Using drivers and transmission pathways to identify SARS-like coronavirus spillover risk hotspots

Renata L. Muylaert[1*], *David A Wilkinson*[2], *Tigga Kingston*[3], *Paolo D'Odorico*[4], *Maria Cristina Rulli*[5], *Nikolas Galli*[5], *Reju Sam John*[6], *Phillip Alviola*[7], *David T. S. Hayman*[1]

[1] *Massey University, Palmerston North, New Zealand*

[2] *UMR ASTRE, CIRAD, INRAE, Université de Montpellier, Plateforme Technologique CYROI, Sainte-Clotilde, La Réunion France*

[3] *Department of Biological Sciences, Texas Tech University, Lubbock, TX, U.S.A.*

[4] *Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, U.S.A.*

[5] *Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, Italy*

[6] *Department of Physics, Faculty of Science, University of Auckland, Auckland, New Zealand*

[7] *Institute of Biological Sciences, University of the Philippines- Los Banos, Laguna, Philippines*

*Corresponding author**:** R.deLaraMuylaert@massey.ac.nz

Abstract: The emergence of SARS-like coronaviruses is a multi-stage process from wildlife reservoirs to people. Here we characterize multiple drivers—landscape change, host distribution, and human exposure—associated with the risk of spillover of SARS-like coronaviruses to help inform surveillance and mitigation activities. We consider direct and indirect transmission pathways by modeling four scenarios with livestock and mammalian wildlife as potential and known reservoirs before examining how access to healthcare varies within clusters and scenarios. We found 19 clusters with differing risk factor contributions within a single country (N=9) or transboundary (N=10). High-risk areas were mainly closer (11-20%) rather than far (<1%) from healthcare. Areas far from healthcare reveal healthcare access inequalities, especially Scenario 3, which includes wild mammals as secondary hosts. China (N=2) and Indonesia (N=1) had clusters with the highest risk. Our findings can help stakeholders in land use planning integrating healthcare implementation and One Health actions.

Key-words: zoonotic risk, viral emergence, land conversion, deforestation, host diversity, coronavirus, pandemics, sarbecovirus, One Health, scenario analysis

Introduction

The process of infectious disease emergence from animals begins with the cross-species transmission (spillover) of a microbe (e.g., virus, bacteria, fungus, protozoans) to a new animal host in which it is pathogenic [1–3]. Identifying areas of higher spillover probability is an important strategy for pandemic prevention, and has largely focussed on estimating host distributions [4,5] and modeling frameworks for adding proxies for disease risk and spread in the face of limited data [4–6]. Yet, successful emergence events are complex multi-stage processes with many possible pathways leading from the original wildlife reservoir to sustained transmission in people [7]. The probability of any of these pathways occurring and resulting in infection emergence varies temporally and spatially, so cross-scale mapping of the multiple, diverse drivers of disease emergence is needed to better allow decision-makers to know where to focus surveillance and mitigation strategies[8].

Human infectious diseases almost all came from other species [3]. COVID-19, Ebola virus disease, HIV/AIDS and Zika virus disease are recent examples, whereas those like measles arose after the Neolithic Agricultural Revolution[9]. Zoonotic disease emergence has accelerated in recent decades, likely as a result of diverse interacting drivers such as accelerated land use change [10], human encroachment of natural habitats, increasing and changing contacts among and between wildlife and domestic animals, and has been mostly linked to mammals and birds [11]. Bats are among the natural hosts of viruses in the coronavirus (family *Coronaviridae*) subgenus *Sarbecovirus* (Severe acute respiratory syndrome (SARS)–related coronaviruses), which includes SARS-CoV-1 and SARS-CoV-2 [12,13]. Bat hosts of sarbecoviruses are broadly distributed but the highest diversity is in Southeast Asia [5]. Human infection with *Sarbecovirus* from bats may be more frequent than reported from traditional surveillance [14] and likely includes secondary hosts [15,16]. Viral infection prevalence contributes to the risk of spillover [2], and can be influenced by biological factors such as birthing cycles [17,18] and external stimuli such as human changes to land use [19] (but see [20,21]).

Large scale risk assessments in which areas with similar risk profiles are identified provide invaluable information [12,22] and can be rapid, while the development of local, detailed and intricate spillover and outbreak risk assessments are costly and can take a long time [23,24]. Since detailed and validated data for recent reports on outbreak risk reduction are lacking for most regions of the globe (e.g. the Sendai framework, https://sendaimonitor.undrr.org/), a broad evaluation targeting *Sarbecovirus* emergence can be advantageous to discuss diverse contexts across the region where most natural hosts of sarbecoviruses occur. Human encroachment has led to decreased distances between bat roosts and human settlements [25], so part of the relevant hazard for inferring spillover risk can be spatially quantified from remotely sensed proxies for socioecological risk factors.

Here, we identify where putative drivers for emergence risk overlap, focusing on the biological possibility of the emergence of a *Sarbecovirus*. Our goal is to aid mitigation and

surveillance activities throughout South, East and Southeast Asia, by identifying both where efforts should focus and which risk factors should be prioritized based on where outstanding values overlap the most (hotspots) and based on healthcare access. Specifically, we asked: Where are the hotspots of risk that capture a range of hypothesized transmission scenarios representing different pathways to the emergence of a novel SARS-like coronavirus? Can we identify spatially cohesive clusters of hypothesized risk drivers that, when combined, increase risk of zoonotic spillover [22,26]?; How do the access to healthcare for high-risk areas vary according to transmission scenario?

We predicted that there would be co-occurring hotspots for most risk drivers converging in biodiverse regions with bat hosts in regions with greatest pressure from anthropogenic land use change. The four scenarios evaluated represent different nested transmission pathways. We assume that the risk of emerging new SARS-like outbreaks is associated with social, biological and environmental components and, because there are unobserved dynamics for emerging viruses [27], we evaluated four nested spillover pathway scenarios based on landscape change and potential hosts [28]: Scenario 1 (direct transmission - known bat hosts) represents direct transmission from bats to people, facilitated by the landscape condition, human population, and known bat hosts. Although molecular investigations suggest that direct transmission of sarbecoviruses from bats to humans may be possible [29], it has yet to be better documented [14,30,31]. Rather, the involvement of an intermediary or bridging host appears more likely, perhaps because this allows for recombination and viral evolution, and/or leads to greater exposure to human populations. Consequently, we developed Scenarios 2, 3, and 4 to represent indirect pathways that build on Scenario 1 by adding livestock (Scenario 2, indirect transmission - mammalian livestock) and wild mammals (Scenario 3, indirect transmission - wild mammals). Scenario 4 (indirect transmission - all mammals) is a global scenario comprising landscape condition, human population, known bat hosts, mammalian livestock and wild mammals. We expected clusters to occur across country borders and differing pathways to alter risk hotspot distributions.

