A generic framework for spatial quantitative risk assessments of infectious diseases: lumpy skin disease case study

Rachel A. Taylor[1], Alexander DC Berriman[1], Paul Gale[1], Louise A. Kelly[1,2], Emma L. Snary[1]

[1]Animal and Plant Health Agency (APHA), Weybridge, UK

[2]Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK

Corresponding author email: Rachel.taylor@apha.gsi.gov.uk

Running Title: Spatial risk assessment framework for infectious diseases

Keywords: Communicable Diseases, Lumpy Skin Disease, Risk Assessment, Spatial Analysis, Stochastic Processes

Summary

The increase in availability of spatial data and the technological advances to handle such data allow for subsequent improvements in our ability to assess risk in a spatial setting. We provide a generic framework for quantitative risk assessments of disease introduction that capitalises on these new data. It can be adopted across multiple spatial scales, for any pathogen, method of transmission or location. The framework incorporates the risk of initial infection in a previously uninfected location due to registered movement (e.g. trade) and unregistered movement (e.g. daily movements of wild animals). We discuss the steps of the framework and the data required to compute it. We then outline how this framework is applied for a single pathway using lumpy skin disease as a case study, a disease which had an outbreak in the Balkans in 2016. We calculate the risk of initial infection for the rest of Europe in 2016 due to trade. We perform the risk assessment on 3 spatial scales – countries, regions within countries, and individual farms. We find that Croatia (assuming no vaccination occurred) has the highest mean probability of infection, with Italy, Hungary and Spain following. Including import detection of infected trade does reduce risk but this reduction is proportionally lower for countries with highest risk. The risk assessment results are consistent across the spatial scales, while in addition, at the finer spatial scales, it highlights specific areas or individual locations of countries on which to focus surveillance.

Introduction

The viability and usefulness of spatial quantitative risk assessment has been increasing with the availability of larger datasets, more detailed data, and improved computational power. This has been manifest across human, animal and plant health, following the growth of data on, for example, human movement via air travel (Grais et al. 2004, Tatem et al. 2006a, Tatem et al. 2006b), tracking and communication of spread of diseases using social media tools (Schmidt 2012, Bengtsson et al. 2015), improved climate predictions (Brownstein et al. 2005, Gale et al. 2012) as well as the development of geospatial software (Wardrop et al. 2012) and next generation sequencing of disease strains (Vayssier-Taussat et al. 2013). Assessing risk on a spatial scale allows for active surveillance to be directed to areas deemed most at risk, spatially-varying procedures prepared to prevent infections occurring, and different management plans depending on which spatial locations become infected, should infection occur. Targeting prevention and control spatially, in this manner, can save time, money and resources.

Whilst there are clear outlines and recommendations on how to perform risk assessment for initial infections and spread of infectious diseases generally, which include quantifying entry, exposure and consequence assessments (Murray 2004), there is, however, no generic framework for performing quantitative risk assessments in a spatial setting. A framework allows for standardisation across different

countries and organisations to facilitate policy and decision making. In this paper, we outline a proposed generic framework for completing spatial quantitative risk assessments for risk of infection. The defining feature of the framework will be its emphasis on disease introductions from one area to another. However, the aim is for the framework to be generic across the type of pathogen, method of transmission, species of host(s) and spatial resolutions. To outline what the framework aims to achieve, we formulate the risk question as: *"What is the risk of infection of a pathogen in Area B given the presence of that pathogen in Area A?"* It thus focuses on initial infection by means of introduction.

We introduce the generic framework and undertake the risk assessment for one pathway, namely the risk of initial infection by means of registered movement of hosts, using a case study. In the case of livestock this implies a focus on trade but human movement via airplanes or other ticketed travel between countries is also applicable. Calculation of the risk of infection due to unregistered movement of hosts across borders, such as vectors, wild birds or illegal trade, is not considered in our case study. Through this case study we highlight the degrees of detail possible in the risk assessments using our framework when the risk assessor has access to data at a country level only, country level data with some regional data, and lastly detailed individual farm level data.

Lumpy skin disease (LSD) is used as the case study for assessing the risk of infection by legal trade. LSD virus, which affects cattle and buffalo, is in the Capripoxvirus genus along with sheeppox and goatpox. It causes nodules on the skin, mucus membranes and internal organs; reduction in milk production; fever; oedema; and sometimes death (Davies 1991, Tuppurainen and Oura 2012). Mortality is usually low (<10% Kumar (2011)), but it can cause significant economic losses and hence the World Organisation for Animal Health (OIE) classifies LSD as a notifiable disease. Mechanical transmission by vectors is believed to be the primary method of transmission but direct contact, infected semen, and contaminated feed and water sources are also considered as rare but possible routes of transmission (Carn and Kitching 1995b, Magori-Cohen et al. 2012). There are many different species that have been implicated as mechanical vectors, such as biting flies (e.g. *Stomoxys calcitrans,* Yeruham et al. (1995)), mosquitoes (e.g. *Aedes aegypti,* Chihota et al. (2001)) and ixodid (hard) ticks with evidence for transstadial and transovarial transmission in these ticks (e.g. *Rhipicephalus spp,* Tuppurainen et al. (2011)).

There has been rapid geographical spread of LSD over the last 30 years with emergence in south-east Europe for the first time in 2015. Historically, it had been restricted to sub-Saharan Africa and appeared to be in decline, but a resurgence occurred in the 1980s and subsequently it has been steadily spreading northwards (Hunter and Wallace 2001). Although there had been infrequent incursions before, since 2006 it has become endemic throughout the Middle East (Tuppurainen and Oura 2012). Similarly, it has been present in Turkey since 2013 and is now considered endemic. A few cases of LSD occurred in Greece for the first time during 2015, followed by a widespread outbreak in the Balkan regions in 2016, specifically in

Greece, Bulgaria, The Former Yugoslav Republic of Macedonia, Serbia, Kosovo and Albania, as well as in Russia (Mercier et al. 2017). These cases and those in 2015 predominantly occurred in the summer months, highlighting the seasonal nature of the disease spread, likely shaped by the environmental requirements of arthropod vectors and pastoralism (Thevenin 2011). Given the potential for further spread within the EU, the disease was considered to be a timely case study for the application of our generic framework. Estimation of the risk of initial infection within each EU country would provide information to aid targeted surveillance. In addition, as the disease is notifiable, we anticipated that there would be more data than would be available for non-notifiable diseases.

We demonstrate our generic framework for spatial quantitative risk assessment for risk of initial infection, specifically due to registered movement of animals or people, by considering the potential import of infected animals in 2016 during the outbreak of LSD in the Balkan region. This is outlined by the following specific risk question: *"What is the risk of initial infection of lumpy skin disease in Europe (not including the countries which had notified cases) from legal trade in 2016 due to the presence of lumpy skin disease in the rest of the World?"*

Methods

The Generic Framework

We define the risk of infection as the probability of one or more initial infections in the native susceptible population in Area B. The risk pathway outlining the probabilistic steps involved in risk of infection for the generic risk question is set out in Figure **1**. Infection can only occur if there is incursion of infected species, non-detection of that species, the survival of that species, and subsequent exposure of native susceptible hosts resulting in transmission, as shown in Figure **1**. We use the term "species" but it could be even more generic than this, such as infected products or feed, provided that they could be in contact with native susceptible hosts. The term "contemporaneous survival" in Figure **1** indicates that some species are only active part of the year (e.g. only summer months in Northern EU countries in the case of vector-borne diseases) and the two species need to coincide for infection to be possible. It also includes the time period over which the infected species is infectious. Thus, animals that are imported directly for slaughter are assumed to have a survival length of 0 days. We combine these steps in the pathway together to produce the probability of one or more infections occurring in Area B for each pathway. Different methods of transmission, different infected species entering Area B or even different pathogen strains require different risk pathways that are then combined together to create the total risk of infection for each location within Area B. Locations at highest overall risk are of most interest to policy makers.

