# Using national movement databases to help inform responses to swine disease outbreaks in Scotland: the impact of uncertainty around incursion time

Porphyre, T; Boden, LA; Correia-Gomes, C; Auty, HK; Gunn, GJ; Woolhouse, MEJ

[Link to publication]

1 Using national movement databases to help inform responses to swine

2 disease outbreaks in Scotland: the impact of uncertainty around incursion

3 time

4 Thibaud Porphyre[1*], t.porphyre@ed.ac.uk

5 Lisa A. Boden[2], Lisa.Boden@glasgow.ac.uk

6 Carla Correia-Gomes[3], Carla.Gomes@sruc.ac.uk

7 Harriet K. Auty[3], Harriet.Auty@sruc.ac.uk

8 George J. Gunn[3], George.Gunn@sruc.ac.uk

9 Mark E.J. Woolhouse[1], Mark.Woolhouse@ed.ac.uk

10

11 [1] Centre for Immunity, Infection and Evolution, University of Edinburgh, King's Buildings, Edinburgh, UK.

12 [2] School of Veterinary Medicine, Boyd Orr Centre for Population and Ecosystem Health, College of

13 Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.

14 [3] Epidemiology Research Unit, SRUC, Drummondhill, Stratherrick Road, Inverness, UK

15

16 *Corresponding author: Thibaud Porphyre, University of Edinburgh, King's Buildings, Ashworth

17 Laboratories, Charlotte Auerbach Road, Edinburgh EH9 3FL, Scotland, UK. Tel: +44 (0)131 650 7263.

18 Email: t.porphyre@ed.ac.uk

19

## Abstract

Modelling is an important component of contingency planning and control of disease outbreaks. Dynamic network models are considered more useful than static models because they capture important dynamic patterns of farm behaviour as evidenced through animal movements. This study evaluates the usefulness of a dynamic network model of swine fever to predict pre-detection spread via movements of pigs, when there may be considerable uncertainty surrounding the time of incursion of infection. It explores the utility and limitations of animal movement data to inform such models and as such, provides some insight into the impact of improving traceability through real-time animal movement reporting and the use of electronic animal movement databases. The study concludes that the type of premises and uncertainty of the time of disease incursion will affect model accuracy and highlights the need for improvements in these areas.

## Introduction

The epidemics of bovine spongiform encephalopathy in Europe[1] and of foot-and-mouth disease in the UK[2] showed the importance of using mathematical models of disease transmission in providing key information to design contingency planning for animal disease outbreaks. By providing epidemiological insight that can be considered alongside the complex interactions between social, economic and welfare outcomes of disease incursions and control strategies, models have helped to inform decisions on disease control[2-5], and can also be used judiciously as tools to improve communication with non-expert stakeholders[6]. Models must be based on robust data and assumptions to usefully inform policies and add value to field-based control activities. However, disease control decisions during epidemic responses are made in the context of wide range of uncertainties. Improving our understanding of the impact of these uncertainties on infectious disease models outcomes is therefore a way to improve their capabilities to efficiently inform policy.

Network models, which were once confined to physics and social science problems[7], have proliferated in the field of human[8-10] and animal[4,5,11,12] health and are increasingly used to inform disease control strategies as part of national contingency plans. When applied to animal diseases, these models consider farms as nodes of a network that are linked by the transfer or movement of (potentially infected) animals. Animal movements are increasingly recorded in national databases, informing on the daily number of animals moved between all farms present in an industry. This large volume of data enables models to appropriately capture the dynamic changes in the contact structure between farms, and therefore enables them to directly adjust for the underlying farm-level economic and behavioural variations when moving animals. As such, predictions from dynamic networks models are potentially more accurate than those from models considering the animal movement network as static[13,14].

As movement of animals within the livestock industry carries the risk of transmitting infectious diseases across substantial geographical distances, dynamic network models have been increasingly used prior to disease outbreaks to improve preparedness. Particularly, dynamic network models have been used to assess the potential for pre-detection spread of infection via movements of animals[5,11], identify regional and local movement patterns[4,11], and provide guidance for the design of efficient control and surveillance strategies[4,12]. However, their use may go further, notably by estimating the extent of the disease spread that has already

58 occurred when disease incursions have been detected and restrictions on animal movements are

59 implemented[15]. By quickly and accurately estimating the spatial extent of the pre-detection spread via

60 movements of animals, they potentially offer additional tools to support field-based contact tracing, and

61 increase the efficiency of disease control responses. However, little work has been done to exploit dynamic

62 network models to such effect.

63 The emphasis on using dynamic network models for contingency planning, but not during an outbreak, may

64 be due to an assumption that they are less useful for making predictions of disease spread or identifying high

65 risk farms in scenarios in which disease incursion has already occurred[6,16]. This assumption may be based on

66 two prior beliefs: (*i*) that data quality may be compromised by time-lags in data recording; and (*ii*) that the

67 date of infection, which is critical to appropriate data selection, may be difficult to ascertain with any

68 certainty. Time-lags in data recording would mean that models have to rely on historical data. However, this

69 problem has been minimised by the advent of electronic databases which mean farmers may directly report

70 movements ahead of time. As a result, live animal movements, such as for sheep and pigs, are now available

71 in real-time in Scotland (through the Scottish livestock electronic identification and traceability database

72 ScotEID, https://www.scoteid.com/) to inform epidemiological modelling to predict the dissemination of a

73 pathogen throughout the livestock industry in a timeframe relevant to disease control activities.

