual domains, such as deprivation or urbanisation. To our knowledge, the only prior study to use an approach similar to the methodology presented here is Kiaghandi et al. (2020) who examined the relationship between confirmed COVID-19 cases in Harris County, Texas and 46 variables across five domains, including access to health services, and environmental exposures. However, the study did not distinguish between severe illness and asymptomatic or mild cases. Further, their research used aggregate measures collated from a census to estimate the demographic characteristics of patients rather than individual records. Thus, while granular, the approach is limited for identifying communities vulnerable to severe illness and/or death. The trends identified may more accurately reflect transmission, rather than underlying susceptibility.

Severe illness from COVID-19 requiring emergency medical services reflects the intersection of *exposure* and underlying *susceptibility*. Vulnerability to severe symptoms is the outcome of complex interactions between individual demographic characteristics and community-scale socio-economic and environmental factors. Our approach, identifying and interrogating unusual clusters of severe illness from COVID-19, and investigating associations between unusual clusters and social and environmental features of landscapes offers a methodology for further supporting vulnerable communities and regions in real-time.

6.1 Identifying unusual clusters and predicting cluster membership

Our spatial analysis revealed 13 statistically significant clusters of suspected COVID-19 cases (Figure 2) with rates of severe illness ranging from 951 to 3,417 per 100,000 population. Regression analysis identified 13 factors that predict cluster membership. Overall, the predictive accuracy of our regression model is high, with lower specific accuracy for cases occurring in clusters compared to cases occurring outside of clusters. However, the proportion of cases predicted in clusters and outside of clusters are both acceptable and suggest good model fit.

Compared to the reference condition (urban towns and cities), clusters of severe illness are more likely to occur in urban minor conurbations, urban cities and towns in sparse areas, rural towns and fringe areas, and rural villages and dispersed areas. Clusters occur *closer* to pubs/bars/clubs, off license stores, and Passive Green Space, and *further away* from fast food, tobacconists, GP practices, A&E hospitals, and pharmacies. The strongest predictors of cluster membership are closer location to Passive Green Space (such as commons and wilderness areas) and higher levels of NO_2. Strong associations were also found to PM10 and SO_2 levels[7].

Some landscape scale trends are consistent with wider literature. For example, NO_2 concentrations are associated with respiratory hospital admissions more generally (Pannullo et al., 2017) as well as COVID-19 related mortality (Kiaghadi et al., 2020; Travaglio et al., 2021). Our results also provide support for other research demonstrating increased vulnerability to disease in urban areas

[7] The small range of mean values for SO_2 (.9-1.9 µg m-3) has almost certainly inflated the odds ratio and effect size for this pollutant. Similarly, the wide odds ratio confidence interval for SO_2, compared to all other variables, suggests a high degree of uncertainty. Further, PM10 comprises numerous air pollutants, including organic matter, and has a long-range transport of thousands of kilometres (Malcom et al., 2010). The complex composition and movement of PM10 may explain the trends observed in our study, particularly in more regional and coastal areas where industrial activity and marine aerosols can contribute to concentration levels (Byrd et al., 2010).

compared to rural areas (Paul et al., 2020), and those further away from health services (Daras et al., 2019). Clusters 1 to 3 are entirely urban, while 10, 12, 13 and 7 are located in areas with higher than average proportions of sites in urban areas compared to non-clusters sites, and to national proportions (DEFRA, 2020).

Other findings are less consistent with assumptions about landscape teleconnections and health outcomes. The likelihood of cluster membership simultaneously increases at locations closer to off license stores and pubs/bars/clubs, but more distant from fast food venues and tobacconists. The AHAHI, and associated literature, assumes that access to all 'healthy' assets promotes better health condition while access to 'hazardous' assets facilitates poorer health condition (Green et al., 2018). These trends may be related to the nature of amenities and services in rural areas compared to urban areas, and may explain why some retail environments increase likelihood of cluster membership while others do no. Rural towns often contain local pubs while tobacconists and fast food venues are less common. Thus, distance from both retail outlets (more commonly found in more densely populated areas) *and* health services may reflect poorer access to services more generally, and thus greater vulnerability to illness (Jordan et al., 2004).