Results

Characterization of risk driver hotspots

The study region comprises a 25796-pixel grid for the terrestrial area evaluated. Univariate hotspot areas differ in magnitude (Figure 1) and extent. Most hotspots concentrate at latitudes between 20 and 40 degrees. The univariate hotspots with the largest spatial extent are those obtained for agricultural and harvest land, followed by high integrity forests and areas with high deforestation potential. The majority of the included region comprises coldspots for primary bat hosts. Drivers with the greatest extent of coldspots were livestock (pigs and cattle) followed by known bat hosts. The largest extent of intermediate areas was for built up land, which presented no coldspots due to the ubiquitous nature of

human occupation in terrestrial areas. The largest differences in results for all Bovidae livestock versus cattle-only hotspots (see Methods) are in central China, parts of north (Hebei, Shanxi, and Henan) China and central India (Figure S2). The complete overlap of hotspots considering all ten univariate hotspots at one grid cell never occurred.

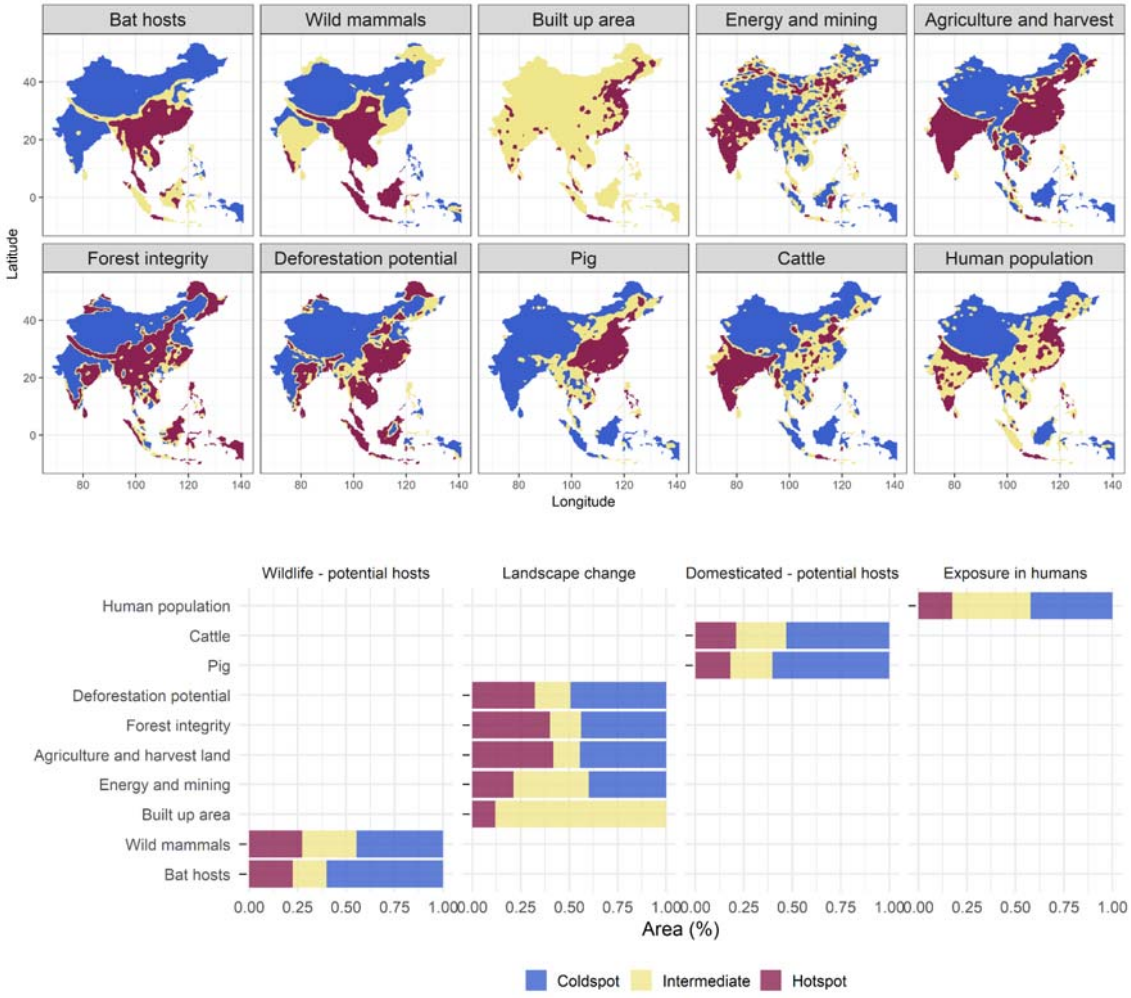

Figure 1. Hotspots of potential factors contributing to emergence of SARS-like coronaviruses. The upper panel shows the spatial distribution of hotspots based on putative drivers of risk of new *Sarbecovirus* emergence evaluated in four scenarios. Bottom panel shows the proportion (%) of areas classified as hotspots, intermediate or coldspots across the study region, including wildlife, landscape change, livestock and exposure in humans. Areas in red represent hotspots, yellow zones are intermediate areas and coldspots in blue, at a 95% alpha error level.

Scenarios

Regardless of scenario, the largest hotspot overlaps occur in central and southeast China, south and northwestern India and Java. Differences between Scenario 1 (direct transmission - known bat hosts) with potential primary known bat hosts and Scenario 4 (indirect transmission - all mammals) are largest in central China (Figure 2). The largest differences between each scenario and Scenario 1 (the scenario with fewest covariates) concentrated in central and southern China. Scenario 3 had the least differences in relation to Scenario 1. Similar to Scenario 1, Scenario 2 shows most hotspot convergences in central and south China. For Scenario 4 (indirect transmission - all mammals) the most important PCA axes show a clear 'natural axis' and an anthropogenic axis, where the pig production layer is intermediate to the influence of both axes (Figure S3). Both main axes explain 58.7% of the total variation (PC1 = 33.5%, PC2 = 24.8%). Maximum overlap for non-human potential primary and secondary hosts occurred across China and Vietnam. While strong commonalities remain, the differences between scenarios (Figure 2) highlight how the involvement or not of domestic or other wildlife hosts alters the risk profile.

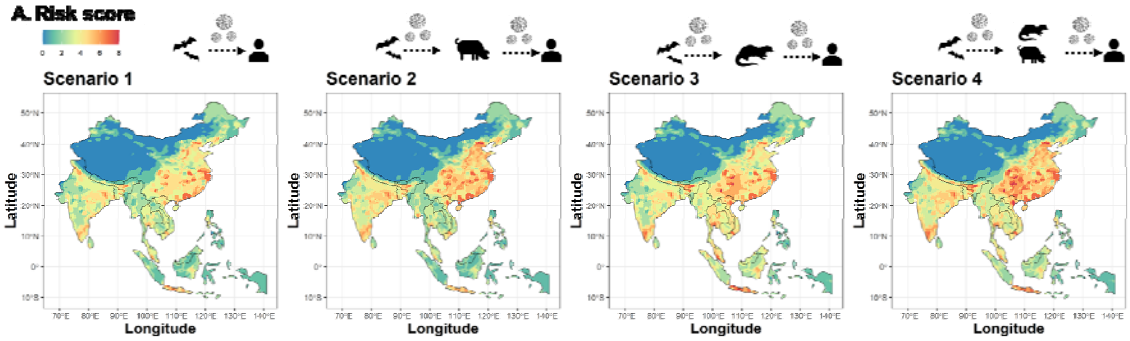

Figure 2. Emergent risk scores for scenarios containing co-occurring drivers associated with landscape change and zoonotic pathogen emergence. Landscape, human population and known bat hosts are included in all models, and are the sole drivers in Scenario 1, representing direct transmission. To incorporate indirect transmission through secondary hosts, mammalian livestock are included in Scenario 2, wild mammals in Scenario 3, and both mammalian livestock and wild mammals in Scenario 4. The internal white area in the continent represents no data values for Lake Qinghai; the largest lake in China.