Risk of Infection

We can mathematically describe and combine the 5 steps of the risk pathway (Figure **1**) and can therefore compute a quantitative probability of initial infection at each location thus giving a spatial risk map. This is adapted from a model that assesses the risk of species jumps in avian influenza (Hill et al. 2015). At first, we describe a single disease pathway from i to j. Inclusion of other pathways resulting in initial infection in Area B from Area A, in which species i and j may be different, is outlined afterwards. As the risk assessment may be calculated on different spatial scales, we use k to denote subregions of Area A and g for locations in Area B. Both of these will be determined by the spatial data available. Step 1 of the risk of initial infection is the estimation of the number of infected hosts (I_k) entering Area B. Based on the prevalence in Area A_k and the total number of hosts exported, the number of infected hosts (I_k) entering location g of Area B during a set time interval is given by:

$$I_k(g) \sim Bin(N_k(g), p_k)$$

Here $N_k(g)$ is the number of hosts imported to or entering location g from Area A_k in a unit time interval and p_k is the prevalence of infected hosts in Area A_k. We use a stochastic representation in the number of imported hosts that will be infected to better describe the potential variability. This requires an assumption of independence and therefore an assumption that infected and non-infected hosts are equally likely to be exported. The number of infected hosts entering location g in Area B from Area A is derived by summing over all sub-regions in Area A, thus

$$I(g) = \sum_k I_k(g)$$

However, some infected hosts may not make it through import control due to detection of symptoms or testing of hosts, Step 2 in Figure **1**. The probability of detection and the sensitivity of the tests can also vary by location g. We assume, however, that the probability of detection is independent of which country is the exporter. Therefore, we denote $p_D(g)$ as the probability of successfully detecting and removing an infected host. The actual number of infected hosts $J(g)$ entering location g in Area B will be given by

$$J(g) \sim Bin(I(g), 1 - p_D(g))$$

We next calculate Steps 3 – 5 (Figure **1**), namely the survival rate of the species, the contact rate between hosts and the probability of transmission leading to initial infection of susceptible hosts in each location g of Area B. We combine these components using the basic reproductive number R_0. The basic reproductive number gives the number of susceptible hosts likely to be infected by the introduction of one infected host at each location g. R_0 is a fundamental metric of disease systems, but the equation to represent R_0 depends on how the transmission of the disease is modelled. Therefore, using R_0 facilitates adaptation to

different methods of transmission due to different interactions between species i and j, e.g. vector-borne transmission, direct transmission or sexual transmission, as well as specific aspects that are only applicable in some cases, such as environmental factors, in determining survival of species. For example, direct transmission would be represented by the equation:

$$R_0(g) = \frac{c\beta S(g)}{r}$$

In this equation, c is the contact rate between hosts (Step 4), β is the probability that contact results in successful transmission (Step 5), $S(g)$ is the population size of susceptible hosts of species j in location g and r is the recovery rate, based on the length of time infected hosts of species i remain present and infectious in location g (Step 3). Additionally, R_0 can change depending on the location g to incorporate differences in transmission in different regions. And if the data are available R_0 could be a function of temperature, changing for each location based on average temperatures. For calculating the risk of initial infection occurring within location g we assume that there is homogeneous mixing between the newly introduced species and the native susceptible population, but this could be adapted to other scenarios in which homogeneous mixing is not a good assumption by changing the contact rate c.

Based on our definition of risk of infection, we calculate the probability of one or more infections occurring in the susceptible population within location g, per unit time by combining the information from Steps 1 – 5. The probability of random events, such as infections, happening is described by a Poisson process with parameter λ where λ is the expected number of events occurring per unit time. For each observation of $J(g)$ we can estimate the expected number of new infections occurring in a unit time in location g by $R_0(g)J(g)$. Hence, in our framework the number of new infections per unit time follows a Poisson process with parameter $\lambda = R_0(g)J(g)$. The probability of no events happening in a Poisson process is $e^{-\lambda}$. Hence, the risk of infection, alternatively the probability of one or more infections occurring in the susceptible population, in location g from introduction of infected hosts from Area A is given by:

$$R_I(g) = 1 - e^{-R_0(g)J(g)}$$

If there are multiple routes of transmission (which could involve different species i and j) then all parameters, including the contact rates, number of susceptible animals and number of imported infected hosts, may be different for each route. To incorporate these different routes, we denote the route with an additional subscript ω and, hence, re-write the above equation as

$$R_{I,\omega}(g) = 1 - e^{-R_{0,\omega}(g)J_\omega(g)}$$

for each route ω. Therefore, the complete risk of infection over all routes of transmission would become:

171 $$R_I(g) = 1 - \prod_{\omega}\left[1 - R_{I,\omega}(g)\right]$$

172 This risk of infection calculation is capable of representing different methods of transmission, a wide range
173 of pathogens with different environmental requirements, any spatial scale and any route of introduction.

174 Data Requirements

175 The data requirements necessary to calculate the risk of initial infection are presented in Table **1**. The data
176 needs to be on the spatial scale for which the risk assessment is to be performed as well as a suitable time
177 scale. The major determinant of the time scale will be the movement data which could be on a daily,
178 monthly or yearly scale. Data on the quantity of susceptible hosts is not likely to be censored as often but
179 can be assumed to stay relatively constant. Prevalence data can be on any time scale depending on
180 availability, although if different from the time scale of the movement data then assumptions either have
181 to be made that the prevalence is constant over the duration of multiple movements, or on how to split the
182 movement data up to be on the same time scale as the prevalence data.

183 Lumpy Skin Disease case study

184 We outline the data that we use to compute the risk of infection for our LSD case study (Table **2**). Since we
185 assess the risk due to the potential import of infected animals in 2016, during the outbreak in the Balkan
186 region, we use 2016 data as much as possible. Our Area B is defined as all countries in Europe excluding
187 those countries which had notified cases in 2016, namely Greece, Bulgaria, The Former Yugoslav Republic
188 of Macedonia, Serbia, Kosovo, Albania, Turkey and Russia. Our Area A is the whole world. To highlight the
189 ability for the framework to cope with different spatial resolutions, we use three different scales of
190 locations in Area B. We compute risk of infection for countries, for regions within each country, and at
191 individual farms within Europe. The spatial regions that we use are based on the NUTS classification
192 (Nomenclature of territorial units of statistics) which is a system for dividing up the European Union (EU)
193 into hierarchical levels in order to collect and harmonise European regional statistics. These regional
194 classifications are created and maintained by Eurostat (Eurostat 2017b), the statistical office for the
195 European Union. There are three NUTS regions below a country level, which decrease in size and socio-
196 economic status, labelled as NUTS 1, NUTS 2 and NUTS 3. Some small countries have NUTS 1 and even
197 NUTS 2 defined to be the whole country. Some non-EU European countries also have NUTS regions and
198 provide some regional data to Eurostat, namely Iceland, Norway and Switzerland. Those non-EU European
199 countries which do not have NUTS regions assigned and which do not give regional data to Eurostat are
200 excluded from our regional and postcode risk assessments, specifically, Belarus, Bosnia and Herzegovina,
201 Moldova and Ukraine.

Trade data ($N_k(g)$)

We use two different sources of trade data which represent the most common resolutions of data available. In the first case, we assume the risk assessor has access to individual farm trade data, i.e. the number of animals each farm in Area B imports from Area A. Obtaining individual farm trade data across the whole of Area B is likely to be difficult in many cases, unless Area B is a single country. Nevertheless, using these data we can calculate the risk of infection for individual farms. In the second case, we assume the much more frequent scenario, that the risk assessor has access to country trade data only. This allows for computation of risk of infection at a country level only. However, based on the assumption that it is often possible to access a higher resolution of data for your own country, we outline a method to infer regional trade data for all countries based on regional trade data for one country and country-level trade data for all countries. For specific details on this method see Appendix A. Thus, we compute risk of infection at a regional level based on country trade data for all countries and regional trade data from one country. For all trade data, we assume that cattle are imported to a farm and are not moved elsewhere during the time they are infectious.