Similarly, health literature suggests that closer proximity to green space is associated with better health outcomes (Daras et al., 2018). We found that clusters are more likely to occur closer to, rather than more distant from Passive Green Space, like commons or conservation areas. This may be related to the nature of Passive, as opposed to Active green spaces. Passive Green Space like commons is likely to reflect urban periphery or rurality while Active Green Space like gymnasiums tend to be located in urban centres. The varied relationships between landscape features and vulnerability in urban compared to more regional areas, deserve more detailed consideration. For example, it is possible that closeness to Passive Green Space reflects social behaviour during the pandemic. In a perspective piece published in this Special Edition[8] we examine the relationship between landscape features and the implications for COVID-19 exposure and underlying susceptibility in more depth. During extended phases of lockdown parks and arboretums became social hubs that were poorly monitored by local authorities. News reports documented continual violations of social distancing rules in public spaces like beaches and common green areas. Thus, improving the monitoring and enforcement of social distancing in these spaces may be a future avenue for mitigating high rates of severe COVID-19 cases.

Below we suggest that the balance of expected and unexpected associations between unusual clusters and landscape features reflect differences in the individual characteristics of clusters, and the nature of vulnerability between more rural and more urban landscapes.

6.3 Characteristics of individual clusters

The characteristics of clusters vary in two ways. Firstly, the degree of relative risk, and secondly in relation to wider geographic context. In order, clusters with the highest relative risk compared to the medium value were east of Rugby (5), east of the Peak District (6), south west of Derby (10), west of Grimsby (7), Leicester (2), and East Grimsby (12) (Figure 3). This analysis gives some indication of regions where communities may be particularly vulnerable.

Spatial analysis (Figure 5-9) revealed several important geographic distinctions between clusters. On this basis we classify clusters in the following categories: inland urban, rural or rural-urban mosaic,

[8] *Rethinking the health implications of society-environment relationships in built areas: an assessment of the Access to Healthy and Hazards Index in the context of COVID-19.*

and coastal urban (Table 7). Category One, 'Inland Urban' including Nottingham, Leicester, Derby, Northampton, and Chesterfield, are predominately or entirely urban and characterized by closeness to healthy and hazardous services and beneficial physical environments. Clusters located closer to city centres (*central* Urban Inland: 1, 2, 3, 11) are more deprived, while clusters located in the periphery (*peripheral* Inland Urban: 13, 10) are more affluent. Category Two, 'Rural and Mosaic' clusters in the Peak District, near Rugby, and south west of Leicester, are either entirely rural or display a rural-urban mosaic with a higher proportion of cases in rural areas compared to areas with randomly distributed cases. Rural and Mosaic clusters are characterized by further distance from healthy and hazardous services, closer proximity to beneficial physical environments, and greater affluence. Category Three, 'Coastal Urban', includes clusters in predominately urban areas near Skegness and Grimsby. These clusters are characterized by deprivation, and further distance from all services and beneficial physical environments. Importantly, while each category includes clusters with higher levels of NO_2, the sources are likely to vary; traffic contributes to poor air quality in large urban centres, such as Nottingham, while the operation of power plants effects air quality in more regional areas, such as the coastal Grimsby clusters.

Table 7. Characteristic of clusters categorized as 'Inland Urban', 'Rural and Mosaic', and 'Coastal Urban', including Index of Multiple Deprivation (IMD), geographic location (inland or coastal), urban and rural dynamics, and Access to Healthy Assets and Hazardous Index (AHAHI).

		Inland Urban		**Rural & Mosaic**	**Coastal Urban**
		Central	*Peripheral*		
IMD		More deprived	More affluent	More affluent	More deprived
Geographic location		Inland		Inland	Coastal
Urban/rural		Entirely or higher than average % urban		Entirely or higher than average % rural	Entirely or higher than average % urban
AHAHI	Retail	Closer		More distant	More distant
	Health	Closer		More distant	More distant
	Physical	Closer		Closer	More distant
	Air pollution	Worse		Better	Variable*

*Skegness cluster has better quality; Grimsby clusters have poorer quality.

6.4 Understanding vulnerability in the social-environmental landscape

Bioecological models suggest that health outcomes are the cumulative result of complex interactions between individual demographic and biological factors, and the social characteristics of wider environments (Bronfenbrenner, 1979; Eriksson et al., 2018). Social factors may reflect both exposure to transmission of a contagious virus, (e.g. poorly designed housing estates), and underlying susceptibility related to pre-existing health conditions (Patel et al., 2020). In addition to social dynamics, our analysis included physical characteristics of the built environment that may explain vulnerability to severe symptoms of infectious disease, such as distance from green space (Green et al., 2018).