Hotspot co-occurrence in clusters

We identified spatially cohesive clusters by including all hypothesized risk drivers (Figure 3). The optimal number of multivariate spatial clusters is nine when 10% of the human population is used as a minimum bound variable and 19 for 5% of the human population. There is an incremental benefit reduction from adding clusters, from nineteen groups on (Figure S4). The clusters from the cut-off value of 5% are nested within the 10% clusters (Figure S5), and we present the clusters for 19 areas.

The detected clusters were commonly transboundary (Table 1), located across a maximum of six countries, and a median of 2 countries. Besides 10 transboundary clusters, nine clusters are restricted to a single country, 6 in China, 2 in India, and 1 in Indonesia (Java). The top ranking clusters in terms of hotspots were all located in a single country (China or Indonesia): Beijing (cluster 19), Java (cluster 17), and Sichuan and Yuzhong District, Chongqing (cluster 16). The clusters with highest scores were among the smaller clusters in geographical extent. Inner-West China (cluster 1), South Lhasa and Arunachal Pradesh (cluster 15), and Philippines, East Timor, West Papua (cluster 9) had the highest scores for coldspots. Areas with the highest scores for the Intermediate class were Assam, West Burma block, Steppe and Sri Lanka (cluster 2), followed by Southwest Indochina (cluster 11) and North India (cluster 14). Clusters with the all Bovidae livestock version are in Figure S6, and they were very similar to the cattle-only versions from the main text, except for the Beijing area and the division of the two larger clusters in India, West India and East India.

Table 1. Identification of spatial clusters and the number of variables for which the median were coldspots, intermediate or hotspots (n=190). The top three values for each hotspot, intermediate and coldspot are in boldface.

Cluster ID	Countries	Number of countries	Coldspot	Intermediate	Hotspot
1	Bhutan, China, India, Nepal	4	**9**	1	0
2	Bangladesh, Bhutan, China, India, Sri Lanka, Myanmar	6	0	**8**	2
3	Brunei, Indonesia, Lao PDR, Myanmar, Malaysia, Thailand	6	3	4	3
4	India	1	4	3	3
5	China	1	0	5	5
6	China	1	6	3	1
7	China, Cambodia, Lao PDR, Myanmar, Thailand, Vietnam	6	0	5	5
8	India	1	2	3	5
9	Philippines, Indonesia, East Timor	3	**7**	2	1
10	China	1	5	0	5
11	Cambodia, Lao PDR, Philippines, Thailand, Vietnam	5	0	**6**	4
12	China	1	5	4	1
13	Bangladesh, Bhutan, India, Nepal	4	4	2	4

14	Bhutan, China, India, Nepal	4	1	**6**	3
15	Bhutan, China, India, Myanmar, Nepal	5	**8**	1	1
16	China	1	2	2	**6**
17	Indonesia	1	1	2	**7**
18	Bangladesh, India	2	2	4	4
19	China	1	1	1	**8**
		Total	60	62	68

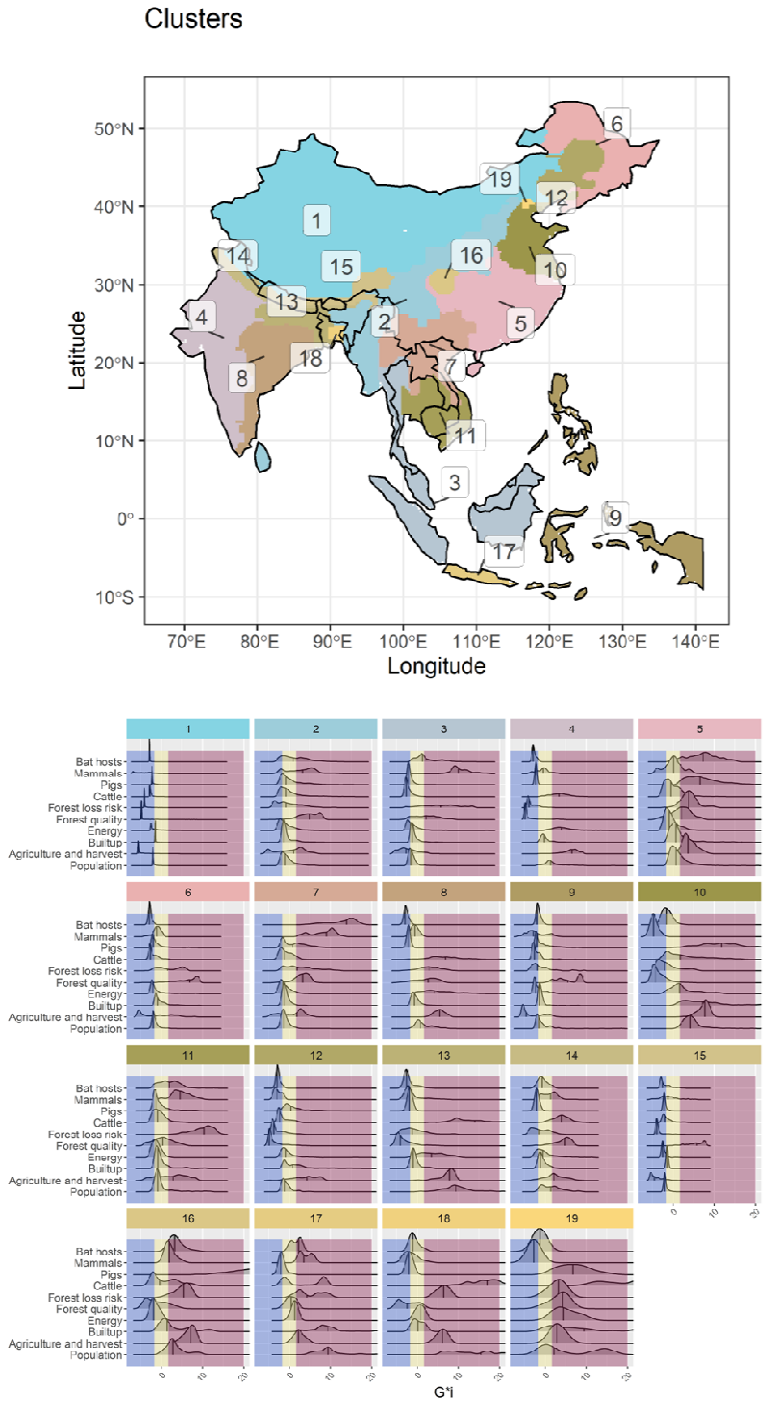

Figure 3. Distribution of clusters of risk factors associated with potentially new emerging SARS-like coronaviruses. The values include all potential mammalian hosts, land use change and human exposure density distributions (Scenario 4). Areas located in the red zone represent hotspots, yellow zones are intermediate areas and coldspots in blue, at a 95% alpha error level.