For the risk of initial infection at individual farm level, we use the Trade Control and Expert System (TRACES 2017), a primarily EU-based trade system. On request TRACES kindly provided us with data on all cattle trade registered through their system in 2016 and included the postcode of the final destination. The country trade data which we use for the country and regional risk assessments is the freely available dataset COMEXT, also provided by Eurostat (Eurostat 2017a), which denotes all trade of any product from any country in the world to any other. We subset the data by product code so that we only include trade in live cattle not for slaughter. Using the methodology described in Appendix A, we infer the distribution of the imported cattle amongst regions in each country. To do this, we use trade data at the regional level for the UK, from 2012-2015 obtained from TRACES, to estimate predictors for determining the proportion of cattle going to each region. This is under the assumption that UK regional trade data is similar to regional trade data in the rest of Europe. Based on data availability at a regional level, we analysed the number of cattle farms in each region and the proportion of those which are dairy as predictors. Our model in Appendix A identifies both as important predictors for predicting the proportion of cattle trade to each region. Data on cattle farms across Europe, including how many are dairy, are provided by Eurostat (Eurostat 2017a). Thus, we use this COMEXT data, alongside the model for distribution of cattle imports across regions, in our calculations for risk both at the country and regional level.

Since we have the TRACES data at a farm level for the whole of Area B, we can use it to calculate regional risk instead of inferring distribution of cattle amongst regions based on the country trade data. However, as it is much more likely for the risk assessor to only have access to the country level data, we highlight in the main text the use of country level data for computing regional risk. It also highlights the differences that occur between the two datasets. We plot the country risk based on TRACES data in Appendix B and

compare this to country risk using COMEXT data to assess whether these differences impact significantly on the risk calculations.

Prevalence of LSD around the world (p_k)

We use 2016 data from the EU-funded SPARE project (Simons et al. 2017), which estimates prevalence of disease around the world using OIE data on the number of outbreaks and the number of cases per outbreak of the disease in the past 10 years. These data were provided to us in the form of distribution parameters, thus giving a distribution of prevalence of LSD for each country. We use these distributions to represent our uncertainty in the prevalence in one year for each country. The data do not assume that a country is free of the disease if it has not notified the OIE of any cases because it takes into account under reporting and the occurrence of notified cases in bordering countries.

Probability of import detection ($p_D(g)$)

Although our locations g can be either countries, regions or even farms, it is unlikely to know the probability of import detection of infected hosts on the finer scales. Even on a country level it may be difficult to determine, especially for diseases which do not have a specific test on import. Therefore, for this case study we set $p_D(g) = 0$ to indicate no testing of imports occurs and perform a scenario analysis for $p_D(g) = 0.5$ for all countries. Approximately 30-50% of animals with LSD will show severe clinical signs, with more showing mild symptoms, and thus may be detected by physical examination (Weiss 1968, Ali et al. 1990, Carn and Kitching 1995b). A health certificate signed by an official veterinarian is the only health requirement for movement between different EU countries, although individual countries within the EU may have their own regulations. For example, the UK tests on import (and quarantines animals until results are confirmed negative) for LSD from high-risk countries. However, we do not know the procedures of other countries in Europe and therefore we maintain that an import detection of 0.5 is in a realistic range for LSD.

Number of susceptible hosts ($S(g)$)

For the country level assessment of risk, the data is based on cleaned data from the OIE from 2014 which gives the numbers of cattle and farms in each country (World Organisation for Animal Health (OIE) 2017). This is to be consistent with the prevalence data which uses OIE data in its methodology. Since OIE data on number of cattle is rare at the regional level, we use data from Eurostat for the numbers of cattle and farms in each region in 2016 (some countries are for 2015 due to a lack of 2016 data) for the regional risk assessment (Eurostat 2017a). Most countries that provide data to Eurostat do so at a NUTS 2 level (approximately the size of counties), whilst the UK and Germany provide their data on the NUTS 1 level (larger regions consisting of multiple counties). We, therefore, calculate regional risk of infection depending on the scale of regional data on susceptible hosts provided to Eurostat. We calculate the average number of cattle on a farm in each country/region to represent the number of susceptible animals the imported

animals will be in contact with. Although we have data on trade to individual farms (determined by their postcode), we do not have detailed information about those farms, such as the number of cattle. Thus, we use the regional average number of cattle per farm for the relevant region. We do not include differences in the types of farming and husbandry systems that may occur throughout the various countries, however, if data is available for LSD or in a different case study, this could be included by changing the underlying formula for R_0.

Lumpy Skin Disease data (c, β, r)

There is a great deal of uncertainty over which transmission routes are most important for LSD, the contact rates along those routes, the minimum infective dose required for each route and the probability that the infective dose would be met. Therefore, it would be exceedingly difficult to produce reliable estimates for contact rates (c) and the probability of transmission (β) separately. Instead, we combine the contact rate between animals and probability of transmission into a transmission rate (ξ) and use results from a statistical analysis of an outbreak on a single farm (Magori-Cohen et al. 2012) which estimated transmission rates. This study, followed up by personal communication with the authors, determined that vector (mechanical) and direct transmission were key transmission routes, with rates of 0.026 and 0.006 per day, respectively. In Magori-Cohen et al. (2012) the mechanical transmission term from cow to cow is represented by the formula ξSI rather than using frequency-dependent transmission therefore we use the same formula to represent mechanical transmission. Magori-Cohen and authors (personal comm.) provided us with estimates of the uncertainty for the two transmission rates through 95% confidence intervals, [0.013, 0.052] and [0.003, 0.012] for mechanical and direct transmission, respectively. Since mechanical vector transmission is thought of as the most important mode of transmission for LSD (Carn and Kitching 1995b) we use the upper and lower confidence interval bounds for this parameter in a sensitivity analysis to estimate our uncertainty in the risk calculations due to our uncertainty in transmission rate. We assume that across the whole of Europe there are suitable vectors which are able to transmit LSD virus. This is a fair assumption considering the wide range of species that have been implicated as potential mechanical vectors. The number of days that cattle will exhibit viremia, as well as shedding of virus from nasal, oral and conjunctival secretions, has been estimated to be between 6-18 days (Carn and Kitching 1995a, Tuppurainen et al. 2005, Babiuk et al. 2008). However, virus has been detected by PCR in skin nodules on cattle up to 42 days post inoculation (Babiuk et al. 2008) and up to 159 days in the semen of experimentally infected bulls (Irons et al. 2005). We assume a 15 day infectious period (r) to represent the shorter viremic period and perform a sensitivity analysis for a 42 day infectious period to incorporate the effects of potential longer skin nodule infectivity.

All of our calculations are performed in R (R Core Team 2016). We take random draws from the prevalence distribution to estimate the prevalence in each country but reject those samples which fall outside the 5

and 95 quantiles due to long tails of this distribution. This is based on the methodology from the SPARE project which provides the prevalence data. We then randomly draw the number of infected animals entering a country, as described earlier. In total, we use 10,000 iterations. Executing the calculations described earlier, the model outputs a distribution of the risk of infection. To present the results the mean and variance of this distribution are provided. The default scenario considered is a 15 day infectious period, a mechanical transmission rate of 0.026 and no detection on import. Modifications to the default scenario are considered in the sensitivity and scenario analyses and are clearly stated; otherwise it can be assumed that the parameters are those of the default scenario.