Our analysis suggests that unusual clusters occur at the nexus of individual susceptibility and exposures in the built environment. However, the dynamics of vulnerability vary between geographic locations. For example, Inland Urban clusters are located *closer* to all services while Coastal Urban clusters are located *further* from all services. Except Peripheral Inland Clusters, these regions are more deprived than areas with cases occurring randomly. Thus, the *cumulative* effect of exposure in high density urban areas and susceptibility associated with deprivation may be more important determinants of vulnerability than distance from specific healthy and hazardous features of built environments.

The characteristics of clusters in more affluent areas, including two with very high relative risk (10, 6) suggests that the dynamics of vulnerability vary markedly from clusters in poorer regions. Firstly, affluent clusters tend to be in more regional locations including urban peripheries and rural areas which are typically occupied by older communities (ONS, 2020c). In the U.S., rural communities with high rates of severe COVID-19 symptoms are characterised by aging populations and greater distance from health services (Lakhani et al., 2020). Similar characteristics may explain high relative risk in more peripheral and rural clusters in the East Midlands. With one exception (Skegness), the average age of patients located in the Peripheral Inner Urban clusters is higher than all other clusters, and Rural and Mosaic clusters are located further from health services. These dynamics indicate a 'rural paradox'; lower risk of transmission, greater susceptibility to severe symptoms, and less access to the medical services required to meet the needs of susceptible communities. Taken together, these observations suggest that the relative contribution of demographic, socio-economic, and environmental factors to vulnerability varies depending on wider geographic location. Factors that influence underlying health susceptibility, like older age and distance from health services, may be stronger predictors of severe illness than socio-economic status in regional locations that are less exposed to transmission risks. In contrast, deprivation and high-density urbanism may outweigh the benefits of closeness to, or distance from, physical features of the built environment. In these cases, it is likely that susceptibility related to deprivation, and exposure related to urbanization, are more powerful drivers of overall vulnerability than access to health services or retail outlets.

Similar to closeness to passive green spaces discussed above, our findings about vulnerability in rural areas suggest some policy responses for future pandemics and phases of lockdown. News reporting during the first national phase of lockdown suggests that the public viewed rural areas as less vulnerable to contagion and mortality related to COVID-19 compared to urban areas (e.g., McCarthy, 2020). Further, rural communities reported the phenomenon of people from urban and peri-urban areas 'flocking' to rural regions for recreation during phases of lockdown when only essential travel was legally permitted (Asquith, 2020). In the event of future phases of lockdown, mitigating high rates of severe illness in rural areas with aging populations may require more stringent policing of travel between urban and rural areas.

In summary, factors and processes that explain vulnerability to severe COVID-19 illness and or death are complex and highly location-specific. Bioecological models traditionally focus on the interaction between complex social systems while urban theories emphasise distal associations within physical environmental landscapes. We suggest that social *and* physical landscape factors rightly belong in a theoretical space akin to Bronfenbrenner's Mesosystem, which includes processes and interactions that occur within homes, communities, and neighbourhoods (Bronfenbrenner, 1979). Figure 11 visualises what this Mesosystem might look like.

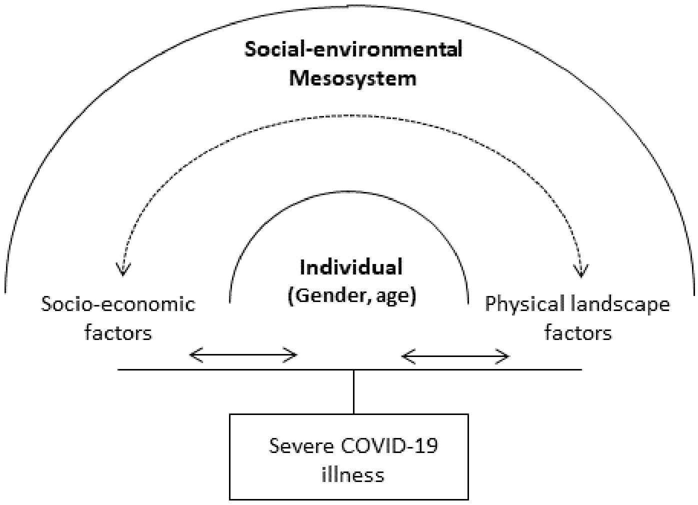


Figure 11. Schematic showing the social-environmental Mesosphere demonstrating the multi-level factors associated with severe illness from COVID-19. The dotted arrow indicates the interaction between socio-economic factors and physical landscape factors within the Mesosphere.