74 Establishing an accurate date of infection is crucial for identifying which data should be included in the

75 model. This can be difficult, as it depends on factors such as clinical presentation and the success of field-

76 based contact tracing procedures, both of which can vary widely. The impact of this uncertainty around date

77 of infection may depend on the temporal dynamism in the pattern of animal movements between farms, and

78 differences in farm trading behaviour, in a given livestock industry. This may affect model predictions (and

79 the uncertainty around them) of the patterns of disease spread.

80 The objective of this study is to assess the usefulness of dynamic network models for predicting the spatial

81 extent of the pre-detection spread via movements of animals, when there may be considerable uncertainty

82 surrounding the time of incursion of infection. In order to achieve this objective, we have focused on

83 diseases of pigs (e.g. swine fevers such as classical swine fever (CSF) or African swine fever (ASF) viruses)

84 which have non-specific clinical signs as well as a high potential to be transmitted through animal

85 movements[17,18]. These characteristics provide a useful model scenario because of the challenging nature of

86 disease detection and the increased potential for silent spread within the pig population. We then explored

87 the usefulness and limitations of using pig movement data (using ScotEID as an exemplar) to inform models

88 when attempting to respond to an infectious disease incursion. Thus the results of this study should also

89 provide insight into the impact of improving traceability through real-time animal movement reporting and

90 the use of electronic animal movement databases.

## Results

### Impact of uncertainty in infection time

93 We looked at the extent to which inaccuracy in defining the disease incursion date may impact on the

94 accuracy of predictions of pre-detection spread of acute swine diseases via movements of pigs. A premises-

95 based model was developed to simulate their spread through the Scottish swine industry via movements of

96 pigs, in which gathering places (such as markets, and collection centres) were explicitly modelled together

97 with pig producers. In the first instance, we considered the extreme case where infection occurs if at least 1

98 animal from an infected premises is received by a susceptible one. In this situation, the "infection paths" $\Gamma_{t,i}$

99 of farms that were infected via movements of animals from a single pig producer $i$ was computed for each

100 Monday of the year 2012. Here, we considered all $i^{th}$ producers that were active during the period $[t_0, t_0+T]$

101 eligible to be an index-case, where $t_0$ is the incursion date and $T$ is the "pre-detection period" (that is, the

102 period between the date of the incursion $t_0$ and the date of the first detection of the disease). We then

103 compared the infection paths $\Gamma_{t,i}$ with those $\Gamma_{t+\delta,i}$ generated when time of infection $t_0$ is inaccurately estimated

104 by an error $\delta$ ranging from $-7\delta_0$ to $7\delta_0$. In this study, infection path $\Gamma_{t,i}$ refers to the "correct" full epidemic

105 tree that is generated by a single infection event at time $t_0$ and left freely spreading for the pre-detection

106 period $[t_0, t_0+T]$, while $\Gamma_{t+\delta,i}$ refers to the "predicted" full epidemic tree when the incursion date is

107 inaccurately estimated and for which the pre-detection period is $[t_0+\delta, t_0+T-\delta]$. We considered, $\delta_0=7$ days

108 and $T=60$ days[19].

109 In Figure 1, we show how increasing uncertainty around the time of incursion may affect one's ability to

110 accurately predict not only the number of premises involved in the full epidemic tree but also their identity.

111      Overall, progressively increasing the error $\delta$ around the time of the incursion from $\delta_0$ to $7\delta_0$ yielded a marked

112      reduction in the correlation between sizes (i.e., the number of premises involved in) of infection paths $\Gamma_{t,i}$

113      and $\Gamma_{t+\delta,i}$ (Figures 1a-b). Although this reduction was consistent across paths of all sizes (Figures 1a), it was

114      more pronounced for paths of larger sizes (Figure 1b). Also, there was a clear divide between infection paths

115      generated from commercial producers and those generated from non-commercial producers (Figure 1c).

116      Despite a wide uncertainty on the time of the incursion, the correlation remained high between infection

117      paths generated by commercial producers (Spearman's correlation coefficient $\rho$>0.60), whether assured or

118      non-assured, for errors ranging from -$7\delta_0$ to $4\delta_0$. In contrast, correlation between infection paths becomes

119      weaker for incursions in non-commercial producers, with $\rho$ <0.60 for errors of ±$3\delta_0$.

120      In order to see if we could accurately predict which individual premises would be involved in epidemics

121      despite some inaccuracy in the incursion time, we compared the concordance between infection paths $\Gamma_{t,i}$ and

122      $\Gamma_{t+\delta,i}$ generated from the same index-case $i$, by calculating the Jaccard similarity index $J(\Gamma_{t,i},\Gamma_{t+\delta,i})$. The

123      Jaccard index measures the fraction of common premises within paths $|\xi_{t,i} \cup \xi_{t+\delta,i}|$ among the total number of

124      premises $|\xi_{t,i} \cap \xi_{t+\delta,i}|$ involved in both paths. Here, we only focused on infection paths involving more than

125      10 infected premises.