Results

We reiterate here that we define the risk of infection as the probability of one or more initial infections in the native susceptible population in Area B. In nearly all simulations, we find that virtually all locations will have a probability very close to or indistinguishable from 1 if an infected animal is imported to the farm and obviously a probability of 0 if no infected animal is imported. Therefore, each simulation of the risk is essentially a Bernoulli distribution. Combining all the simulations together produces a distribution akin to a scaled Binomial distribution. Hence, the mean and variance of the risk can be interpreted similarly to the mean and variance of a Binomial distribution. The variance is a representation of our total uncertainty arising from the model and input parameters and is only driven by the distribution for prevalence.

The mean annual probability of initial infection per location, at a country level assessment for the default scenario is plotted in Figure 2. Many countries in Europe import only from countries which have an estimated zero prevalence, according to our prevalence data within each country. Therefore, they have negligible risk of having one or more native susceptible animals becoming infected due to importing an infected animal. Croatia, with a mean probability estimated at 0.87, has the highest annual risk, followed by Italy (mean risk 0.72), Hungary (mean risk 0.62), Spain (mean risk 0.6) and Slovenia (mean risk 0.448). This indicates a probability of 87% for at least one infected native host in Croatia in 2016. We also plot the variance in this estimate of risk of infection. There is little uncertainty in our estimate of risk for Croatia, indicating that the high risk assessment holds true regardless of the stochastic nature of prevalence within countries. On the other hand, uncertainty is high for Italy, Spain, Hungary and Slovenia, with variance between 0.2 and 0.25.

The mean probability of infection due to trade at a regional level also highlights Croatia, Italy and Hungary as countries with the highest risk (Figure 3). However, now it is possible to observe that this risk is focussed in specific regions of these countries. Croatia has high risk across the whole country whereas Italy has highest risk in the northern regions, and in Hungary the highest risks are in the southern part of the country. Similarly, the variance of this risk is plotted on a regional scale. Some countries, for example

Romania and Germany, have regions with both low probability of infection and low variance in this probability, as well as other regions which have low mean probability but a higher variance. This allows for better understanding of where to focus surveillance activities.

Risk of initial infection at an individual farm level (Figure 4) indicates that the majority of trade within Area B with countries that have non-zero prevalence occurs in Croatia, hence why Croatia had such a high risk at a country and regional level. The two farms with the highest probability of infection are in Croatia including one farm with a risk of 0.65. However, Spain follows closely behind with the next 3 highest risk farms with probabilities of infection between 0.51 – 0.54. Some countries which are assessed as having negligible risk due to not importing from infected countries (according to the COMEXT dataset, Figure 2), have individual farms in the TRACES data doing so, e.g. France and UK, Figure 4. This is due to differences that occur between the two different trade datasets (see Appendix B). However, the regional and individual farm risk maps agree on the regions with highest risk in Italy, Germany, Poland and Hungary. The plot of variance in the risk of initial infection at an individual farm level indicates much uncertainty in the risk assessment centred in Croatia, in contrast to the country and regional level risk assessments which indicated lower uncertainty in Croatia's risk. Although each farm has high uncertainty as to whether infection is likely to occur in a native host, the combination of multiple farms with high risk culminates in more certainty that infection would occur. Hence, the result of many farms with high risk and high uncertainty results in high risk with little uncertainty at a regional or country level.

Sensitivity Analysis

We performed sensitivity analysis on the two main parameters that have uncertainty, the length of the infectious period in cattle (15 or 42 days) and the mechanical (vector) transmission rate (in the range 0.013-0.052). Varying either of these parameters within these values does not make a noticeable difference to the results. The high transmission rates, the ability for hosts to infect any of the susceptible hosts on the farm and the long infectious period, lead to high R_0 values at each location, even when these parameters vary.

Scenario Analysis

A risk assessment at a country level when the import detection probability is increased from 0 to 0.5 for all countries indicates, as expected, a decrease in the overall probability (Figure 5). Although import detection approximately halves the number of imported infected animals, it does not halve the value of risk. This is because in many simulations no infected animals enter due to low prevalence, which will not change when import detection does occur. In general, the values for risk are reduced most for countries with low to medium risk with, for example, Germany, the Netherlands, Romania and Poland reducing their probability of infection by 30 – 50%. In contrast, import detection is not as successful for countries with higher risk, with Italy and Croatia only reducing their risk by 13% and 7% respectively. This is due to high import rates

from countries with non-zero prevalence, so that with 50% detection many infected animals could still enter the country.

Discussion

Our risk assessment was performed at various spatial scales, both to show the flexibility of the framework, and to understand more clearly how LSD risk is distributed across Europe. In comparison to the risk of infection calculation at a regional or country level, the individual level informs whether higher risk arises due to a large number of farms with low to medium risk, or due to a small number of farms with high risk. For example, only a few individual farms in Spain import but they have a high risk of LSD, whereas in Italy and Hungary many farms import from countries with non-zero prevalence but they have low to mid risk (Figure 4). This is particularly pertinent for Spain, in this case, as at a country level its risk is lower than a few other countries. This could lead to less overall surveillance than those countries when in reality there are a few farms with very high risk. This level of detail is not possible in the regional and country level risk assessment. We reiterate that our regional risk assessment involves the assumption that the distribution of animal imports amongst regions will be similar between the UK and the rest of Europe. Our model for trade distribution (see Appendix A) found that the number of farms and the proportion that are dairy are significant determinants of cattle being imported into different regions of the UK but this may not be the case for other European countries. Comparing Figure **3** and Figure **4** we can see that while most countries have similar areas of risk predicted at a regional level and at an individual farm level, this is not true for Spain. The regional areas at risk are predicted to be in the north west of Spain but according to the TRACES dataset, Figure **4**, the individual farms at risk are in the north east. This is likely due to Spain not importing according to our model for trade distribution amongst regions. However, for most countries our results are consistent across the spatial scales. Whilst individual farm risk provides the most detailed information, when time is short or data are not available, country and regional risk assessments provide a useful and relevant measure of risk.

This risk assessment for LSD is focussed on 2016, coinciding with the outbreak in the Balkans, and highlights Croatia as the country with the highest risk. However, our risk assessment did not take into account any control measures, other than import detection, which countries may have implemented during this outbreak. In fact, Croatia started to vaccinate its cattle population for LSD in August 2016, achieving 100% coverage by November 2016 (European Food Safety Authority (EFSA) 2017). This risk assessment, alongside Croatia's close location to infected countries, suggests that Croatia was wise to vaccinate to avoid infection. In fact, the FAO have released a position paper (FAO 2017), following a number of confirmed cases of LSD in 2017, suggesting full-scale vaccination policies in countries with high risk in Eastern Europe, regardless of whether they have had infected cases, in order to reduce the likelihood of another outbreak. This is to

avert the spread of the disease and reduce the need for a total stamping-out policy, which can significantly affect farmers, especially smallholders. As in other models of the spread of infectious diseases (Keeling and Rohani 2007), our model can include the role of vaccination by reducing the number of susceptible animals in Area B that may come into contact with the imported infected species. This reduction would be based on available data on the proportion of animals that are vaccinated at each location in Area B. Vaccination could also lower the transmission rates, as vaccinated cattle may be less susceptible to the disease, or it could reduce infectivity of cattle by shortening the length of the infectious period, both of which can be easily changed in our model.

However, vaccination against LSD may not be suitable or recommended for all countries due to the fact it is a live vaccine with no test to distinguish infected from vaccinated cattle (Tuppurainen and Oura 2012). As an alternative, increasing the probability of detection on import does reduce the risk from trade imports. However, the effects of import detection are not equivalent across countries, and the countries with highest risks would need higher probabilities of successful detection to be able to reduce their risk by the same proportion.

Italy, Hungary, Spain and Slovenia had higher risk than many other countries, demonstrating the potential for LSD to lead to local infections in cattle populations across Europe due to trade. However, this risk assessment assumed that a suitable active vector is always present in all locations – this may not be the case in more northern countries in Europe, or during certain times of the year. This could significantly reduce the risk estimates in more northerly countries as the transmission rates through direct contact are significantly smaller than through mechanical vector contact (Carn and Kitching 1995b, Magori-Cohen et al. 2012). Croatia, Italy and Spain have similar Mediterranean climates to Greece, where infection has occurred and spread, indicating disease suitability in that climate for at least part of the year provided vectors are present.