The granularity of our analysis, facilitated by the high resolution of the data, offers some important insights for supporting the most vulnerable communities in real-time during the early phase of a pandemic when laboratory testing is limited and public policy is informed by cases in the community. In the case of COVID-19, those vulnerable communities include deprived urban neighbourhoods *and* more affluent regional neighbourhoods.

6.5 Strengths and limitations

There are three limitations of the research. Firstly, big data does not capture individual behaviour; distance from green space and other amenities does not reflect use. Secondly, factors beyond the scope and scale of this research may affect ambulance use. People within close proximity to hospitals with A&E services are more likely to access those services directly rather than calling an ambulance. Similarly, willingness to call an ambulance may vary between communities. Poor health literacy, including ability to recognize symptoms of illness, is often associated with deprivation (Niksic et al., 2015). As a result, it is likely that our data does not represent all severe cases of suspected COVID-19 in the study region. Thus, qualitative community scale research is needed to ground truth the trends and associations reported here. Finally, without data linkage, suspected COVID-19 cases cannot be confirmed. However, the preliminary diagnosis of suspected COVID-19 is based on the assessment of trained medical professionals following the guidelines and algorithms that were widely employed by medical services in the early phases of the pandemic before rapid testing was available. Further, measures taken by ambulance paramedics, including blood oxygen levels (Soltan et al., 2021) and self-reported OTD (Wee, 2020; Patterson, 2020; Printza & Constantinidis, 2020), have been demonstrated to predict positive cases with a high degree of accuracy. The need for rapid response is paramount. The spatial accuracy of our approach, using a novel routinely collated dataset demonstrates a methodology for identifying vulnerable communities in real-time, as well as understanding the demographic, socio-economic, and environmental characteristics of vulnerability across dynamic geographic landscapes.

7. Conclusions:

Vulnerability to severe illness from contagious disease occurs at the intersection of exposure and underlying susceptibility. The effect of biological, social, and environmental risk factors is cumulative. Thus, single characteristics of built environments like deprivation or air pollution do not explain severe symptoms that require emergency medical attention. Our analysis builds on advancements in public health, epidemiology and urban planning by integrating features of the built environment with more traditional bioecological frameworks that tend to focus on complex social interactions.

The analysis of ambulance attendance data for monitoring the progress of the pandemic is a novel approach in the UK, and to our knowledge, has not been used to identify clusters of COVID-19 elsewhere. We acknowledge that analysing suspected COVID-19 cases is an imperfect science. However, we offer some insights that may be of benefit for rapid response as well as longer-term urban planning:

- Joining ambulance data to publicly available big datasets like the IMD and AHAHI could identify vulnerable communities in real-time;
- Understanding the social *and* environmental characteristics of vulnerability may help policy makers to mitigate the impact of a new EID on communities;
- Identifying vulnerable communities in real-time could inform earlier localised lockdowns to mitigate transmission and reduce rates of severe illness. Targeting areas where contagion is likely to result in high rates of hospitalisation would also reduce burden on emergency medical services;
- Opportunities for mitigating transmission also include more effective monitoring and enforcement of social distancing rules in Passive Green Space, including parks, commons and arboretums, as well as for urban-rural travel during lockdown;
- The dynamics of vulnerability vary between urban centres and more peripheral or rural regions, and between more deprived compared to more affluent communities. The opportunities for minimising the impacts of a pandemic include reducing the underlying susceptibility of communities as well as minimising transmission. In part, this involves urban planning to enhance opportunities for health behaviours. Improving the safety of green spaces for cost-free exercise, and increasing infrastructural and financial access to healthy food would promote healthier lifestyles in deprived communities. Further, improving access to health services in more affluent and isolated communities may help to mitigate the most severe outcomes of a pandemic. However, in both cases this requires top-down financial investment to encourage healthy retail outlets to locate in deprived neighbourhoods, and health services to locate in low-density neighbourhoods.

At the time of writing, twelve months has elapsed since the declaration of the COVID-19 pandemic and the introduction of national responses to contain transmission. Some approaches have proven more successful than others. Identifying unusual clusters of suspected COVID-19 cases and the factors that predict the location of clusters offers a way forward for the UK to adopt more targeted physical distancing approaches that have been effective for preventing further outbreaks and reducing the economic burden of nation-wide lockdown elsewhere. Unequal health outcomes and severe illness in the UK reflects decades of systemic disadvantage and accumulated vulnerability (Marmot et al., 2020). Addressing underlying susceptibility will require long-term investment in areas including neighbourhood quality, educational attainment, and closing income gaps. Mitigating the impact of future pandemics necessarily involves ‘levelling up’ health across the UK, including between rural and urban spaces, coastal and inland spaces, and deprived and affluent communities.