126      Progressively increasing the error around the infection time up to $7\delta_0$ revealed a reduction in the median

127      degree of overlap between paths (Figure 1$d$). The rate of this reduction differed, however, whether the

128      incursion time is believed earlier (i.e. $\delta$<0) or later (i.e. $\delta$>0) than the true one. Overall, an error of -$4\delta_0$ in the

129      infection time yielded 77% (95% CI 0.76 - 0.79) overlaps between the true and predicted paths, whereas an

130      error of >$2\delta_0$ is enough to create completely different paths with paths involving, on average, less than half

131      of common premises.

132      Unsurprisingly, variations between producer types were observed in the degree of overlap between $\Gamma_{t,i}$ and

133      $\Gamma_{t+\delta,i}$. While the degree of overlap between predicted and the true paths generated by commercial producers

134      followed closely the general trend, it differs greatly when considering paths generated by non-commercial

135      small producers. This was expected, because most paths of >10 infected premises have been generated by

136      commercial producers. However, differences between the degree of overlap for paths generated by

137    commercial and those by non-commercial producers depends on the direction of the error $\delta$: when $\delta<0$,

138    predicted paths generated by non-commercial producers have a greater number of common premises with

139    the true path, whereas paths would show a completely different pattern (i.e. $J(\Gamma_{t,i},\Gamma_{t+\delta,i})<10\%$) from $>4\delta_0$

140    (Figure 1d). These results suggest that if incursion occurs in non-commercial producers, conservative

141    estimates in incursion times would be preferential. However, this may not be true for incursion occurring in

142    commercial producers as a trade-off may exist between optimising the proportion of premises that are truly

143    on the infection path (true positives) and minimising the proportion of premises that are not (false positives).

144    Figure 2 explores how these two epidemiological measures vary with $\delta$ for paths generated by the different

145    producer types. Over-estimating incursion times for outbreaks generated from commercial producers

146    (whether assured or not), would increase the risk of misclassification. For example, inferences generated for

147    outbreaks from non-assured and from assured commercial producers when $\delta =-5\delta_0$ would involve 24% (95%

148    CI 22% - 29%) and 39% (95% CI 37% - 41%) of false positives, respectively (Figure 2b).

149    So far in this analysis, the potential for spread of infection via movements of animals has been evaluated

150    considering that any movement from infected premises during the pre-detection period would result in

151    disease transmission to susceptible farms. In reality, the prevalence of disease within infected premises will

152    determine what proportion $\beta$ of its livestock becomes infectious. This, together with the number of animals

153    that are being moved off, will determine what proportion of movements will contain infectious animals. To

154    gain general insight and ensure robustness of the results to variation in $\beta$, 10,000 simulations for each

155    Monday of the year 2012 with a random index-case per simulation were carried (i.e., total of 570,000

156    simulations). For each incursion date $t_0$, the infection paths $\Gamma'_{t,i}$ of farms that were infected via animal

157    movements from a single pig producer $i$ was then computed and compared to the infection paths $\Gamma'_{t+\delta,i}$ that

158    were predicted when an error $\delta$ around the time of the incursion is made. As above, $\Gamma'_{t,i}$ and $\Gamma'_{t+\delta,i}$ are the

159    "correct" and "predicted" partial epidemic tree, respectively, and correspond to all farms that have a non-null

160    probability of being infected via animal movements from a single pig producer $i$. Figure 3 shows that,

161    whether comparison is made with the "correct" full epidemic tree $\Gamma_{t,i}$ (i.e. when $\beta=1$) or with the "correct"

162    partial epidemic tree $\Gamma'_{t,i}$ (i.e. when $\beta<1$), qualitatively similar results as in Figure 1 are obtained. However, it

163    further appears that decreasing the value of $\beta$ would reduce the effect of $\delta$ when predicting the size of the

164    infection path (Figure 3a). It is to note, however, that this result may give a false sense of security as the

165    degree of overlaps between correct and predicted paths still sharply decreases with increasing error $\delta$ around

166    the time of the incursion from $\delta_0$ to $7\delta_0$ (Figure 3b).

**Intrinsic structure of infection paths**

168    Although our findings suggest that inferring the spread of an epidemic from dynamic network models is

169    precarious when the date of the disease incursion is unknown, infection paths may have some intrinsic

170    structure which may still guide contact tracing procedures. Previously, such a structure was found in the

171    Italian cattle industry by comparing epidemic trees and regrouping index-cases which generated similar

172    trees, thereby providing critical information to optimize surveillance systems and define rapid containment

173    strategies[4]. Applying a similar method for the Scottish swine industry, however, would only result in

174    regrouping producers that belong to the same business or are part of the same breeding pyramid. Instead, we

175    looked at the producer type of both the index case and all farms that have been infected via the movement of

176    animals when considering $\beta=1$, and determined, for all full epidemic tree $\Gamma_{t,i}$ that gave rise to at least 10

177    cases from the year 2012, the proportion of producers of each type that were involved in each infection path.

178    The results are summarized in Figure 4.

179    If disease incursion occurs in the herd of a small producer, the mean risk of disease spillover into assured

180    producers is low (0.011); and similar to the mean risk of disease spillover from assured producers to small

181    producers (0.032). Epidemics which start in a small producer spread into at least one assured producer in

182    only 1.9% of the incursions. However, once an assured producer is infected, 60% (Q1-Q3: 17% - 71%) of

183    the premises in the generated infection paths would belong to assured producers. In contrast, epidemics

184    generated from assured producers would spread into small producers in 39% of the incursions, but would not

185    involve many of them, with only 8% (Q1-Q3: 2% - 27%) of premises in these infection paths belonging to

186    small producers. These findings are the consequence of producers adhering to quality assurance scheme

187    guidelines on risks associated with animal trading[20], confirming that excluding interactions with producers

188    that have lower biosecurity standards is a good biosecurity practice[21]. Such a result may constitute a basis for

189    the development of qualitative rules modulating surveillance activities in the face of an epidemic.