This case study did not assess the risk of initial infection through vector movement across borders or other unregistered trade/movement, such as illegal trade or movement of people, including refugees who may bring cattle with them. Undoubtedly, this would increase the risk in those countries which are neighbouring infected countries, such as Croatia, Bosnia and Herzegovina, Hungary and Romania, depending on the political climate, wind events, presence of vectors and temperature requirements of vectors. Furthermore, if transtadial and transovarial transmission of LSD is shown to occur in many tick species (Tuppurainen et al. 2011), this may be an important transmission pathway to consider for spread of LSD.

The difference in risk for LSD across Europe is primarily driven by prevalence in export countries and the number of cattle imported, whereas specific disease parameters and the differences between countries/regions in the average number of susceptible animals on a farm did not have much impact on risk

estimates. This finding is expected to be similar for most diseases, although different formulas for R_0 to estimate disease transmission within each area could affect it, especially if the R_0 parameters are dependent on environmental factors, which vary between areas. The lack of consensus on which vectors of LSD are the most influential in spreading the disease (via mechanical transmission) mean that we are not able to vary R_0 with location based on how the vectors respond to environmental variables. Therefore, our LSD case study is not able to include details surrounding a key element of transmission. However, as the difference in risk is primarily driven by the import trade and as the greatest risk is located in countries which are likely to have similar climates to those which had outbreaks in 2016, we believe our results provide a robust estimation of risk across Europe for 2016.

Our generic framework can be adapted for different modes of transmission by changing the formula used for R_0. This provides flexibility to consider other case studies which involve different transmission pathways, such as vector-borne transmission, environmental and contaminated feed and water sources. Furthermore, the generic framework can also be used for different types of movement in to Area B, such as unregistered movement. Movement of terrestrial wildlife, migratory birds and local and windborne travel of vectors can be included in to the framework through the parameter $N_k(g)$ in the same way as the import of animals by trade. The difference is that the estimation of $N_k(g)$ will likely be based on a model of animal movement instead of a global database, such as we used for trade in the LSD case study. Similarly, we estimated the number of susceptible animals using databases on the number of animals on farms in each country, but for other pathways a more useful measure could be the number of animals in a 10km^2 area extracted from density maps of animals, depending on the data used for movement of animals. Importantly, however, the framework itself remains unchanged for these different pathways of disease introduction.

The accuracy of our framework to predict the risk of initial infection for different diseases will depend greatly on the quality and quantity of data available for the disease and relevant animal species. Even though the framework is applicable to all disease introductions as outlined above, it may not always be possible to calculate the risk of initial infection if there is insufficient data available. If exact data required by the model are not available, it is possible to use proxy data, although this increases the uncertainty associated with the results. For example, it is possible to use cattle density as a proxy for vector density but there may be greater uncertainty over the reliability of the proxy data at low and high densities. In general, a focus on a risk ranking of countries is likely to produce more reliable predictions rather than concentrating on the risk estimates themselves. Missing or inaccurate data can always bias the results, even if under-reporting factors or models are used to estimate the true values. This can be seen in Appendix B, Figure B2, in which the risk estimates at a country level are compared using either the Comext dataset or the TRACES dataset, indicating that there are clear differences between these datasets. However, it is not possible to state with any certainty which is the better dataset to use. Similarly, reliable

estimates for prevalence of a disease in all countries throughout the world are difficult to obtain. The data we use for prevalence from Simons et al. (2017) is based on OIE notified cases, the best freely-available source for worldwide animal disease outbreaks. But it is judicious to remember that the resulting data on prevalence is the result of a model with uncertainty and assumptions. We provide a risk assessment for LSD based on the best available data. Conversely, the generic framework we present is applicable for all data sources and can be re-used as better data becomes available.

A major advantage of a framework that promotes the computation of risk at various spatial scales is that it allows for the identification of hotspots of disease, and hence it can guide policy decisions regarding the implementation of more specific and directed surveillance of potential infection. Enhancing surveillance methods has the potential to reduce time to detection of infection thus reducing the likelihood of widespread outbreaks, as well as decreasing the costs of, and time spent, on surveillance. Clearly countries with a range of low, medium and higher risks should target surveillance within those regions appropriately. On the other hand, countries with low risk may also want to have surveillance in place but this may be too costly to implement across the whole country. In this case, the variance in the risk can provide additional information for directing surveillance. Additionally, the generic framework can also be expanded to include a method for risk of spread from the hotspots of infection. Combining the risk of initial infection with risk of spread would allow the risk assessor to determine not only which locations have highest risk of infection occurring from outside sources but also, of those locations, which are most likely to spread the disease further within Area B. Our generic framework provides a method for the calculation of risk of initial infection for the introduction of any pathogen from any Area A to Area B, and specifically it allows these calculations to be made across many different spatial scales depending on the question in mind and the data available to the risk assessor. This is a powerful tool that can be used to determine not only which diseases are of most concern for different countries, but also where to focus surveillance within countries for different pathogens.

Acknowledgements

The authors wish to sincerely thank Dr Robin Simons and Dr Amie Adkin, APHA, (funded through the Animal Health and Welfare ERA-NET consortium (https://www.anihwa.eu/) under SPARE 'Spatial risk assessment framework for assessing exotic disease incursion and spread through Europe') for providing the prevalence distribution parameters for lumpy skin disease in 2016. The authors also wish to thank Dr Rachel Jinks, APHA, for useful discussion on the statistical methods used in Appendix A. This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 643476.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

Ali, A., M. Esmat, H. Attia, A. Selim and Y. Abdel-Hamid (1990). "Clinical and pathological studies of lumpy skin disease in Egypt." Veterinary Record **127**(22): 549-550.

Babiuk, S., T. Bowden, G. Parkyn, B. Dalman, L. Manning, J. Neufeld, C. Embury-Hyatt, J. Copps and D. Boyle (2008). "Quantification of lumpy skin disease virus following experimental infection in cattle." Transboundary and Emerging Diseases **55**(7): 299-307.

Bengtsson, L., J. Gaudart, X. Lu, S. Moore, E. Wetter, K. Sallah, S. Rebaudet and R. Piarroux (2015). "Using mobile phone data to predict the spatial spread of cholera." Scientific reports **5**.

Brownstein, J. S., T. R. Holford and D. Fish (2005). "Effect of climate change on Lyme disease risk in North America." EcoHealth **2**(1): 38-46.

Carn, V. and R. Kitching (1995a). "The clinical response of cattle experimentally infected with lumpy skin disease (Neethling) virus." Archives of virology **140**(3): 503-513.

Carn, V. and R. Kitching (1995b). "An investigation of possible routes of transmission of lumpy skin disease virus (Neethling)." Epidemiology & Infection **114**(1): 219-226.

Chihota, C., L. Rennie, R. Kitching and P. Mellor (2001). "Mechanical transmission of lumpy skin disease virus by Aedes aegypti (Diptera: Culicidae)." Epidemiology & Infection **126**(2): 317-321.

Davies, F. G. (1991). "Lumpy skin disease of cattle: a growing problem in Africa and the Near East." World Animal Review **68**(3)): 37-42.

European Food Safety Authority (EFSA) (2017). Lumpy skin disease: I. Data collection and analysis. EFSA Journal. **15(4):** 4773.

Eurostat. (2017a). "Eurostat Bulk Download Listing." Retrieved 9th May, 2017, from http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing.

Eurostat. (2017b). "NUTS - Nomenclature of territorial units for statistics - Overview." Retrieved 8th May, 2017, from http://ec.europa.eu/eurostat/web/nuts.