As a global society, we have entered an indeterminate phase of uncertainty and trial-and-error in combating the pandemic. In the wake of the most immediate threat to human life, policy makers face the challenge of redefining the relationship between societies and their urban habitats. Utilizing big-data to identify hot-spots of vulnerability could be used as a method to inform current mitigation policy, as well as longer-term transitions towards healthier urban landscapes.

References:

Asquith, J., 2020. *People have been flocking to rural areas during COVID-19 lockdowns*, viewed 24 June, 2021, https://www.forbes.com/sites/jamesasquith/2020/03/29/people-have-been-flocking-to-rural-areas-during-covid-19-lockdowns/?sh=69382a576578.

Balasegaram, S., Ogilvie, F., Glasswell, A., Anderson, C., Cleary, V., Turbitt, D. and McCloskey, B., 2012. Patterns of early transmission of pandemic influenza in London–link with deprivation. *Influenza and Other Respiratory Viruses*, *6*(3), pp.e35-e41.

Beevers, S.D., Westmoreland, E., de Jong, M.C., Williams, M.L. and Carslaw, D.C., 2012. Trends in NO_x and NO_2 emissions from road traffic in Great Britain. *Atmospheric Environment*, *54*, pp.107-116.

Byrd, T., Stack, M. and Furey, A., 2010. The assessment of the presence and main constituents of particulate matter ten microns (PM10) in Irish, rural and urban air. *Atmospheric Environment*, *44*(1), pp.75-87.

Calabrese, R., 2014. Optimal cut-off for rare events and unbalanced misclassification costs. *Journal of Applied Statistics*, *41*(8), pp.1678-1693.

Coleman, M., Coleman, M., Mabuza, A.M., Kok, G., Coetzee, M. and Durrheim, D.N., 2009. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malaria Journal*, *8*(1), p.68.

Connolly, C., Keil, R. and Ali, S.H., 2021. Extended urbanisation and the spatialities of infectious disease: Demographic change, infrastructure and governance. *Urban Studies*, *58*(2), pp.245-263.

Cuadros, D.F., Tomita, A., Vandormael, A., Slotow, R., Burns, J.K. and Tanser, F., 2019. Spatial structure of depression in South Africa: A longitudinal panel survey of a nationally representative sample of households. *Scientific Reports*, *9*(1), pp.1-10.

Daras, K., Davies, A., Green, M. and Singleton, A., 2018. Developing indicators for measuring health-related features of neighbourhoods. *Consumer data research*, pp.167-77.

Daras, Konstantinos., Green, Mark A., Davies, Alec., Singleton, Alex., Barr, Benjamin., 2019. Access to Healthy Assets and Hazards (AHAH). https://doi.org/10.6084/m9.figshare.8295842.v1, viewed 5 August 2020.

Daras, K., Alexiou, A., Rose, T.C., Buchan, I., Taylor-Robinson, D. and Barr, B., 2021. How does vulnerability to COVID-19 vary between communities in England? Developing a Small Area Vulnerability Index (SAVI). *J Epidemiol Community Health*.

Department of Environment, Food, and Rural Affairs., 2020. Rural Population 2014/2015. https://www.gov.uk/government/publications/rural-population-and-migration/rural-population-201415#:~:text=In%202015%2C%2011.4%20million%20people,20.7%25%20of%20the%20England%20population., viewed 6 September 2020.

Durand, C.P., Andalib, M., Dunton, G.F., Wolch, J. and Pentz, M.A., 2011. A systematic review of built environment factors related to physical activity and obesity risk: implications for smart growth urban planning. *Obesity reviews*, *12*(5), pp.e173-e182.

European Commission., 2020. *East Midlands - Internal Market, Industry, Entrepreneurship And Smes - European Commission*, https://ec.europa.eu/growth/tools-databases/regional-innovation-monitor/base-profile/east-midlands, Accessed 21 September 2020.

Frank, L.D. and Wali, B., 2021. Treating two pandemics for the price of one: Chronic and infectious disease impacts of the built and natural environment. *Sustainable Cities and Society*, p.103089.

Green, M.A., Daras, K., Davies, A., Barr, B. and Singleton, A., 2018. Developing an openly accessible multi-dimensional small area index of 'Access to Healthy Assets and Hazards' for Great Britain, 2016. *Health & Place*, *54*, pp.11-19.