Potential outbreak detection and spread

When we cross the risk factor spatial information with healthcare access measured as travel time, the largest differences between combinations of quantiles of the two covariates are in the lowest and highest quantiles of both variables (Figure 4A, GIF 1). We calculated the areas with high-risk values that are far or close to healthcare for all scenarios (Figure S7) within the spatial clusters from the skater analysis. From the entire study region, areas closer to healthcare that had high hotspot overlap (areas in yellow in Figure 4, Figure S7) covered an area ranging from 11.96% in Scenario 1, to 20.28% in Scenario 2, 14.66% in Scenario 3, and 13.67% in Scenario 4. Areas far from healthcare that present high hotspot overlap (in red Figure 4, Figure S7) were much rarer and varied according to scenarios, always covering less than 1% of the studied region, ranging from 0.1% in Scenario 1, to 0.30 in Scenario 2, 0.91% in Scenario 3—with significantly higher travel times to healthcare (Figure 4B, $p < 0.05$)— and 0.22% in Scenario 4 (Table S1, Figure 4B). The relationship between travel time to healthcare and human population counts (Figure S8) shows that areas far from healthcare tend to have lower population counts (out proxy for exposure), but the relationship is non-linear.

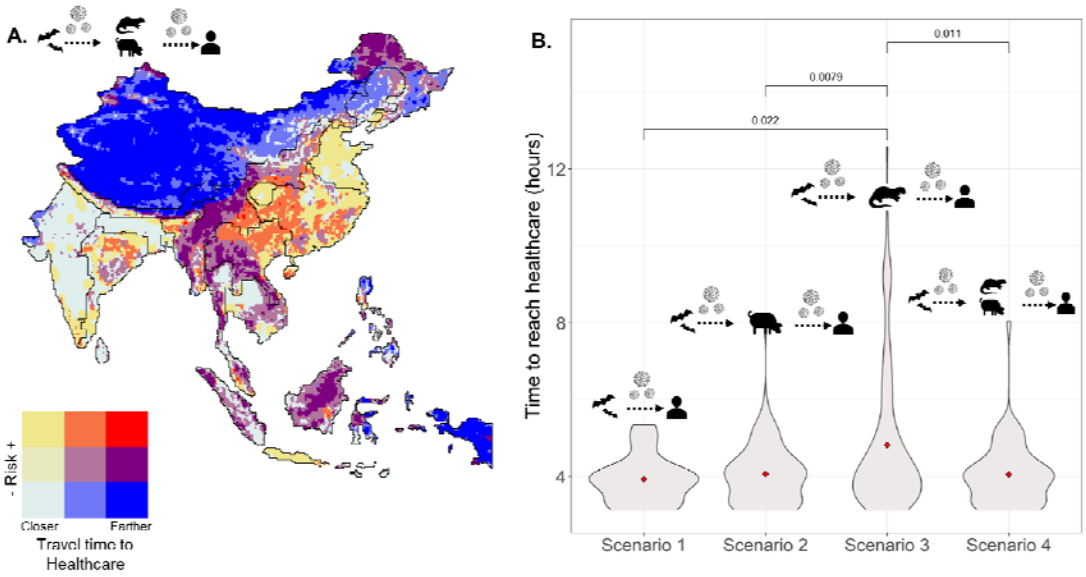

Figure 4. Bivariate maps crossing emergent risk scores from hotspot data on risk quantiles and access to healthcare. A. Black lines divide the limits for the 19 clusters identified by the multivariate spatial cluster analysis; Scenario four is represented in the map. B. Time to reach healthcare in areas where high emergent risk co-occurred far from healthcare for each of the four scenarios. Average values in red, p-value for differences is shown.

Discussion

Urgent actions are needed to decrease disease emergence risk [32,33]. Using a macroscale approach, we assessed the distribution of locations with a greater risk of experiencing *Sarbecovirus* spillover events using landscape conditions and exposure of potential hosts (wildlife, domestic, human). Landscape conditions coupled with predictions of the distribution of known hosts and proxies for potential hosts and processes linked to human exposure to novel viruses can be a powerful tool for sample prioritization when limited viral spillover information is available, such as for sarbecoviruses [14].

The overlap of risk factor hotspots represents pressure points on natural ecosystems that have been extensively altered in terms of agriculture, deforestation, and livestock production. In some cases these clusters still have high values for forest quality and known host diversity (for instance, cluster 5 – central China, and cluster 17 – Java, Indonesia). Areas where outstanding values of different risk factors converge can pose a severe risk to disease emergence and biological conservation. In Sichuan, China – cluster 16 – values of livestock production are extremely high and largely extensive farming takes place concomitantly with the presence of hotspots for mammal diversity (including higher values for known bat hosts) and very high deforestation risk. Unfortunately, deforestation rates and increasing demand for livestock are evident in our top-rated clusters [4], within biodiversity-rich areas, with high forest loss risk and a very large human population (in the case of Beijing - cluster 19 and Java - cluster 17).

We assume that intermediate areas in proximity to hotspots, and where socio-ecological transitions such as those related to the livestock revolution, are at the greatest risk of transitioning to hotspots [4]. Even without complete transition to hotspots, clusters with mostly intermediate values for stressors have had zoonotic spillovers in the past [15,30,31], notably those in central China on cluster 2 and edges with cluster 7 (north Lao PDR, north Vietnam, south China). Further, there is overlap of several identified clusters with areas that concentrate hosts of other viruses with pandemic potential, such as Nipah virus [34]. The intermediate and high-risk areas within clusters need a multidimensional approach to mitigation that combines targeted surveillance of human populations, other animals and the environment with One Health approaches. These approaches emphasise nature-based mitigation strategies, looking at the socio-economic drivers that shape local landscape conditions. Our analyses also show that risk factor clusters are commonly multinational, and action plans are a complex task to implement. However, transboundary, coordinated action between nations that share territorial limits is paramount if configuration of hotspots is taken into account when managing, protecting and restoring land to mitigate disease emergence risk.

Remote areas that present little spatial overlap in risk factor hotpots (blue, Figure 4) may represent conditionally safer areas. In those areas, priority should be assessing and reducing other

disaster and disease risks. In areas of high potential assessed risk (khaki, orange and red, Figure 4), actions should be focused on the drivers of spillover. Recent literature suggests three broad, cost-effective actions to minimize pandemic risk: better surveillance of pathogen spillover, better management of wildlife trade, and substantial reduction of deforestation (i.e. primary prevention) [33]. Landscape planning should have priority, as these can have other benefits [35,36] and can include preventive measures to reduce levels of contact between people and potential wild and domestic animal hosts. Biosecurity measures and surveillance and integrated wildlife monitoring [37] are also key where multi-component risk levels are higher [38]. Syndromic, virological, serological, and behavioral risk surveillance of people with regular proximity with known reservoir or potential amplifier hosts [38] can be of great value in these hotspots, but the ultimate prevention should be in primary prevention. Beyond viral monitoring and discovery, prevention can be achieved by reducing deforestation, intensive concentrated livestock production, wildlife trade and increasing sustainable management of agricultural areas [33].