8

190  Non-assured commercial producers appear to have a totally different epidemiological profile (Figure 4).

191  Non-assured commercial producers have a consistently high probability (> 95%) of being on an infection

192  path and make up, on average, 17% (Q1-Q3: 9% - 22%) of premise in these paths, regardless of the producer

193  type of the incursion. In addition, epidemics generated by non-assured producers show a high likelihood of

194  infecting both small producers (0.62) and assured producers (0.85). This result highlights that Scottish swine

195  producers who are commercially driven but do not belong to assurance schemes may potentially represent

196  "epidemiological" bridges between non-commercial and commercial partners, likely because they

197  implement lower biosecurity, particularly with regard to sourcing and sending pigs, compared to assured

198  commercial producers. Therefore, improving biosecurity and targeting surveillance to non-assured producers

199  may be particularly beneficial to optimise responses to disease incursions.

## Discussion

201  In order to improve preparedness for disease incursion, it is critical to have some understanding of model

202  resiliency to uncertainties which fundamentally underlie the stochastic nature of disease control activities. In

203  this study, we evaluated the resilience of dynamic network models in predicting disease spread after disease

204  incursion, when there may be considerable uncertainty surrounding the timing of infection. A model which

205  predicts the spread of swine fevers was chosen as an exemplar because of the characteristics of the disease

206  and its parameterisation using pig movement data from an electronic database. This has particular relevance

207  and potential policy impact because ASF virus has recently spread within the eastern European region[22] and

208  the middle east[23], and now poses an imminent threat to the European swine industry[24,25]. Although there are

209  measures in place to reduce the risk of introduction of disease, such as restrictions on the movement of live

210  pigs and animal products in affected areas, and regulations on animal swill feeding (which has been banned

211  in the European Union since 2002), further incursions and spread of these diseases throughout Europe are

212  considered likely[24,26,27].

213  Our analysis not only confirmed that increasing the uncertainty around the incursion date significantly

214  reduced the ability of dynamic network models to predict epidemic characteristics, such as epidemic size, or

215  specific premises that become infected, but also quantified the magnitude of the loss of accuracy of

216  predictions. For example, erroneously estimating the time of incursion more than three weeks earlier appears

217  to generate a low accuracy of predicting cases (i.e. less than 60%, Figure 1d), which would miss between

218  30% to 50% of the potentially infected farms (Figure 2a). Although such a measurement bias may

219  potentially generate longer and more severe epidemics, it may be preferable to the alternative

220  misclassification error. A prediction that a farm is potentially infected, when it is not likely to be because of

221  the true absence of contact with an at-risk farm, may have unintended negative consequences for resource

222  allocation (of veterinarians which may be needed more urgently elsewhere) and farmer welfare and

223  behaviour (in response to the fear for potential loss of livestock and livelihood).

224  The type of premises where the incursion occurs can drastically impact on the scale of both of these biases

225  and, therefore, on the resilience of predictions to temporal uncertainties. In the Scottish swine industry, the

226  predictability of the number of premises infected via animal movement (Figure 1c) and of specific premises

227  that become infected (Figure 1d) differ whether epidemics are generated by commercial or non-commercial

228  producers. While our results indicate that all inferences produced from dynamic network models clearly

229  suffer when the time of infection is estimated earlier (Figures 1d and 2), more conservative estimates of time

230  of infection appear only preferable when incursion occurs in small producers. In this situation, widening the

231  time window considered for the incursion would ensure that the incursion is included while not losing

232  performance. Although this may be counterintuitive, it could be explained by the frequency of movements

233  occurring from small producers. It has been previously shown that the rate of movement from and to small

234  producers in Scotland is four to ten times lower than commercially-driven producers[20], with an average of a

235  movement every 29 weeks. It is therefore likely that increasing the time window for the incursion would

236  include most of the movements that may be infectious while avoiding the inclusion of a large number of

237  farms that are not infected. These results suggest that widening the time window considered for the incursion

238  would provide a cost-efficient strategy when responding to incursion of infectious diseases in small

239  producers, avoiding wasting resources that would be required to establish a precise incursion date.

240  In the model, we have first assumed that the trade of at least one animal between infected and susceptible

241  premises was sufficient to allow infection to occur. It is obvious that this assumption may overestimate the

242  extent of disease spread via movements of pigs (although bearing in mind that this model did not consider

243  the potential for spread by other routes), as the infection process between farms is stochastic and depends on

10

244 the within-farm prevalence as well as the virulence of the relevant outbreak strain. However, these

245 assumptions seem appropriate because they not only increase the ease of the comparison between epidemic

246 trees, but also enable (1) robust estimates of the potential geographical extent of disease spread that is

247 consistent with contact tracing procedures and (2) communication of the general implication of temporal

248 uncertainties in model inferences to policy makers (and model users in general). Nevertheless, varying the

249 probability of transmission did not change the qualitative outcome of our analysis (Figure 3).