Gomez, J.E., Johnson, B.A., Selva, M. and Sallis, J.F., 2004. Violent crime and outdoor physical activity among inner-city youth. *Preventive Medicine*, *39*(5), pp.876-881.

Goryakin, Y., Rocco, L. and Suhrcke, M., 2017. The contribution of urbanization to non-communicable diseases: Evidence from 173 countries from 1980 to 2008. *Economics & Human Biology*, *26*, pp.151-163.

Guilmoto, C.Z., 2020. COVID-19 death rates by age and sex and the resulting mortality vulnerability of countries and regions in the world. *MedRxiv*.

Hu, H., Nigmatulina, K. and Eckhoff, P., 2013. The scaling of contact rates with population density for the infectious disease models. *Mathematical Biosciences*, *244*(2), pp.125-134.

Jordan, H., Roderick, P., Martin, D. and Barnett, S., 2004. Distance, rurality and the need for care: access to health services in South West England. *International Journal of Health Geographics*, *3*(1), p.21.

Kasda, E., Robson, C., Saunders, J., Adadey, A., Ford, B., Sinha, N., Teter, J., Warner, N. and Paine, L., 2020. Using event reports in real-time to identify and mitigate patient safety concerns during the COVID-19 pandemic. *Journal of Patient Safety and Risk Management*, *25*(4), pp.156-158.

Kelly, S., Martin, S., Kuhn, I., Cowan, A., Brayne, C. and Lafortune, L., 2016. Barriers and facilitators to the uptake and maintenance of healthy behaviours by people at mid-life: a rapid systematic review. *PloS One*, *11*(1), p.e0145074.

Khalatbari-Soltani, S., Cumming, R.C., Delpierre, C. and Kelly-Irving, M., 2020. Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. *J Epidemiol Community Health*, *74*(8), pp.620-623.

Kiaghadi, A., Rifai, H.S. and Liaw, W., 2020. Assessing COVID-19 risk, vulnerability and infection prevalence in communities. *Plos One*, *15*(10), p.e0241166.

Kim, G.U., Kim, M.J., Ra, S.H., Lee, J., Bae, S., Jung, J. and Kim, S.H., 2020. Clinical characteristics of asymptomatic and symptomatic patients with mild COVID-19. *Clinical Microbiology and Infection*, *26*(7), pp.948-e1.

Kontopantelis, E., Mamas, M.A., Deanfield, J., Asaria, M. and Doran, T., 2021. Excess mortality in England and Wales during the first wave of the COVID-19 pandemic. *Journal of Epidemiol Community Health*, *75*(3), pp.213-223.

Kulldorff, M., 1997. A spatial scan statistic. *Communications in Statistics-Theory and methods*, *26*(6), pp.1481-1496.

Kulldorff, M., Heffernan, R., Hartman, J., Assunçao, R. and Mostashari, F., 2005. A space–time permutation scan statistic for disease outbreak detection. *Plos Med*, *2*(3), p.e59.

Krieger, J. and Higgins, D.L., 2002. Housing and health: time again for public health action. *American Journal of Public Health*, *92*(5), pp.758-768.

Lachowycz, K. and Jones, A.P., 2011. Greenspace and obesity: a systematic review of the evidence. *Obesity Reviews*, *12*(5), pp.e183-e189.

Lakhani, H.V., Pillai, S.S., Zehra, M., Sharma, I. and Sodhi, K., 2020. Systematic review of clinical insights into novel coronavirus (CoVID-19) pandemic: Persisting challenges in US rural population. *International Journal of Environmental Research and Public Health*, *17*(12), p.4279.

Low, C.T., Lai, P.C., Tse, W.S.C., Tsui, C.K., Lee, H. and Hui, P.K., 2013. Exploring tuberculosis by types of housing development. *Social Science & Medicine*, *87*, pp.77-83.

Marmot, M., Allen, J., Boyce, T., Goldblatt, P. and Morrison, J. (2020). Health Equity in England: The Marmot Review 10 Years On. London, UK: Institute of Health Equity.

McCarthy, J., 2020. *Coronavirus: 'People think rural areas are clear – they aren't'*, viewed 24 June, 2021, https://www.bbc.co.uk/news/uk-wales-52230209.

Mehra, M.R., Desai, S.S., Kuy, S., Henry, T.D. and Patel, A.N., 2020. Cardiovascular disease, drug therapy, and mortality in Covid-19. *New England Journal of Medicine*, *382*(25), p.e102.