Surveillance effort correlates with detecting infections and where human populations intersect with wildlife, risk increases [39,40]. In addition, evidence from Brazil suggests zoonotic risk increases with remoteness (along with increased wild mammal species richness) and decreases in areas with greater native forest cover [41]. Our results suggest high-risk areas are often (11-20%) associated with faster travel times to healthcare, compared to remote areas (<1%) (yellow and red respectively, Figure 4). The problem posed by remote sites for emergence mitigation is that while spillover probability and initial ease of spread may be lower, so too is detection probability [39] because of the distance to healthcare. This may allow localized, remote outbreaks to establish and spread in human populations before detection [42–44]. Our findings can be helpful in allocating efforts for surveillance, sustainability and conservation actions and long term plans for ecological intervention, including in areas with high emergent risk scores. Importantly, additional layers of prioritisation could be added to implement mitigation actions on hotspots, for instance, where climate change vulnerability is also high, such as in Java [45]. Also, regions of China, in terms of mobility are outstandingly connected, which highlights the need to reduce pressures arising from multiple hotspots.

Scenario 2 (indirect transmission through livestock) had the highest number of regions with high-risk areas close to healthcare (yellow, Figure 4). These areas are extensive across the study region in all scenarios, and should be prioritised for temporal screening for viruses in livestock, the understanding of known hosts, and investments in improving public health responses to spread. High-risk areas far from healthcare (red) represent small regions of our study area (<1%) in all scenarios, where Scenario 1 had the fewest and Scenario 3 had the highest areas. These are areas with higher possibilities for spillover, that would also be likely to go undetected during the early stages of human-to-human transmission and spread. In those regions, urgent action to prevent contact, reduce deforestation, and enhance biodiversity protection should take place, as well as improvements in

healthcare access. Human populations that are more vulnerable to risks could be targets for equitable distribution of promising solutions, such as pan-coronavirus vaccines [46].

Our findings are a snapshot of macroscale spatial trends that can be used for prioritising more detailed analysis depending on the context and policy priorities. The United Nations Development Programme (UNDP) recommends the creation of 'Maps of Hope' for maintaining essential life support areas [47], but the relationship between biodiversity loss, fragmentation, and zoonotic disease is seldom considered in the designation of such areas. We advocate for a One Health approach in which the risks of pathogen emergence are explicitly integrated into initiatives addressing habitat management, restoration and protection [47], and have demonstrated that this risk can be mapped at large scales with insights into variability in the distribution of key drivers [48].

We acknowledge the complexity of pathogen responses to land use modification [49], and important limitations of our findings. The datasets used here are all static yet global and accessible. However, hotspots may change in response to changes in economic and agricultural policies at national and subnational levels, international agreements such as Agenda 2030, and climate change adaptation [50]. There are also several empirical data limitations. Cryptic diversity in bats [51] and uneven sampling occur for sarbecoviruses and their bat hosts [5] create uncertainty regarding host-pathogen interactions that is difficult to account for. Ecological analyses at finer spatial and temporal scales than used here can elucidate cascading events that result in zoonotic spillover. For example, Hendra virus spillover from bats to horses in Australia seems to be driven by interactions between climatic change altering the flowering phenology of important nectar sources, exacerbating food shortages resulting from native habitat loss and degradation, and nutritional stress in bats that may increase Hendra virus shedding. Native resource declines have concurrently promoted urbanization of many bat populations, increasing the human-bat interface and potential for spillover events to horses, which can act as intermediary hosts, or even potentially direct to humans [52]. Our analyses may capture the macroscale processes, but not these local events.

Similarly, while knowing that the top-priority traded mammals [53] are correlated with total mammalian diversity, local analyses should evaluate how factors that cannot be easily mapped or tracked deviate from the large scale trend, such as animal trade and hunting, which is currently not feasible using a macroscale approach. Our workflow can, however, be easily coupled with detailed local data for spillover 'barriers' and host characteristics to bring insights and customize action plans, such as data on reservoir density, pathogen prevalence, pathogen shedding, and data on spillover recipients, such as susceptibility and infection [54]. This is especially important when macroscale and subnational level risk assessments are neither complete or validated for most nations (accessed in September 2022, https://drmkc.jrc.ec.europa.eu/inform-index/INFORM-Subnational-Risk).

To date, the role of domestic intermediate hosts for sarbecoviruses is unclear, with numerous species able to be infected by SARS-CoV-2 [55]. Here we include cattle and all Bovidae livestock evaluations, leading to similar overall results for clusters but with some univariate hotspots less intense, especially in central India and south China, while making them more intense around Beijing, highlighting how uncertainties around host susceptibility and potential pathways leads to uncertainty regarding risk. The emergence of a novel coronavirus and re-emergence of a known *Sarbecovirus* through spillback is also possible [55] and may change risk profiles. Other factors that play a large role in outbreak response such as conflict [56] and other societal challenges associated with health and the environment might also be considered in local contexts.

The use of remote sensing layers can bring insights for land use planning when considering complex processes such as disease emergence. This process may benefit not only the understanding of risks but also local actions informed by broad patterns [6]. Recent models suggest that the implementation of smaller-scale land-use planning strategies guided by macro-scale patterns may help to reduce the overall burden from emerging infectious diseases [57], while also taking into account biodiversity conservation. This could be evaluated from multiple perspectives, including in the context of other planetary boundaries and how zoonotic disease risk inserts within it [58], considering we have already passed the 1-degree warmer planet threshold [59].

In conclusion, we found evidence that hotspots of multiple conditions that contribute to risk of emerging *Sarbecovirus* are commonly transboundary, but areas scoring the highest values of risk occurred in China and Indonesia. We also identify, for each transmission scenario, that most high risk areas are not far from healthcare, but in all scenarios there are several that are far away from healthcare, especially Scenario 3 that includes wild mammals as secondary hosts. This work contributes to strengthening evidence of spatial clusters of multiple risk factors for disease emergence. We use a reproducible workflow based on hotspot analysis from broad-scale data that is accessible through open software and maps for easy interpretation. This can enable agencies to discuss and engage in new land-use planning actions by including stakeholders (academia, government, local communities and non-governmental organisations) under a One Health perspective. The need to reduce inequalities in the access to healthcare[60] without promoting encroachment into natural areas is a challenge. Efforts should focus on comprehensive land use planning, including the placement of healthcare facilities and other infrastructure [61]. Biodiversity provides essential ecosystem services, so primary prevention of spillover can benefit sustainability at multiple scales, sustaining life on earth and human health [54]. Our findings can help stakeholders when evaluating multiscale policies, land use planning and considering integrating community health programmes to universal healthcare implementation [62] into transboundary, national or subnational levels.