250 It is clear from this study that on detection of an incursion, effort should be focused on obtaining an accurate

251 incursion date. Improved accuracy of this estimate will improve the validity of epidemiological outputs from

252 dynamic network models at early stages of an epidemic, and therefore will optimise the identification of the

253 sources of infection and any presumed susceptible in-contact animals. However, quick detection of disease

254 incursions is also critical. While the role of small producers in the spread of swine diseases has been

255 previously shown[28,29], routine surveillance activities (i.e. surveillance conducted not during an outbreak)

256 mostly target assured commercial producers (for example abattoir inspection, veterinary/health scheme

257 monitoring). Superficially, this risk-based surveillance strategy is reasonable because of the important

258 influence of commercial producers on the sustainability of pig products (and the pig industry) and thus, food

259 security[30]. However, Figure 4 suggests that exclusively targeting assured commercial producers during

260 routine surveillance activities will likely miss incursion events in backyard producers. Simulation studies

261 looking at the spread of CSF in Bulgaria, where small producers are believed to play a role in the persistence

262 of the disease[31], have shown that infections from small producers to assured producers were rare[32]. Although

263 consistent with our findings (Figure 4), our results also indicated that non-assured commercial producers

264 may constitute a bridge of infection between the non-commercial and commercial sectors of the swine

265 industry in Scotland. With regards to improving surveillance for incursions of emerging swine diseases in

266 Scotland, non-assured commercial producers may represent a sentinel population which would allow the

267 detection of incursions in the non-commercial sector of the industry.

268 In this study, we assessed the usefulness of national electronic animal movement databases as a tool for

269 traceability by examining the degree to which uncertainty around incursion time may affect predictions on

270 the pre-detection spread of emerging swine diseases such as CSF and ASF in Scotland. Our results on

271     movement patterns of swine in Scotland are also important for other exotic diseases of swine (e.g. foot-and-

272     mouth disease) and may have relevance for other swine industries. Although the pig industry in Scotland is

273     small, commercial production is well organised and focuses on assured production of high quality farrow-to-

274     finish pigs. The pig industry in Scotland also shows a relative high diversity of producer types, with a large

275     proportion of non-commercial pig holdings[20]. The Scottish swine industry may then represent a good

276     example for similar industries, where non-commercial pig farming has an important place.

277     In Scotland,  movements of swine shows a lack of seasonality[20], similar to what has been reported in other

278     countries[12,33]. It may therefore be possible to extrapolate these results to other similarly structured pig

279     populations. In contrast, more work is required to determine whether these findings are applicable to other

280     livestock sectors. The magnitude and directionality of movements of cattle and sheep in Scotland are highly

281     seasonal. As such, these patterns will likely have an impact on the probability of epidemic take off[5,11], and

282     therefore are also likely to affect the predictability of the network structure in these sectors.

283     In conclusion, the type of premises and the uncertainty of the time of disease incursion will affect dynamic

284     network model accuracy and thus, usefulness. Cursorily, it may appear that if the incursion time is uncertain,

285     using conservative estimates of incursion time (i.e. covering a wider time window) would increase the

286     probability of detecting all potentially infected farms. However, this approach also generates a larger number

287     of premises that would require field-based investigation (of which a higher proportion would be negative),

288     which would be challenging when resources are limited. Resources may be better placed trying to more

289     accurately determine the incursion time, since dynamic network models can make valuable predictions to

290     help with disease control and resource allocation if the incursion time is known. In such situation, efforts in

291     improving surveillance prior to disease incursion are critical to optimise responses to disease incursions.

## Methods

### Data

294     All movement data were extracted from the Scottish livestock electronic identification and traceability

295     database (ScotEID) which came into use in November 2011. We refer to [20] for further details on the data

296     collection, process and quality as well as some preliminary descriptive analyses.

297   Briefly, under Scottish (SSI 2011/351) and European legislation (Commission Decision 2000/678/EC), all

298   pig keepers moving animals are required to register online with ScotEID and electronically record any

299   movements ahead of time. To avoid selection bias due to inevitable missing or non-reported movements in

300   the early stages of implementation of the database, we restricted our analysis to all movements recorded

301   from January 1st 2012 to May 31st 2013. We used January 1st 2012 for the start of the study period, on the

302   basis that (1) it corresponds to the time when the previous movement database (the Scottish Animal

303   Movement System, SAMS) recording Scottish animal movements ended (i.e. November 2011), and (2) there

304   has been a stabilisation of the movement pattern since December 2011.

305   The database provides a comprehensive picture of all movements of pigs in Scotland at the batch level. As

306   such, each movement record reports the County Parish Holding (CPH) identifier and postcode for departures

307   and destinations, the number of animals involved, and the date of the movements. Details of premises type

308   for departures and destinations are recorded in the movement database, allowing slaughterhouses, markets,

309   show-grounds and ferry collection centres to be differentiated from agricultural holdings. Note that all

310   markets recorded in ScotEID operate as auctioneers holding dedicated sales/collections of pigs for onward

311   consignment to a slaughterhouse, also named "red markets". Collections of animals that are destined to be

312   slaughtered are therefore regularly carried out in these markets, but remain separated from the other

313   activities of such premises, particularly activities dedicated to sales of pigs between producers.