Miller, F.W., Alfredsson, L., Costenbader, K.H., Kamen, D.L., Nelson, L.M., Norris, J.M. and De Roos, A.J., 2012. Epidemiology of environmental exposures and human autoimmune diseases: findings from a National Institute of Environmental Health Sciences Expert Panel Workshop. *Journal of autoimmunity*, *39*(4), pp.259-271.

Namosha, E., Sartorius, B. and Tanser, F., 2013. Spatial clustering of all-cause and HIV-related mortality in a rural South African population (2000–2006). *PloS one*, *8*(7), p.e69279.

Neiderud, C.J., 2015. How urbanization affects the epidemiology of emerging infectious diseases. *Infection Ecology & Epidemiology*, *5*(1), p.27060.

Niksic, M., Rachet, B., Warburton, F.G., Wardle, J., Ramirez, A.J. and Forbes, L.J., 2015. Cancer symptom awareness and barriers to symptomatic presentation in England—are we clear on cancer?. *British Journal of Cancer*, *113*(3), pp.533-542.

Norris, J.M., Johnson, R.K. and Stene, L.C., 2020. Type 1 diabetes—early life origins and changing epidemiology. *The Lancet Diabetes & Endocrinology*, *8*(3), pp.226-238.

Oddo, P.C. and Bolten, J.D., 2019. The value of near real-time earth observations for improved flood disaster response. *Frontiers in Environmental Science*, *7*, p.127.

Office of National Statistics., 2011. Usual Resident Population, https://www.nomisweb.co.uk/query/asv2htm.aspx, accessed 6 August 2020.

Office for National Statistics,. 2020a. Population Estimates For The UK, England And Wales, Scotland And Northern Ireland, Provisional, <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019, Accessed 19 September 2020.

Office for National Statistics, 2020b., 2011 Census: Aggregate Data, [data collection], UK Data Service, Accessed 21 September 2020. SN: 7427, http://doi.org/10.5257/census/aggregate-2011-2.

Office of National Statistics, 2020c. *Rural Population 2014/2015*, https://www.gov.uk/government/statistics/rural-population-and-migration/rural-population-201415, accessed 5 March 2021.

Ortolan, A., Lorenzin, M., Felicetti, M., Doria, A. and Ramonda, R., 2020. Does gender influence clinical expression and disease outcomes in COVID-19? A systematic review and meta-analysis. *International Journal of Infectious Diseases*, *99*, pp.496-504.

Pannullo, F., Lee, D., Neal, L., Dalvi, M., Agnew, P., O'Connor, F.M., Mukhopadhyay, S., Sahu, S. and Sarran, C., 2017. Quantifying the impact of current and future concentrations of air pollutants on respiratory disease risk in England. *Environmental Health*, *16*(1), p.29.

Patel, J.A., Nielsen, F.B.H., Badiani, A.A., Assi, S., Unadkat, V.A., Patel, B., Ravindrane, R. and Wardle, H., 2020. Poverty, inequality and COVID-19: the forgotten vulnerable. *Public Health*, *183*, p.110.

Paul, R., Arif, A.A., Adeyemi, O., Ghosh, S. and Han, D., 2020. Progression of COVID-19 From Urban to Rural Areas in the United States: A Spatiotemporal Analysis of Prevalence Rates. *The Journal of Rural Health*.

Peric, S. and Stulnig, T.M., 2020. Diabetes and COVID-19. *Wiener Klinische Wochenschrift*, *132*(13), pp.356-361.

Public Health England, 2018. *East Midlands Profile A Summary Of Public Health In The Region 2018*. PHE publications, p.7.

Romero Starke, K., Petereit-Haack, G., Schubert, M., Kämpf, D., Schliebner, A., Hegewald, J. and Seidler, A., 2020. The age-related risk of severe outcomes due to COVID-19 Infection: a rapid review, meta-analysis, and meta-regression. *International Journal of Environmental Research and Public Health*, *17*(16), p.5974.

Rushton, S.P., Goodfellow, M., O'Donnell, A.G. and Magee, J.G., 2007. The epidemiology of atypical mycobacterial diseases in northern England: a space–time clustering and Generalized Linear Modelling approach. *Epidemiology & Infection*, *135*(5), pp.765-774.

Seo, S., Choi, S., Kim, K., Kim, S.M. and Park, S.M., 2019. Association between urban green space and the risk of cardiovascular disease: A longitudinal study in seven Korean metropolitan areas. *Environment international*, *125*, pp.51-57.