Materials and Methods

We use South, East and Southeast Asia (including western New Guinea) as our study region, where most *Sarbecovirus* hosts are concentrated [5,14] and where many unknown sarbecoviruses are estimated to exist [26]. We define our study region as the terrestrial area of the following countries: Bangladesh, Bhutan, Brunei, Cambodia, China, India, Indonesia, Lao PDR, Malaysia, Myanmar, Nepal, Philippines, Singapore, Sri Lanka, Thailand, Timor-East, and Vietnam.

Characterization of univariate risk indicator hotspots

We identified spatial clusters of components of risk. We assume our inferred risk arises not from individual factors having outstanding high values (hotspots), but instead it arises when they are combined, facilitating conditions for viral spillover. In that sense our inference of risk is an emergent property of the system (Emergent risk). We adapted a broad-scale risk estimation framework (https://mcr2030.undrr.org/quick-risk-estimation-tool) focusing on the potential for sarbecoviruses to emerge. The broad risk factors were five landscape-level conditions and five biological layers, according to four scenarios (see data sources in Table S2). The analysis is naive about the influence of individual drivers on the risk of spillover in the sense that all factors were weighted equally in our scenario evaluations. We selected the following factors for land use change and landscape conditions: Intensity of 1) built-up land, 2) mining and energy, 3) agricultural and harvest land, 4) forest quality, and 5) local forest loss risk. As a measure of human or animal exposure, we used wildlife (known bat hosts and all other wild mammals), livestock (pigs and cattle), and human populations. We use the average of known sarbecovirus bat hosts as our primary host layer [5,14,63,64]. In all scenarios we also include human population counts because this increases the risk of contact and therefore transmission [65]. We included built-up area, energy and mining and agriculture and crop harvest landscape changes in all scenarios because these have been linked to increased coronavirus shedding in bats putatively due to stress and human contact rates [26,66]. We similarly used forest quality and risk of cover loss because emerging infectious disease risk is elevated in forested tropical regions experiencing land-use changes and where wildlife biodiversity (mammal species richness) is high [4,7,67,68]. We then used different potential secondary (aka intermediate) hosts to understand different interactions and transmission pathways, where we included pigs (Scenario 2, Scenario 4) because while SARS-CoV-2 transmission is unlikely [69], coronaviruses cause a variety of diseases in pigs [70], including infections related to bat coronavirus [71]. We also included cattle (Scenario 2, Scenario 4), as coronaviruses caused infection [72] and disease in cattle [73], and coinfection can lead to recombination events [74]. We have also included other bovid livestock (Scenario 2, Scenario 4), because there are several regions in Southeast Asia where carabao (*Bubalus bubalis*) and other Bovidae livestock are more common than cattle (*Bos taurus*) so we provide results for Bovidae livestock instead of cattle-only in the Supplementary materials. We also included wild mammals as potential secondary (aka intermediate) hosts, minus the

known bat hosts (Scenario 3, Scenario 4) because EID risk is elevated in forested tropical regions experiencing land-use changes and where wildlife biodiversity (mammal species richness) is high. SARS-CoV-2 has also been detected in wildlife (spillback events) [55,67].

To avoid collinearity, we only selected variables with product-moment correlation coefficient (r) values < ±0.7 (Figure S1). The study region was divided into a spatial grid composed of 0.25 decimal degrees-sized tiles (~27 km). All variables were resampled to match this resolution. For data layers that were counts from shapefiles (other mammal species numbers), we applied median values for resampling. We ran a univariate hotspot analysis based on Getis-ord G*i values considering each factor individually at 95% alpha error cut-off. We created a list of closest neighbors considering all data and n=25 for the closest neighborhood. Local G assumes a two-sided alternative hypothesis, where high-positive values indicate hotspot regions and low negative values indicate coldspots. Pixels located in-between the alternative hotspot or coldspot hypothesis values are referred as intermediate regions, where the value may reflect random spatial process, i.e. no spatial clustering detected. Critical values for defining univariate hotspots followed the critical values for 95th percentile [75].

Scenarios

Detected hotspots for all landscape condition components were used in combination with biological components in the scenario analyses. Scenario 1 considers direct transmission from bats to humans, where the biological risk is composed of the average number of bat species in which sarbecoviruses have been reported as the known primary hosts. For Scenario 2, we then considered the components of scenario one in combination with potential intermediate hosts using: pig counts, cattle-only or Bovidae livestock counts. Scenario 3 considered bat hosts and the number of other wild mammal species present. For Scenario 3, we used the wild mammals layer (minus known bat hosts) and known bat hosts as the potential intermediate hosts. We considered using a traded mammal layer instead of an all wild mammal layer in Scenario 3, because of evidence the first COVID-19 cases identified were linked to the Huanan Seafood Wholesale Market in Wuhan [16]. An available high priority traded mammal layer [53], however, is highly correlated (r = 0.864) with the wild mammal layer. Because of this correlation in addition to high uncertainty regarding trade, we kept only the mammal layer and bat hosts layer in Scenario 3. A fourth scenario including all of the previous mammalian layers, be it wild or livestock, was constructed.

Hotspot convergence in clusters

We evaluated the spatial clustering among hotspots including all ten selected drivers (Scenario 4, five landscape descriptors, five potential host components). We opted for doing a single cluster analysis because we cannot weigh the importance of the single variables for influencing an ultimate spillover

event. The variables comprised here describe landscape condition, human population, cattle, pig, bat hosts and all other wild mammals. We assume areas that contain most hotspots or that are on the verge of becoming hotpots (intermediate areas) for the components evaluated are at higher risk of emerging new sarbecoviruses. A multivariate spatial cluster analysis was applied to Getis-ord G*i values for every variable after the univariate hotspot analysis using rgeoda 0.0.9 [76]. We used the multivariate skater (Spatial `K'luster Analysis by Tree Edge Removal) hierarchical partitioning algorithm [77] to infer contiguous clusters of similar values in the region based on the optimal pruning of a minimum spanning tree. Spatial clusters represent emergent, cohesive risk combinations distributed in space. Contiguity was assessed by a queen weights matrix after transforming pixels to geographical coordinates. Distance functions were set to euclidean. We evaluated the k number of clusters from 1 to 40. To find the optimal number of clusters, we evaluated the total within-cluster sum of squares variation, visually inspecting the point of inflection in the curve towards stabilization. As the reduction in increment was very smooth, we present the number of clusters for skater informed by the max-p algorithm. We used max-p to find the solution for the optimal number of spatially-defined clusters setting as a bounding variable (a variable that allows for a minimum value summed for each cluster) the human population amounts at 5% and 10%. The algorithm was computed at 99 interactions with 123456789 as a random seed.

To interpret variation of hotspots between clusters, we counted the number of variables for which the median of the distribution is a hotspot (i.e. falling within the hotspot interval at 95% Gi*). We then discuss the clusters based on the number of drivers that are already hotspots and the median values that fall in intermediate zones, so closer to becoming hotspots, and coldspots for each cluster. Finally, to understand the overall variation (and among clusters) we provide a principal component analysis (PCA) biplot through Scenario 4 to discuss major axes of variation between optimal number of clusters. We ran the hotspot analyses with cattle-only and with the summed values for Bovidae livestock (presented in the Supplemental Material). All geographical coordinates were warped to World Mercator (EPSG: 3395) and World Geodetic System 1984 datum before spatial analysis.