314   **Pig producer types**

315   Through the CPH identifier, the ScotEID movement database was linked to the 2011 Scottish Agricultural

316   Census, and to the 2010 GB Agricultural Census, to obtain information on the total number of pigs and sows

317   present on farm. We further link the data to the 2013 Quality Meat Scotland (QMS) register (for Scottish

318   premises) and to the 2013 Red Tractor register (for non-Scottish premises) to identify if producers were

319   members of a health quality assurance scheme. Pig producers were then classified according to their pig

320   population size, movement activity and the health quality assurance scheme membership[20]:

1. "Small pig producers": agricultural holdings with an unknown number of pigs; or less than five sows, and/or less than 10 finishers; and showing no records of movements of more than 50 pigs within the study period.

2. "Non-assured commercial producers": agricultural holdings with more than five sows and/or more than 10 finishers; or showing records of movements of more than 50 pigs during the study period, but do not belong to a quality health assurance scheme from QMS or Red Tractor, the main British assurance schemes.

3. "Assured commercial producers": agricultural holdings with more than five sows and/or more than 10 finishers; or showing records of movements with more than 50 pigs during the study period but also belong to a quality health assurance scheme from QMS and/or Red Tractor.

**Infection path**

The spread of disease within the Scottish swine industry was modelled using a simple stochastic discrete-time SI model. Our model treated each premises involved in the movement of Scottish pigs as a single unit. In this model, all premises are susceptible (S) to the infection at the start of the epidemics, except for a single premises, chosen at random, that would initially be at the infected, and infectious, state (I). During the course of an epidemic, disease passes from infected premises $i$ to susceptible ones $j$ via movements of pigs with a probability $M_{ij,t}$ such as $M_{ij,t} = 1 - (1 - \beta)^{N_{ij,t}}$, where $\beta$ is the probability that a single pig from $i$ may carry the disease and potentially transmit it to $j$ (somewhat corresponding to the within-herd prevalence), $N_{ij,t}$ the number of pigs that moved from $i$ to $j$ per time-step $t$ and with movements of pigs synchronously updated at each time-step. Although other transmission routes have also been implicated in the spread of swine fevers (such as spread via fomites, wild boar, semen or pig products), only infection through live pig movements was considered as it the most common transmission route [17,18]. Here, the model is seeded at incursion time $t_0$, progresses in discrete time steps $t$ of one day, and runs for a fixed period $T$.

In addition to swine producers, gathering places (e.g. markets, show grounds, and ferry collection centres) were considered in the spread of diseases. Regulations are in place in Scotland, as in most EU countries, to ensure that the spread of pathogens via movements of animals through gathering places is limited. Gathering places should not keep pigs overnight and have cleaning and disinfection implemented after each day of

348 activity (Council Directive 97/12/EC). As such, the model considers that all infected gathering places would

349 go back to the susceptible state after one day (thereby following a SIS process), whereas infected swine

350 producers would remain infected for the remaining of the simulation period $T$. As a consequence, epidemics

351 were considered starting by a swine producer only. The model was used only to look at the spread of disease

352 before detection. Therefore the control measures that would be initiated on identification of the disease (such

353 as culling of pigs on infected premises, movement restrictions) were not included in the model.

354 To ensure that only the heterogeneity and the structure of the dynamic network formed by the movements of

355 pigs were driving the modelled epidemics, all swine producers involved in the movements of pigs were

356 considered identical, such that their producer type or herd size would not have any effect on the transmission

357 dynamics. Unless otherwise stated, we considered the extreme case where infection occurs if at least one

358 animal from an infected premises is received by a susceptible one, i.e. when $\beta = 1$. It is obvious that, given

359 such a model structure, the characteristics of simulated epidemics would be overestimated and would not

360 reflect the intrinsic potential of disease spread in the Scottish swine industry. However, such a model

361 provides information on the maximum infection tree generated by each index-case via movements of

362 animals, which not only provides an estimate of the maximum epidemic size generated by the movement of

363 animals for the considered $T$, but also identifies all premises that are likely to be infected. Furthermore, such

364 a model structure provides an estimate of infection trees for each incursion location that is easily comparable

365 between time periods.

366

# References

1    Anderson, R. M. *et al.* Transmission dynamics and epidemiology of BSE in British cattle. *Nature* **382**, 779-788 (1996).

2    Keeling, M. *et al.* Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813 - 817 (2001).

3    Backer, J. A., Hagenaars, T. J., van Roermund, H. J. W. & de Jong, M. C. M. Modelling the effectiveness and risks of vaccination strategies to control classical swine fever epidemics. *J R Soc Interface* **6**, 849-861, doi:10.1098/rsif.2008.0408 (2009).

4    Bajardi, P., Barrat, A., Savini, L. & Colizza, V. Optimizing surveillance for livestock disease spreading through animal movements. *Journal of the Royal Society Interface* **9**, 2814-2825, doi:10.1098/rsif.2012.0289 (2012).

5    Tildesley, M. J. *et al.* Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature* **440**, 83-86, doi:10.1038/nature04324 (2006).

6    Taylor, N. *Review of the use of models in informing disease control policy development and adjustment.*, 98 (A report for Department for Environment Food & Rural Affairs (DEFRA), 2003).

7    Newman, M. The structure and function of complex networks. *Society for Industrial and Applied Mathematics Review* **45**, 167-256 (2003).

8    Eames, K. T. D. & Keeling, M. J. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences* **99**, 13330-13335, doi:10.1073/pnas.202244299 (2002).

9    van Bunnik, B. *et al.* Efficient national surveillance for health-care-associated infections. *BMC Public Health* **15**, 832 (2015).