Seto, K.C., Reenberg, A., Boone, C.G., Fragkias, M., Haase, D., Langanke, T., Marcotullio, P., Munroe, D.K., Olah, B. and Simon, D., 2012. Urban land teleconnections and sustainability. *Proceedings of the National Academy of Sciences*, *109*(20), pp.7687-7692.

Shen, Y.S. and Lung, S.C.C., 2016. Can green structure reduce the mortality of cardiovascular diseases?. *Science of The Total Environment*, *566*, pp.1159-1167.

Smith, C.M., Lessells, R., Grant, A.D., Herbst, K. and Tanser, F., 2018. Spatial clustering of drug-resistant tuberculosis in Hlabisa subdistrict, KwaZulu-Natal, 2011–2015. *The International Journal of Tuberculosis and Lung Disease*, *22*(3), pp.287-293.

Steinberg, E., Wright, E. and Kushner, B., 2020. In young adults with COVID-19, obesity is associated with adverse outcomes. *Western Journal of Emergency Medicine*, *21*(4), p.752.

Stier, A., Berman, M. and Bettencourt, L., 2020. COVID-19 attack rate increases with city size. *Mansueto Institute for Urban Innovation Research Paper Forthcoming*.

Sze, S., Pan, D., Nevill, C.R., Gray, L.J., Martin, C.A., Nazareth, J., Minhas, J.S., Divall, P., Khunti, K., Abrams, K.R. and Nellums, L.B., 2020. Ethnicity and clinical outcomes in COVID-19: a systematic Review and Meta-analysis. *EClinicalMedicine*, p.100630.

Tanser, F., Bärnighausen, T., Dobra, A. and Sartorius, B., 2018. Identifying 'corridors of HIV transmission'in a severely affected rural South African population: a case for a shift toward targeted prevention strategies. *International Journal of Epidemiology*, *47*(2), pp.537-549.

The Health Foundation. 2020. *NHS Test and Trace: the journey so far*. Published 23 September, 2020, https://www.health.org.uk/publications/long-reads/nhs-test-and-trace-the-journey-so-far, accessed 28 February 2021.

Tomita, A., Cuadros, D.F., Mabhaudhi, T., Sartorius, B., Ncama, B.P., Dangour, A.D., Tanser, F., Modi, A.T., Slotow, R. and Burns, J.K., 2020. Spatial clustering of food insecurity and its association with depression: a geospatial analysis of nationally representative South African data, 2008–2015. *Scientific RepoRtS*, *10*(1), pp.1-11.

Travaglio, M., Yu, Y., Popovic, R., Selley, L., Leal, N.S. and Martins, L.M., 2021. Links between air pollution and COVID-19 in England. *Environmental Pollution*, *268*, p.115859.

Viegi, G., Maio, S., Pistelli, F., Baldacci, S. and Carrozzi, L., 2006. Epidemiology of chronic obstructive pulmonary disease: health effects of air pollution. *Respirology*, *11*(5), pp.523-532.

World Health Organization., 2020. Media Statement: Knowing the risks of COVID-19. https://www.who.int/indonesia/news/detail/08-03-2020-knowing-the-risk-for-covid-19#:~:text=Most%20people%20(about%2080,are%20at%20greater%20risk.. Viewed, October 1 2020.

Wu, T., Perrings, C., Kinzig, A., Collins, J.P., Minteer, B.A. and Daszak, P., 2017. Economic growth, urbanization, globalization, and the risks of emerging infectious diseases in China: a review. *Ambio*, *46*(1), pp.18-29.

Author statement of contributions

Dr Moore undertook the SatScan™ and regression analysis and wrote the first draft of the manuscript. Mr Hill completed all data linkage and prepared all figures and maps. Professor's Gussy, Siriwardena, Tanser, Law and Thomas were involved in study design and conception, as well as contributing substantive editing of multiple drafts. Professor's Gussy, Siriwardena and Tanser contributed substantively to manuscript revisions. Mr Spaight oversaw data cleaning, analysis and interpretation, and provided access to data from the East Midlands Ambulance NHS Trust.

Click here to access/download
Supplementary Material
supplimentary material.docx

Click here to access/download
Supplementary Material
LSOA_IMD_RUC_AHAHI.shx

Click here to access/download
Supplementary Material
LSOA_IMD_RUC_AHAHI.xlsx

Supplementary Material

Click here to access/download
Supporting File
LSOA_IMD_RUC_AHAHI.prj