Emergent risk and its relationship with access to healthcare

After identifying the hotspots within the scenarios, we match their proximity to detection by matching the emergent risk score (i.e. number of hotspots) for every pixel with the level of motorized access to healthcare (hospitals and clinics). Access to healthcare measured as travel time was considered as both a proxy for connectivity and an indicator of the likelihood of detection, following infection spillover and spread. We built bivariate maps and three-by-three quantile (N=9) combinations considering the intensity of hotspots from their overlay and the values for access to healthcare, all rescaled from zero to one. We compared travel time in high-risk areas among scenarios using

Wilcoxon's test. All analyses were done in QGIS 3.10.7 [78], R 4.1.3 [79] and bash [80]. Code for the analyses can be found at https://github.com/renatamuy/hotspots/**.**

References and notes

1. Antia, R., Regoes, R. R., Koella, J. C. & Bergstrom, C. T. The role of evolution in the emergence of infectious diseases. *Nature* **426**, 658–661 (2003).
2. Lloyd-Smith, J. O. *et al.* Epidemic dynamics at the human-animal interface. *Science* **326**, 1362–1367 (2009).
3. Wolfe, N. D., Dunavan, C. P. & Diamond, J. Origins of major human infectious diseases. *Nature* **447**, 279–283 (2007).
4. Rulli, M. C., D'Odorico, P., Galli, N. & Hayman, D. T. S. Land-use change and the livestock revolution increase the risk of zoonotic coronavirus transmission from rhinolophid bats. *Nature Food* **2**, 409–416 (2021).
5. Muylaert, R. L. *et al.* Present and future distribution of bat hosts of sarbecoviruses: implications for conservation and public health. *Proc. Biol. Sci.* **289**, 20220397 (2022).
6. Wilkinson, D. A., Marshall, J. C., French, N. P. & Hayman, D. T. S. Habitat fragmentation, biodiversity loss and the risk of novel infectious disease emergence. *J. R. Soc. Interface* (2018) doi:10.1098/rsif.2018.0403.
7. Plowright, R. K. *et al.* Land use-induced spillover: a call to action to safeguard environmental, animal, and human health. *Lancet Planet Health* **5**, e237–e245 (2021).
8. Snedden, C. E. *et al.* SARS-CoV-2: Cross-scale Insights from Ecology and Evolution. *Trends Microbiol.* **29**, 593–605 (2021).
9. Düx, A. *et al.* Measles virus and rinderpest virus divergence dated to the sixth century BCE. *Science* **368**, 1367–1370 (2020).
10. Hansen, M. C. *et al.* Global land use extent and dispersion within natural land cover using Landsat data. *Environ. Res. Lett.* **17**, 034050 (2022).
11. Gibb, R. *et al.* Zoonotic host diversity increases in human-dominated ecosystems. *Nature* **584**, 398–402 (2020).

12. Lytras, S. *et al.* Exploring the Natural Origins of SARS-CoV-2 in the Light of Recombination. *Genome Biol. Evol.* **14**, (2022).

13. Latinne, A. *et al.* Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* **11**, 4235 (2020).

14. Sánchez, C. A. *et al.* A strategy to assess spillover risk of bat SARS-related coronaviruses in Southeast Asia. *Nat. Commun.* **13**, 4380 (2022).

15. World Health Organization. WHO-convened global study of origins of SARS-CoV-2: China part. *World Health Organization* (2021).

16. Worobey, M. *et al.* The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science* **377**, 951–959 (2022).

17. Joffrin, L. *et al.* Seasonality of coronavirus shedding in tropical bats. *R Soc Open Sci* **9**, 211600 (2022).

18. Montecino-Latorre, D. *et al.* Seasonal shedding of coronavirus by straw-colored fruit bats at urban roosts in Africa. *PLoS One* **17**, e0274490 (2022).

19. Eby, P. *et al.* Pathogen spillover driven by rapid changes in bat ecology. *Nature* (2022) doi:10.1038/s41586-022-05506-2.

20. Wacharapluesadee, S. *et al.* Longitudinal study of age-specific pattern of coronavirus infection in Lyle's flying fox (Pteropus lylei) in Thailand. *Virol. J.* **15**, 38 (2018).

21. Seltmann, A. *et al.* Seasonal Fluctuations of Astrovirus, But Not Coronavirus Shedding in Bats Inhabiting Human-Modified Tropical Forests. *Ecohealth* **14**, 272–284 (2017).

22. Ruiz-Aravena, M. *et al.* Ecology, evolution and spillover of coronaviruses from bats. *Nat. Rev. Microbiol.* **20**, 299–314 (2022).

23. Djalante, R., Shaw, R. & DeWit, A. Building resilience against biological hazards and pandemics: COVID-19 and its implications for the Sendai Framework. *Prog Disaster Sci* **6**, 100080 (2020).

24. Pearson, L. & Pelling, M. The UN Sendai Framework for Disaster Risk Reduction 2015–2030: Negotiation Process and Prospects for Science and Practice. *J. of Extr. Even.* **02**, 1571001 (2015).

25. Pretorius, M., Markotter, W. & Keith, M. Assessing the extent of land-use change around important bat-inhabited caves. *BMC Zoology* **6**, 1–12 (2021).
26. Anthony, S. J. *et al.* Global patterns in coronavirus diversity. *Virus Evol* **3**, vex012 (2017).
27. Heesterbeek, H. *et al.* Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**, aaa4339 (2015).
28. Grange, Z. L. *et al.* Ranking the risk of animal-to-human spillover for newly discovered viruses. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
29. Guo, H. *et al.* ACE2-Independent Bat Sarbecovirus Entry and Replication in Human and Bat Cells. *MBio* e0256622 (2022).
30. Wang, N. *et al.* Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virol. Sin.* **33**, 104–107 (2018).
31. Li, H. *et al.* Human-animal interactions and bat coronavirus spillover potential among rural residents in Southern China. *Biosaf Health* **1**, 84–90 (2019).
32. Carlson, C. J. *et al.* Climate change increases cross-species viral transmission risk. *Nature* **607**, 555–562 (2022).
33. Bernstein, A. S. *et al.* The costs and benefits of primary prevention of zoonotic pandemics. *Sci Adv* **8**, eabl4183 (2022).
34. Albery, G. F. *et al.* The science of the host-virus network. *Nat Microbiol* **6**, 1483–1492 (2021).
35. Lu, Y. *et al.* Spatial variation in biodiversity loss across China under multiple environmental stressors. *Sci Adv* **6**, (2020).
36. Jaureguiberry, P. *et al.* The direct drivers of recent global anthropogenic biodiversity loss. *Sci Adv* **8**, eabm9982 (2022).
37. Barroso, P. *et al.* Using integrated wildlife monitoring to prevent future pandemics through one health approach. *One Health* **16**, 100479 (2023).
38. Keusch, G. T. *et al.* Pandemic origins and a One Health approach to preparedness and prevention: Solutions based on SARS-CoV-2 and other RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2202871119 (2022).
39. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).