10   Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences* **106**, 21484-21489, doi:10.1073/pnas.0906910106 (2009).

11   Green, D. M., Kiss, I. Z. & Kao, R. R. Modelling the initial spread of foot-and-mouth disease through animal movements. *Proceedings of the Royal Society B: Biological Sciences* **273**, 2729-2735, doi:10.1098/rspb.2006.3648 (2006).

12   Ciccolini, M., Dahl, J., Chase-Topping, M. E. & Woolhouse, M. E. J. Disease transmission on fragmented contact networks: Livestock-associated Methicillin-resistant Staphylococcus aureus in the Danish pig-industry. *Epidemics* **4**, 171-178, doi:http://dx.doi.org/10.1016/j.epidem.2012.09.001 (2012).

13   Vernon, M. C. & Keeling, M. J. Representing the UK's cattle herd as static and dynamic networks. *Proceedings of the Royal Society B: Biological Sciences* **276**, 469-476, doi:10.1098/rspb.2008.1009 (2009).

14   Fefferman, N. H. & Ng, K. L. How disease models in static networks can fail to approximate disease in dynamic networks. *Physical Review E* **76**, 031919 (2007).

15   Danon, L. *et al.* Networks and the Epidemiology of Infectious Disease. *Interdisciplinary Perspectives on Infectious Diseases* **2011**, 28, doi:10.1155/2011/284909 (2011).

406  16  Dubé, C. *et al.* in *Compendium of technical items presented to the 75th World Organisation for*
407      *Animal Health (OIE) General session.*  13-23.

408  17  Sánchez-Vizcaíno, J. M. in *Disease of Swine*   (eds Barbara E. Straw, Jeffrey J. Zimmerman, Sylvie
409      D'Allaire, & David J. Taylor)  291-298 (Blackwell Publishing, 2006).

410  18  Le Potier, M.-F., Mesplède, A. & Vannier, P. in *Disease of Swine*   (eds Barbara E. Straw, Jeffrey J.
411      Zimmerman, Sylvie D'Allaire, & David J. Taylor)  309-322 (Blackwell Publishing, 2006).

412  19  Elbers, A. R. W. *et al.* The classical swine fever epidemic 1997–1998 in the Netherlands: descriptive
413      epidemiology. *Preventive Veterinary Medicine* **42**, 157-184, doi:http://dx.doi.org/10.1016/S0167-
414      5877(99)00074-4 (1999).

415  20  Porphyre, T. *et al.* How commercial and non-commercial swine producers move pigs in Scotland: A
416      detailed      descriptive      analysis.      *BMC      Veterinary      Research*      **10**,      140,
417      doi:http://www.biomedcentral.com/1746-6148/10/140 (2014).

418  21  Toma, L., Stott, A. W., Heffernan, C., Ringrose, S. & Gunn, G. J. Determinants of biosecurity
419      behaviour of British cattle and sheep farmers—A behavioural economics analysis. *Preventive*
420      *Veterinary Medicine* **108**, 321-333, doi:http://dx.doi.org/10.1016/j.prevetmed.2012.11.009 (2013).

421  22  Costard, S. *et al.* African swine fever: how can global spread be prevented? *Philosophical*
422      *Transactions    of    the    Royal    Society    B:    Biological    Sciences*    **364**,    2683-2696,
423      doi:10.1098/rstb.2009.0098 (2009).

424  23  Rahimi, P. *et al.* Emergence of African Swine Fever Virus, Northwestern Iran. *Emerging Infectious*
425      *Disease* **16**, 1946-1948, doi:DOI: 10.3201/eid1612.100378 (2010).

426  24  Nigsch, A., Costard, S., Jones, B. A., Pfeiffer, D. U. & Wieland, B. Stochastic spatio-temporal
427      modelling of african swine fever spread in the European Union during the high risk period. *Prev Vet*
428      *Med* **108**, 262-275, doi:http://dx.doi.org/10.1016/j.prevetmed.2012.11.003 (2013).

429  25  De Vos, C. J., Saatkamp, H. W., Nielen, M. & Huirne, R. B. M. Scenario tree modeling to analyze
430      the probability of classical swine fever virus introduction into member states of the European Union.
431      *Risk Anal* **24**, 237-253, doi:10.1111/j.0272-4332.2004.00426.x (2004).

432  26  Costard, S. *et al.* Introduction of african swine fever into the European Union through illegal
433      importation of pork and pork products. *PLoS One* **8**, e61104, doi:10.1371/journal.pone.0061104
434      (2013).

435  27  Wooldridge, M., Hartnett, E., Cox, A. & Seaman, M. Quantitative risk assessment case study:
436      smuggled meats as disease vectors. *Revue Scientifique et Technique de l'Office International des*
437      *Epizooties* **25**, 105-117 (2006).

438  28  Lupulovic, D. *et al.* First Serological Study of Hepatitis E Virus Infection in Backyard Pigs from
439      Serbia. *Food Environ Virol* **2**, 110-113, doi:10.1007/s12560-010-9033-6 (2010).

440  29  Martínez-López, B., Alexandrov, T., Mur, L., Sánchez-Vizcaíno, F. & Sánchez-Vizcaíno, J. M.
441      Evaluation of the spatial patterns and risk factors, including backyard pigs, for classical swine fever
442      occurrence in Bulgaria using a Bayesian model. *Geospatial Health* **8**, 489-501 (2014).