FAO (2017). Sustainable preventon, control and elimination of Lumpy Skin Disease - Eastern Europe and the Balkans. FAO Animal Production and Health Position Paper. No. 2. Rome, Italy.

Gale, P., B. Stephenson, A. Brouwer, M. Martinez, A. de la Torre, J. Bosch, M. Foley-Fisher, P. Bonilauri, A. Lindström and R. Ulrich (2012). "Impact of climate change on risk of incursion of Crimean-Congo haemorrhagic fever virus in livestock in Europe through migratory birds." Journal of applied microbiology **112**(2): 246-257.

Grais, R., J. H. Ellis, A. Kress and G. Glass (2004). "Modeling the spread of annual influenza epidemics in the US: The potential role of air travel." Health care management science **7**(2): 127-134.

538 Hill, A., T. Dewé, R. Kosmider, S. Von Dobschuetz, O. Munoz, A. Hanna, A. Fusaro, M. De Nardi, W. Howard
539 and K. Stevens (2015). "Modelling the species jump: towards assessing the risk of human infection from
540 novel avian influenzas." Royal Society open science **2**(9): 150173.

541 Hunter, P. and D. Wallace (2001). "Lumpy skin disease in southern Africa: a review of the disease and
542 aspects of control." Journal of the South African Veterinary Association **72**(2): 68-71.

543 Irons, P., E. Tuppurainen and E. Venter (2005). "Excretion of lumpy skin disease virus in bull semen."
544 Theriogenology **63**(5): 1290-1297.

545 Keeling, M. J. and P. Rohani (2007). Modeling infectious diseases in humans and animals, Princeton
546 University Press.

547 Kumar, S. M. (2011). "An outbreak of lumpy skin disease in a Holstein dairy herd in Oman: a clinical report."
548 Asian Journal of Animal and Veterinary Advances **6**(8): 851-859.

549 Magori-Cohen, R., Y. Louzoun, Y. Herziger, E. Oron, A. Arazi, E. Tuppurainen, N. Y. Shpigel and E. Klement
550 (2012). "Mathematical modelling and evaluation of the different routes of transmission of lumpy skin
551 disease virus." Veterinary research **43**(1): 1.

552 Mercier, A., E. Arsevska, L. Bournez, A. Bronner, D. Calavas, J. Cauchard, S. Falala, P. Caufour, C. Tisseuil and
553 T. Lefrançois (2017). "Spread rate of lumpy skin disease in the Balkans, 2015–2016." Transboundary and
554 Emerging Diseases.

555 Murray, N. (2004). Handbook on import risk analysis for animals and animal products: quantitative risk
556 assessment, Office international des épizooties.

557 R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical
558 Computing. Vienna, Austria https://www.R-project.org/.

559 Schmidt, C. W. (2012). "Trending now: using social media to predict and track disease outbreaks."
560 Environmental health perspectives **120**(1): a30.

561 Simons, R. R. L., V. Horigan, M. De Nardi, G. Ru, A. E. Pena and A. Adkin (2017). Mighty models from little
562 data grow: Estimating animal disease prevalence. Proceedings of the Society for veterinary epidemiology
563 and preventative medicine. Inverness, Scotland**:** 166.

564 Tatem, A. J., S. I. Hay and D. J. Rogers (2006a). "Global traffic and disease vector dispersal." Proceedings of
565 the National Academy of Sciences **103**(16): 6242-6247.

566 Tatem, A. J., D. J. Rogers and S. I. Hay (2006b). "Estimating the malaria risk of African mosquito movement
567 by air travel." Malaria journal **5**(1): 57.

568 Thevenin, M. (2011). "Kurdish Transhumance: Pastoral practices in South-east Turkey." Pastoralism:
569 Research, Policy and Practice **1**(1): 23.

570 TRACES. (2017). "TRACES - Trade Control and Expert System." Retrieved 12th July, 2017, from
571 https://ec.europa.eu/food/animals/traces_en.

572 Tuppurainen, E. and C. A. Oura (2012). "lumpy skin disease: an emerging threat to Europe, the Middle East
573 and Asia." Transboundary and Emerging Diseases **59**(1): 40-48.

574 Tuppurainen, E. S., W. Stoltsz, M. Troskie, D. Wallace, C. Oura, P. Mellor, J. A. Coetzer and E. H. Venter
575 (2011). "A potential role for ixodid (hard) tick vectors in the transmission of lumpy skin disease virus in
576 cattle." Transboundary and emerging diseases **58**(2): 93-104.

577 Tuppurainen, E. S., E. Venter and J. Coetzer (2005). "The detection of lumpy skin disease virus in samples of
578 experimentally infected cattle using different diagnostic techniques." Onderstepoort Journal of Veterinary
579 Research **72**(2): 153-164.

580 Vayssier-Taussat, M., S. Moutailler, L. Michelet, E. Devillers, S. Bonnet, J. Cheval, C. Hébert and M. Eloit
581 (2013). "Next generation sequencing uncovers unexpected bacterial pathogens in ticks in western Europe."
582 PloS one **8**(11): e81439.

583 Wardrop, N. A., E. M. Fèvre, P. M. Atkinson, A. S. Kakembo and S. C. Welburn (2012). "An exploratory GIS-
584 based method to identify and characterise landscapes with an elevated epidemiological risk of Rhodesian
585 human African trypanosomiasis." BMC infectious diseases **12**(1): 316.

586 Weiss, K. (1968). Lumpy skin disease virus. Cytomegaloviruses. Rinderpest Virus. Lumpy Skin Disease Virus,
587 Springer**:** 111-131.

588 World Organisation for Animal Health (OIE). (2017). "OIE - WAHIS Interface." Retrieved March, 2017, from
589 http://www.oie.int/wahis_2/public/wahid.php/Wahidhome/Home.

590 Yeruham, I., O. Nir, Y. Braverman, M. Davidson, H. Grinstein, M. Haymovitch and O. Zamir (1995). "Spread
591 of lumpy skin disease in Israeli dairy herds." Veterinary Record **137**: 91-91.

592

593 Tables

594

595 Table 1 Data required to calculate the risk of initial infection for the generic framework.

Parameter	Specific Data	Further Details	Potential Data Sources
Movement from Area A to Area B ($N_k(g)$)	Trade/Registered Movement	Trade in livestock or registered movement of hosts from regions in Area A to locations in Area B. Reason for movement helpful to determine final location of host. For illegal trade an underreporting factor is necessary or estimates for locations with no reported data.	UN Comtrade data Eurostat COMEXT data Trade Control and Expert System (TRACES)
	Import Detection ($p_D(g)$)	The probability that locations in Area B will detect infection in animals imported through trade.	Published literature
	Movement of wild animals	Average home range size or distance routinely travelled by wild animals. This may be affected by weather events. For birds, this would include migration routes together with time of year.	Published literature
	Location and abundance of	Approximate spatial distribution and numbers of wild animals in Area A.	Published literature

	wild animals in Area A		Global Biodiversity Information Facility (GBIF)
Prevalence (p_k)	Prevalence of the disease in Area A	Preferably for the same regions in Area A as the movement data.	OIE Animal Disease Notification System FAO EMPRES-i
Susceptible Hosts ($S(g)$)	Size of farms in Area B	When importing to farms this determines how many susceptible hosts will be in contact with the infected imports. Depending on spatial scale, this could be the number of hosts on a specific farm, or the average number of hosts on farms in a region.	OIE Eurostat
	Spatial distribution of wild animals or people	To determine where and how many susceptible hosts the imported hosts will have contact with	Published literature Global Biodiversity Information Facility (GBIF)
Specific host and disease data (c, β, r)	Length of the infectious period (r)	This could be affected by whether disease is detected at locations in Area B once symptoms appear and if they perform culling or eradication measures.	Published literature OIE for control measures
	Average lifespan of species i and j in Area B (r)	This could be affected by climate data.	Published literature
	Probability of transmission (β) between species i and j		Published literature
	Contact rates (c) between species i and j	This could be affected by climate data.	Published literature

596

597 Table 2 Parameter values and data sources used in the lumpy skin disease case study for the default
598 scenario. Different values used in the sensitivity and scenario analyses are provided in square brackets.