40. Guernier, V., Hochberg, M. E. & Guégan, J.-F. Ecology Drives the Worldwide Distribution of Human Diseases. *PLoS Biol.* **2**, e141 (2004).

41. Winck, G. R. *et al.* Socioecological vulnerability and the risk of zoonotic disease emergence in Brazil. *Sci Adv* **8**, eabo5774 (2022).

42. Giles, J. R. *et al.* Trip duration drives shift in travel network structure with implications for the predictability of spatial disease spread. *PLoS Comput. Biol.* **17**, e1009127 (2021).

43. Riley, S. *et al.* Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* **300**, 1961–1966 (2003).

44. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).

45. Green, R. *et al.* Growing health: global linkages between patterns of food supply, sustainability, and vulnerability to climate change. *The Lancet Planetary Health* **6**, e901–e908 (2022).

46. Dolgin, E. Pan-coronavirus vaccine pipeline takes form. *Nat. Rev. Drug Discov.* **21**, 324–326 (2022).

47. Maps of Hope – UN Biodiversity Lab. https://unbiodiversitylab.org/maps-of-hope/.

48. One Health High-Level Expert Panel (OHHLEP) *et al.* One Health: A new definition for a sustainable and healthy future. *PLoS Pathog.* **18**, e1010537 (2022).

49. Mader, A. D. *et al.* Messaging Should Reflect the Nuanced Relationship between Land Change and Zoonotic Disease Risk. *Bioscience* biac075 (2022).

50. Transforming our world: The 2030 agenda for sustainable development. in *A New Era in Global Health* (Springer Publishing Company, 2018).

51. Chornelia, A., Lu, J. & Hughes, A. C. How to Accurately Delineate Morphologically Conserved Taxa and Diagnose Their Phenotypic Disparities: Species Delimitation in Cryptic Rhinolophidae (Chiroptera). *Front. Ecol. Evol.* **0**, (2022).

52. Becker, D. J., Washburne, A. D., Faust, C. L., Mordecai, E. A. & Plowright, R. K. The problem of scale in the prediction and management of pathogen spillover. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190224 (2019).

53. Cronin, M. R., de Wit, L. A. & Martínez-Estévez, L. Aligning conservation and public health

goals to tackle unsustainable trade of mammals. *Conservation Science and Practice* **n/a**, e12818 (2022).

54. Sokolow, S. H. *et al.* Ecological interventions to prevent and manage zoonotic pathogen spillover. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 1–10 (2019).
55. Nerpel, A. *et al.* SARS-ANI: a global open access dataset of reported SARS-CoV-2 events in animals. *Scientific Data* **9**, 1–13 (2022).
56. Gayer, M., Legros, D., Formenty, P. & Connolly, M. A. Conflict and emerging infectious diseases. *Emerg. Infect. Dis.* **13**, 1625–1631 (2007).
57. Baker, R. E. *et al.* Infectious disease in an era of global change. *Nat. Rev. Microbiol.* **20**, 193–205 (2022).
58. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
59. Armstrong McKay, D. I. *et al.* Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science* **377**, eabn7950 (2022).
60. Tao, Z. & Wang, Q. Facility or Transport Inequality? Decomposing Healthcare Accessibility Inequality in Shenzhen, China. *Int. J. Environ. Res. Public Health* **19**, (2022).
61. Toh, K. B. *et al.* Guiding placement of health facilities using multiple malaria criteria and an interactive tool. *Malar. J.* **20**, 455 (2021).
62. Garchitorena, A. *et al.* Geographic barriers to achieving universal health coverage: evidence from rural Madagascar. *Health Policy Plan.* **36**, 1659–1670 (2021).
63. Forero, N. *et al.* The coevolutionary mosaic of bat betacoronavirus emergence risk. *EcoEvoRxiv* (2022) doi:10.32942/osf.io/8mgv6.
64. Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
65. Zhang, X. *et al.* The effect of population size for pathogen transmission on prediction of COVID-19 spread. *Sci. Rep.* **11**, 18024 (2021).
66. Warmuth, V. M., Metzler, D. & Zamora-Gutierrez, V. Human disturbance increases coronavirus prevalence in bats. *Sci Adv* **9**, eadd0688 (2023).

67. Allen, T. *et al.* Global hotspots and correlates of emerging zoonotic diseases. *Nat. Commun.* **8**, 1–10 (2017).

68. Shah, H. A., Huxley, P., Elmes, J. & Murray, K. A. Agricultural land-uses consistently exacerbate infectious disease risks in Southeast Asia. *Nat. Commun.* **10**, 1–13 (2019).

69. Sikkema, R. S. *et al.* Experimental and field investigations of exposure, replication and transmission of SARS-CoV-2 in pigs in the Netherlands. *Emerg. Microbes Infect.* (2021) doi:10.1080/22221751.2021.2011625.

70. Thakor, J. C. *et al.* Swine coronaviruses (SCoVs) and their emerging threats to swine population, inter-species transmission, exploring the susceptibility of pigs for SARS-CoV-2 and zoonotic concerns. *Vet. Q.* **42**, (2022).

71. Zhou, P. *et al.* Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* **556**, 255–258 (2018).

72. Wernike, K. *et al.* Serological screening suggests single SARS-CoV-2 spillover events to cattle. *bioRxiv* 2022.01.17.476608 (2022) doi:10.1101/2022.01.17.476608.

73. Hodnik, J. J., Ježek, J. & Starič, J. Coronaviruses in cattle. *Trop. Anim. Health Prod.* **52**, 2809–2816 (2020).

74. Ulrich, L., Wernike, K., Hoffmann, D., Mettenleiter, T. C. & Beer, M. Experimental Infection of Cattle with SARS-CoV-2. *Emerg. Infect. Dis.* **26**, 2979–2981 (2020).

75. Ord, J. K. & Getis, A. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* **27**, 286–306 (2010).

76. Li, X. & Anselin, L. rgeoda: R Library for Spatial Data Analysis. Preprint at (2022).

77. AssunÇão, R. M., Neves, M. C., Câmara, G. & Da Costa Freitas, C. Efficient regionalization techniques for socio□economic geographical units using minimum spanning trees. *Int. J. Geogr. Inf. Sci.* **20**, 797–811 (2006).

78. QGIS Development Team. *QGIS Geographic Information System.* (Open Source Geospatial Foundation, 2009).

79. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.R-project.org/ (2020).

80. Gnu, P. Free Software Foundation. *Bash (3. 2. 48)[Unix shell program]*.

Acknowledgments

We thank Dr Jinhong Luo and Marco A. R. Mello for comments on earlier versions of the draft. Massey University's subscription to New Zealand eScience Infrastructure (NeSi) enabled us to use high-performance computing facilities. Te Pūnaha Matatini gave us support during the 2022 Mahia Te Mahi workshop.

Funding

RLM, DTSH, RSJ: Bryce Carmine and Anne Carmine (née Percival), through the Massey University Foundation; DTSH: Royal Society Te Apārangi, grant number MAU1701.