443  30  Stark, K. *et al.* Concepts for risk-based surveillance in the field of veterinary medicine and
444      veterinary public health: Review of current approaches. *BMC Health Services Research* **6**, 20
445      (2006).

446   31    Alexandrov, T., Kamenov, P. & Depner, K. Surveillance and control of classical swine fever in
447          Bulgaria, a country with a high proportion of non-professional pig holdings. *Epidémiol. et santé*
448          *anim.*, 140-142 (2011).

449   32    Martínez-López, B. *et al.* Evaluation of the risk of classical swine fever (CSF) spread from backyard
450          pigs to other domestic pigs by using the spatial stochastic disease spread model Be-FAST: The
451          example      of      Bulgaria.      *Veterinary      Microbiology*      **165**,      79-85,
452          doi:http://dx.doi.org/10.1016/j.vetmic.2013.01.045 (2013).

453   33    Bigras-Poulin, M. B., K; Mortensen, S & Greiner, M. Relationship of trade patterns of the Danish
454          swine industry animal movements network to potential disease spread. *Prev Vet Med* **80**, 143-165,
455          doi:http://dx.doi.org/10.1016/j.prevetmed.2007.02.004 (2007).

456

457

# Acknowledgements

# Authors' contributions

TP designed the study, carried out the modelling and statistical analyses and drafted the manuscript. LAB and HKA participated in the design of the study and contributed to final manuscript. CCG provided information on the Scottish swine industry and contributed to final manuscript. GJG and MEJW contributed to final manuscript. All authors gave final approval for publication.

# Additional Information

The authors declare no competing financial interests.

19

## Figure legends

Figure 1. Comparison between the correct infection path and predicted paths generated when the error $\delta$ in the time of the incursion ranges from $-7\delta_0$ to $7\delta_0$. (a) Lines plot showing the smoothed size of the predicted full epidemic tree $\Gamma_{t+\delta,i}$ as a function of the size of the correct full epidemic tree $\Gamma_{t,i}$. (b) Changes in the Spearman correlation coefficient between the size of $\Gamma_{t,i}$ and that of $\Gamma_{t+\delta,i}$ as a function of the error $\delta$ in the time of the incursion. Correlation coefficients are computed either upon all generated infection paths or upon infection paths of >10 infected premises. (c) Changes in the Spearman correlation coefficient between the size of $\Gamma_{t,i}$ and that of $\Gamma_{t+\delta,i}$ as a function of $\delta$ and stratified by the producer type of the index-case. (d) Quality of infection path prediction, as measured by the median Jaccard similarity index, as a function of $\delta$ and stratified by the producer type of the index-case. Shaded areas around each line shown in (a)-(d) represent their respective confidence interval. Here, $\delta_0=7$days. Diagonal solid line in (a) indicates perfect concordance between the true and predicted length of infection paths. The vertical solid line in (b)-(d) indicates the time of the correct incursion time.

Figure 2. Proportions of true (a) and false (b) positives between the correct full infection path and predicted paths generated when the error $\delta$ in the time of the incursion ranges from $-7\delta_0$ to $7\delta_0$. Here, $\delta_0=7$days. Points/lines represent the observed median proportions, stratified as a function of the producer type of the index-case, whereas shaded areas represent their respective 95% confidence interval. Only infection paths of >10 infected premises are used. The vertical solid line indicates the time of the correct incursion time. The proportion of true positives measures the fraction of common premises within paths $|\xi_{t,i} \cup \xi_{t+\delta,i}|$ among the number of premises $|\xi_{t,i}|$ that are on the correct path. The proportion of false positives measures the fraction of uncommon premises within paths $1 - |\xi_{t,i} \cup \xi_{t+\delta,i}|$ among the number of premises $|\xi_{t+\delta,i}|$ that are on the wrong path.

Figure 3. Comparison between the correct infection path and paths generated when the error $\delta$ in the time of the incursion ranges from $-7\delta_0$ to $7\delta_0$ and when imperfect transmission occurs. (a) Changes in the Spearman correlation coefficient between the size of the correct infection path and paths generated when both the error $\delta$ in the time of the incursion and the transmission probability $\beta$ vary. Quality of infection path prediction, as

20

502  measured by the median Jaccard similarity index (b), proportion of true (c) positives and proportion of false

503  positives (d) between the correct infection path and paths generated when both $\delta$ and $\beta$ vary. Solid and dotted

504  lines indicate how measures may change when comparing predicted partial infection path $\Gamma''_{t+\delta,i}$ to either the

505  correct partial epidemic tree $\Gamma''_{t,i}$ generated with $\delta=0$ and $\beta<1$ (solid) or the correct full epidemic tree $\Gamma_{t,i}$

506  generated with $\delta=0$ and $\beta=1$ (dotted). Shaded areas around each line shown in (b)-(d) represent the

507  confidence interval around the median. Here, $\delta_0=7$days. The vertical solid line in (a)-(d) indicates the time of

508  the correct incursion time.

509  Figure 4. Proportion of non-assured commercial, assured commercial and small non-commercial producers

510  involved in infection paths of >10 infected premises generated by each producer type. Here, columns

511  indicate the producer type of the index-case, whereas rows indicate the type of the producers that are

512  involved in each infection path. The thickness of the shapes is proportional to the density of data points

513  along the x-axis.