Parameter	Description	Value	Source
Movement from Area A to Area B ($\boldsymbol{N_k(g)}$)	Legal trade	-	Country & Regional level: Eurostat Comext data (Eurostat 2017a) Farm level: TRACES data (TRACES 2017)
Prevalence ($\boldsymbol{p_k}$)	Prevalence of the disease in Area A	-	EU-funded SPARE project (Simons et al. 2017)
Susceptible Hosts ($\boldsymbol{S(g)}$)	Average (Country or regional level) or specific (farm level) number of animals on a farm	-	Eurostat data (Eurostat 2017a)
Import Detection ($\boldsymbol{p_D(g)}$)	Probability of detecting infected animals on import	0 [0.5]	-
Infectious period ($\boldsymbol{r}$)	Length of infectious period in days	15 [42]	(Carn and Kitching 1995a)
Transmission rate ($\boldsymbol{\xi}$)	Direct transmission Mechanical transmission via vectors	0.006 0.026 [0.013 – 0.052]	Magori-Cohen et al. (2012)

599

600

Figure Legends

Figure 1 The 5 steps of the risk pathway for the generic spatial risk question “What is the risk of infection of a pathogen in Area B due to the presence of that pathogen in Area A?” The term “unit” in Step 1 refers to a generic source of infection such as species, products or feed, provided that they could be in contact with native susceptible hosts.

Figure 2 The mean annual risk of infection (A) and the variance of this risk (B) are plotted in shades of purple across Europe, calculated at the country level. Countries in yellow have negligible risk due to only trading with countries that have zero prevalence, according to our prevalence data. Countries which had notified cases in 2016 are in red. Comext trade data is used.

Figure 3 The mean annual risk of infection (A) and the variance of this risk (B) are plotted in shades of purple across Europe, calculated at the regional level. Countries in yellow have negligible risk due to only trading with countries that have zero prevalence, according to our prevalence data. Countries which had notified cases in 2016 are in red. Countries in grey have insufficient data for calculating risk. Comext trade data is used.

Figure 4 The mean annual risk of infection (A) and the variance of this risk (B) are plotted in shades of purple across Europe, calculated at the individual farm level. Individual farms are only plotted if they trade with a country that has non-zero prevalence. In (C) and (D), the mean and variance, respectively, of the annual risk of infection are again plotted, but zoomed in to the areas outlined by rectangles in (A) and (B). Regions in yellow have negligible risk due to farms within those regions only trading with countries that have zero prevalence, according to our prevalence data. Countries which had notified cases in 2016 are in red. Countries in grey have insufficient data for calculating risk. TRACES trade data is used.

Figure 5 The percent that mean risk is reduced by when the probability of detection is increased from 0 to 0.5 is plotted in shades of purple across Europe, calculated at the country level. Countries in yellow have negligible risk due to only trading with countries that have zero prevalence, according to our prevalence data. Countries which had notified cases in 2016 are in red. Comext trade data is used.

A generic framework for spatial quantitative risk assessments of infectious diseases: lumpy skin disease case study

Supplementary Material

Rachel A. Taylor[1], Alexander DC Berriman[1], Paul Gale[1], Louise A. Kelly[1,2], Emma L. Snary[1]

[1]Animal and Plant Health Agency (APHA), Weybridge, UK

[2]Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK

Appendix A

Distributing cattle amongst regions

The country trade data that we use only indicates which country is importing the cattle, and not the region within that country. This hinders us from performing risk assessment at the regional level. To estimate the distribution of cattle amongst each region given the number that enter the country, we extrapolate based on Great Britain (GB), in which we are able to access a higher resolution of data. We use the Trade Control and Expert System (TRACES) database, which collects data on trade of live animals from both within and outside Europe (TRACES 2017). Importantly, it provides detailed information, such as addresses, when the import country is the country through which you have access to the database. Therefore, we collate this data from 2012-2015 to calculate the proportion of imported animals to each region relative to the total number of imports to GB overall. We perform a generalised linear regression with a logit transformation on the number of imports into each region versus the number which are imported into all GB regions (hereafter referred to as the proportion of imports). The null hypothesis is that cattle imported to GB are randomly distributed amongst regions. We use the following predictors: the number of farms in each region, and the proportion of farms in each region which are dairy farms, both of which are accessed on a NUTS 1 level through Eurostat (Eurostat 2017). Pearson's correlation coefficient indicates no correlation between the number of farms in each region and the proportion which are dairy farms. We compare the models with each and all of the predictors using Akaike's Information Criterion (AIC) to choose the best fitting model.

Including the two predictors, the number of cattle in each region and the proportion of cattle farms that are dairy farms, as well as the interaction term is the best model by AIC to describe how imported cattle to GB are distributed amongst regions of GB, with all coefficients statistically significant. An increase in the proportion of dairy farms in a region always leads to an increase in imports, but this is not true for the number of farms (Table A, Figure A 1). If the number of farms increases but none of them are dairy, imports go down but this relationship changes when the dairy proportion increases. Regions with large numbers of farms and high proportion of dairy farms will import the most cattle.

Table A Model results from the generalised linear model with a logit transformation. In the model, PropImport indicates the proportion imported to each region, Farms and Dairy and the number of farms and the proportion of dairy farms in each region, respectively. *** indicates a p-value less than 0.001.

Model: PropImport ~ logit($\beta_0 + \beta_1 * Farms + \beta_2 * Dairy + \beta_3 * Farms * Dairy$)			
Variable	Estimate	Std. Error	P-value
Intercept (β_0)	-5.61	2.23 x10^{-2}	***
Farms (β_1)	-3.52 x10^{-4}	1.15 x10^{-5}	***
Dairy (β_2)	3.87	6.94 x10^{-2}	***
Farms*Dairy (β_3)	2.44 x10^{-3}	3.54 x10^{-5}	***

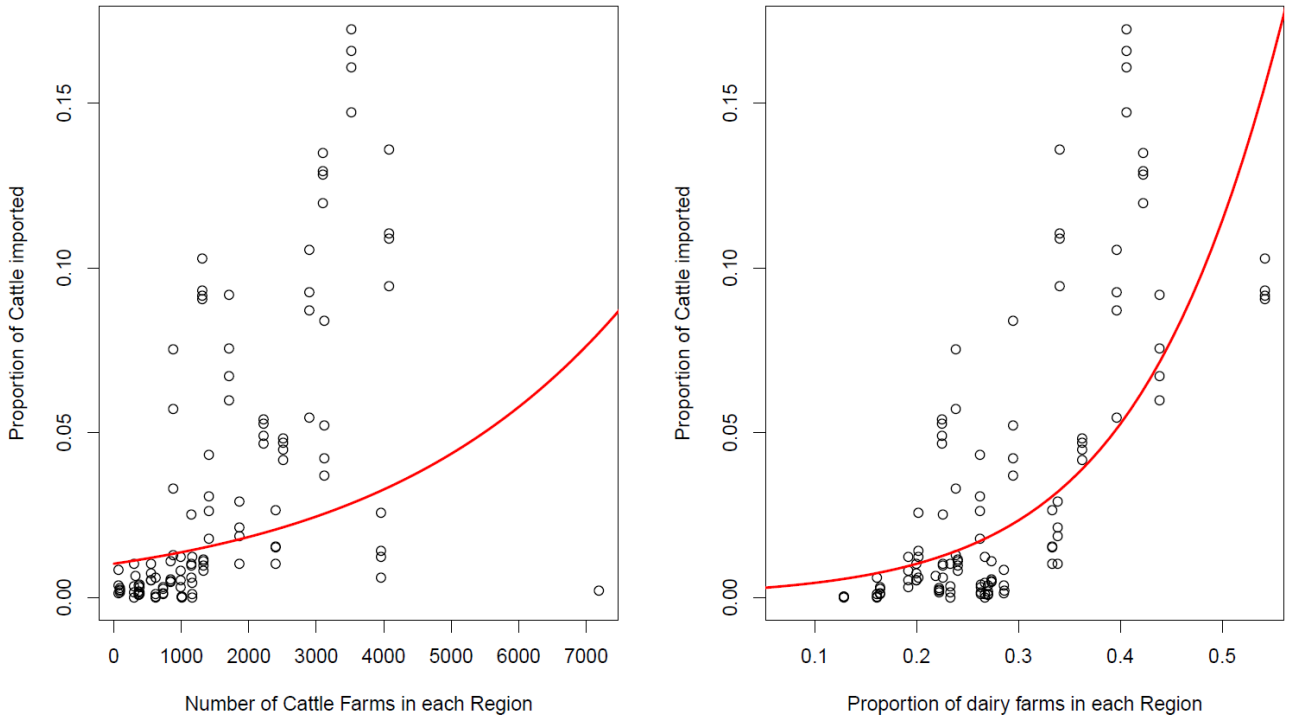

Figure A 1 The proportion of cattle imported in to each region plotted against each predictor. The curve of best fit is plotted in red against each predictor while the other predictor is held at its mean value.

Now we have this model for how to distribute cattle amongst regions for GB, we use it to estimate the proportion of imports into each region of each country across Europe in 2016. To calculate the predicted proportion, we input the number of cattle farms and the proportion of dairy farms for each region in each country and then normalise the predicted proportions, as the sum of all the regional predicted proportions of a country should sum to 1. We then multiply each proportion by the total number of imports to the country to get the number of imports to each region. Finally, we round these numbers off to the nearest integer for whole animals, ensuring that the total is still equal to the total imported to the country as a whole.

Appendix B

Comparing Data Sources

The trade data that we use in the main text to compute country and regional risk is COMEXT, a source of freely available trade data listing all products traded between countries in 2016. However, we also had access to the TRACES data, a primarily EU-based trade system, which provided us with data on all cattle traded in 2016 to the EU, including the postcode of the final destination. We used this for our calculation of risk at the individual farm level. However, as these are different data sources, it is not clear how accurately they match up and whether they would lead to very similar results. We consolidate the TRACES data into the total imports into each country as a whole in order to compute the risk assessment at a country level with TRACES data (Figure B 1A). This can then be compared with the risk assessment using the COMEXT data (Figure 2) to produce Figure B1B.

Figure 4 and Figure B 1 demonstrate the presence of data in the TRACES (postcode) data that is not in the COMEXT (country) data. In particular, some countries now have a risk estimate whereas in the COMEXT risk map they were recorded as having negligible risk due to not trading with any infected partner. Mostly this arises from a single farm (see Figure 4, e.g. France, UK, Finland). However, the most notable country that changes from negligible risk is Austria, with a probability of infection now of 0.45. Other countries also have higher probabilities, most notably Spain, while Germany's risk estimate is significantly smaller. The Netherlands has negligible risk according to the TRACES data whereas COMEXT predicts a probability of infection in the range 0.1-0.2.

Clearly, there are differences between the two different data sources in terms of how much trade occurs with partner countries. However, looking at the level of risk throughout Europe overall, both datasets predict a very similar ordering of countries with the highest risk. Both outline Croatia as the only country in the highest risk bracket, with Italy, Hungary and Spain having the next highest risks. It is not possible to determine which of the two data sources is more reliable. We choose to focus on the COMEXT dataset in the main text due to its freely available nature, with the use of TRACES restricted to when we have no other dataset available at that resolution.

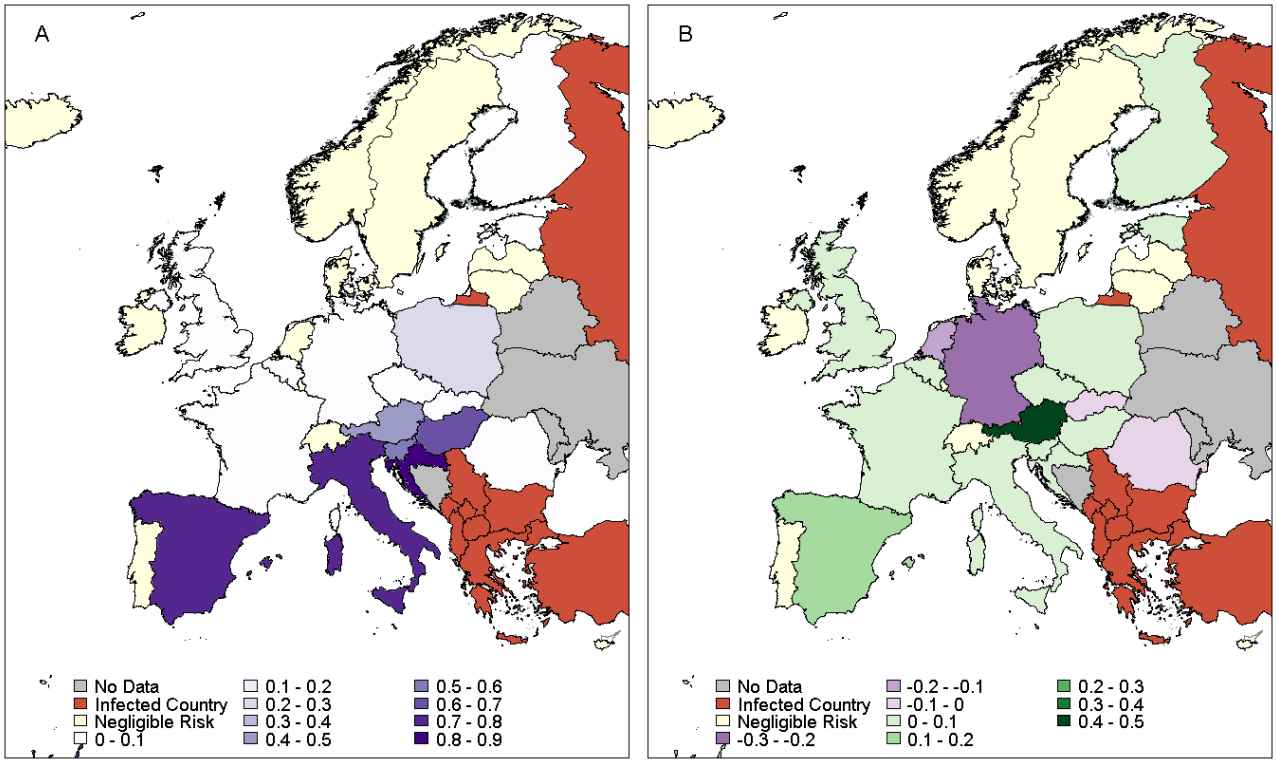

Figure B 1 In (A) the mean annual risk of infection is plotted in shades of purple across Europe, calculated at the country level, when the trade data is from TRACES. In (B) the difference between the mean risks calculated using TRACES (Figure B1A) or COMEXT data (Figure 2A) is plotted across Europe. Positive values (in green) indicate TRACES predicts a higher risk. Negative values (purple) indicate COMEXT predicts a higher risk. Countries in yellow have negligible risk due to only trading with countries that have zero prevalence, according to our prevalence data. Countries which had notified cases in 2016 are in red. Countries in grey have insufficient data for calculating risk.

References

Eurostat. (2017). "Eurostat Bulk Download Listing." Retrieved 9th May, 2017, from http://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing.

TRACES. (2017). "TRACES - Trade Control and Expert System." Retrieved 12th July, 2017, from https://ec.europa.eu/food/animals/traces_en.